THE EFFECT OF EARLY LIFE EVENTS ON THE BURDEN OF DIABETES MELLITUS AMONG COSTA RICAN ELDERLY: ESTIMATES AND PROJECTIONS

By
Gilbert Brenes

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Sociology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2008

Abstract

THE EFFECT OF EARLY LIFE EVENTS ON THE BURDEN OF DIABETES MELLITUS AMONG COSTA RICAN ELDERLY: ESTIMATES AND PROJECTIONS

By

Gilbert Brenes

Under the supervision of Professor Alberto Palloni

At the University of Wisconsin-Madison

Current cohorts of Latin American elderly have witnessed the fast demographic and epidemiologic transition and the process of economic modernization through which the region underwent during the 20th century. During their childhood years, they experienced an environment characterized by economic deprivation, malnutrition, and high prevalence of infectious diseases, but survived to old age due to the introduction of medical technology and public health measures (Palloni, 1981). In light of the thrifty phenotype hypothesis (Hales and Barker, 1992; 2001), people that were undernourished during gestation and early childhood are more likely to develop Type II Diabetes Mellitus (DM). Given the high prevalence of these conditions among current Latin American elderly cohorts, Palloni *et al* (2006) and Prentice and Moore (2005) have hypothesized that there might be an epidemic of DM.

The goal of this dissertation is to estimate the effect of early childhood conditions on DM burden among the population 60 years old and older in Costa Rica, and to project the prevalence

of DM among this population for the period 2005-2030. I use short knee height (KH) and the level of childhood mortality (CMI) in the respondents' place of birth as surrogate markers of malnutrition during gestation and infancy.

I find that there is weak but significant effect of KH on DM incidence, but only when having short KH interacts with obesity. Being born in high CMI counties is positively associated with having high levels of glucose or glycosylated hemoglobin. The projected DM burden among Costa Rican elderly shows that the population age 60 years old and older with DM in Costa Rica is going to quadruple in the next 25 years, but this increase is basically due to the growth in the total elderly population.

I conclude that Costa Rica is going to have a sharp increase in the burden of DM among its senior population, but that this increase is related to population increase rather than to adverse health conditions during childhood. I argue that for other Latin American countries this increase in DM might happen if the same cohorts in those countries experienced adverse early childhood conditions and malnutrition, as well as obesity. As a broader conclusion, these results suggest that the factors that produce the mortality decline in Latin America –public health measures rather than an improvement in socio-economic conditions– do not necessarily lead to a lagged increase in the burden of chronic diseases (like DM) that are associated with deleterious conditions early in life.

ACKNOWLEDGMENTS

I would like to acknowledge the funding as a Trainee from the National Institute of Health's (NIH) Fogarty International Center (FIC) training program (5D43TW001586) at the Center for Demography and Ecology.

I want to thank my family: my mother Jenny Camacho, my sisters Tatiana, Jessica, and Cynthia, and my nephews Gabriel and Camilo. Their support and love were always a guiding light throughout my graduate studies.

I want to thank my committee members for all their work and advice. I will always be grateful to my adviser, Alberto Palloni, who has always trusted in me. Having a conversation about my dissertation topic and about demography in general in his office or in a coffee shop is one of the best experiences that a demography student can ever have. Michel Guillot, my professor in several classes, has also been a key figure in my development as a demographer and I will always thank him for his dedication and good will. Katherine White is a role model for all of us in the CDE; her interesting research projects and intelligent remarks in class or at the DemSem are perfectly complemented with her perennial smile and joyfulness: Thank you for saving me at the last minute. I want to express gratitude for Maureen Smith's support, too. Her keen understanding of research methods and disease etiology was very valuable for finishing this dissertation. I want to thank her as well for her course on "Health Services Research", one of the most interesting classes I took during my graduate studies. I want to thank Burt Olson, too, because his clear understanding of biological processes helped me comprehend the dilemmas I faced with the effect of risk factors on diabetes and with biomarker information that I used in this research project. His class is also one of the most illuminating courses I have ever taken.

I want to thank all my friends and fellow students at the CDE and the Department of Sociology. Many have showed me their support during these years, and I will always be grateful to those with whom I shared joys and sorrows: Martín, Beatriz, Julieta, Carolina, Flavia, Ana Cristina, Marcelo, Teresa, So-Jung, and many others. The good moments at room 7106 will remain in my memory forever.

I am grateful to all the staff at the CDE: Vicki, Carol, Happy, Janet, and Jeff. They all helped me in making my life easier in Madison. Special thanks to Vicki, a light that is always shining due to her cheerfulness and diligence. I also want to thank the staff at the Sociology Department, who helped me a lot especially during the final steps of the process.

I also want to include in the acknowledgments special thanks to my Soc361 students during the Fall of 2007 because their joyfulness and their complains about statistics helped me overcome the very stressful period that precedes the PhD Oral Examination. I want to thank Rick Nordheim for helping me to understand statitistical techniques I used.

I want to thank all the people from Centro Centroamericano de Población CCP in Costa Rica, because their support and knowledge have always convinced me that the path I started here in Madison is worth while.

I want to express special gratitude to Luis Rosero-Bixby, one of the most interesting and genuine scholars I have ever met. Thank you, Don Luis, for initiating me in this amazing and fruitful field called demography and believing that I can be a good demographer; thank you for showing me how interesting population problems are in Costa Rica; thank you for respecting me: This whole PhD enterprise would have been impossible without you.

Finally, thank you, Costa Rica, the most inspiring country on earth.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADA	American Diabetes Association
CCP	Centro Centroamericano de Población
CCSS	Caja Costarricense del Seguro Social
CMI	Child Mortality Index
CRELES	Costa Rica Estudio de Longevidad y Envejecimiento Saludable
CRP	C-reactive Protein
DM	Diabetes Mellitus
GDP	Gross Domestic Product
HbA_{1C}	Glycosylated hemoglobin
IGF	Insulin-like Growth Factors
INEC	Instituto Nacional de Estadística y Censos
KH	Knee height
MHAS	Mexican Health and Aging Study
OGTT	Oral Glucose Tolerance Test
OLS	Ordinary Least Squares
PREHCO	Puerto Rican Elderly: Health Conditions
SABE	Salud, Bienestar y Envejecimiento en América Latina
SG	Serum glucose
U.S.A.	United States of America
UK	United Kingdom
WHO	World Health Organization

Chapter I: Introduction

From an economic perspective, as measured by the Gross Domestic Product (GDP), Latin America includes very productive economies (e.g., Mexico, Chile and Venezuela), as well as some of the poorest nations in the world (e.g., Nicaragua, Honduras and Bolivia). From a demographic perspective, some countries are well advanced in the so-called Demographic Transition (Notestein, 1945), while others are still in the process of reducing their fertility and mortality levels (Chackiel, 2004). Nonetheless, all countries in the region have experienced sharp mortality declines. The average decline between 1950 and 1980 was the fastest in the world, comparable only to the one achieved by certain Asian countries (Palloni, 1981). As predicted by both the demographic and the epidemiological transition frameworks (Notestein, 1945; Omran, 1971), most of the gain in life expectancy was produced by a diminution in cause-specific death rates due to communicable diseases, particularly at early ages. Some of these illnesses, such as Chagas disease, malaria and tuberculosis, have not been entirely eliminated. However, according to assessments using the Global Burden of Disease measures (Murray and Lopez, 1997), their burden in the region is lower than that of chronic diseases, such as cardiovascular diseases, diabetes mellitus (DM), and cancer. The lowering of death rates at all ages, accompanied by a decline in fertility, has contributed to the aging of the population. During the next quarter-century (Guzman, 2002), the segment of the population that is 60 years-old or above will increase in the region at a rate of 3.5% per year. Because of its age, this is the group that will be most affected by chronic and degenerative diseases mentioned above.

Residents of Latin American countries aged 60 or above in 2005 were young during an epoch characterized by persistent and widespread poverty and deficient development of public

infrastructure. Infectious diseases were the most common cause of infant mortality and of considerable health impairments for those who survived (Palloni, 1981; McQuestion, 2000). Malnutrition, as signaled by limited stature, appeared to be highly prevalent during this epoch, affecting both children and adults, especially indigenous and mestizo populations (Bogin and Keep, 1999). However, these cohorts also witnessed the introduction of medical technology and public health measures, such as new vaccines, medications, and oral rehydration therapy, among others (Preston, 1980; Palloni, 1981). In addition, they experienced the economic boom that followed World War II, with both its beneficial and harmful effects. The "developmentalist" approach of the governments during the fifties, sixties, and seventies diversified the countries' productive schemes, increased employment in both the public and a modernized private sector, and expanded public services, such as education and health care (Kaufman, 1990; Villarreal, 1990). However, at the same time, this same group of people was the one that triggered the urbanization process that transformed the region's main cities into overcrowded megalopolis surrounded by circles of shanty towns (Benitez-Zenteno, 2000).

What are the implications of these historical circumstances on the burden of disease in Latin America? The "life-course approach" to chronic disease etiology stresses the importance of "insults," which occur through the life course and increase the likelihood of developing certain diseases (Kuh and Ben-Shlomo, 1997). One of the milestones of this approach is the so-called "Barker Hypothesis" (Kuh and Ben-Shlomo, 1997; Barker, 1998; Barker *et al.*, 2002), which has resulted in a large amount of research focusing on the impact of early life conditions on adult disease. According to this hypothesis, some chronic conditions are "programmed" *in utero* or during early life as the result of circumstances, such as limited caloric intake by the mother or the child, or infectious diseases during infancy. The strongest supporting evidence is

in the area of cardiovascular and metabolic diseases, such as ischemic heart disease, and Type II Diabetes Mellitus (Type 2 DM). During their early lives, the present elderly population of Latin America typically experienced diets deficient in both quantity and quality. In conjunction with this there is evidence that the overall prevalence of chronic diseases among this populations is increasing (see Aschner, 2002, for projected DM trends).

The aim of the present research is to: (1) analyze the possible link between early-life insults and the later-life incidence of Type 2 DM, and (2) to project the future toll of DM due to the persistence in the population of insults in early life. Initially, the effect of other risk factors on the projection will be held constant. Subsequent projections will include the simulation of changes in prevalence produced by modifications in other risk factors. The projection and associated simulations will provide an estimate of how persistent across age certain chronic conditions will be, given the social and epidemiological history of the region. From a theoretical perspective, the projections might show whether a possible decline in the incidence of certain non-communicable diseases (i.e., the transition from the third stage, "Age of Degenerative and Man-Made Diseases," to the fourth stage, "Age of Delayed Degenerative Diseases," of the epidemiologic transition) can be explained by deferred effects of past conditions. From a more pragmatic view, the projections might be useful in predicting future needs and for programming the needs that the public health care system in Costa Rica might have.

In this dissertation, I will utilize micro-level data to analyze the relationship between Type 2 DM and early life conditions, as measured by surrogates such as knee height and the level of child mortality in the respondent's place of birth. The main analysis will be performed

with the dataset from the project “Costa Rica Estudio Longitudinal de Envejecimiento Saludable” (CRELES)[1].

[1] The translation of project name is Costa Rica: Study of Longevity and Healthy Aging.

Chapter II: Theoretical and Historical Background

A. The epidemiologic transition and the "Life course approach" to epidemiology

In 1971, Omran published his breakthrough article, "The epidemiologic transition". Its main idea is that societies sequentially undergo three stages or transitions: (1) The Age of Pestilence and Famine, (2) The Age of Receding Pandemics, and (3) The Age of Degenerative and Man-Made Diseases. The populations experiencing the first stage typically have high mortality across ages and almost null growth in their sizes. In the early phase of the second stage, there are slight improvements in nutrition and sanitation, and mortality due to epidemics and famine declines; however, child death rates are still high. During the later phase, preventive measures (e.g., vaccination, nutritional improvements, hygienic practices, and a potable water and sewage system) are reinforced by new scientific ideas (Preston and Haines, 1991) and help to diminish the incidence of infectious and parasitic diseases; consequently, child and maternal mortality diminish, generating a dramatic increase in life expectancies. In the last stage, degenerative diseases, particularly cardiovascular and cerebro-vascular diseases, become the main causes of death.

In recent years, Omran's framework seemed incomplete for explaining demographic patterns that appear in Europe and East Asia during the late 20th century. These new trends, including declines in heart disease and certain cancer mortality, led Olshansky and Ault (1986) to propose an additional stage in the epidemiologic transition within the framework: The Age of Delayed Degenerative Diseases. This stage is characterized, not only by the decline in the incidence of the causes of death mentioned above, but also by a strong population aging process in which the declines in death rates are occurring at older ages (i.e., after 75 years old). Horiuchi

(1999) synthesizes the framework and adds new concerns such as the emergence of new infectious diseases (e.g./i.e., HIV/AIDS, ebola) and new forms of violence.

Despite these concerns, as well as new declines in cardiovascular disease in some European countries, chronic maladies are the main causes of disability, morbidity and mortality in the industrialized world. Heart disease, stroke, DM, and certain types of cancer (i.e., lung, breast and prostate) remain as the main "killers." The "life course perspective" in epidemiologic research is an application of the epidemiologic transition model to trends within industrialized nations. In particular, it is focused on explaining the risk factors associated with these chronic degenerative diseases. Kuh and Ben-Shlomo (2002) define it as:

> "…(T)he study of long-term effects on chronic disease risk of physical and social exposures during gestation, childhood, adolescence, young adulthood and later adult life. It includes studies of the biological, behavioural and psychosocial pathways that operate across an individual's life course, as well as across generations, to influence the development of chronic diseases…" (p.285).

According to these authors, there are two variants to this approach: (1) the critical period model, and (2) the accumulation of risk model. The latter variant suggests that risk factors, occurring at different times of exposure, accumulate over the life course and culminate in the likelihood of developing a disease. The model might assume that risk factors are independent (i.e., uncorrelated with one another); that they cluster at certain periods of time (e.g., childhood); or that they act together in successive/cumulative "chains." This last mechanism assumes that the presence of a certain risk factor (e.g., poverty in childhood) might increase the likelihood of other subsequent risk factors (e.g., smoking and/or a job with bad environmental conditions).

Kuh and Ben-Shlomo (Kuh and Ben-Shlomo, 1997; Ben-Shlomo and Kuh, 2002) do not directly endorse one of these models above the others. However, their call for more studies that can test the different pathways defined by combinations of risk factors strongly suggests that they prefer this third variant.

The critical period model is clearly represented by research performed by Barker and colleagues, who came to be widely known for their Hertfordshire study (Barker *et al.*, 1989). This team has emphasized the importance of "biological programming", which implies that the insults an organism experiences during gestation or infancy will have an effect on health later in life. They explain that a main "programming mechanism" is developmental phenotypic plasticity; i.e., one genotype can produce different phenotypes depending on environmental conditions during the developmental period (in this case, gestation and early infancy). A fetus or an infant may grow less than his capacity if his mother is small, or if he is exposed to conditions constraining its growth. This smaller phenotype might be well suited for conditions in early life, but not for those in a later stage of life, thus, generating health problems in adulthood and/or later life (Barker *et al.*, 2002). Ben-Shlomo and Kuh (2002) call this framework the "critical period model," because the main risk factors occur during a "critical period": gestation or infancy. In recent versions of the theory, these researchers incorporate the effect of later life risk factors, describing biological interactions that might predispose individuals for illness at later stages in life. In explaining the etiology of cardiovascular disease or DM, the "Southampton group" recalls the effect of accelerated or compensatory growth. They argue that this mechanism can produce cell death, organ degradation, or atypical responses to poor living standards (Barker *et al.*, 2002). Barker and colleagues have published numerous scientific articles on the relationship between intrauterine growth retardation (IUGR) and two chronic diseases in particular: heart

disease and Type 2 DM (Barker and Osmond, 1986; Hales and Barker, 2001; Barker *et al.*, 2002). With regards to DM, Hales, Barker and colleagues (Hales *et al.*, 1991; Hales and Barker, 1992; Hales and Barker, 2001) have proposed the "thrifty phenotype hypothesis" to explain how insufficient fetal and infant nutrition plays an important role in developing illness in adulthood. The hypothesis acknowledges that the effect of early undernutrition might also be mediated by other factors (e.g., compensatory growth and later obesity among previously thin children). The biological mechanism is the following:

(i) poor fetal nutrition affects the development and function of pancreatic beta-cell mass and islets of Langerhans;

(ii) poor fetal nutrition predisposes to insulin resistance; i.e., organs and metabolism are accustomed to low energy intake, becoming adapted to survival under adverse conditions[2];

(iii) maternal hyperglycaemia can also induce fetal malnutrition;

(iv) poor fetal nutrition also leads to low birth weight or thinness at birth;

(v) low protein intake during infancy also hinders the development of adult beta-cells;

(vi) damaged beta-cells impair insulin secretion and increase insulin resistance;

(vii) insulin resistance and impaired insulin secretion occur earlier in the natural history of Type 2 DM (McKeigue, 1997; Hales and Barker, 2001).

Hales and Barker (2001) advance the possibility that early nutrition may have an effect on both insulin secretion and insulin resistance. Citing the results of studies conducted by Phillips *et al.* (1994) and Taylor *et al.* (1995), McKeigue (1997) argues that insulin resistance is the main

[2] Hales and Barker (2001) give the example of offspring of rats that were fed with low protein diets; they develop livers with larger but fewer lobules; this mechanism of "organ sparing" has been observed for other organs.

mechanism underlying/responsible for Type 2 DM. According to Phillips *et al*., the ponderal index at birth (a more valid measure of undernutrition than birth weight) was related to insulin resistance. Taylor *et al.* showed that women who were thin at birth exhibited limited capacity for anaerobic glycolisis, which is part of the process of transforming glucose into insulin. These are examples of phenotypic plasticity, which – as mentioned above – is a type of gene-environment interaction (Barker *et al*., 2002).

According to the thrifty phenotype hypothesis, Type 2 DM is more likely to occur in persons who were undernourished at birth, but overnourished in childhood and adulthood; this is in contrast to persons who remain thin throughout their lives. Researchers base this argument on the rationale that increasing food intake and decreasing energy expenditure might trigger glucose intolerance (Hales and Barker, 2001). In this framework, Type 2 DM might also arise from an imbalance produced by the abnormal growth of organs. Here, the rationale is that organs were initially small in size at birth due to fetal malnutrition and inadequate to manage the increased food intake in later years. This latter idea has been extrapolated from animal experiments showing that poor nutrition *in utero* reduces growth of other organs with the aim of protecting brain growth (Hales and Barker, 2001).

There are two main criticisms of the thrifty phenotype hypothesis. The first suggests that the relationship between phenotype and Type 2 DM might be confounded by the effect that adiposity has on DM (Paneth and Susser, 1995, mentioned by McKeigue, 1997). The second criticism suggests that the phenotype might be produced by a selection process through which persons that are genetically susceptible to DM are born with low birth weight or low ponderal index (Zimeth, 1995, cited by McKeigue, 1997). McKeigue refutes the second argument, concluding that the evidence provided by the critics does not clearly support their premise.

Concerning the first argument, McKeigue agrees that ponderal index is better than low birth weight as an indicator of early malnutrition; however, given the results reported by Phillips *et al.* (1994) and Taylor *et al.* (1994), he concludes that the mechanism through which early malnutrition affects Type 2 DM is most likely to be independent of the effect of adiposity.

Other lines of research linking early health experiences with adult disease emphasize the effect of early infections on late life chronic maladies (Martyn, 1991; Leon and Smith, 2000; Finch and Crimmins, 2004; Cimmins and Finch, 2006). In this literature, early infections are not studied in direct relation to a critical period in the development of organs. Rather, they are analyzed in terms of the latency period of a certain disease agent within an organism, or of the slow pace at which an agent progressively causes degeneration within an organ, as with chronic inflammation (Finch and Crimmins, 2004; Crimmins and Finch, 2006). An example of the former is the relationship between *human papilloma virus*, contracted at younger ages, and the development of cervical cancer after age 50 (Koutsky, Galloway and Holmes, 1988). An example of the latter is the effect of streptococcal infection, which produces rheumatic fever after a few days of the infection, but may also slowly cause rheumatic heart fever.

As mentioned above, Latin America as a region has a high prevalence of DM. Using stunted growth as an indicator, a large proportion of children born in the region during the first part of the 20th century were affected by chronic malnutrition (Bogin and Keep, 1999). Several researchers have already studied the relationship between stunted growth/chronic malnutrition in early life and later onset of DM in Latin American countries (Conlisk *et al.*, 2004; González-Barranco y Ríos-Torres, 2004; Martorell *et al.*, 1998, 2001; Palloni *et al.*, 2005, 2006; Ramakrishan *et al.*, 1999). Posing the thrifty phenotype hypothesis as a possible explanation for these data was straightforward.

B. Aging and disease in Latin America: Results of the demographic and epidemiological transition

1. The situation in Latin America

Before applying a "life course perspective" to the health of Latin American elderly, it is useful to understand the economic and demographic context that these cohorts experienced during their lifetime.

The demographic transition in Latin America has occurred at a faster pace than in industrialized countries. As Chackiel (2004) explains, during the second half of the 20th century, life expectancy at birth in the region rose from 52 years to 70 years; the Total Fertility Rate (TFR) dropped from 6 to 2.8 children per woman; and the mean annual growth rate dropped from 2.7% to 1.6%. There was, nonetheless, considerable heterogeneity across countries in the occurrence of this process. While Argentina, Uruguay, and Cuba already had low fertility and mortality during the 1950s, most Latin American countries were only beginning their demographic transition at that time. Furthermore, people in countries such as Honduras, Guatemala, Haiti, and Bolivia continue to experience low probabilities of survival and moderately high numbers of births per woman. In the traditional demographic literature, it is common to find several classifications of Latin American countries according to stage of demographic transition. From the 1970s to the 1990s, Costa Rica was classified within the group that had reached "Full demographic transition"; however, in light of the most recent data, Chackiel (2004) classified Costa Rica within the "Advanced demographic transition" group (along with Chile, Brazil, and Argentina[3]).

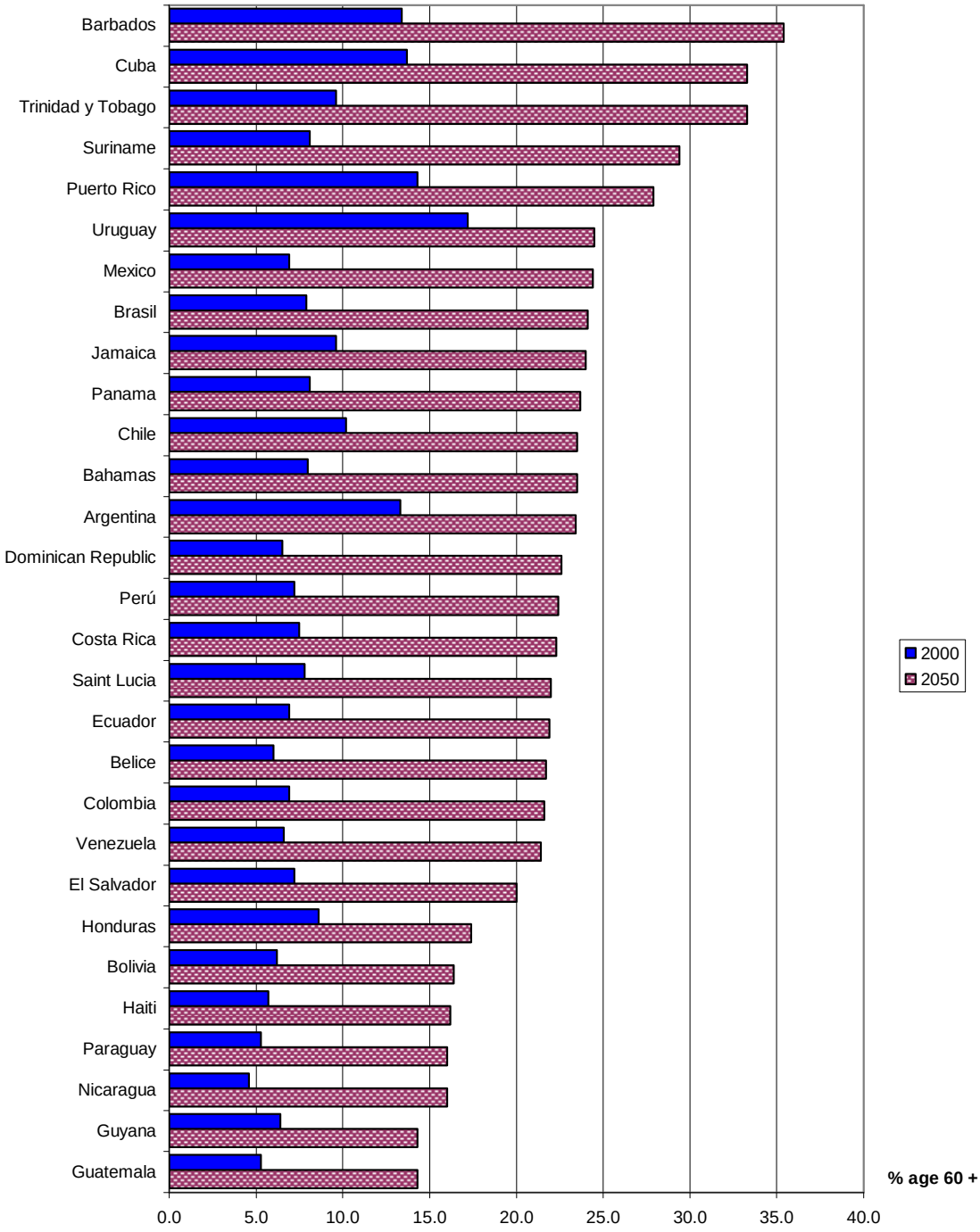
[3] Uruguay and Cuba are classified under the label "Very advanced demographic transition".

Fertility and mortality declines are the main factors that underlie the aging of a population. Having been at some point the leaders in low fertility and mortality, Argentina, Cuba, and Uruguay have already age structures that are very similar in shape to some Western European countries (Guzman, 2002). In the two latter nations, strong emigration flows have accentuated this process, because the majority of emigrants have been young adults. This pattern is also observed in Anglo-Caribbean countries such as Barbados and Trinidad and Tobago. In 2000, the proportion of the population that was 60 years old or older in these countries was higher than 10%. In Uruguay, this proportion has risen to 17% and, although the process is more recent than in the other countries in this group, in Chile this percentage is 10%. For countries in this first group, the proportion of the population 60 years old or above is estimated to increase to 22% or higher by 2050. In Barbados and Trinidad and Tobago, it might exceed 30%.

For a second group of countries this proportion was observed to be relatively low in 2000. However, because the pace of fertility decline in these countries has been quite rapid, the proportions of the population 60 years old or above will be similar to that of countries in the South Cone. Among the countries in this second group, there are Brazil, Mexico, Colombia, Costa Rica, Panama, and the Dominican Republic. The population aging process in Costa Rica has been slowed by the effect of immigration, given that most immigrants are young adults.

There are a number of countries – most of the Central American nations, as well as Bolivia, Paraguay, and Haiti -- that are still going through earlier stages of the demographic transition. They continue to experience relatively high fertility and mortality rates. Since their aging process has been slower than that in countries within the previously mentioned groups, it is projected that even in 2050 the elderly will constitute less than 20% of their populations (Guzman, 2002) (See Figure II.1).

Figure II. 1. Latin American and the Caribbean: Percentage of population that is 60 years old or above, 2000 and 2050.



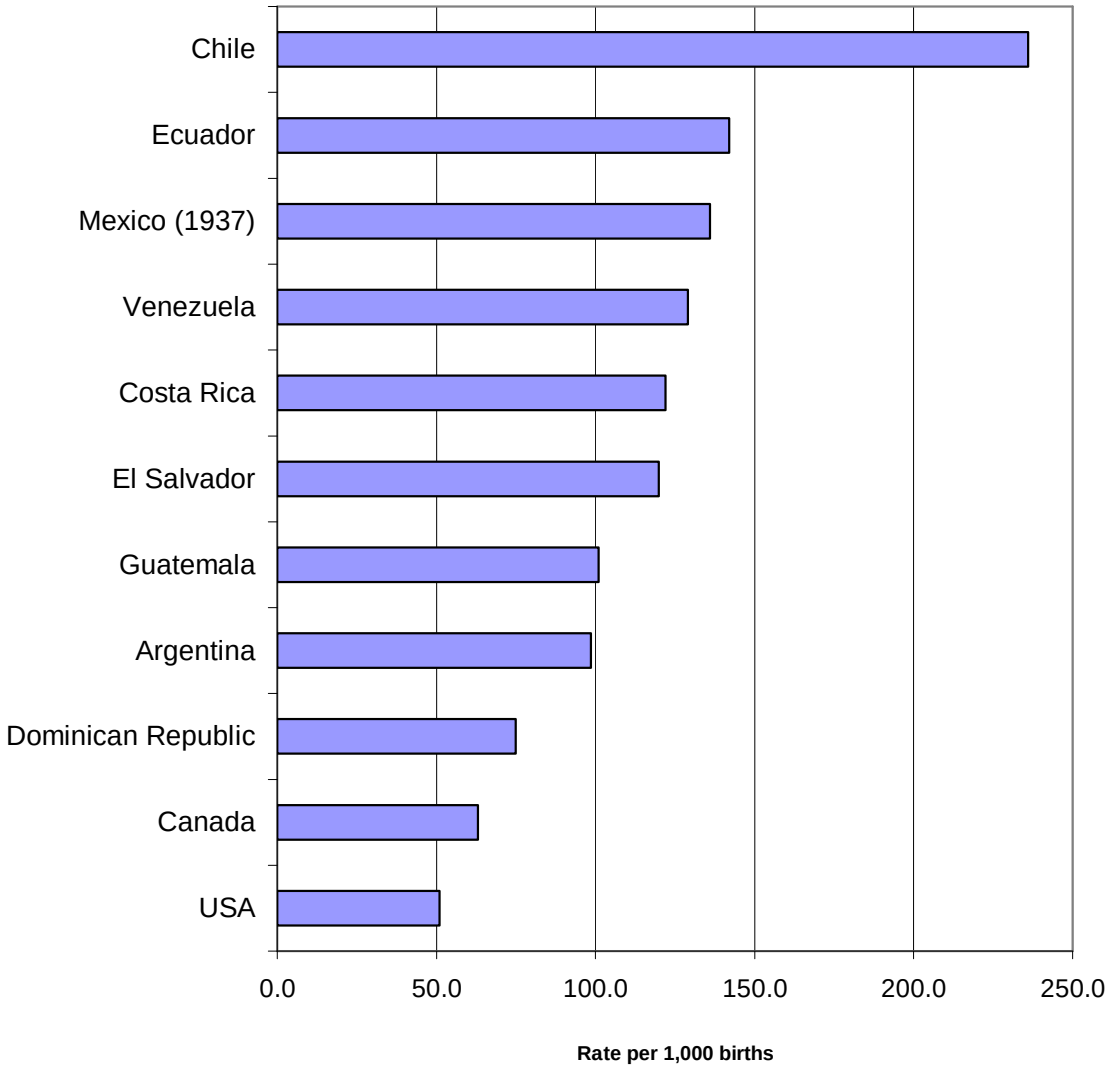
Source: Guzman (2002)

These changes in age structure are closely related to the epidemiological profile of the region, where considerable heterogeneity remains across countries. Those that are more advanced in their demographic transition have achieved lower infant mortality by controlling transmissible diseases. Given that these same countries are the ones with a more advanced aging process, most deaths are caused by chronic and degenerative diseases typical of adulthood or old age. While Uruguay is the nation with the highest proportion of mortality due to cardiovascular diseases, cancer, and external causes, Chackiel (2004) notes that Mexico, Costa Rica, Chile, and Venezuela are the countries with the most rapid change in the composition of their mortality pattern. During the sixties and seventies, the proportion of deaths in these countries due to transmissible diseases was higher than 30%. Around 1995, this percentage dropped to less than 10% (Chackiel, 2004). In all of these countries, the proportion of deaths that occurred to the elderly (65 years or above) is higher than 40%. In contrast, countries such as Guatemala still have a high burden due to infectious diseases, and the proportion of deaths occurring among the elderly (around 30%) is similar to the proportion occurring among children (0 to 14 years old) (Chackiel, 2004).

These epidemiological profiles were not as heterogeneous when cohorts of the current elderly population were young. During the first half of the 20th century, communicable diseases and malnutrition accounted for the majority of deaths in the region (Palloni, 1981; Palloni and Wryck, 1981; McQuestion, 2000). Infant mortality rates were very high (as seen in Figure II.2, Costa Rica was located in the middle of the range). Also during this time, the health care systems were not well prepared to control the high prevalence of infectious agents. The Latin American

elderly population at the beginning of the 21st century is constituted by the survivors of that disease environment.

Figure II. 2. Selected countries of the Americas: Infant Mortality Rate (per 1,000 births), circa 1938



Source: Luros (1942).

As with the demographic transition, there has also been heterogeneity in the pace of economic development across countries. The link between both social processes (economic development and demographic transition) has been debated throughout the 20th century. Some authors have stressed the interdependence between the two (Coale and Hoover, 1958; Coale, 1973; Notestein, 1945; Omran, 1971), while others have contested the premise especially by noting that, in Latin America, industrialization occurred in countries that kept high fertility rates for a considerable period of time (Alba and Potter, 1986; Benitez-Zenteno, 2000; Caldwell, 1976; Palloni, Hill and Pinto-Aguirre, 1996; Zavala de Cosio, 1992). Regardless of whether or not a link exists, history shows that Latin American elderly at the beginning of the 21st century experienced a very particular socio-economic context during their life course. Between 1930 and 1970, the Governments of most of the countries in the region implemented a set of economic policies known as import-substituting industrialization (ISI). This development strategy was promoted by the Economic Commission for Latin America and the Caribbean (ECLAC, or in Spanish, CEPAL, name that generated the adjective "cepalino" in reference to this strategy) (Haggard, 1992; Kaufman, 1990; Villarreal, 1990). It was a reaction to the problems generated by the 1930s economic crisis. Its main logic was to diminish dependence on: (1) the prices of primary products in which these countries were specialized and (2) imported goods from developed countries. Governments promoted national manufacturing companies, the establishment of subsidiaries of transnational companies in their territories, or a wider public involvement in the production of State-owned companies, such as the oil companies in Mexico (PEMEX) and Brazil (PETROBRAS) (Kaufman, 1990). Latin American countries enjoyed steady economic growth during the post-war decades. On average, aggregate output grew 5.5%

a year and GDP per capita, around 2.5% (Grilli, 2005). Without trying to imply any causal relationship, it is a fact that the population forming households and entering the job market during those years enjoyed an economic boom and the advantages of industrial diversification. These developments resulted in an influx of employment opportunities for an increasing number of young adults. They also ushered in the so-called "Western diet," linked to a greater variety of consumer products, and lifestyles characterized by less physical activity (Popkin and Gordon-Larsen, 2004).

These years constituted an epoch, during which important public health measures were introduced (e.g., vaccination campaigns, the use of DDT for malaria eradication, oral rehydration therapy to treat childhood diarrheas, improvement of water potability and sewage systems, and the development of primary health posts). Half of the mortality decline occurring in Latin America during the period 1950-1970 can be attributed to changing medical technology (Palloni, 1981; Palloni and Wryck, 1981). In addition, by the seventies, the region had the most extensive health care infrastructure of the developing world (McQuestion, 2000), with an active participation of the State, either through the Ministries of Health or Social Insurance programs (Mesa-Lago, 1992).

It is necessary to outline this historical context in order to highlight the fact that, as young adults, the elderly in Latin America experienced the postwar economic boom and improvement of health standards. These circumstances provided them with more employment options, better health care for their families, and increased access to consumer goods (Kaufman, 1990). However, as children, they faced a disease environment, characterized by a high prevalence of infectious agents and a health care system that was not well prepared to control them. In other

words, Latin American elderly population is constituted by the survivors of that disease environment.

During their middle–age adult years, Latin American elders saw the breakdown of the ISI program. Most of the countries in the region underwent a severe economic crisis, the so-called "lost decade", starting in the seventies with the increase in oil prices and reaching its peak during the 1980s. During this period, the region's GDP per capita fell by an average of 8%. In some countries, the annual inflation rate was higher than 100%. In others, like Bolivia, it was higher than 1000%. There was a steep increase in poverty and unemployment rates (Cardoso and Fishlow, 1992; Grilli, 2005; Remmer, 1991). Due to the crisis, Latin American countries were unable to pay interest on foreing debts. Governments implemented structural adjustment programs in order to diminish their macro-economic deficits. Social investment decreased, affecting public services, such as health care. While public health services in some countries "survived" the crisis (the Costa Rican public system is one of these cases), in others they clearly deteriorated (Mesa-Lago, 1992). During the nineties, income inequality increased in almost all the countries of the region. This scenario has important implications for the well-being of the elderly in Latin American. Due to the social and economic environment of the region, they are likely to have faced unemployment and poverty in the years approaching retirement, poorly funded pension programs in retirement, and unequal access to health care services in the years when chronic conditions start to take their toll.

Throughout the last decade, Palloni and colleagues (Pelaez *et al.*, 2000; Palloni *et al.*, 2002; Palloni *et al.*, 2005) have been discussing what makes the process of aging in Latin America so distinct from aging in other parts of the world. Summarizing the social,

demographicl and historical outline posed above. They highlight –with some variants over time– the following main peculiarities:

(i) Rapid rate of population aging: The aging of populations in Latin American countries is occurring in less than half the time that it took in industrialized countries (Pelaez, Palloni and Ferrer, 2000; Palloni, Soldo and Wong, 2002; Palloni *et al.*, 2005, 2006);

(ii) The effect of exogenous variables –particularly medical interventions– on mortality decline (Pelaez, Palloni and Ferrer, 2000; Palloni *et al.*, 2006);

(iii) Poverty and social inequality as the context of aging (Pelaez, Palloni and Ferrer, 2002; Palloni, Soldo and Wong, 2002): Whereas aging occurred in the industrialized world during a period of large improvements in living standards, in Latin America it is occurring after periods of severe economic crises, which hindered the achievements of the post-World War II era, in a continent where income distribution is the most unequal in the world (as compared to other regions as a whole);

(iv) Weak institutional and informal safety nets (Pelaez, Palloni and Ferrer, 2000; Palloni, Soldo and Wong, 2002): Societies are not prepared to address the changing needs of an aging population; i.e., a significant proportion of the population no longer contributing to the economy, due to retirement or inable to work, and experiencing chronic diseases and disability. Access to health care services, social insurance, social provision, and welfare services is unequally distributed. In addition, changes in fertility and living arrangements might increase the number of elderly living by themselves, making them more vulnerable to hazardous events such as heart attacks, geriatric falls, and crime.

(v) The "stickiness" of early health status (Palloni, Soldo and Wong, 2002; Palloni *et al.*, 2005, 2006): In early life, the current elderly population in Latin America was exposed

to under-nutrition and an infectious disease environment and is, therefore, at increased risk for developing chronic diseases (e.g., DM, heart disease, and stomach cancer) in later life.

2. Diabetes Mellitus in Costa Rica

As explained above, most Latin American countries are now in advanced stages of the epidemiological transition. Non-communicable diseases are the main causes of death and disability in the region, and DM is one of these highly common maladies. According to official figures published by the World Health Organization (WHO) (Mackay and Mensah, 2004), in 2000, the prevalence of DM in Costa Rica among people aged 20 years and above was 3.3%. However, survey results estimate higher prevalence rates. In the capital, San Jose, a recent study indicated that 26% of the population 20 years old or older is diabetic or borderline diabetic (Ministerio de Salud-CR, 2005). In 2002, death rates due to DM for the entire population of Costa Rica were around 15 per 100,000 (Mackay and Mensah, 2004).

Another concern associated with DM is the economic burden it imposes on those who suffer from it. In a study using information from several published articles, Barcelo and colleagues (2003) estimated that the direct per capita cost of DM was US$ 624 in Costa Rica and as high as US$ 703 in Latin America. These estimates included costs for medication, hospitalization, inpatient and outpatient consultation, and complications. Due to the underlying assumptions, Barcelo's computations might be questioned. In particular, they disregard the differences between countries with national health care systems, which are favored by economies of scale (e.g., Costa Rica), and countries with mixed or private systems. Morice and Achio (2003) estimated that, for the public hospital and clinics system, DM was the illness with the

highest total cost for hospitalization, and with the second-highest cost (after hypertension) for outpatient services. Given the role of the Government in delivering health services, these articles highlight the magnitude of the economic burden imposed by this chronic disease on public investment.

Research focusing on DM risk factors in Costa Rica is not very sizable, although it has been increasing during the last decades. Rosello-Araya (2003a, 2003b) collected data from patients between the ages of 15 and 75, who visited the community health center in a semi-urban county East of the capital. Findings indicated that age, obesity (as measured by BMI), hypertension, and family history of DM were significant risk factors of elevated serum glucose (SG) levels. In a study conducted in Costa Rica in 2000, Lacle and Peralta (2006) found that implementing routine DM screening in primary health centers reduced the prevalence of undiagnosed DM, without affecting the daily work of primary health centers, or increasing its operations costs. DM screening has become one of the regular examinations conducted routinely in primary health centers in Costa Rica, and its coverage is one of the measures used by the public health institute (CCSS) to evaluate the quality of care provided by primary health centers. In a community in the Western part of the Central Valley, Goldhaber-Fiebert *et al.* (2003) implemented an intervention consisting of a 90-minute weekly nutrition class (11-week duration), plus occasional walking sessions (12-week duration). They reported that the intervention achieved a significant reduction in weight, fasting plasma glucose levels, and glycosylated hemoglobin for the experimental group, when compared to the control group. Another similar intervention in a community in the Eastern part of the Central Valley accomplished comparable results (Arauz *et al.*, 2001). With a non-representative sample of diabetic patients in San Jose, Firestone *et al.* (2004) found that the levels of DM-specific

knowledge were greater than in a sample of Spanish-speaking U.S. Latinos in Starr County, Texas. Greater levels of DM knowledge were inversely related to age, and directly associated with schooling, years since diagnosis, and the practice of measuring blood glucose with a glucometer at home. None of these six studies conducted in Costa Rica focused specifically on the elderly population.

DM is not only a severe disease by itself, but it is also a risk factor for other illnesses, particularly cardiovascular and cerebro-vascular diseases. Heart attacks and strokes are common complications of Type 2 DM. At the beginning of the 21st century in several Latin American countries, these diseases were the primary cause of death, especially among adults. Cubillos-Garzon *et al.* (2004) even argue that coronary artery disease could be catalogued as an epidemic in Latin America. In Costa Rica, the non-standardized mortality rate due to heart disease is 72 per 100,000 and due to stroke, 40 per 100,00 (Mackay and Mensah, 2004). Although cardiovascular diseases as a group constituted the major cause of death during most of the second part of the 20th century, mortality rates for these diseases among older subjects (50-79 in Costa Rica) decreased steeply during the period from 1970-1990 (Rosero-Bixby, 1996) Paradoxically, this period coincides with that of the severe economic crisis that occurred during the 1980s. During more recent decades, mortality rates for cardiovascular diseases have shown slight increases, especially among people 65 and over. From 1995 to 2002, in Costa Rica, the cause-specific death rate for this age group changed from 863 per 100,000 to 1000 per 100,000 (CCP, 2005).

The impact of other well known risk factors for cardiovascular diseases (aside from DM) has also been studied by several researchers. Most of the research in this area that has been conducted in Costa Rica has focused on how deleterious dietetic changes, such as obesity

(Nuñez-Rivas *et al.*, 2003) and hyperhomocysteinemia (Holst-Schumacher *et al.*, 2005), augment the prevalence of cardiovascular risk factors, or increase the risk of myocardial infarction (Baylin *et al.*, 2003; Baylin and Campos, 2004; Kabagambe *et al.*, 2003; Kabagambe *et al.*, 2005a; Kabagambe *et al.*, 2005b; Martínez-Ortiz *et al.*, 2006).

Chapter III: Study Objectives

There are two main objectives for this dissertation:

A) Estimate the effect of early childhood conditions on the burden of DM among the population 60 years old and older in Costa Rica, particularly by:

   A.1) Comparing the incidence rates of: a) been diagnosed with Type 2 DM and b) dying given a previous diagnosis of Type 2 DM, between persons with short knee height (KH), and persons with longer knee height (KH), controlling for other risk factors; using knee height (KH) as a marker of adverse early nutritional status.

   A.2) Comparing the incidence rates of: a) been diagnosed with Type 2 DM and b) dying given a previous diagnosis of Type 2 DM, between persons born in cantones with high child mortality levels (CMI), and persons born in cantones with lower high child mortality levels (CMI), controlling for other risk factors; using child mortality rates (CMI) in the canton of birth as a marker for adverse early nutritional status.

B) Project minimum plausible prevalence of DM in relative and absolute figures among the population 60 years old and older in Costa Rica for the period 2005-2030, given the prevalence of adverse early life conditions (as represented by short KH or high CMI) among the cohorts born between 1945 and 1970.

Among other risk factors that will be taken into account in the analyses are: sex, age, socio-economic status (SES; i.e., schooling, occupation), region of residence, family history of DM, smoking, alcohol, physical activity, Body Mass Index BMI, self-reported weight and scale at age 25, self-reported maximum weight, self-reported weight change, and self-

reported caloric intake (Ballesteros *et al.*, 2005; Bonora *et al.*, 2004; Duran-Varela, *et al.*, 2001; Mackay and Mensah, 2004; Pagan and Puig, 2005; Rivera *et al.*, 2004; Rosas-Peralta *et al.*, 2004; Rosello-Araya, 2003a, 2003b; Velazquez-Monroy *et al.*, 2003).

Chapter IV: Data and Methods

A. Data

For the purpose of constructing the models for estimating the effects of early life conditions on adult disease, my primary source of information was the dataset from the Costa Rica Estudio Longitudinal de Envejecimiento Saludable (CRELES). CRELES is a research project, which is focused on the health of the elderly in Costa Rica. For the estimates and projections of disease prevalence, I used existing life table series. For the projections of disease prevalence, I also use existing life table series. I use historical information from statistical reports published during the first part of the 20th century to collect information on child mortality by canton.

1. CRELES: Costa Rica Estudio de Longevidad y Envejecimiento Saludable

a) Study description

The CRELES sample consists of individuals, who were born in Costa Rica in 1945 or earlier and who were alive during the period 2004-2006. It is derived from a random sample of 8,000 individuals ages 55 and over, who were interviewed in the 2000 Census. The sample was stratified by 5-year age groups, and within each stratum, individuals were selected randomly using a systematic procedure. Census information for individuals in the sample was linked to the Vital Registration System database in order to study their mortality patterns. With the

expectation of producing a subsample of 3,000 survivors, a sample of 4,000 individuals was selected using a two-stage cluster sampling design. Clusters represented the official Health Areas established by the Ministry of Health and the Social Security Institution (CCSS, for Caja Costarricense del Seguro Social)[4].

CRELES interviews the selected person only, and not the spouse. The fieldwork schedule is very particular. There are only two fieldwork teams, each composed of a driver, a laboratory technician, and two interviewers. This means that, on average, project members can expect to complete only 32 interviews per week, and that each wave will be concluded two years after its initiation. The first wave started in November 2004 and completed in September 2006. The second wave was launched in November 2006 (and is scheduled for completion in during mid-2008). Respondent's vital status is followed up by linking the dataset with the National Vital Registration System (the Death Index); therefore, only their mortality will be traced before the second wave starts. This means that mortality is the only characteristic that can be studied in this dissertation using a cohort study design.

The non-response rate for most items was very small: 0.5% did not respond to the question concerning a previous diagnosis of DM, and 0.18% did not respond to the item regarding their place of birth. However, because 4.7% of respondents are foreign-born, the proportion of missing values in the variable CMI is higher. Also, some respondents were constrained to bed due to health problems, making measurements impracticable; therefore, 4.6% are missing data for one or more of the following: KH, height, weight, and body mass index (BMI).

[4] Social Security is a concept that is understood differently in the U.S.A. and in Costa Rica. While in the former, Social Security refers to the pension system that transfers money to retirees, in Costa Rica it also includes the public health care system; it comprises primary health care clinics and public hospitals.

By design, the questionnaire used in CRELES utilizes items similar to those found on questionnaires from other projects on aging in Latin American; e.g., the Mexican Health and Aging Study (MHAS) and the "Salud, Bienestar y Envejecimiento en América Latina" (SABE) project. CRELES introduced a noteworthy innovation, in that the data is gathered by Computer Assisted Interviews (CAIs), using PDAs or palms. This technology allows controlling for inconsistencies in the data produced during the fieldwork and for continuous generation of information (Rosero-Bixby, 2005).

b) Main outcome

The main outcome variable is self-report of previous DM diagnosis by medical personnel. The question in CRELES is:

Has a doctor or medical personnel ever told you that you have diabetes or high blood sugar levels?

CRELES gathers clinical information –biomarkers– derived from blood and urine tests, and assessment of blood pressure. There are two biomarkers that CRELES can use to determine DM: glycosylated hemoglobin levels ($HbA_{1C} \geq 7\%$) and fasting serum glucose levels ($SG \geq 200$ mg/dL). Both biomarkers are used to assess both "total" DM prevalence and what is called "DM unawareness" or "undiagnosed DM" in this document. High cutoff points are selected to diminish the chances of false positives.

Serum glucose (SG) is the criterion recommended by a World Health Organization (WHO) Consultation Group (WHO, 1999), while HbA_{1C} has been used to control DM treatment. Glycosylated hemoglobin (HbA_{1C}) has been proposed as an alternative to the Oral Glucose

Tolerance Test (OGTT, the so-called "gold standard" for DM diagnosis which is also recommended by the WHO) because patients do not need to fast, to drink the glucose solution, or to wait for 2 hours before blood samples are drawn, and because it is considered to be a better biomarker for daylong blood glucose concentrations (Peters *et al.*, 1996).

The drawbacks of using SG are:

- Interviewers can not verify that respondents were really fasting when the blood sample was drawn;
- The test is less accurate for elderly populations;

Concerning HbA_{1C}, the main drawback is that laboratory standards for using it as a diagnosis tool are not completely uniform. In 2003, an Expert Committee on the Diagnosis and Classification of Diabetes Mellitus suggested that this biomarker should not be used for diagnosis, because of the lack of unified standards across laboratories (ECDCDM, 2004). However, in recent years, unified standards have been established in the United States, as well as in Costa Rica.

Biomarker levels were determined by laboratories in the University of Costa Rica (UCR) and in Caja Costarricense del Seguro Social (CCSS, in Hospital San Juan de Dios). The first laboratory used the glucose oxidase method, while the other used glucose oxidase and oxygen consumption methods. The survey was applied and blood samples were drawn after an informed consent form was read and signed by the interviewees. Statistical adjustments were conducted to ensure comparability across laboratories (see Mendez Chacon *et al.*, 2007 for a description of laboratory analyses). The informed consent text was approved by the University of Costa Rica's Institutional Review Board (IRB).

c) Main explanatory variable

The main risk factor is knee height (KH), measured as the distance in centimeters from the knee to the ankle. KH is a marker for early growth retardation among the survivors within the cohorts under study. The rationale for using KH –or related anthropometric measures like height or leg length– derives from the literature that finds an association of stature with DM or with metabolic disorders. Researchers in different countries have found that women with gestational DM (GDM) are on average shorter than women who do not have the disease: Meza *et al.*, 1995 in Mexico; Anastasiou *et al.*, 1998 in Greece; and Jang *et al.*, 1998 in South Korea. Only the Korean study controls for BMI, but fails to control for any measure of socio-economic status (SES). In Brazil, after controlling for a relatively long list of confounders, Branchtien *et al.* (2000) report the association, but only among women with high skinfold thickness. In the United Kingdom, Kousta *et al.* (2000) found the association among women of European and South Asian origin, but not among Afro-Caribbean women. They did not take BMI and social class into account. In Hungary, Tabak *et al.* (2002) found an association between height and GDM, but only when the two subsamples under study were combined. In addition, when year of birth (a sign of a birth cohort effect) was included in the model, the association was no longer significant.

An association between height and diagnosed DM, plasma glucose concentration or a positive result in an OGTT has also been found by researchers in the United Kingdom (UK; Brown *et al.*, 1991; Riste, Khan and Cruickshank, 2001), Nigeria (Olatunbosum and Bella, 2000), Norway (Njolstad, Arnesen and Lund-Larsen, 1998), Spain (Guerrero-Igea *et al.*, 2001), Taiwan (Pan *et al.*, 2001), Mexico (Lara-Esqueda *et al.*, 2004; Sanchez-Castillo *et al.*, 2005), Guatemala (Conlisk *et al.*, 2004), Russia (Stanner *et al.*, 1998), and the U.S.A. (Schmidt *et al.*, 2005). Most of these studies controlled for age, sex, obesity (BMI or Waist-to-Hip Ratio), and

social class, although this last confounder was included least often in multivariate analyses. The association was weak or non-existent in certain subgroups: Njolstad, Arnesen and Lund-Larsen (1998) reported a significant relationship among women, but not among men. Conlisk *et al.* (2004) found it only among men who were born small, but not among those who were of normal size at birth. Stanner *et al.* (1998) observed that, when adult social class and education were added to the multivariate analysis, the degree of association was reduced. Olatunbosum and Bella (2000) commented that stature is associated with a dichotomous variable that differentiates people with normal glucose levels from subjects with abnormal levels, but it is not associated with the plasma glucose levels as a continuous variable.

Given these results, additional research has been done to analyze what component of human stature is more closely related to Type 2 DM. In the UK, Smith *et al.* (2001) found that leg length, but not trunk length, is related to insulin resistance, operationalized according to the concept of Homeostasis model assessment (HOMA). In their analyses of gestational DM in Australia, Moses and Mackay (2004) reported that leg length and the leg-to-height percentage were correlated with 2-hour glucose concentration from an OGTT, but not with fasting glucose levels. In Puerto Rico and 5 Latin American cities, Palloni *et al.* (2005, 2006) found that an indicator variable that refers to KH under the first quartile (in Puerto Rico) or under the first quintile (in the SABE project cities) has a small but significant coefficient in a logistic regression of prevalence of self-reported diabetes. However, Gunnell *et al.* (2004) did not find a significant association between leg length and non-fasting plasma glucose in the UK.

The rationale behind the use of leg length or KH is that either of them is a good marker of limited nutrient intake during gestation, infancy and early childhood, and therefore is used to study the hypothesis that early life events are risk factors for chronic conditions (Hales and

Barker, 2002; Gunnell, 2002). The idea that limited energy intake during early life is associated with anthropometric measures during childhood was originally proposed by Leitch (1951), along with the concept of "growth potential" and the observation that stunted mammals have disproportionately short legs relative to the their trunk size. Mitchell (1962) noted that, in Japan, the mean stature increase from 1949 to 1959 was mostly due to an augment in leg length and suggested that this might be related to an increase in animal protein intake. More recently, Wadsworth *et al.* (2002) reported that leg length at age 43 was positively associated with breastfeeding in infancy and energy intake at age 4. They controlled for parental height, subject birth weight, and weight at age 4, but they controlled for neither education, nor ponderal index at birth (possible confounders). In response to Wadsworth *et al.*, Gunnell (2002) argued that there was not enough evidence to determine whether leg length or bodily disproportion (leg-to-trunk ratio) was a better marker for health disadvantage during childhood. Palloni *et al.* (2005) favor KH because it appears to be a good predictor of current height among the elderly (given that it is not affected by bone disease as much as other measures) and because they consider it a marker of early malnutrition. Expanding on the possible explanations for the relationship, Gunnell *et al.* (2004) analyzed Insulin-like Growth Factors (IGF). They found that leg/trunk ratio was associated with the molar ratio IGF-I/IGFBP-3, while both leg length and trunk length are related to IGFBP-2.

Despite calls for more corroborating research on the topic, Smith *et al.* (2001) suggested that "leg length seems to serve as an indicator of nutritional status in childhood. It is the component of overall height that grows proportionately more in the years up to puberty" (p.871). Gunnell (2002) argues that the study by Wadsworth *et al.* (2002) "provides further evidence of the potential use of leg length as a measure of pre-pubertal exposures" (p.393).

There is no clear cutoff point for KH as an indicator of poor versus adequate nutrition in early life. Since it was necessary to group respondents for subsequent projections, I used quartiles as cutoff points in order to transform KH into four categories rather than mean values with their corresponding standard deviations. This operationalization has been utilized by Palloni *et al.* (2005) for Puerto Ricans and other Latin Americans, as well as by Lara-Esqueda *et al.* (2004) in the study of height among diabetics in Mexico. Nonetheless, some of the statistical models presented below were estimated using KH as a continuous variable.

It is important to acknowledge that the ideal setting would be to have direct measurements of poor nutritional status during childhood, such as birth weight, ponderal index, or weight-for-age percentiles during infancy (Barker, 1998). However, this information is very rare in Latin American sources. Fortunately KH and leg length are already available in MHAS and CRELES.

2. Historical official statistical publications in Costa Rica.

The other main explanatory variable is the level of child mortality in the canton of birth of CRELES respondents at the time of their birth. Since 1907, the Government institution in charge of official statistics has been publishing statistical reports. These publications were scanned and posted on-line by the Central American Center for Population (Centro Centroamericano de Poblacion) in its virtual library: http://ccp.ucr.ac.cr/bvp/pdf/anuariocr/index.htm. The publications present data on number of births by place of residence of the mother, and number of deaths under age 5 by place of residence. This information was used to compute the variable named Child Mortality Index, or CMI, for every canton and year in which the information was available. Then, this set of figures

was merged with the CRELES dataset, based on the answers to the question: "Where were you born?" CMI is equivalent to:

$$CMI = \frac{\text{Number of deaths of children age <5 years old}}{\text{Number of births}} * 100 \qquad (1)$$

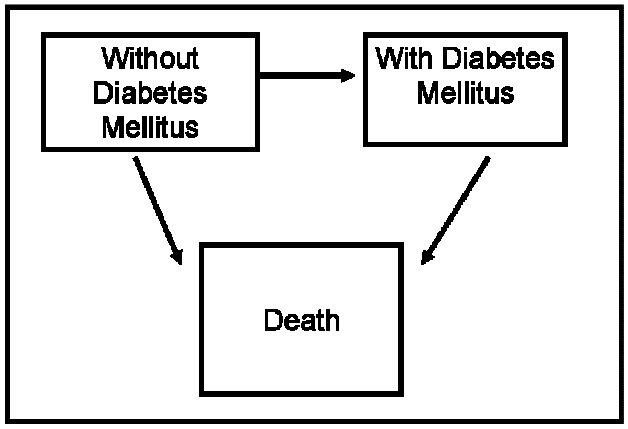
The variable was included in the models as a continuous variable, and as a categorical variable operationalized as dummy variables, where the cutoff points were defined based on quintiles. It was not possible to compute the infant mortality rate (IMR) because some of the publications do not disaggregate child deaths by single year of age.

B. Methods

1. The ideal setting: A multi-state model.

The topic studied in this research can be understood as a multi-state system. There are three possible states of interest: i) not having DM, ii) having DM, and iii) death. The system can be represented by the following diagram:

Figure IV. 1.Multistate system of disease and death.



As the diagram shows, this study assumes that once an individual has DM, the person remains in the state of "With Diabetes Mellitus". Acknowledging this assumption is important because it restricts the magnitude of the "denominator" for the hazard rates. A multi-state hazard model (also known as multi-state survival analysis) is the ideal statistical technique to analyze the system because it allows studying a process with transitions to two or more states, which are not all absorbing (e.g., as death) and not all proper[5] (Palloni, 2002). Multi-state survival models are better suited than common "two-state" hazard models for taking into account the correlation between mortality and disease incidence. However, the total number of deaths is not sufficient to provide adequate power for an analysis of differences between deaths among respondents with DM versus deaths among those without DM. Therefore, this dissertation uses separate "two-state" event history models for the transition to disease and to death.

2. Statistical methods for estimating incidence of DM and death rates and for determining inter-group differences

Before any figure is projected, it is necessary to determine whether DM incidence and mortality are statistically different between the groups defined by early life conditions: exposed to early life insults to health (short KH, or high CMI) vs not exposed to early life insults to health (longer KH, or lower CMI). In order to perform this task, several statistical methods are used.

a) Logistic regressions

Logistic regressions are used for two purposes. The first purpose is to analyze DM prevalence. The analysis of prevalence is important to compare the estimated relationships with

[5] Proper means that everyone in the system will experience each and every one of the states.

the articles written by Palloni *et al.* (2005, 2006), which have similar datasets, goals, and analytical techniques. The outcome variable DM is operationalized in the following way:

Diabetes Mellitus =	1 if:	-Self-report of previous DM diagnosis -Serum Glucose (SG) ≥ 200 mg/dl -Glycosylated Hemoglobin $HbA_{1C} \geq 7\%$
	0 if:	Otherwise

This logistic regression infers to the total population of Costa Rican elderly.

Another logistic regression model is used to analyze "DM unawareness", also called "undiagnosed DM". These concepts refer to the population that does not have a previous DM diagnosis, but their biomarker levels are above certain threshold levels. Therefore, the model infers to the population ages 60 and above without a DM diagnosis. The variable is operationalized in the following way:

Diabetes Mellitus Unawareness =	1 if:	-Serum Glucose (SG) ≥ 200 mg/dl -Glycosylated Hemoglobin $HbA_{1C} \geq 7\%$
(Conditional on not having a previous DM diagnosis)	0 if:	Otherwise

Incidence probabilities are estimated assuming that persons had their last glycemia within the last year. This decision helps to approach the concept of 1-year incidence, since I assume that persons develop high levels of biomarkers during a 1-year period. This decision is not perfect, but it has been the only alternative for computing "incidence estimates" with cross-sectional information.

Different transformations of the variable "age" are tried in order to inspect non-linearities in the age-schedule of DM prevalence and DM unawareness.

b) Kaplan-Meier Life tables for risk of DM and risk of death

Kaplan-Meier estimates are used to inspect the bivariate relationships between early life conditions and the risk of DM or of death. The Kaplan-Meier methodology is a set of techniques for calculating estimates based on traditional demographic and actuarial life tables by opening up the possibility of computing standard errors for the survival and hazard functions. According to the Kaplan-Meier methodology, the non-parametric maximum-likelihood estimate of the survivor function is:

$$\hat{S}(t) = \prod_{j|t_j \le t} \left(\frac{n_j - d_j}{n_j} \right) \qquad (2)$$

where t_j is the time at failure for event j, n_j is the number of units at risk (which is equivalent to the l(x) function in standard demographic terminology), and d_j is the number of failures at time t_j. The Kaplan-Meier is a non-parametric technique because it does not impose any parametric distribution to the parameters to be estimated. It has the limitation that it does not allow controlling for a relatively large set of covariates, as parametric or semi-parametric (Cox) survival regressions do.

c) Parametric survival regressions for risk of DM and risk of death

The multivariate analysis of the hazards of DM or death is performed with parametric survival models that control for the confounding effect of other covariates. Parametric models comprise one type of event history techniques. Event history models take into account the number of events or failures (numerator of a hazard rate) and the time of exposure for the number of individuals at risk (denominator of a hazard rate). The risk is more accurately estimated than if using a logistic regression because it directly assumes an ordering in the times

at risk for the individuals under analysis and it allows different parametric distributions for the hazards over time.

A parametric proportional hazard model can be represented by the following formula:[6]

$$\mu_{i.j}(t_k) = \exp[\alpha X_{i.h} + \beta Z_{i.h} + f(t_k) + \varepsilon_{i.j}] \qquad (3)$$

where:

i :	State of origin
j:	State of destination
t_k:	Observed time to the occurrence of the event
$X_{i.h}$	Represents the vector of "early life" independent variables for individual h in state i
$Z_{i.h}$	Represents the vector of control variables for individual h in state i
α	Vector of coefficients for "early life" covariates
β	Vector of coefficients for control variables
$f(t_k)$	Parametric function that describes the relationship between time and the hazard
$\varepsilon_{i.j}$	Error term

A hazard is "the rate at which spells (being in one state for a certain time) are completed after duration *t*, given that they last until *t*" (Greene, 2003: p.792). A hazard function, thus, gives a continuous representation of a curve that describes incidence rates through time (age, duration since diagnosis, etc.). The time under analysis will be assessed as the time between an initial age

[6] This is a variation from the formulation provided by Spittel (1999).

(25 years old) and time of diagnosis; the other time spans in the models are times to death, starting either at the initial age or at age at diagnosis.

Parametric survival models are preferred over semi-parametric regressions (e.g., Cox regressions) for two reasons:

- the estimated hazard curve is smoother than the one produced by a semi-parametric model, which is a useful characteristic for a projection input;
- Parametric distributions (e.g., the exponential, Gompertz, Weibull, and log-normal) are common in demographic theory and analysis; therefore parameters can be interpreted or used for other purposes.

The parametric model for analyzing mortality is estimated with prospective information generated from the linkage of the CRELES dataset with the Costa Rican Death Index. For individuals who died, the time of exposure is computed as the time from first interview to time of death. For right censored observations (those who do not experience the event before the period of observation is closed), the time of exposure is estimated as the difference from the date of interview to a fixed date: April 30th, 2007. It would have been possible to include deaths reported up to June 30th, 2007. However, due to late registration of deaths, Costa Rica's Civil Registry is problematic in this regard. Therefore, it is preferable to be conservative in the final date, so as to reduce the effect of late registration. Notice also that individuals have different times of exposure, even if they did not die, because most of them have different dates of first interview. The first wave was conducted from November 2004 to October 2006. This means that on average respondents who did not die have 1.5 years of exposure, but ranging from 2 years and 5 months to just 6 months (people who were interviewed at the end of the first wave).

The model for analyzing DM incidence is estimated with retrospective information on age at diagnosis. Estimates using retrospective data are affected by selection bias, since persons within the same cohorts who had DM and died before the survey was conducted are excluded from the analysis. In this sense, the hazard estimates are biased downwards, since there are fewer cases in the numerator and longer exposure than would be if these individuals were included. A correction for this selection bias is utilized. The correction is based on methods described by Keiding (Keiding, 1991, 2006; Keiding, Holst and Green, 1989). However, instead of operationalizing the correction as a model with Horvitz-Thompson weights (Horvitz and Thompson, 1952), it is operationalized as an equation with an offset term. The offset term corresponds to the natural logarithm of a survival probability over 15 years. Survival probabilities for the hazard model are estimates from the mortality event history model.

Offsetting an equation is equivalent to adding a covariate to the equation and the regression coefficient associated to this covariate is constrained to be 1. The function of the offset is to divide the estimated hazard by the person-time at risk. The equation with the offset can be expressed in the following way:

$$\log[\mu(Y)] = \ln(S(d)) + f\left[\sum_I \beta_i X_i\right] + \varepsilon \qquad (4)$$

where:

$\mu(Y)$: Hazards of the event Y.

$S(d)$: Survival probability since beginning time at risk until interview.

β_i: Coefficient for the i-th explanatory variable

X_i: The i-th explanatory variable.

ε: Error term.

d) Controlling for unobserved heterogeneity

In the estimation of hazard rates via statistical event-history models, there is the concern that if not all covariates are introduced in the model specification, the estimates for the effects of the observed covariates are going to be biased (Heckman and Singer, 1982, cited by Trussel and Richards, 1985). There are two approaches for inspecting the possible effects of unobserved heterogeneity. One is modeling an additional scalar parameter that might have a pre-defined distribution (normal, gamma, log-normal, log-gamma), while the other is a non-parametric method suggested by Heckman and Singer (1982, cited by Trussell and Richards, 1985). The analyses in this dissertation rely on the first method, which is available in STATA.

e) Biases in analysis of differences of incidence rates.

(1) Selection bias.

The present study is affected by selection bias, particularly survivorship bias. The main source of the bias comes from the fact that the sample refers only to survivors to age 60 or above in Costa Rica. Thus, the sample is excluding those who were diagnosed with Type 2 DM at younger ages and did not survive to the minimum age defined in the study. There might also be differential mortality by KH or CMI. This selection bias can affect the estimated coefficients in the parametric models. Selection bias is even stronger in the estimation of incidence of diagnosed DM, given that I am using retrospective rather than prospective information to estimate them. These limitations imply that the study can not determine the full effect of the exposure to the risk factors on the two diseases of interest. The study can not address these limitations with the existing datasets. This means that the analyses in this dissertation can not clearly make a statement about the full causal relationship between the risk factors and the

diseases under study, but among the possible effects of adverse early life conditions among the population of survivors to age 60.

Selection bias may also arise by differential non-response in the main outcome questions. Fortunately, missing cases are very few (less than 1%) for the CRELES questions regarding diagnosis of DM. This means there is a relatively low risk for increasing the bias.

Another selection bias may arise in analyses that exclude subjects whose information was given by a proxy informant; i.e., if the differences between direct interviewees and cognitively-impaired subjects (who require proxy interviews) are statistically significant and substantively meaningful. However, if data from proxy interviews are included, results might be influenced by information bias, since information provided by proxies might be less accurate than that provided directly by respondents. All the models are estimated for the full sample and for the non-proxy sample. I am not claiming that the bias due to proxy informants can be eliminated, but that the comparison between the models can shed light on the problem.

(2) Recall bias

The analyses might also be affected by recall bias, which is typical of retrospective information. Recall bias occurs when a certain subset of individuals report the occurrence of an event or the time of its occurrence differentially than another subset. For example, if subjects with short KH are sicker than subjects with longer legs, they might very well have more difficulties in reporting accurately when DM diagnosis occurred. Recall bias on DM can not be controlled for in the study. The models for DM incidence are estimated using only the persons that report having had their DM diagnosis 15 years ago at most. This generates a selection

sample bias, but I claim that the selection bias produced by this decision is smaller than the effect that recall bias may have on the final estimates of DM incidence.

f) Measurement error due to self-reporting

Measurement error due to self-reporting may arise if only the question from the interview is available. However, this kind of measurement error will be analyzed using biomarker criteria given that CRELES collects this kind of information. Another source of measurement error arises from the fact that adverse early life conditions are operationalized from surrogate measures (CMI, KH) rather than from direct measures, such as birth weight or ponderal index at birth. This research has no means of measuring the possible measurement error introduced in the analysis due to the utilization of surrogate measures rather than direct measures.

3. Strengths and Limitations in the analysis of difference of incidence.

a) Strengths

- Few studies in Latin America provide information on both chronic disease prevalence and anthropometric measures, such as KH, that can be used as markers of early health disadvantages. CRELES is one of these few.
- CRELES provides both self-reported measures and biomarkers. Few aging studies in Latin America have both types of measures.
- The survey dataset will be linked to the National Death Registry, so prospective information on deaths will be available. The problem with this Registry is that there is a delay in recording new cases.

b) Limitations

- The main exposure variables for DM (KH and CMI) are only surrogate measures for the main risk factors, which according to the thrifty phenotype hypothesis are adverse nutritional conditions during gestation and infancy. Although other authors have previously used KH or leg length (see citations above), or have shown their relationship with adverse nutritional status during early life, there is measurement error associated with using proxies instead of direct measures.
- The study population of CRELES represents persons ages 60 years old and above in Costa Rica in 2004-2006. However, DM onset can occur earlier in life and, when retrospective information is used, this information is missed or affected by selection bias.
- It will not be possible to use second-wave data from CRELES, because the fieldwork was initiated in September 2006 and has not yet been completed.

4. Demographic methods for projecting prevalence of DM.

Population projections have always been an important part of demographic work (Preston *et al.*, 2002). Projecting the prevalence of DM in Costa Rica will allow for the study of a key step in advanced stages of the epidemiologic transition; i.e., the decline of certain chronic maladies (Olshansky and Ault, 1986; Horiuchi, 1999).

a) The Cohort component method

The method derives its name from the fact that it models and projects the behavior of the main demographic components –fertility, mortality, migration–, and uses these results to

estimate the future size of the population. Normally, the rates or probabilities that describe the components are distributed by age because age is an important factor of variation in these demographic components and users frequently ask for projections for specific age groups. A full explanation of this method can be found in United Nations (1958) and in Preston *et al.* (2002). For the present document, and assuming a closed population, it is sufficient to show that the method can be summarized by the following formula (a variation of formula (6.7) in Preston *et al*, 2001: 126):

$$ {}_nN_x^S(t+5) = \left({}_nN_{x-5}^S(t)\right) * \frac{{}_nL_x^S}{{}_nL_{x-5}^S} \qquad (5) $$

where:

x:	Age
S:	Subgroups, defined by the combination of the covariates categories: sex, obesity categories, knee height categories, and CMI categories.
t:	Reference time (usually year or quinquennium)
${}_nN_x^S(t+n)$	Projected population of sex S in the age group [x, x+n[at time t+n
${}_nN_x^S(t)$	Population of sex S in the age group [x, x+n[at time t
$\frac{{}_nL_x^S}{{}_nL_{x-n}^S}$	Probability of people of sex S ages [x, x+n[to survive to ages [x, x+n[

There is an additional formula for the first age group, which depends on the number of births. However, because this study analyzes disease at old ages, this step in the procedure has a negligible importance[7].

b) Multi-state setting in the cohort component method.

As Preston *et al.* (2002) explain: "The cohort component method cannot be simply adjusted to project population by characteristics that are changing during the life course. Transitions among subpopulations are best handled by a multistate methodology that explicitly recognizes patterns of transition by age and sex." (p.129). This is the case for the projections pursued in this analysis, since a healthy person can either die without having DM or can contract the disease before dying. This multistate system has another particularity: death is the ultimate example of an absorbing state. A state is absorbing if the probability of moving from that state to another state is 0. The multistate system has been graphically illustrated in a previous subsection (see Figure IV.1). As explained in that subsection, the model is simplified because it will not consider being cured from DM, which in the diagram means no reverse arrows from "with DM" to "without DM". To make further formulas easier, states are going to be referred to by number: state 1 is "without the disease", state 2 is "with the disease", and state 3 is "death".

The existence of an absorbing state and non-reversals from certain states make the projection easier, because there are only two pathways from state 1: to state 2 or to state 3. Therefore, this stage can be considered as a competing-risk situation. The last pathway –from state 2 to state 3- can then be considered as a simple projection between two states, which can be described by formula (6) but in a particular subgroup whose size can increment or decrement depending on age.

[7] For further details, read Preston et al (2002: chapter 6).

If migration is not considered in the model, then projections will be based on three life tables representing each of the three transitions: 1→2, 1→3, and 2→3. The three life tables can be estimated with the following system of equations, derived from formula 12.3 in Palloni (2002: 263):

$$l^1(x+1) = l^1(x) - {}_1d_x^{12} - {}_1d_x^{13} \quad (6.1)$$

$$l^2(x+1) = l^2(x) - {}_1d_x^{23} \quad (6.2)$$

$$l^3(x+1) = l^3(x) + {}_1d_x^{13} + {}_1d_x^{23} \quad (6.3)$$

} Equations of type I

$${}_1\mathrm{d}_x^{12} = {}_1\mu_x^{12} * {}_1\mathrm{L}_x^1 \quad (6.4)$$

$${}_1\mathrm{d}_x^{13} = {}_1\mu_x^{13} * {}_1\mathrm{L}_x^1 \quad (6.5)$$

$${}_1\mathrm{d}_x^{23} = {}_1\mu_x^{23} * {}_1\mathrm{L}_x^2 \quad (6.6)$$

} Equations of type II

$${}_1L_x^1 = .5 * \left[l^1(x) + l^1(x+1) \right] \quad (6.7)$$

$${}_1L_x^2 = .5 * \left[l^2(x) + l^2(x+1) \right] \quad (6.8)$$

$${}_1L_x^3 = .5 * \left[l^3(x) + l^3(x+1) \right] \quad (6.9)$$

} Equations of type III

where:

$l^i(x)$: Number of individuals in state i at exact age x;

${}_1d_x^{ij}$: Number of individuals moving from state i to state j between ages x and x+1;

${}_1L_x^i$: Number of person-years lived in state i between ages x and x+1;

${}_1\mu_x^{ij}$: Rates of moving from state i to state j (estimated by methods mentioned in the previous subsection).

Given that state 3 is an absorbing state, equations (6.3) and (6.9) can be eliminated. Thus, the system of 7 equations has 7 unkowns: $l^1(x)$, $l^2(x)$, ${}_1d^{12}{}_x$, ${}_1d^{13}{}_x$, ${}_1d^{23}{}_x$, ${}_1L^1{}_x$, and ${}_1L^2{}_x$.

Having estimated the three life tables (although, actually only two of them are needed), projections can be made with an equation that is derived from equation (6):

$$ {}_1N_x^i(t+1) = \left[{}_1N_{x-1}^i(t) \right] * \frac{{}_1L_x^i}{{}_1L_{x-1}^i} \qquad (7) $$

A slight difference in the logic of the equation is that the expression $\frac{{}_1L_x^i}{{}_1L_{x-1}^i}$ cannot be interpreted as a survival probability.

c) The variation of the cohort component method used in this dissertation

The population projection estimated in this dissertation is based on a variation of the cohort component method. In this methods chapter, I will describe the basic premises of the method, while the projections chapter contains a more detailed description of the decisions made for computing the projection.

(1) Basic formulas for the projection

The projection used as inputs: diagnosed DM hazard rates, death rates, and incidence probabilities of DM unawareness. The basic formulas used for projecting the population are:

For unaware diabetic population:

$$ {}_1U_x(t+1) = \left[{}_1U_{x-1}(t)\right] * \left(1 - {}_1\mu(d)_{x-1}\right) + \left[{}_1ND_{x-1}(t)\right] * ({}_1\delta_{x-1}) * \left(1 - {}_1\lambda_{x-1}\right) * \left(1 - {}_1\mu(d)_{x-1}\right) \quad (8) $$

For diagnosed diabetic population:

$$ {}_1DD_x(t+1) = \left[{}_1DD_{x-1}(t)\right] * \left(1 - {}_1\mu(d)_{x-1}\right) + \left[{}_1ND_{x-1}(t) + {}_1U_{x-1}(t)\right] * \left({}_1\delta_{x-1}\right) * \left(1 - {}_1\mu(d)_{x-1}\right) \quad (9) $$

For total diabetic population:

$$ {}_1D_x(t+1) = \left[{}_1U_x(t+1)\right] + \left[{}_1DD_x(t+1)\right] \quad (10) $$

For non-diabetic population:

$$ {}_1ND_x(t+1) = \left[{}_1ND_{x-1}(t)\right] * \left(1 - {}_1\delta_{x-1}\right) * \left(1 - {}_1\lambda_{x-1}\right) * \left(1 - {}_1\mu(\bar{d})_{x-1}\right), \quad (11) $$

Where:

${}_1U^S_x(t+1)$ is the unaware diabetic population age x, with characteristics S, at time (t+1),

${}_1DD^S_x(t+1)$ is the diagnosed diabetic population age x, with characteristics S, at time (t+1),

${}_1D^S_x(t+1)$ is the total diabetic population age x, with characteristics S, at time (t+1),

${}_1ND^S_x(t+1)$ is the non-diabetic population age x, with characteristics S, at time (t+1).

$\mu(d)$ is the death rate for diabetic population

$\mu(\bar{d})$ is the death rate for non-diabetic population,

λ is the DM incidence rate, and

δ is the incidence rate of developing DM but being unaware of it (DM unawareness).

All formulas are applied to different sub-populations defined by covariate patterns.

d) Splitting the projections by two groups of cohorts

The projection was performed in two groups of cohorts: cohorts born before 1945 and cohorts born in 1945 or after. I made this decision because the source of the information about the size of the population is different for these groups:

- The cohorts born before 1945 are represented by the persons interviewed in CRELES. Therefore the total population in 2005 is equal to the sum of the sampling weights, which is equivalent to the size of the population age 60 and over in that year.
- The cohorts born in 1945 or after are estimated by multiplying the relative distribution by KH, canton of birth, obesity categories, and sex of people age 60 in CRELES by the size of the population age 60 for each year, according to Costa Rica's

official population projections. In this way, it is more likely to get a closer reproduction of the official population projections.

Notice therefore that the size of the projected population is achieved by introducing the absolute size of each incoming cohort (from 1945 to 1970) and transforming it with the DM and mortality hazards, and also by subsequently transforming the sampling weights of the population interviewed in the CRELES project (thus, by "killing" them).

C. Study Design from an epidemiological perspective: A summary of methods and data for the study of the relationship between risk factors and disease incidence.

1. Target Population:

The target population is comprised of the Costa Rican population ages 25 and older[8].

2. Study Population:

This dissertation will study Costa Rican persons born before 1946 (ages 60 and over in the period 2004-2006). The study population was defined based on the study population of the main data source: CRELES (See methods section for more detail).

3. Type of study:

The study can be classified as a retrospective cohort study. According to Gordis (2004), in this type of studies, "exposure is ascertained from past records and outcome (…) is ascertained at the time of the study" (p.153). However, when mortality is considered, the analysis with

[8] Clearly, the data sources for this dissertation refer to people 50 years old or older. However, the retrospective information will be used from the time that respondents were 25 year old.

CRELES might be seen as a combination of prospective and retrospective designs, since the first outcome (been diagnosed with diabetes) is recorded with retrospective questions, while time to death is recorded prospectively with the linkage to the Vital Registration System.

4. Main outcome

The main outcome is the incidence of DM, by age, and age-specific death rates in Costa Rica. In the population projections, DM burden is operationalized as the difference in the relative and absolute prevalence of DM between a scenario that retains the characteristics analyzed with statistical techniques and another scenario that assumes that no member of the cohorts under study experienced adverse early childhood conditions (undernutrition, in this case). Therefore, DM burden in the projections can be expressed as:

DM Burden due to early life conditions	=	Prevalence (observed scenario)	-	Prevalence (hypothetical scenario with no adverse early childhood conditions)

5. Exposure

The main exposure variables are knee height (KH) and the level of child mortality in respondents' respective canton of birth in their birth's year (CMI). These two variables are considered good surrogate markers of impaired growth during gestation and childhood among survivors.

6. Potential confounders

For DM, the other risk factors that will be taken into account in the analyses are: sex, age, socio-economic status SES (schooling, occupation), region of residence, family history of diabetes, smoking, alcohol, physical activity, body mass index (BMI), self-reported weight (reporting values and comparing drawings in a scale) at age 25, self-reported maximum weight (reporting values and comparing drawings in a scale), self-reported weight change, and self-reported diet patterns (Ballesteros *et al.*, 2005; Bonora *et al.*, 2004; Duran-Varela, *et al.*, 2001; Mackay and Mensah, 2004; Pagan and Puig, 2005; Rivera *et al.*, 2004; Rosas-Peralta *et al.*, 2005; Rosello-Araya, 2003a, 2003b; Velazquez-Monroy *et al.*, 2003).

Since CRELES is a cross-sectional survey, some covariates are measured at the same time as the dependent variables. However, most independent variables are measured with retrospective questions, and therefore it is most likely that the risk factors occurred prior to the outcome variable. For example, regarding DM risk factors, smoking (year when started to smoke), alcohol (age when started to drink), year of diagnosis of heart attack, year of diagnosis of cholesterol, self-reported weight at age 25, education, region of residence, and type of social insurance since 1970 (which might be a proxy for health care access). Genetics is an important risk factor that will not be included in the analysis, because the data source does not contain this information. Nonetheless, respondents' reports of a relative (parents or siblings) with diabetes will be used as a proxy for genetic endowment.

7. Sample size needed for statistical analyses

The Costa Rican study has sample size of 2,827 people. Although CRELES principal investigators aimed to study the distribution of chronic diseases in Costa Rica and their effects

on longevity, the project was not specifically designed to study the main objectives of this dissertation. Therefore, instead of planning for the adequate sample size for the present analysis, I computed approximate statistical power for the expected estimates in the analyses, given the pre-defined sample sizes. I use Collet's simplest formula (1994, p.255) for calculating statistical power for a relative risk. Assuming a significance level of 0.05 and having nearly 400 persons with DM diagnosis in the analysis, the statistical power for an expected relative risk of 1.5 for people with short KH is 86%. This implies that, if the real relative risk is less than 1.5, the power for detecting such magnitude in the association would be lower than 86%. In the mortality analysis, I have less than 250 deaths. This implies that the statistical power for detecting a real relative risk of 1.5 for people with short KH is a little less than 80%.

8. Statistical methods for estimating incidence rates and determining inter-group differences

For a detailed explanation of each method, read the appropriate description in the corresponding section of this chapter. The intention of mentioning the methods again is to provide a summary for the description of the Study Design.

a) Logistic regressions

b) Kaplan-Meier estimates for bivariate life tables.

c) **Parametric survival analysis**

d) Other statistical techniques (Demographic methods for projecting the population).

Chapter V: Estimating the hazard of diabetes and death controlling for early life conditions.

A. Introduction.

The aim of this chapter is to test whether there are statistically significant differences in DM hazard rates across the values of the main independent variables: categorized knee height (KH) and CMI operationalized as both the natural logarithm of the CMI (lnCMI) and in three categories. The chapter starts with a review of the methodological approaches used to test and estimate differences in hazard rates. The information on which the analyses are based is better suited to estimate DM prevalence rather than hazard rates because the study was designed and the questions were constructed in such a way that is easier to collect retrospective or biomarker information rather than prospective data. For this reason, the initial results in this chapter refer to estimates of DM prevalence, and its association with risk factors. The chapter then contains a discussion about measurement error in the CMI variable and other data problems. Next, a statistical approach is used to test differences in hazard rates. This approach requires information about mortality risks. An analysis of mortality using event history models precedes the models for DM hazard rates, given that death hazard rates are also needed for projection purposes. A special section of the chapter explores whether it is Barker's "thrifty phenotype hypothesis" or Finch and Crimmins's "inflammatory exposure hypothesis" the theoretical background that seems to fit better to the results. The chapter concludes with an analysis of DM

unawareness (people with high levels of serum glucose –SG– or glycosylated hemoglobin - HbA_{1C}– that have not been diagnosed with DM).

B. Statistical methods for assessing the effect of early life conditions on diabetes and death hazard rates

As described in the Methods Chapter, three statistical techniques are used to assess the impact of early life conditions on DM hazard and mortality risks: logistic regressions, Kaplan-Meier life tables controlling for KH and CMI categories, and parametric survival regressions. Logistic regressions are also utilized to estimate the effect of several covariates on DM prevalence. Prevalence is estimated before computing hazard rates because prevalence is the main outcome for the projections and because it is convenient to produce indicators comparable to previous articles that have dealt with the topic, particularly Palloni *et al.* (2005, 2006). DM hazard rates and mortality rates are used as inputs for the projections in the next chapter rather than direct prevalence proportions because the procedure allows analyzing the impact of both hazard and death risks separately and controlling for possible improvements in Costa Rica's mortality level. A logistic regression model is also used to estimate incidence probabilities of becoming an unaware diabetic. A logistic regression is used because the outcome variable is binary and we do not know the exact time between having SG levels≥200 mg/dl or HbA1C≥7%, and being interviewed in the survey.

Parametric survival regressions are estimated to determine whether DM and mortality hazard rates are statistically different across the groups that define early life conditions: KH categories and CMI. The models for DM hazard rates are estimated from retrospective

information on age at diagnosis. In order to partially correct for selection produced by the use of cross-sectional data and differential mortality, especially across groups defined by the outcome variable –DM– and the main explanatory variables, the estimated equations for DM hazard contain also an offset term. This offset term corresponds to survival probabilities estimated with mortality models.

It is necessary to highlight here that multi-state survival analysis would have been an ideal method to estimate DM hazard and mortality. However, this method was discarded because there are not enough deaths among people with diabetes at wave 1 so as to being able to estimate a separate equation for the transition "DM → death". In other words, there is a problem of limited statistical power in using multi-state estimation techniques.

1. Logistic regressions

The logistic regression aimed to analyze prevalence has a dichotomous variable as its dependent variable: it is equal to 1 if the person has DM defined using either self-reported information or biomarker information: SG≥200 mg/dl or HbA_{1C}≥7%. It is equal to 0 if the respondent does not report to have DM and at the same time has low levels of SG and HbA_{1C}. These cutoff points are typically higher than the cutoff points for diagnosis recommended by the American Diabetes Association (ADA, 2007). However, given that only one blood sample was drawn from each individual, a more conservative approach seems more adequate.

The second logistic regression is used to estimate the prevalence of being unaware of own's diabetic condition. It is conditional to all respondents that self-report <u>not</u> to have a DM diagnosis. Its dependent variable is equal to 1 if SG≥200 mg/dl or HbA_{1C}≥7%, and it is equal to 0 otherwise. Incidence probabilities are estimated using a demographic equation that contains

the DM unawareness prevalence as one of its inputs. The logistic regression allows having a smoothed prevalence curve across ages. The drawback of this approach for estimating DM unawareness incidence probabilities is that statistical tests from the regression model are performed on prevalence rather than on incidence.

Additional linear and logistic regressions are estimated to analyze the relationship between CMI and several health outcomes in order to test whether the data bring evidence for the Finch and Crimmins's framework of the "cohort-morbidity phenotype" ((Finch and Crimmins, 2004; Crimmins and Finch, 2006)).

2. Kaplan-Meier Life tables for risk of DM and risk of death

An initial exploratory technique to assess the impact of early life conditions on the risk of developing DM and the risk of death is computing life tables with the Kaplan-Meier methodology, controlling by categories from the main explanatory variables (KH and CMI). The Kaplan-Meier method is also known as the actuarial method. It is a non-parametric approach to estimate risk curves. It is a useful method to explore the bivariate relationship between independent variables and the outcome variable: risk of death or risk of developing DM. It has the limitation that it does not allow controlling for a relatively large set of covariates, as parametric survival regressions do. Moreover, the Kaplan-Meier estimates are derived from retrospective information that does not take into account differential mortality during the period of observation.

3. Parametric survival regressions for hazard rates of DM and death

Survival models are the usual techniques to analyze the effect of factors on the risk of getting a disease or dying because this type of models take into account not only the number of events of the main occurrence (numerator of a hazard rate) but the time of exposure of the individuals at risk (denominator of a hazard rate); therefore, the hazard rates is more accurately estimated than with a logistic regression with a fixed period of observation[9]. Two parametric survival equations are estimated in this chapter: one for the risk of dying and another for the risk of getting DM. The first model uses retrospective data; the model only uses DM diagnosis that occurred during the last 15 years with the aim of reducing the impact of recall bias. Persons who had the diagnosis 16 years ago or more are excluded from the analysis. Estimates using retrospective data are affected by selection bias, since persons of the same cohorts that had DM and died before the survey was conducted are excluded from the analysis. As mentioned before, the model is corrected using methods suggested by Keiding (Keiding, 1991, 2006; Keiding, Holst and Green, 1989). The correction consists on estimating the parametric equation with an additional offset term. The offset term corresponds to the natural logarithm of a survival probability over 15 years. Survival probabilities for the hazard model are estimates from the mortality event history model.

The survival model for mortality is not affected by this selection bias because it is based on prospective data. The CRELES dataset is linked to the Costa Rican Civil Registry Database (the Costa Rican Death Index) to determine which individuals died after they were interviewed in the first wave but before an established date of observation: April 30th, 2007. The time of exposure is computed as the time from first interview to time of death for individuals who died;

[9] This means that if an individual is observed during a certain period of time (e.g., two years), incidence rates computed with a logistic regression of the number of events at the end of that period of time implies that all individuals had the same amount of exposure to risk.

for those who did not die, the final date of observation is established as April 30th, 2007. On average respondents who did not die have 1.5 years of exposure, but ranging from 2 years and 5 months to just 6 months (people who were interviewed at the end of the first wave).

As described before, parametric survival models are preferred over semi-parametric regressions (like Cox regressions) because the estimated hazard curve is smoother than the one produced by a semi-parametric model, and it is better to have a smoothed curve for population projections. The parametric distribution of the DM hazard rates is selected comparing Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). Parametric distributions for mortality hazards are selected by comparing the estimated curve for a constant-only model with the official national life table for the period 2000-2005 (CCP-INEC, 2002; CCP, 2005). The purpose of selecting the parametric distribution in this way is to have a curve that resembles the official projected mortality schedule of Costa Rica for the same period.

4. Controlling for unobserved heterogeneity

Unobserved heterogeneity is the condition that arises when not all relevant covariates are introduced in the model specification, and hence the estimates for the effects of the observed covariates are going to be biased (Heckman and Singer, 1982, cited by Trussel and Richards, 1985). The method used to assess possible effects of unobserved heterogeneity in this dissertation is modeling an additional scalar parameter that might have a pre-defined distribution (normal, gamma, log-normal, log-gamma).

5. Demographic model for estimating incidence probabilities of DM unawareness from a prevalence model

In order to compute a set of incidence probabilities of DM unawarness, I will use the estimates from a logistic regression on DM prevalence of DM unawareness and from the mortality model, and evaluate these estimates in the following formula[10]:

$$\Pi(x+n,t)(1-q^{D}) = \Pi(x,t)(1-q^{HU}-q^{HD}) \qquad (1)$$

where:

q^{HU}: Unknown rate from transition Healthy-Unhealthy

q^{HD}: Rate from transition Healthy-Death

q^{D}: General mortality rate.

$\Pi(x+n, t)$: Proportion of healthy people (without DM unawareness) at age x+n, in time t.

$\Pi(x,t)$: Proportion of healthy people at age x, in time t

Mortality rates are obtained from the Gompertz mortality model estimated before, while the proportions of healthy people are estimated from a logistic regression. Its outcome variable is equal to 1 if the person has no DM diagnosis but have SG≥200 mg/dl or HbA_{1C}≥7%, and is equal to 0 if the person has no DM diagnosis and low levels in the two biomarkers. Prevalence of healthy people are estimated by substracting one from the estimates drawn from the logistic regression.

[10] Thanks to Michel Guillot for providing this formula.

The procedure has the following assumption:

- Mortality and DM unawareness hazard rates have remained constant over time. Although this assumption might be questionable, it is very likely that most of the new cases of undiagnosed or unaware DM have developed since a short time ago.
- Prevalence of healthy people has also remained constant over time. This can be restated in the following way: prevalence at age x in time t is equal to the prevalence of healthy people at age x in time t-1.

C. Results

1. DM prevalence and descriptive information

According to CRELES data, one in every four (25%) Costa Ricans ages 60 and over has DM (Table V.1) if both SG and HbA_{1C} are used as a complement to previous diagnosis information. If only SG is used, prevalence falls to 23%. This proportion rises to 25% if HbA_{1C} is used as the biomarker criterion. It is somehow surprising that the prevalence estimate is slightly lower with SG than with HbA_{1C} because HbA_{1C} is considered less sensitive than SG (Hanson *et al.*, 1993; Modan *et al.*, 1984; Wiener and Roberts, 1998). Only 21% report to have a DM diagnosed by a health care professional (mainly a physician); therefore, around 3% to 4% of the elderly can be classified as potentially undiagnosed diabetic patients or diabetic persons that are unaware of their condition. DM unawareness might become a serious problem if it means that a certain proportion of the diabetic population is not controlling well their disease. Notice

however, that this proportion is relatively low. Based on this criterion, only 22% of the diabetic senior population is not aware of their disease.

I decide to use both HbA_{1C} and SG because the data show discrepancies between the two biomarkers, even though they are measuring similar but not equal constructs. According to Table V.2, the correlation between the two biomarkers is just about 0.312 among those persons with no diagnosis and, consequently, no medication for controlling the disease. The zero-order correlation is also lower among those that have a diagnosis and are taking medication for controlling the disease. Among people under medication, HbA1C is a more reliable marker of DM control, given that SG is more volatile to momentary changes in diet and metabolism. The strongest correlation is among diagnosed diabetic patients that are not taking medication. This result was expected given that high levels of SG might be steadier in this population, and this steadiness is represented well by HbA_{1C}.

Glyosylated hemoglobin has been proposed as an alternative to the Oral Glucose Tolerance Test (OGTT, the so-called "gold standard" for diabetes diagnosis which is also recommended by the WHO) because patients do not need to fast, to drink the glucose solution, or to wait for 2 hours before blood samples are drawn, and because it is considered to be a better biomarker for daylong blood glucose concentrations (Peters, Davidson, Schriger and Hasselblad, 1996). However, in 2003, an Expert Committee on the Diagnosis and Classification of Diabetes Mellitus suggested not to use this clinical exam for diagnosis because the lack of unified standards across laboratories (ECDCDM, 2004). Although in the U.S.A., standardization in laboratory processing of this biomarker has been achieved lately, the American Diabetes Association recommends against using HbA_{1C} for diagnosis purpose (ADA, 2007). In Costa Rica, laboratory standardization in HbA_{1C} processing has not been achieved yet.

On the other side, the drawbacks of using SG are:

- Interviewers can not verify that respondents were really fasting when the blood sample was drawn;
- The test is less accurate for elderly populations;
- The study does not analyze Impaired Fasting Glycaemia (IFG), which is defined as having FSG concentrations greater than 110 mg/dL but lower than 126 mg/dL (WHO, 1999). The limitation consists on having patients whose physicians might have told them that they had high blood sugar levels, as a simpler way to tell them that they had IFG.

In this dissertation, I decide to use high cutoff points in SG and HbA_{1C} so to be conservative in using biomarker information and reduce possible false positives.

The next phase is to describe confounding covariates that the analyses for DM prevalence control for, as well as the association between each of them and the main dependent variable. Table V.3 contains frequency distributions of these variables for the total population, the population with DM, and the population without the disease. These tabulations are also useful for understanding the characteristics of the diabetic old age population in Costa Rica. The data show that there is a considerably high prevalence of obesity and overweight among Costa Rican elderly: 26% can be classified as obese, whereas 42% have a BMI between 25 kg/m^2 and less than 30 kg/m^2. Overweight is as common among diabetic patients as among people without the disease, but obesity (BMI$\geq$ 30 kg/m^2) is more frequent among the former. Nonetheless, only 15% of the older age population eats 400 g of carbohydrates or more daily, and this carbohydrate intake level is more frequent among people without DM (although the difference is not statistically significant). Obesity prevalence among Costa Rican seniors is similar to figures observed in Bridgetown (Barbados) and Mexico for the same age group, but is lower than among

elderly urban dwellers in Santiago (Chile), Montevideo (Uruguay), and Mexico City, though it is higher than in Havana (Cuba) and Sao Paulo (Brazil) (Andrade, 2006; Monteverde *et al.*, 2007).

Fifty two percent of the population age 60 and over are women. This figure is worth commenting because, even though female life expectancy is greater than male life expectancy in Costa Rica, the difference is not as large as observed in other countries, and male excess mortality has remain roughly constant over recent decades (Rosero-Bixby, 1996, 2005; Rosero-Bixby, Brenes-Camacho and Collado Chaves, 2004). This means that the sex ratio in the Costa Rican elderly population is not as high as in other countries. Concerning other sociodemographic characteristics, around half of the elderly population lives in the Metropolitan Area (where the capital is) called GAM (Gran Area Metropolitana in Costa Rica), and half is retired; schooling is not very common among these cohorts, and only 49% have 6 years or more of formal education. Regarding health habits, alcohol drinking is more frequent than smoking. Four out of every 5 persons that have ever smoked have already quit, but among people with an alcohol drinking history, half are still drinking. Additionally, a third of the sample reports doing some kind of physical activity regularly. Almost all of the elderly have ever worked, and 42% of the older age population can be considered to have low income (less than $100 per month). This latter figure is important because cross-country comparisons have shown that Costa Rica is one of the few countries in Latin America where poverty levels are larger among the elderly than among the rest of the population (del Pópolo, 2001). Forty percent of the sample report that either their parents or siblings have been diagnosed with DM, which shows how prevalent this disease is among surviving older cohorts. The bottom part of the table contains the distribution of female respondents by parity. This distribution shows how high fertility was among these cohorts of women, since more than 80% had 3 children or more and almost half had 6 children or

more. Finally, this table includes the percentage of interviewees that needed a proxy respondent: 12% (weighted percentage) needed another person to respond to the questionnaire given cognitive limitations. This proportion is not different among people with and without DM.

Although differences across distributions can give an idea of the association between DM and this set of covariates, it is better to assess whether there is any association with χ^2 homogeneity tests for differences in DM prevalence across the subgroups defined by the covariates. Table V.4 shows these figures, as well as p-values for the corresponding statistical tests. As expected from descriptive statistics in Table V.3, DM prevalence is strongly associated with obesity. It is worth noticing that the proportion of people having DM is larger among respondents with missing BMI –usually interviewees who can not stand up because of disability or illness– than among people with overweight. Other risk factors with significant associations (at a significance level of 0.05) are: being a woman, not doing regular physical activity, having being hospitalized (which is not a risk factor of DM, but hospitalization increases the probability of having a DM diagnosis because of routine exams), and self-reported family history of DM. At a significance level of 0.05, it seems that nulliparous women have a smaller DM prevalence than women with children. This relationship might be due to multiparous women being fatter than average but, given that this variable is based on information about children ever had, the causal association might be the reverse: women with DM are more likely to experience miscarriages because of the metabolic consequences of DM (Dunne, 2005).

As mentioned above, the main independent variables in the analysis are KH and CMI in the respondent's canton (or county) of birth. Prevalence figures, as well as the respective χ^2 test, are presented in Table V.5. This table also shows mean SG and HbA_{1C} levels and p-values from One-way Analysis of Variance F-tests to assess whether the mean of at least one of the

categories is statistically different to the rest. Before describing the results, it is important to remark that the categories for CMI include the category foreign born which was created due to missing values: for foreign-born, there is no child mortality information for their place of birth. There were other non-randomly distributed missing values due to limited availability of statistical information for the year when some of the respondents were born. Values for CMI were imputed estimating them from time-series linear regression models that accounted for serial autocorrelation (equations not shown). The effect of these imputed values is analyzed when appropriate. Regarding the results about prevalence, the χ^2 test does not let rejecting the null hypotheses of homogeneity of proportions across groups defined by either CMI or KH. There is no statistical difference in SG or HbA_{1C} means across KH or CMI groups.

2. Measurement error in CMI

Before continuing with the substantive analyses, it is important to acknowledge that the variable CMI could be affected by underreporting in Vital Statistics. Nowadays, under-registration in birth and death records in Costa Rica is very rare (CCP-INEC, 2002), but the most effective measures to improve Vital Statistics and the Civil Registry occurred during the 1940's and the 1960's decades. There is considerable measurement error in the variable CMI since it was constructed from Vital Statistics information from the first part of the 20^{th} century. Underreporting of child deaths relative to birth reports are calculated by subregion, based on information from the 1927 and 1950. The procedure followed is to reconstruct the number of expected deaths in a subregion by comparing the enumerated population younger than 1 year old

to the mean annual number of births occurred in the census year and 1 year after, in such a way as to have the equivalent cohort size observed in each census; the difference between the two quantities should be equivalent to the expected number of deaths. Bear in mind that birth figures might also be affected by underreporting, but if births and child deaths have the same proportion of underreporting, the CMI is not biased by the measurement error. Therefore, this comparison of observed to expected deaths signals the proportion of death underreporting relative to birth underreporting. The magnitude of the measurement error is quite large especially in 1927 (Table V.6). In most of the subregions, underreporting was larger than 30%. However, in 1950, underreporting was lower than 10%, except in three very rural subregions: Guapiles-Sarapiqui, San Carlos, and Eastern Cartago. Two strategies are selected to study the effect of this measurement error: (a) categorizing the CMI after a geographical-historical inspection, and (b) estimating the same statistical model several times, but in each time a different subset of cases corresponding each to a different subregion. The actual procedures are explained with more detail when their results are presented.

In order to implement the first strategy, several maps are drawn illustrating CMI in cantones during the first half of the 20^{th} century (Figure V.1). Four years are selected: 1915, 1925, 1935, and 1945, so as to have equal time spacing between one year and the other. CMI is categorized in three groups: CMI$\leq$ 18.0 deaths per 100 births, 18.0<CMI<31.0, and CMI$\geq$31.0 deaths per 100 births. The cutoff points 18.0 and 31.0 are roughly equivalent to the 20^{th} and 80^{th} percentile of CMI among those with no missing values on it.

One of the most obvious and expected results in the maps is that cantones in the lowest CMI categories started to be more frequent through time. In 1945, there were no cantones in the upper CMI category any more. One of the most important results observed in the maps is that

there were certain cantones that were in the lowest CMI category during several years in the period. Among them, the ones around Santa Cruz and Carrillo in the Northern Pacific, as well as certain mountainous towns in the outskirts of the Central Valley (Poás, Alfaro Ruiz, Dota). These places had in common that they had settlements established before the 20th century, but they had not been densely populated. Another possible explanation regarding the Northern Pacific is that it is located in the Dry Tropical Forest region, which has a very mild climate.

On the other hand, regions with high child mortality varied throughout the period. However, particularly before the 1940s, there were two areas where high child mortality was more common during the period. The first area was comprised by province capital towns and some other more urban cantones (Goicoechea, Tibás, Alajuelita). Although Costa Rica did not have densely populated cities at the beginning of the 20th century (like Mexico City or Lima, Peru), towns in the Central Valley were denser than the rest of the country. Population density favored the spread of infectious diseases before vaccination and other public health measures were carried on. Also, a report from the Health Secretariat (Luros, 1942) noticed that water pollution due to coffee processing (a very important industry in the Central Valley) increased the incidence of diarrheas and other infectious diseases. The second area with high CMIs grouped the cantones in the Turrialba-Limón axis. This subregion was the main way of communication between the Central Valley and the Port of Limón. It has a very rainy climate, which then favored certain infectious agents (especially, water-born or mosquito-transmitted, like malaria).

From a methodological point of view, the mapping roughly matches historical processes that can explain why child mortality due to infectious diseases was high or low in certain areas. Given these results, the CMI categorization is transformed into two dummy variables –one for low CMI and the other for high CMI–, and the models are estimated twice: the first time with the

continuous CMI (lnCMI), and the second time, with these two dummy variables. Results are compared to see if this strategy to overcome measurement error in CMI produces similar results. Using the two dummy variables diminishes the explanatory richness of the continuous variable, but it apparently reduces the measurement error given that the new variables have not only a quantitative explanation, but also a substantive justification.

The second strategy of excluding respondents from each subregion sequentially and observing how coefficients change is implemented after each of the most important models is estimated.

3. Confounding effects in the relationship between DM prevalence and early life conditions

Returning to the main topic of this chapter, it is important to notice that, although the underlying curve of DM prevalence by KH categories presented in Table V.5 has the expected shape (some kind of U- or V- shape), the bivariate analyses does not yield evidence of a strong association between the main explanatory variables and DM. Concerning KH, Palloni *et al.* (2005, 2006) found a weak but statistically significant association between short KH (defined by length under the first quartile or first quintile) only in Puerto Rico, Santiago (Chile), and Mexico City, but not in Bridgetown (Barbados), Havana (Cuba), Montevideo (Uruguay), and Sao Paulo (Brazil). In this study, the authors controlled for sex, age groups, years of schooling, bad health before age 10, and socio-economic status in a logistic regression. I use the same model specification with the Costa Rican data with the aim of estimating comparable odds ratios (Table

V.7). Two models are estimated: one that defines DM with self-reported information only –so to have a comparable model to the one estimated by Palloni and colleagues– and the other with self-reported data plus biomarker information (SG≥200 mg/dl or HbA_{1C}≥7%). The size of the effect, though is not significant at a 0.05 level, is very similar to the one found in Puerto Rico (OR=1.35), and smaller than the ones found for Mexico City (OR=1.51) and Santiago (OR=1.67). The odds ratio gets smaller (OR=1.23) and remains non-significant when DM is defined using both self-reported and biomarker information.

Odds ratios computed with these logistic models do prove that there is considerable confounding effect on the bivariate relationship between KH and DM prevalence. The following table (Table V.8) contains coefficients and standard errors of two other logistic equations that control for more independent variables. The first model includes the three dichotomous variables used to operationalize KH, while the second equation adds lnCMI to the set of independent variables, as well as the foreign born indicator. These equations can be understood as models for studying the association between known risk factors and the prevalence of DM. The first result to notice is that none of the coefficients associated to the KH variables are significantly different from zero, not even at a 0.10 level. Their values suggest the V-shape that was expected according to the literature. In the second model, the coefficient for lnCMI is not significant, neither does the coefficient for foreign born.

The two models (with and without the CMI set of variables) have the same covariates with statistically significant coefficients (Table V.8):[11] obesity (OR=2.46), overweight (OR=1.51), missing BMI (OR=2.50), physical activity (OR=0.65), known DM family history (OR=2.33), and parity (OR=1.11, per additional child). A change that is important is the

[11] All OR refer to the ones in the second model that include the two sets of main explanatory variables.

models' constant, since it changes from -1.55 in the first model to -1.19, in the second model; this change is important when the model is used for estimating hazard rates for the projection in the next chapter.

4. What is CMI measuring?

An important result in the models presented in Table V.8 is that including lnCMI and the "foreign-born" indicator variable in the model neither makes the coefficient for short KH to change much, nor produces large changes in the regression coefficients of the other covariates. The model is then robust to the new specification. The coefficient for lnCMI does not vary much either whether the model is specified with the set of KH dummy variables or without them (Table V.9). This means that each of the two sets of variables (KH and lnCMI) contributes separately to explaining the variability in the dependent variable, but it can also mean that both variables are not measuring the same construct. The Pearson zero-correlation coefficient between KH (as measured as a continuous variable) and the logarithm of CMI is significantly different to zero at a 0.05 level, but its value is rather small: 0.124. What this value signals might be a problem because it was expected that the two variables were both surrogates for bad health or undernutrition during early childhood.

But, if KH is a good marker of early nutrition, what does CMI represent? As explained in the "Theoretical Background" chapter, Barker's "thrifty phenotype hypothesis" is the main framework on which this dissertation is based. This means that the aim of this chapter is to link early undernutrition –as marked by either KH or CMI– with DM hazard rates and prevalence.

However, the "thrifty phenotype hypothesis" is not the only theoretical explanation that links early life events with chronic disease. The "life course approach" to chronic disease etiology is broader than just the "thrifty phenotype hypothesis". A new "life course" framework that links early life conditions with chronic disease is what Finch and Crimmins (Finch and Crimmins, 2004; Crimmins and Finch, 2006) call the "hypothesis of inflammatory exposure". It has been developed to explain the relationship between child mortality and adult mortality levels within the same cohort. They argue that the decrease in cardiovascular mortality in old age is positively associated with the decrease in exposure to infectious diseases early in life and subsequent decrement in infant and child mortality due to these causes; they call this association the "cohort morbidity phenotype" (Crimmins and Finch, 2006: 498). The authors try to prove their main argument with historical life tables from four European countries. Their data are age-specific mortality probabilities for the 18^{th} and 19^{th} centuries. They study historical populations because "…the inflammatory mechanisms that we describe can only work when mortality from infection is high" (Crimmins and Finch, 2006: 499) and because they can parse out the effect of the popularization of smoking, a major factor of inflammation.

The main mechanism in this relationship relies on the idea that infections, even when latent or cured, produce chronic inflammation, as marked by C-reactive protein (CRP), interleukin 6 (IL-6), and/or fibrinogen. Chronic inflammation accelerates atherosclerosis, a risk factor for cardiovascular disease. The long-term role of infectious diseases in mortality change has been discussed by other authors as well (Elo and Preston, 1992; Leon and Smith, 2000); it is also a key factor in Gersten and Wilmoth's (2002) "Cancer Transition" framework, since death rates due to tumors linked with infections (gastric cancer, cervical cancer, liver cancer) have been declining in industrialized countries over time during the last 100 years. All these

hypotheses add to the increasing number of literature inspired on the "life course perspective in epidemiology" (Kuh and Ben-Shlomo, 1997). According to one of its variants, conditions early in life have deleterious effects on adult health (Barker, Eriksson, Forsén and Osmond, 2002; Doblhammer, 2004).

A way of testing whether CMI is a marker of early undernutrition or of early exposure to infectious diseases is to estimate the association of CMI with three variables: SG, HbA_{1C}, and and C-reactive protein (CRP). There is research that has pointed out the association between CRP in one hand, and FSG or DM in the other (Barzilay *et al.*, 2001; Festa *et al.*, 2002; Nakanishi *et al.*, 2005; Wang and Hoy, 2007), although the evidence is not yet clear about the role that CRP plays in the pathogenesis of DM (Thorand *et al.*, 2003). Thus, it can be stated that the inflammatory exposure hypothesis might predict elevated SG and HbA_{1C} levels. However, if CMI is related to the two DM markers but not to CRP, or if CRP does not mediate or modify the relationship between the two DM biomarkers and CMI, then the results would suggest stronger evidence for the thrifty phenotype hypothesis rather than for Finch and Crimmins's framework. The three biomarkers (SG, HbA_{1C}, and CRP) are estimated from fasting blood samples analyzed in clinical laboratories, although around 2% of respondents reported that they were not fasting. As response variables, they are operationalized in two ways: as continuous variables (in their own units of measurement), and as categorical variables that indicate high levels: FSG $\geq$200 mg/dl, $HbA_{1C} \geq 7\%$, and CRP$\geq$10 mg/L.

In this sense, results should resemble one hypothesis more than the other if the following positive associations are found:

Condition predicted by framework	Framework	
	Finch and Crimmins's hypothesis of inflammatory exposure	Barker's thrifty phenotype hypothesis
ln of CRP	Yes	No
ln of SG after controlling for CRP	No	Yes
ln of HbA1C after controlling for CRP	No	Yes

As with the other equations in this project, data are analyzed with logistic and linear regressions that take into account the correlation across individuals born in the same canton and in the same year. The regressions control for other covariates related to the chosen biomarkers. Most control variables are the same utilized in the models about DM prevalence. Additionally, variables about saturated fat intake, carbohydrate intake, and the ratio total cholesterol/HDL (High Density Lipoproteins) –which is assessed in clinical laboratories from fasting blood samples– are also included in the models for CRP.

In the first lines of Table V.10, there are regression coefficients for lnCMI and categorized CMI in models that have CRP, SG, and HbA_{1C} as dependent variables. The three variables are logged because they have skewed distributions. The coefficient for CRP is not significant at a 0.10 level. The coefficients for the SG equations are significant at a 0.05 level. The coefficient remains significant and with the same magnitude even after controlling for CRP levels. The coefficients in the HbA_{1C} equations are not significant at a 0.10 level. The same conclusions are drawn if CMI dummy variables are used as independent variables instead of lnCMI. It is worth noticing that the magnitude of the coefficient for the regression of lnCRP is larger than for the equation of lnSG. Given that both variables are logged and the coefficients refer to the same explanatory variable, they can be compared. That the coefficient for the lnCRP equation is not significant even though it is larger means that lnCRP has larger variability.

Equivalent logistic regressions are estimated. The dependent variables are dicothomies that are equal to one if the levels of the biomarker are higher than cutoff points commonly used in clinical research: SG≥200 mg/dL, HbA_{1C}≥7%, and CRP≥10 mg/L. None of the odds ratios for CMI, either as a continuous variable or as a set of dummies, is significantly different to one, except the odds ratio for the variable CMI≤18 deaths per 100 births in the logistic regression for CRP≥10 mg/L. That this coefficient is statistically significant rather than the one for CMI above 31.0 deaths per 100 births possibly means that there is a threshold that indicates improvement in child health, rather than a linear association or a "dose response. Again, the magnitude of the odds ratios for SG or for HbA_{1C} does not change after controlling for CRP.

These findings are contradictory, but they seem to lean more towards Barker's hypothesis rather than Finch and Crimmins's as an explanation for the association between CMI and health among Costa Rican elderly. The thrifty phenotype hypothesis does predict the association between early childhood conditions and DM (in this case, with high SG levels), and the association is not explained by CRP levels, the main mechanism suggested by the "inflammatory exposure hypothesis". Therefore, as a conclusion, CMI seems to be a marker of early life undernutrition rather than of early life exposure to infections (even though both are theoretically linked) because the observed associations resemble the "thrifty phenotype hypothesis" more than the "inflammatory exposure" hypothesis.

5. The timing in the occurrence of independent variables and the pictographic scales

Before estimating relative risks, an additional problem in the data analysis is the timing in the occurrence of characteristics operationalized as independent variables. It was mentioned in the Data and Methods section that the data come from a cross-sectional survey, and the only longitudinal information is the date of death in a period of approximately two years after the interview. Given that the data are cross-sectional, some of the independent variables refer to the present, while the diagnosis of DM occurred in the past. This is a clear limitation if any causal inference wants to be drawn from the analyses. The covariates that have this problem are: obesity/overweight, place of residence, retirement status, income, smoking, alcohol drinking, hospitalization, and physical activity. The problem can not be solved in a straightforward fashion because of data characteristics. Among the covariates, the effect of obesity might be seen as the most troublesome. The effect of obesity on DM etiology occurs throughout the life course, rather than simultaneously (Jeffreys *et al.*, 2006). However, CRELES data allow a way to approach investigating the effect of body weight over the life course. There are pictographic scales that represent different body sizes, and the respondent judges how she/he looks currently, and how he/she looked at age 25 and at age of maximum weight. The answers to these scales are affected by subjectivity. However, assuming that the same perception bias is constant in the three questions, changes between scales can offer a good idea of weight increase from previous to current age, and therefore these changes can be considered as an acceptable surrogate for weight change through the life course. The problem with this set of variables is that they are asked only to individuals that did not need a proxy to answer the questions; in other words, the pictographic scale questions are only answered by direct respondents, and not by proxy respondents. Therefore, Table V.11 contains the basic logistic model, but estimated in three ways: the original one (as presented in Table V.8), the same model estimated only among direct

(non-proxy) respondents, and the model including the set of variables about perceived weight change among direct respondents.

As it can be seen in the middle columns of Table V.11, the effect of short knee height and lnCMI grow slightly stronger in the non-proxy subsample, although the coefficients remain non-significant. The other coefficients do not change much. Contrary to what was expected, there is no effect of perceived weight change on DM prevalence, and the coefficients for the obesity and overweight dummy variables increase only slightly. From a positive point of view, this can be interpreted as that the inclusion of the variables for current obesity and overweight is accounting for the bulk of the association between body mass index (BMI) and DM. From a negative point of view, this might mean that the subjective weight change measure is not representing the real effect of becoming obese over time on DM. Even though there is no effect of the set of perceived weight change variables on the dependent variable, all of the following statistical models are estimated twice: one with the total sample and the other only with the non-proxy subsample.

Results do not vary much either, if CMI is incorporated to the model as a set of dummy variables rather than as a continuous variable, as shown in Table V.12. The same coefficients have similar magnitudes and are similarly significant as in the previous set of models. The coefficient for high CMI's (CMI≥31) and the one for low CMI (CMI≤18) are small and non-significant.

As explained before, the other strategy for controlling for the measurement error in CMI is to estimate each model excluding all the respondents from each subregion sequentially, one subregion at a time. Figures I.2 and I.3 show the confidence intervals for the variable KH<Q1 and for CMI≤18.0, without the respondents from the subregion indicated in the graph; e.g., the

first confidence interval corresponds to the estimation that excludes respondents from the subregion called San Jose (basically, the capital city and surrounding cantones). The estimate is very stable across the different specifications for both variables. Sizes of the sub-samples are obviously smaller, reducing the statistical power of the test even more. The coefficients for KH<Q1 and CMI≤18 are not significant for the total sample, thus it was expected that its corresponding sets of confidence intervals would intersect the value 0 anyway.

As a conclusion to this analysis, the measurement error does not appear to have an effect on the estimates, according to the two approaches used to overcome it. However, it is important to acknowledge that this measurement error may still be affecting the estimates in ways that are captured by neither of these approaches.

6. Analyzing and estimating mortality hazard rates

As explained several times before, the analysis of differential mortality by DM status and early life conditions has two purposes: to generate estimates of one-year survival probabilities needed for the DM projection approach and to estimate 15-year survival probabilities that are going to be used as offset terms in the equations for estimating DM hazard. The statistical technique that is used to analyze mortality is a parametric survival analysis.

In the mortality analysis, the parametric distribution was chosen so as to resemble the curve of official death rates for Costa Rica as closely as possible. This selection criterion is needed in order to get an adequate set of mortality hazard rates for the projection, so it will roughly fit Costa Rica's official population projections. Therefore, hazard curves are estimated

using constant-only parametric survival regressions with several distributions: Gompertz, Weibull, log-normal, and log-logistic. The exponential distribution is excluded because the hazard curve is a flat line. The curves are compared graphically with the curve of mortality rates in the official Costa Rica life table for the period 1995-2005 (Rosero-Bixby, Brenes Camacho and Collado Chaves, 2004). The Gompertz and Weibull curves are the ones that most resemble the official life table (Figure V.4). A closer inspection to the graphs in the logarithmic scale (second panel in Figure V.13) reveals that the Gompertz curve is more similar to the official death rate curve than the Weibull curve. Therefore, the Gompertz distribution is selected for the analysis. It is also important to clarify that the curve is also estimated using a Cox proportional hazards regression. The estimated curve is compared to the official death rate curve (Figure V.5). The Gompertz curve has still a better fit. Nonetheless, the important issue to observe is that the curve estimated with Cox regression is less smooth than the ones estimated with parametric regressions. This is a characteristic of the Cox model which does not fit well into one of the main purposes of this dissertation, since the estimated hazards are used as inputs in the projection of DM prevalence. Projections with "smooth" inputs have the advantage of facilitating the analysis and extrapolation of demographic trends and the setting of assumptions (Congdon, 1993).

According to the first set of regressions, using lnCMI as an independent variable rather than the set of dichotomous variables (Table V.13), people with DM have a higher risk of dying (RR=1.44) although the relative risk is not significantly different to one. The same value is observed in the equation with the total sample and in the equation with the total non-proxy sample. The most striking result is that people with short KH have a lower mortality risk than people with KH between the second and the third quartiles (significant at a 0.05 level). On the

contrary, people born in cantones with high CMI have higher chances of dying than people born in cantones with lower child mortality. Results are very similar if CMI is categorized and incorporated in the model as a set of dummy variables rather than as a continuous variable (Table V.14). Only the dummy variable that corresponds to CMI≥31 deaths per 100 births is significant at a 0.10 level, but a linear trend test (not shown) suggests that the positive association between CMI and mortality is not due to chance ($p<0.05$).

It is important to highlight than, in the model with a continuous measure for CMI, the regression coefficient for the "foreign-born" indicator is considerably large and significant. In the case of the foreign born, this might be due to few but influential observations given that they represent a very small fraction of the sample. The coefficient is no longer significant in the equation with the set of CMI dichotomous variables. This means that the significant coefficient in the first regression is an artifact of the construction of the variables since this indicator variable was created to assess the missing values observed in the CMI variable.

Given the importance of early childhood condition measures in the analysis, an additional model was estimated. This model includes interactions of the dummy variable "Diabetes" with each of the early childhood variables: KH and CMI (Table V.15). The only model that is shown has CMI as a continuous variable[12]. None of the interaction terms has hazard ratios that are significantly different to one; therefore a model with no interactions is the one that was chosen for explanatory and estimation purposes. However, it is worth noticing that the relative risk for the interaction between short KH (KH<Q1) and DM is considerably far from one. This suggests that with a larger sample size, people with short KH might even have lower death rates, especially those who have DM compared to those that have both longer KH and DM.

[12] The model with CMI operationalized as a set of dummy variables is not shown, but the conclusions are the same.

The patterns observed in Tables I.13 and I.14 have strong implications for the prevalence models presented above. Lower mortality means longer life span and, among ill people, longer duration of a disease. If people with short KH have had on average lower mortality than taller people, the higher odds ratios of DM prevalence in this group might be due to longer duration rather than due to higher hazard rates.

This survival model also suggests that there is a real process of mortality-driven selection in the hazard model for DM based on retrospective data. The significant relative risks for the CMI variables also matches with analyses by Finch and Crimmins (Finch and Crimmins, 2004; Crimmins and Finch, 2006) that link mortality levels of a cohort at early ages with mortality levels of the same cohort at older ages. However, as shown above, the link between CRP and CMI is not significant, suggesting that the infection-related mechanism proposed by Finch and Crimmins is not the pathway that is explaining the association between mortality levels at younger and older ages.

Regarding the other control variables, obese Costa Rican elderly do not have a higher risk of dying than people with BMI under 25 kg/m^2, but overweight people have significantly lower chances of dying (RR=0.50) than leaner people. People with missing BMI have higher death hazards, but this is a completely expected result, since people who do not have measures of height and weight were typically people that were too sick to be measured during the first wave fieldwork. Current alcoholic drinkers are less likely to die than non-drinkers but the hazard ratio is not significantly different to one. In a similar pattern, current smokers are as likely to die as non-smokers, but those who quit smoking are less likely to die. Females have lower mortality hazard rates than males, and retired seniors are more likely to die than those that have not retired yet. Finally, it is worth noticing that the coefficients for low income and for education are small

but significant at a 0.10 level. This finding confirms previous articles that have found a weak negative SES gradient in Costa Ricans' mortality schedule (Rosero-Bixby, Dow and Laclé, 2004). Results are consistent if the model is estimated with the CMI categories instead than with the continuous CMI variable.

As mentioned before, this hazard model is used to estimate death rates that are utilized in the next chapter projections. Figure V.6 contains a set of estimated mortality hazards by age that is compared to the official Costa Rican life table death rates. Like the curves in Figure V.4, the estimated Gompertz curve is below official death rates, after controlling for the effect of other confounding variables, especially at older ages. This means that in the next chapter, it is very likely that the projections are overestimated.

Now that the mortality model is selected, its estimates are going to be used for the statistical technique chosen for analyzing and estimating DM hazard rates: a parametric survival model with an offset term.

7. Analyzing and estimating DM hazard rates

a) Kaplan-Meier Life tables for DM hazards

The second methodological approach is based on a parametric survival regression. Before estimating a full parametric equation to model DM hazard rates controlling for different covariates, it is useful to estimate hazards with a more parsimonious model in order to inspect the general pattern. The actuarial model or Kaplan-Meier method allows estimating hazard rates without implying any structural distribution for the "failure rates" (like the equations in the

parametric models). The Kaplan-Meier estimates are derived from retrospective information, but that only takes into account diagnoses that have occurred during the last 15 years. Therefore, if a 70 year old person reported having been told by a doctor that she had DM when she was 40 years old (30 years ago), she is excluded from the estimation.

Instead of presenting tables with all the estimates, it is more eloquent to have graphs controlling for the two main independent variables. According to the curves by KH (Figure V.7), the tallest people (tallest KH) have the largest DM hazard. People with the shortest KH have the smallest hazard curve. The curves seem to be proportional one to the other, except at very old ages (80 years old or more).

According to the curves by CMI categories (Figure V.8), people that were born in cantones with the highest CMI (above 31 deaths per 100 births) have consistently lower DM hazards, except at very old ages (above 80 years old). People born in cantones with CMI under 18 have very similar hazards to those people that were born in cantones with middle range CMIs. The proportional hazards assumption might hold except at very old ages.

In the Kaplan Meier estimates using STATA, it is possible to adjust by differential survival probabilities, following the variation to Keiding's method that I am using in this project. Figures I.9 and I.10 contain the corresponding graphs by KH and CMI categories. After adjusting for survival probabilities, the general pattern observed in Figures I.7 and I.8 remain, although DM hazards for people with the shortest KH are now greater than the hazards for the people with KH between the second and the third quartiles. In the graph by CMI categories (Figure V.10), DM hazard rates are leveled off at very old ages.

In general, the curves show patterns that are completely opposite to what is expected according to the theoretical framework, even after adjusting for differential survival

probabilities. People with short knee height or people born in cantones with high child mortality are not more likely to have DM. The parametric models allow analyzing whether there are confounding effects in the patterns observed for the Kaplan Meier estimates.

b) Parametric survival models on retrospective age at diagnosis.

The estimation of the parametric survival model is the next step. The Weibull distribution is selected for describing the hazard function because Aikaike Information Criteria (AIC) show the best fit among a set of hazard distributions (See Table V.16, where the Weibull equation has the smallest AIC and Bayesian Information Criterion BIC). An unexpected result is that the Weibull model seems to be more appropriate than the log-normal or the log-logistic distributions. These two last distributions are better in describing curves with non-monotonic trends. With the aim of inspecting why this pattern is found, two log-normal curves are estimated: one with no covariates, and the other with the covariates utilized throughout this model. While the graph for the only-constant model has a non-monotonic curvilinear trend, the graph with all the controls shows a curve that increases and then stays at around the same level at older ages (Figure V.11). This is a curve that can be modeled using a Weibull equation.

Table V.17 shows the coefficients for the early condition variables with and without the correction for differential survival. In the model without the correction, neither the KH set of variables nor the CMI set of variables has statistically significant coefficients. These results are observed in the model with the total sample and the model with the non-proxy subsample. Results do not change much after the correction is introduced into the model.

The coefficients for the full models with the correction are presented in Tables I.18 and I.19. Obesity, overweight, not doing regular physical activity, and known family history of DM remain as important risk factors with relative risks significantly different to one. Having low income increases the risk of having DM but the size of the hazard ratio is rather small (RR=1.27, $p<0.10$). It is worth noticing too that the coefficients for the set of perceived weight change variables are not significant in these models either. The same results are obtained if the model is estimated with the dummy variables of the CMI categorization, rather than with the operationalization of CMI as a continuous variable (Table V.19).

A possible explanation for the statistically non-significant coefficients is that some of these covariates are mediating the effect between early childhood conditions and DM risk. Figure V.12 shows how the coefficients for the short KH dummy (KH<Q1) and for lnCMI change after sequentially adjusting the model by the most important DM risk factors. None of the coefficients is ever statistically significant to zero, but the graph shows that adjusting by obesity categories, the size of the two coefficients increases (in absolute value). This means that, rather than mediating, not accounting for obesity in this kind of models would yield coefficients affected by confounding effect. Physical activity does seem to mediate the association of short KH and DM, given that the size of the coefficient shrinks slightly after controlling for it. This might be expected because people affected by stunting might be less likely to perform regular physical activity. The size of the lnCMI coefficient diminishes the most after controlling for the rest of the variables. Given that canton of birth is closely associated with current region of residence (one of the variables in the full model), controlling for the latter condition absorbs part of the effect of CMI on DM risk.

Another possible explanation in finding small and statistically non-significant effects of early childhood on the hazards of DM is that the model is not testing for possible interactions between these early childhood conditions and relevant DM risk factors, particularly obesity. Several models are estimated testing interactions. The only model with statistically significant interactions is one in which KH and BMI are operationalized as continuous variables, and an interaction term between these two variables is added to the model (Table V.20). The interaction term and the coefficient for BMI are the ones that are statistically significant ($p<0.05$). How big this effect is? Table V.21 has estimated hazard ratios for the 12 categories resulting from combining the 4 KH categories and the 3 obesity categories. In the first scenario, the mean values of KH and BMI observed in each of the 12 cells are evaluated into the hazards equation to estimate relative risks for KH categories. In the second scenario, extreme values for BMI and KH are evaluated into the equation. If the observed mean values are taken as reference, the hazard ratios are not very large, not even in the obese category, where the disadvantage for people with short KH is the largest. The hazard ratios are more sizable when the extreme values are used in the hazards equation. This means that for certain values of BMI (particularly among the most obese), people with short KH are more likely to develop DM, but this is not the case among people with normal weight. However, this scenario is very unlikely.

Additionally, it is also possible to test the effect of unmeasured heterogeneity on the estimated coefficients for the main explanatory variables. This analysis tests unmeasured heterogeneity effects by introducing a gamma frailty distribution to the models and observing whether this latent variable changes the magnitude of the estimates (Table V.22). The coefficients for the equations with KH dummy variables remain non-significant, although it is worth noticing that, after introducing the gamma frailty distribution, the pattern described by the

coefficients of the KH set of variables has more resemblance to the expected V-shape form. In the equation with the BMI-KH interaction, the sizes of the two coefficients increase in absolute value, suggesting that accounting for unmeasured heterogeneity might affect the shape of the association between KH and BMI.

Finally, I test the effect of differential underreporting in the estimates of CMI on the coefficients for lnCMI and for CMI<18. For the sake of consistency, the equations are estimated excluding one region of birth at a time so to test the effect of bias in the canton of birth's CMI's due to differential underreporting, as performed in the logistic regressions for DM prevalence. Differential underreporting does not seem to have any effect on the estimates for the coefficients of lnCMI (Figure V.13) or of CMI<18 (Figure V.14).

Even though the only equation that shows any significant effect of early childhood conditions on DM hazard rates is the one that evaluates the interaction between BMI and KH, both this model and the equation with the categorized KH are used to compute estimates for the projections of the following chapter. Hazard curves estimated from the parametric model are monotonically increasing if all the rest of the variables remain constant over time, and reflect the estimated relationships across KH and CMI categories (Figures I.15 and I.16). The monotonically increasing trend in the curves is achieved under the assumption that the values of all the rest of the covariates remain constant over time. However, the statistical analysis showed that some of the risk factors decrease by age and therefore, as it is going to be presented later, the actual curve used in the projections has a monotonically decreasing trend by age, especially after taking into account the association between age and BMI.

There is no graph for the model with the interaction because the graph needs several curves to account for the relationship at different levels of BMI. However, a final curve will be provided showing the final hazard estimates used in the projection.

8. The incidence probabilities of DM unawareness.

This research project operationalizes DM as having been diagnosed with the disease or having SG≥200 mg/dl or HbA_{1C}≥7% in one blood measurement. DM hazard was computed in the previous section from the information about age at diagnosis. However, to keep consistency with the operationalization of DM, it is necessary to compute also the incidence of non-diagnosed DM, or in this case, of unawareness of DM. According to Table V.23, there are no significant differences in the prevalence of DM unawareness across KH or CMI categories.

What are the characteristics of these people without a DM diagnosis but with high levels in the biomarkers? In what way do they resemble the diagnosed diabetic population? I use a logistic regression to determine what these differences are. The outcome variable is equal to one if the person has high levels in the biomarkers (SG≥200 mg/dl or HbA_{1C}≥7%) but not a DM diagnosis. The variable is equal to zero if the person has a DM diagnosis regardless of biomarker levels. This variable is named as DM unawareness.

According to the equation, unawareness is positively associated with being currently a smoker and it is negatively associated with parity, knowing about relatives that have had DM (DM family history), and having intense thirst and fatigue. These last two associations are of particular interest because they might mean that, among persons classified as diabetic,

asymptomatic people are –as expected– more likely to be unaware of their disease. On the contrary, the negative association with parity might reflect that children look after their elderly parents, and therefore they might be helping their parents to have more access to diagnostic services.

As explained before incidence probabilities are computed from the estimates drawn from a logistic regression on prevalence of DM unawareness and from the mortality model, and evaluate these estimates in formula (1) of this chapter.

In the logistic regression for prevalence of DM unawareness, the independent variables are the same that have been used throughout the analysis, plus a dichotomous variable that is equal to one if a respondent had their last glycemia exam more than 1 year ago, and 0 if the glycemia test was conducted within the last year. This question was asked only to the people that did not have a DM diagnosis at the time of the interview. Do Costa Rican elderly really know when their last glycemia was? Preventive service utilization in Costa Rica is relatively high among the population age 60 and above. According to CRELES, 75% of the people without a diagnosis reported to have a glycemia during the last 12 months. This figure can be compared to the official figures reported by the Health Insurance Institution in Costa Rica. According to this institution, during 2005, 61% of the elderly covered by the institution have had a list of several clinical examinations at their primary health centers during the last year (CCSS, 2006). A glycemia test is one of the examinations included in this list. The public health care system managed by this institution covers more than 90% of the elderly in the country. However, this figure published by CCSS can also be considered as a lower bound because it does not include clinical examinations performed in the private health care sector. Results of the

logistic regression for prevalence of DM unawareness are presented in Table V.24 (with CMI operationalized as a continuous variable) and Table V.25 (with CMI in categories).

The effect of KH seems to have a reverse V-shape but the coefficients for KH are not significantly different from zero. The effect that is large and significant is the coefficient for lnCMI, but the OR has the opposite direction. The OR is 0.57 ($p<0.10$), which means that the incidence probability of DM unawareness diminishes around 43% for each additional child death per 100 births. The coefficient is smaller and non-significant for the non-proxy sample (OR=3.52). The same conclusion can be drawn when CMI is operationalized as two dummy variables (Table V.25), although none of the odds ratios are significantly different to one.

Notice that the model is controlling for income, education, and current region of residence, therefore the strong effect of CMI is not related to differences in socio-economic conditions between the people that were born in cantones with low CMI and those born in cantones with high CMI (typically, places with poorer conditions, and with lags in educational development). The odds ratio for the foreign born is very far from one (OR=0.23) in the equation with the continuous CMI variable, but it is not so in the other equation. This is again reflecting the particular operationalization of CMI in this analysis. Neither of the coefficients for the foreign born variable is statistically significant from zero at a 0.10 level. The other variables with significant coefficients are: age, being a current alcoholic drinker (which seems to be a protective factor rather than a risk factor), and living in the Metropolitan Area, but only among the proxy sample. The coefficient for having the last glycemia test more than one year ago is significant at a 0.10 level and has the expected positive direction, which implies that people that have waited longer for having a glycemia test are more likely to be unaware of their condition.

The analysis of excluding one subregion of birth at a time is also performed as with the DM prevalence logistic regression model and the Weibull event history model. The coefficients are stable regardless of which subregion is excluded. The graphs are not shown this time. The next step is to estimate incidence rates using the formula. A way to summarize them is to present the estimated incidence curves: one for diagnosed DM and the other for total DM, including diagnosed DM and DM unawareness. Figure V.17 compares these curves with data observed from other countries. Notice that the curves tend to decrease as age increases, but the decrement is not too strong.

D. Chapter conclusions

The models for DM hazard via age at diagnosis do not show that there is any strong effect of early life adverse conditions on the development of DM, regardless of which marker of early life conditions is used: knee height (KH) or a child mortality index in the canton of birth of the respondent (CMI). The size of the relative risks is small compared to the size of the effects in the logistic regressions for DM prevalence. Recalling that under stability assumptions, prevalence comes from the product of hazards (or more accurately, incidence) and duration, it is not a surprise that the most important effects in this chapter's analyses are found in the mortality model. According to these models, people with DM have higher chances of dying than people without the disease, and individuals born in cantones with high CMI have shorter lives on average. On the contrary, Costa Rican elderly with short KH appear to live longer than their taller peers. This might be making the effect of KH on prevalence be larger than what was observed for the hazard rate.

The significant effects of the variables of early childhood conditions on mortality suggested that there was a strong selection effect if retrospective information is solely used to estimate DM hazard rates. However, after controlling for differential survival in the model for DM hazard rates, the size of the effects of KH and CMI on the DM event history model remains roughly unchanged.

The model that contained fairly strong early life condition effects is the model of DM unawareness prevalence. In this case, people born in high CMI cantones have a higher probability of having DM and being unaware of it. The model that compares this group to the diagnosed diabetic population suggests that these persons that are unaware of their high biomarker levels might actually have the disease, given that there are no significant differences in the main risk factors for DM between these two groups.

There are also characteristics of access to health care that are determining DM unawareness, particularly the duration since the last glucose screening test. However, the data also show that among Costa Rican seniors, preventive behaviors are common since 3 out of 4 people without a DM diagnosis have had a glucose screening examination during the last 12 months. This suggests that a high percentage of the persons that are unaware of their high SG or HbA_{1C} levels have developed this condition fairly recently.

Finally, an important result from this chapter is that the addition of either one of the markers of early childhood conditions does not modify the estimated coefficient for the other variable. In other words, the association between KH and CMI is not large enough to change their respective regression coefficients. This suggests that these variables are markers of different early conditions. Theoretically, KH is a marker of fetal and infant undernutrition. CMI could have been a marker of exposure to infectious diseases early in life. However, when this

hypothesis is tested using Finch and Crimmins's hypothesis of inflammatory exposure (Crimmins and Finch, 2006; Finch and Crimmins, 2004), the statistical associations match with Barker's "thrifty phenotype hypothesis" (Hales and Barker, 1992; 2001) in a more consistent way than with Finch and Crimmins's framework. This implies that CMI might also be a marker of early undernutrition. But, in conclusion, it is not clear why the correlation between the two variables is so small.

E. Tables and Figures

Table V. 1 Costa Rica: Prevalence of Type 2 DM among people age≥60, according to different definitions.

Diabetes Mellitus definition	Prevalence (%)
(n=2655)	
SG≥200 mg/dl or self-report	23.2
HbA_{1C}≥7% or self-report	24.7
SG≥200 mg/dl or HbA_{1C}≥7% or self-report	25.1
Only self-report	21.0

Table V. 2 Pearson correlations between SG and HbA1C by diagnosis condition and intake of medication for DM control in CRELES sample under analysis. Costa Rica, people age age≥60.

Diagnosis condition and medication intake	r	p-value
Not diagnosed and no medication	0.312	***
Diagnosed and no medication	0.766	***
Diagnosed and medication	0.494	***

Notes: *: p<.10; **:p<.05; ***: p<.01

Table V. 3 Costa Rica. Distribution of risk factors associated with diabetes mellitus (DM) prevalence. People age≥60.

Variables		Total	With DM	Without DM
(n)		2655		
Obesity	(Total)	(100)	(100)	(100)
%	Obese: BMI≥30 kg/m^2	26	36	22
	Overweight: 25≤BMI<29.9	42	40	42
	Missing BMI	3	4	3
	Normal: BMI<35 kg/m^2 (Ref. cat)	30	19	33
% Females		52	59	50
Age	(Total)	(100)	(100)	(100)
distrib.	Age 60-64 (Ref. cat)	30	30	31
%	Age 65-69	24	25	24
	Age 70-74	18	19	17
	Age 75-79	14	15	13
	Age 80-84	8	6	8
	Age 85-89	4	3	4
	Age 90 y +	2	2	2
% with schooling≥6 yrs.		49	47	50
% living in Metropolitan Area		53	52	53
% Retired		53	52	53

Table V. 3 Costa Rica. Distribution of risk factors associated with diabetes mellitus (DM) prevalence. People age≥60. (Continue)

Alcohol	(Total)			
		(100)	(100)	(100)
drinking	Current alcoholic drinker	33	27	35
distrib.	Past alcoholic drinker	32	32	31
%	Never drank (Ref. cat.)	36	40	34
Smoking	(Total)			
		(100)	(100)	(100)
distrib.	Current smoker	10	8	11
%	Past smoker	43	39	45
	Never smoked (Ref. cat.)	47	54	44
% doing physical activity currently		32	24	34
% ever worked		87	86	88
% with mean couple's income<50000 col.		42	45	41
% with known DM family history		40	55	34
% hospitalized during last 12 months		10	13	9
% intaking daily 400 g of carbohydrates		15	13	16
Parity	(Total)			
		(100)	(100)	(100)
(Only for	0 children	7	6	8
females)	1-2 children (Ref. cat.)	12	9	13
(n=1345)	3-5	31	28	33
	6-8	25	31	23
	9 or more	24	26	23
% with Proxy respondent		12.0	12.1	11.9

Table V. 4. Prevalence of DM (self-report or biomarkers), by risk factors. (Chi-square homogeneity tests).

Variables		% DM	χ^2 test	Variables		% DM	χ^2 test
Obesity	Obese: BMI≥30 kg/m^2	36	0.000	Smoking distrib.	Current smoker	27	0.050
	Overweight: 25≤BMI<29.9	24			Past smoker	23	
	Missing BMI	36			Never smoked (Ref. cat.)	19	
	Normal: BMI<35 kg/m^2 (Ref. cat)	16					
Sex	Males	22	0.001	Physical Activity	Not doing physical activity currently	28	0.000
	Females	28			Doing physical activity currently	19	
Age distrib.	Age 60-64 (Ref. cat)	25	0.242	Work History	Never worked	28	0.241
	Age 65-69	26			Ever worked	25	
	Age 70-74	27					
	Age 75-79	27					
	Age 80-84	20		Couple's Income	Mean couple's income>50000 col.	24	0.112
	Age 85-89	21			Mean couple's income<50000 col.	27	
	Age 90 y +	17					
Schooling	schooling<6 yrs.	26	0.351	Carb intake	Daily carb intake< 400g	26	0.192
	schooling≥6 yrs.	24			Daily carb intake>400 g	22	
Place of residence	living outside Metropolitan Area	25	0.699	DM family history	No or unknown	19	0.000
	living in Metropolitan Area	25			Known	35	
Retirement	Not retired	26	0.660	Hospitalized (12 months)	Not hospitalized (last 12 months)	24	0.009
	Retired	25			Hospitalized (last 12 months)	33	
Alcohol drinking distrib.	Current alcoholic drinker	28	0.011	Parity (Only for females)	0 children	23	0.046
	Past alcoholic drinker	25			1-2 children (Ref. cat.)	22	
	Never drank (Ref. cat.)	21			3-5	25	
					6-8	34	
					9 or more	31	

Table V. 5. Prevalence of DM prevalence (self-report or biomarkers) and mean SG and HbA_{1C} levels, by main explanatory variables: knee height categories (by quartiles) and Child Mortality Index of canton of birth CMI (categorized). Chi-square homogeneity tests for differences in prevalence, and ANOVA F-test for differences in SG and HbA_{1C} means.

Variables		Preval. (%)	χ^2 p-value	Mean SG (mg/dl)	F p-value	Mean HbA_{1C} (%)	F p-value
	(n)						
Knee Height (KH) [1/]	KH≤Q1	24	0.396	109.9	0.888	5.72	0.956
	Q1<KH≤Q2	27		110.0		5.75	
	Q2<KH≤Q3	23		109.7		5.76	
	KH>Q3	26		112.1		5.75	
Canton of birth Child Mortality Index (CMI)	CMI<18.2	26	0.470	107.3	0.226	5.74	0.751
	18.2≤CMI<22.6	27		111.4		5.77	
	22.6≤CMI<26.6	27		112.6		5.77	
	26.6≤CMI<31.0	23		112.5		5.68	
	CMI≥31.0	22		110.8		5.78	
	Foreign-born	21		104.2		5.70	

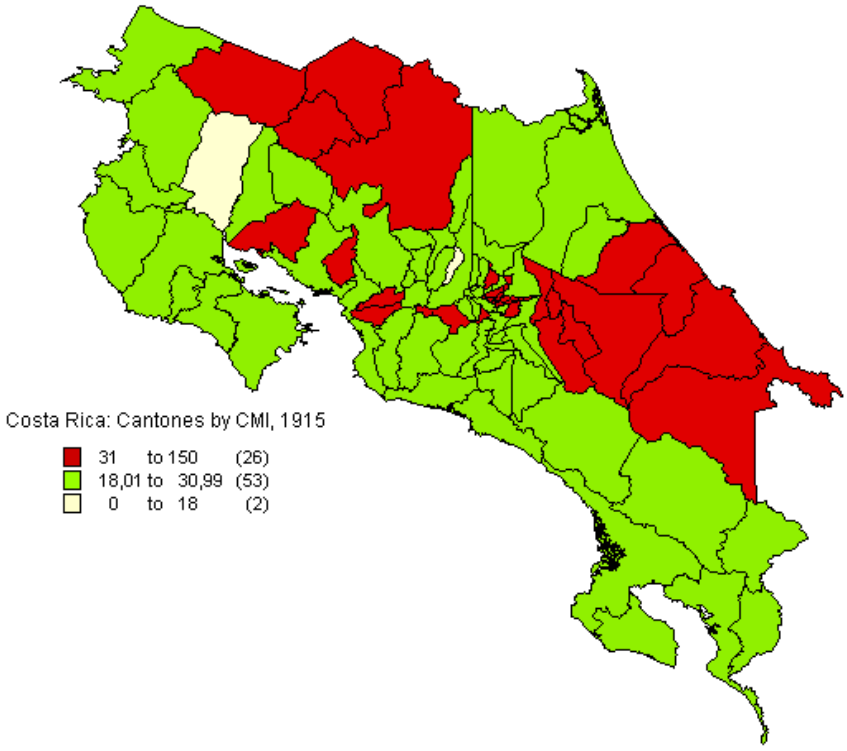
Notes: [1/] For men's knee height, Q1=49.6 cm, Q2=51.3 cm, Q3=53.0cm.
For women, Q1=45.6 cm, Q2=47.2 cm, Q3=48.7

Table V. 6. Costa Rica: Child death underreporting proportion relative to birth underreporting, estimated from census information, by subregion, 1927 and 1950.

Subregion	Relative child death underreporting	
	1927	1950
Los Santos+Perez Zeledón	55.6	1.7
San Carlos	48.2	9.6
Northern Guanacaste	45.7	-4.0
South West Alajuela	44.0	1.2
San Jose	39.0	2.5
South Zone	38.5	-1.8
Guapiles-Sarapiqui	37.8	29.5
Nicoya+Puntarenas	36.2	-0.6
Heredia	24.9	4.9
Alajuela (Central)	19.0	-1.9
Western San Jose	11.1	3.6
Eastern Cartago	-1.9	11.3
Limon	-2.3	2.6
Western Cartago	-8.1	8.0

Figure V. 1 Costa Rica: Child Mortality Index (CMI) by canton: 1915, 1925, 1935 and 1945.

1915



1925

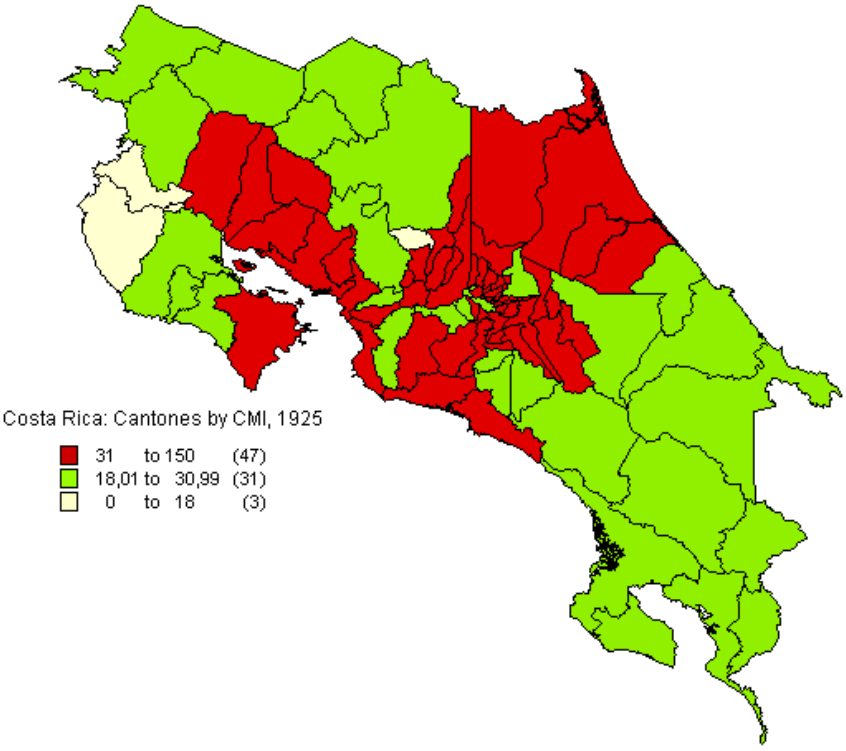
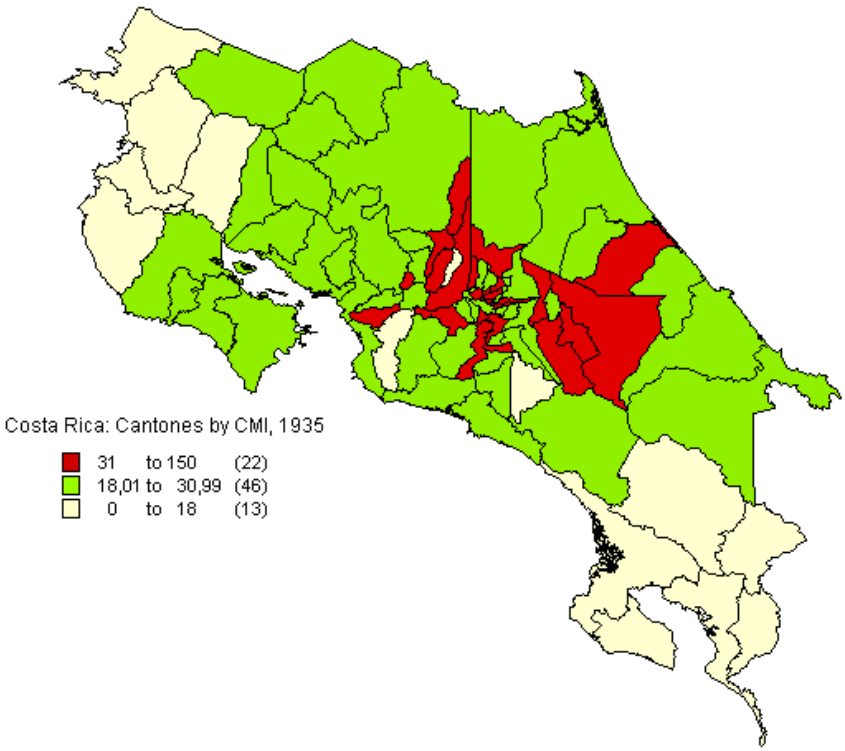


Figure V.1. Costa Rica: Child Mortality Index (CMI) by canton: 1915, 1925, 1935 and 1945 (Continue).

1935



1945

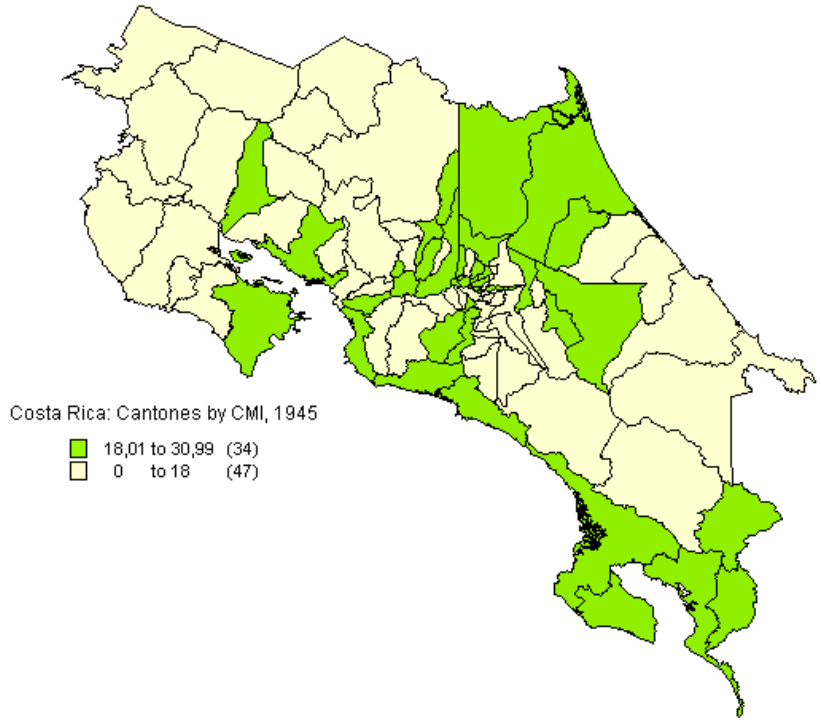


Table V. 7 Coefficients for dichotomous variable indicating knee height<first quartile, in Costa Rica, Puerto Rico, or knee height<first quintil in six cities of the SABE project, controlling by other risk factors. [3/]

Variables	OR	Coeff	SE	
Costa Rica				
--Only self-report	1.35	0.298	0.20	
--Self-report or SG ≥200 mg/dl or HbA1C≥7%	1.23	0.209	0.18	
Puerto Rico [1/]	1.23	0.204	0.09	**
SABE Project Cities [2/]				
--Bridgetown, Barbados	0.95			
--La Habana, Cuba	0.88			
--Mexico City, Mexico	1.51			**
--Montevideo, Uruguay	1.02			
--Santiago, Chile	1.67			**
--Sao Paulo, Brazil	0.92			

Notes: [1/] Palloni, McEniry, Guend, Dávila *et al.* (2005)
[2/] Palloni, McEniry, Guend, Dávila *et al.* (2006)
[3/] All regressions control for sex, age groups, years of schooling, bad health before age 10, socio-economic status before age 10.
*: p<.10; **:p<.05; ***: p<.01

Table V. 8. Costa Rica: Elderly population in 2004-2006. Coefficients and standard errors (SE) for logistic regressions of DM prevalence (self-report or biomarkers).

Variables	Without CMI effects			With CMI effects		
	Coeff	SE		Coeff	SE	
Knee Height≤Q1 (Ref: Q2<KH≤Q3)	0.097	0.170		0.111	0.170	
Q1<Knee Height≤Q2	0.218	0.160		0.228	0.160	
Knee Height>Q3	0.156	0.162		0.163	0.162	
Ln of Child Mortality Index (lnCMI)	-	-		-0.170	0.196	
Foreign born	-	-		-0.633	0.684	
Obese: BMI≥30 kg/m^2	0.904	0.169	***	0.910	0.168	***
Overweight: 25≤BMI<29.9	0.414	0.151	***	0.416	0.151	***
Missing BMI	0.918	0.274	***	0.923	0.274	***
Female	-0.133	0.234		-0.133	0.234	
Age	-0.007	0.007		-0.005	0.008	
Schooling≥6 yrs.	-0.064	0.134		-0.065	0.134	
Living in Metropolitan Area	-0.020	0.122		-0.016	0.122	
Retired	0.057	0.124		0.050	0.125	
Current alcoholic drinker	-0.178	0.171		-0.181	0.171	
Past alcoholic drinker	0.083	0.169		0.084	0.169	
Current smoker	0.021	0.235		0.031	0.234	
Past smoker	-0.094	0.146		-0.092	0.146	
Current physical activity	-0.424	0.143	***	-0.424	0.142	***
Ever worked	0.026	0.167		0.026	0.167	
Mean couple's income<50000 col.	0.166	0.126		0.164	0.126	
Known DM family history	0.845	0.118	***	0.843	0.118	***
Hospitalized during last 12 months	0.325	0.178	*	0.319	0.179	*
Daily carb intake > 400g	-0.072	0.168		-0.077	0.168	
(Interaction with Female)						
Parity	0.108	0.065	*	0.107	0.065	
Constant	-1.550	0.634	**	-1.194	0.761	
(n)	(2655)			(2655)		

Notes: *: p<.10; **:p<.05; ***: p<.01

Table V. 9. Costa Rica: Elderly population in 2004-2006. Coefficients and standard errors (SE) of main explanatory variables for different identifications of logistic regressions of DM prevalence (self-report or FSG ≥129 mg/dl).

Variables	Only w/knee height		Only lnCMI		W/ knee height and lnCMI	
	Coeff (SE)	OR	Coeff (SE)	OR	Coeff (SE)	OR
Knee Height≤Q1	0.097	1.10			0.111	1.12
	(0.170)				(0.170)	
(Ref: Q2<KH≤Q3)			-	-		
Q1<Knee Height≤Q2	0.218	1.24	-	-	0.228	1.26
	(0.160)				(0.160)	
Knee Height>Q3	0.156	1.17	-	-	0.163	1.18
	(0.162)				(0.162)	
Ln of Child Mortality Index (lnCMI)	-	-	-0.152	0.86	-0.170	0.84
			(0.195)		(0.196)	
Foreign born			-0.571	0.57	-0.633	0.53
			(0.681)		(0.684)	

Notes: *: p<.10; **:p<.05; ***: p<.01

Table V. 10. Association between health variables and CMI in continuous (lnCMI) and categorized operationalization. For lnCMI: Odds ratios per 1-unit increase in lnCMI for disease prevalence, regression coefficient for biomarkers, and Gompertz regression hazard ratio for death rate. For categorized CMI: Odds ratios and regression coefficients for dummy variables.

Variables	Sample size	Model with lnCMI †		Model with categorized CMI†		
				CMI ≤18.0 deaths per birth		CMI≥31.0 deaths per birth
(Linear regression coeff)						
ln of C-reactive protein ††	2487	0.092		-0.100		0.006
ln of SG						
-not including CRP in model	2552	0.055	**	-0.050	**	0.011
-including CRP in model	2487	0.040	**	-0.047	**	0.016
ln of HbA1C						
-not including CRP in model	2502	0.011		-0.012		0.011
-including CRP in model	2442	0.011		-0.013		0.011
(Odds ratios)						
C-reactive protein≥10 mg/L	2487	1.39		0.61	**	0.77
SG≥200 mg/dL						
-not including CRP in model	2552	2.06		0.74		1.56
-including CRP in model	2487	2.21		0.71		1.63
HbA1C≥7%						
-not including CRP in model	2502	0.90		1.00		1.29
-including CRP in model	2442	0.97		0.94		1.34

† Confounding variables include: sex, age, living in Metropolitan Area, obesity, overweight, education (6 years of schooling or more), low income, history of smoking, history of alcohol intake, and nutritional variables (More than 3000 daily calories, more than 40 g of daily saturated fat, more than 400 g of daily carbohydrates).
†† Adjusted also by ratio of HDL/total cholesterol and arthritis in the second equation.
Notes: *: p<.10; **: p<.05; **: p<.01

Table V. 11 Costa Rica: Elderly population in 2004-2006. Coefficients and standard errors (SE) of logistic regressions of DM prevalence (self-report or biomarker) in samples with and without proxy, and effect of perceived weight change.

	Total sample			Non-proxy sample			Non-proxy sample, w/ weight change var		
Variables	Coeff	SE		Coeff	SE		Coeff	SE	
Knee Height≤Q1									
(Ref: Q2<KH≤Q3)	0.111	0.170		0.187	0.188		0.167	0.188	
Q1<Knee Height≤Q2	0.228	0.160		0.200	0.178		0.194	0.179	
Knee Height>Q3	0.163	0.162		0.132	0.175		0.117	0.174	
Ln of Child Mortality Index (lnCMI)	-0.170	0.196		-0.292	0.220		-0.283	0.220	
Foreign born	-0.633	0.684		-1.025	0.766		-1.007	0.767	
Obese: BMI≥30 kg/m2	0.910	0.168	***	0.876	0.185	***	0.995	0.209	***
Overweight: 25≤BMI<29.9	0.416	0.151	***	0.399	0.167	**	0.468	0.177	***
Missing BMI	0.923	0.274	***	1.357	0.505	***	1.402	0.513	***
Female	-0.133	0.234		-0.184	0.258		-0.192	0.258	
Age	-0.005	0.008		0.000	0.010		0.000	0.010	
Schooling≥6 yrs.	-0.065	0.134		-0.060	0.144		-0.041	0.145	
Living in Metropolitan Area	-0.016	0.122		-0.044	0.135		-0.054	0.135	
Retired	0.050	0.125		0.106	0.137		0.097	0.137	
Current alcoholic drinker	-0.181	0.171		-0.208	0.185		-0.217	0.185	
Past alcoholic drinker	0.084	0.169		0.097	0.187		0.097	0.188	
Current smoker	0.031	0.234		-0.011	0.254		-0.028	0.256	
Past smoker	-0.092	0.146		-0.086	0.162		-0.089	0.163	
Current physical activity	-0.424	0.142	***	-0.394	0.146	***	-0.393	0.146	***
Ever worked	0.026	0.167		-0.040	0.199		-0.020	0.199	
Mean couple's income<50000 col.	0.164	0.126		0.213	0.139		0.191	0.140	
Known DM family history	0.843	0.118	***	0.881	0.130	***	0.884	0.130	***
Hospitalized during last 12 months	0.319	0.179	*	0.252	0.208		0.239	0.207	
Daily carb intake > 400g	-0.077	0.168		-0.183	0.183		-0.173	0.183	
(Interaction with Female)									
Parity	0.107	0.065		0.107	0.074		0.105	0.074	
Scale perceived change of current weight-weigth at 25	-	-		-	-		-0.042	0.035	
Missing perceived current weight	-	-		-	-		0.115	0.396	
Missing perceived weigth at 25	-	-		-	-		0.384	0.375	
Constant	-1.194	0.761		-1.096	0.886		-1.124	0.886	
(n)	(2655)			(2014)			(2014)		

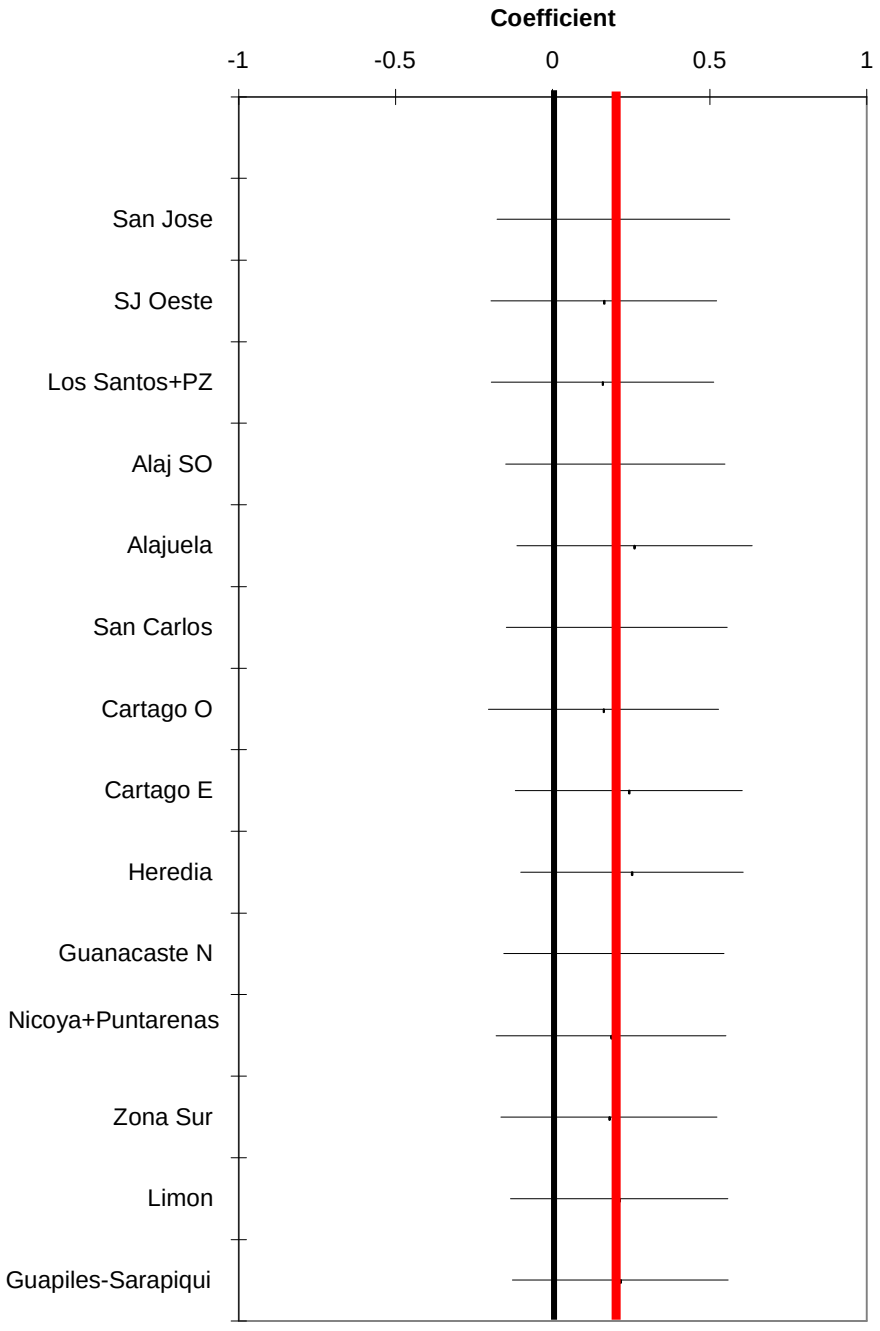
Notes: *: p<.10; **:p<.05; ***: p<.01

Table V. 12. Costa Rica: Elderly population in 2004-2006. Coefficients and standard errors (SE) of logistic regressions of DM prevalence (self-report or biomarker) in samples with and withoug proxy, and effect of perceived weight change, with CMI categorized.

	Total sample			Non-proxy sample			Non-proxy sample, w/ weight change var		
Variables	Coeff	SE		Coeff	SE		Coeff	SE	
Knee Height≤Q1									
(Ref: Q2<KH≤Q3)	0.101	0.170		0.173	0.188		0.154	0.188	
Q1<Knee Height≤Q2	0.221	0.160		0.192	0.179		0.186	0.179	
Knee Height>Q3	0.164	0.162		0.133	0.175		0.118	0.175	
CMI≤18	-0.065	0.159		-0.014	0.169		-0.021	0.169	
CMI≥31	-0.191	0.169		-0.223	0.202		-0.228	0.202	
Foreign born	-0.144	0.292		-0.149	0.329		-0.164	0.326	
Obese: BMI≥30 kg/m2	0.908	0.169	***	0.872	0.185	***	0.996	0.209	***
Overweight: 25≤BMI<29.9	0.414	0.152	***	0.395	0.167	**	0.467	0.178	***
Missing BMI	0.925	0.273	***	1.358	0.498	***	1.404	0.506	***
Female	-0.130	0.234		-0.181	0.259		-0.189	0.258	
Age	-0.005	0.008		-0.002	0.010		-0.002	0.010	
Schooling≥6 yrs.		0.134		-0.060	0.144		-0.041	0.145	
Living in Metropolitan Area	-0.023	0.122		-0.050	0.135		-0.059	0.135	
Retired	0.052	0.125		0.106	0.138		0.097	0.138	
Current alcoholic drinker	-0.183	0.171		-0.206	0.185		-0.216	0.185	
Past alcoholic drinker	0.086	0.168		0.100	0.187		0.101	0.187	
Current smoker	0.024	0.234		-0.020	0.255		-0.037	0.256	
Past smoker	-0.091	0.146		-0.084	0.161		-0.087	0.162	
Current physical activity	-0.425	0.142	***	-0.398	0.146	***	-0.397	0.146	***
Ever worked	0.032	0.167		-0.040	0.198		-0.020	0.199	
Mean couple's income<50000 col.	0.161	0.126		0.212	0.139		0.190	0.140	
Known DM family history	0.842	0.118	***	0.878	0.130	***	0.881	0.130	***
Hospitalized during last 12 months	0.317	0.179	*	0.254	0.209		0.241	0.208	
Daily carb intake > 400g	-0.075	0.168		-0.177	0.183		-0.167	0.183	
(Interaction with Female)									
Parity	0.107	0.065		0.107	0.074		0.104	0.074	
Scale perceived change of current weight-weigth at 25				-	-		-0.043	0.035	
Missing perceived current weight				-	-		0.126	0.393	
Missing perceived weigth at 25				-	-		0.376	0.376	
Constant	-1.639	0.657	**	-1.828	0.780	**	-1.831	0.777	**
(n)	(2655)			(2014)			(2014)		

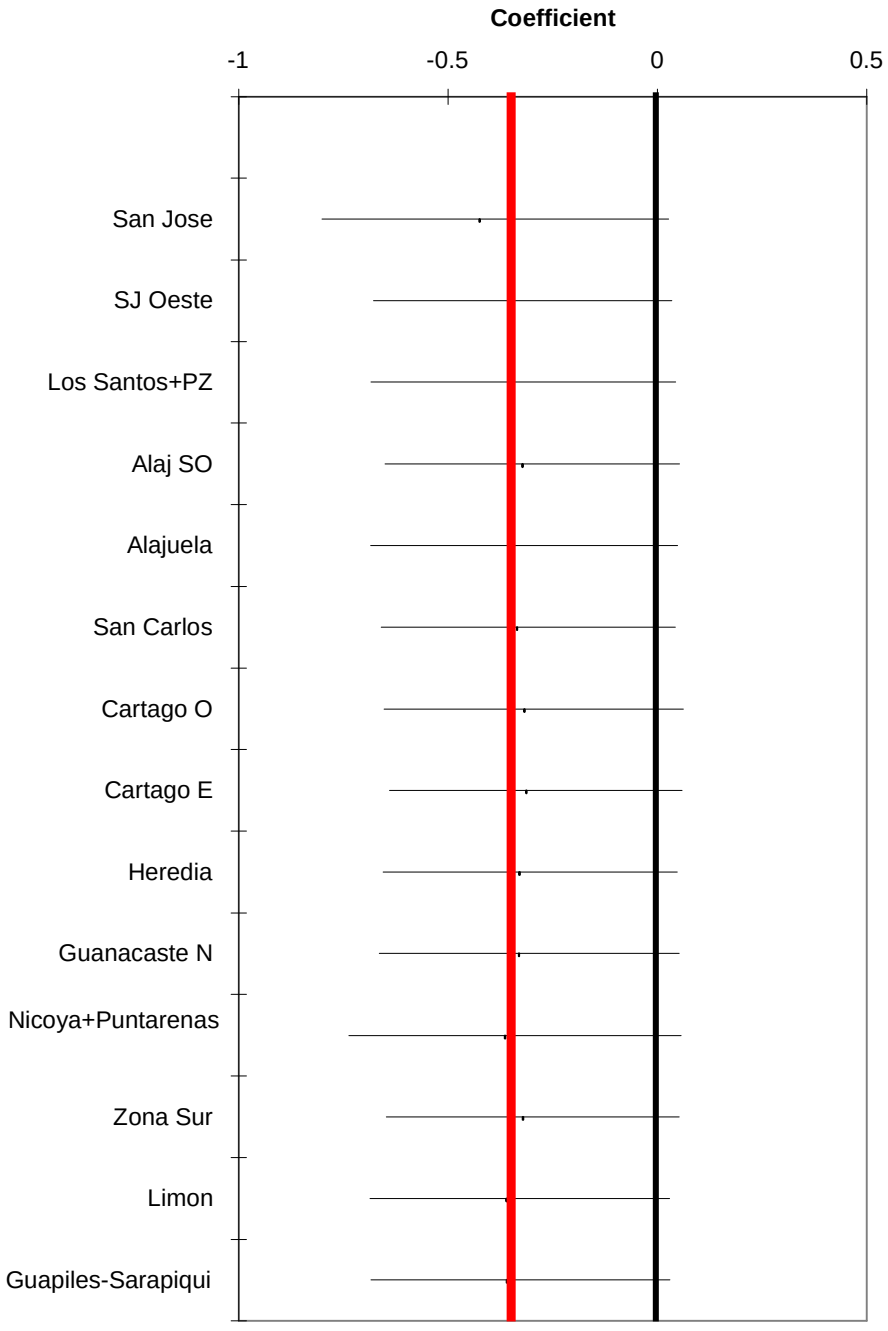
Notes: *: p<.10; **:p<.05; ***: p<.01

Figure V. 2. Point and interval estimates for coefficient of KH<Q1, in logistic regression of DM prevalence, excluding cases from Costa Rica's subregions one at a time.



Note: Red line denotes the coefficient estimated with all regions.

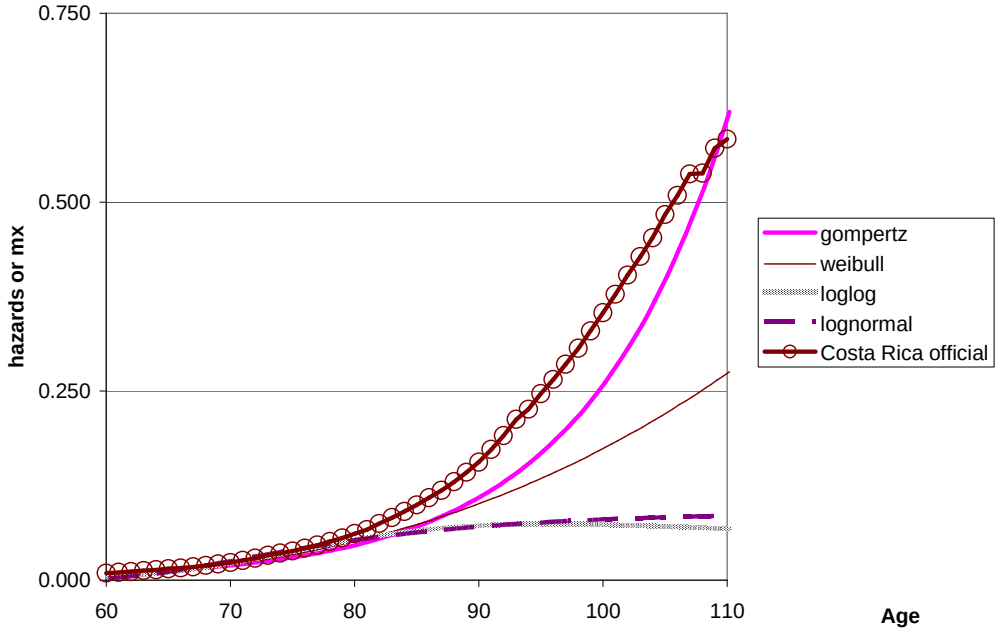
Figure V. 3. Point and interval estimates for coefficient of CMI≤18 deaths per 100 births, in logistic regression of DM prevalence, excluding cases from Costa Rica's subregions one at a time.



Note: Red line denotes the coefficient estimated with all regions.

Figure V. 4. Comparison between death rates in official Costa Rica Life Table 1995-2005, and hazard estimates using no covariates and several parametric distributions. In arithmetic and logarithmic scale.

Arithmetic scale



Logarithmic scale

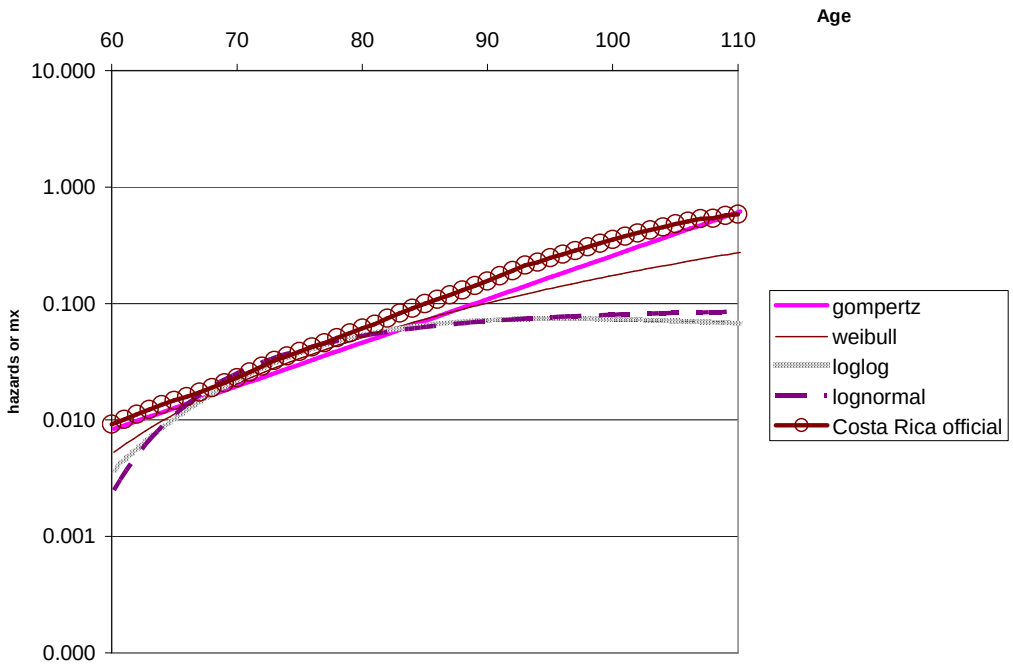


Figure V. 5. Comparison between death rates in official Costa Rica Life Table 1995-2005, and hazard estimates using no covariates and Cox non-proportional hazards regression.

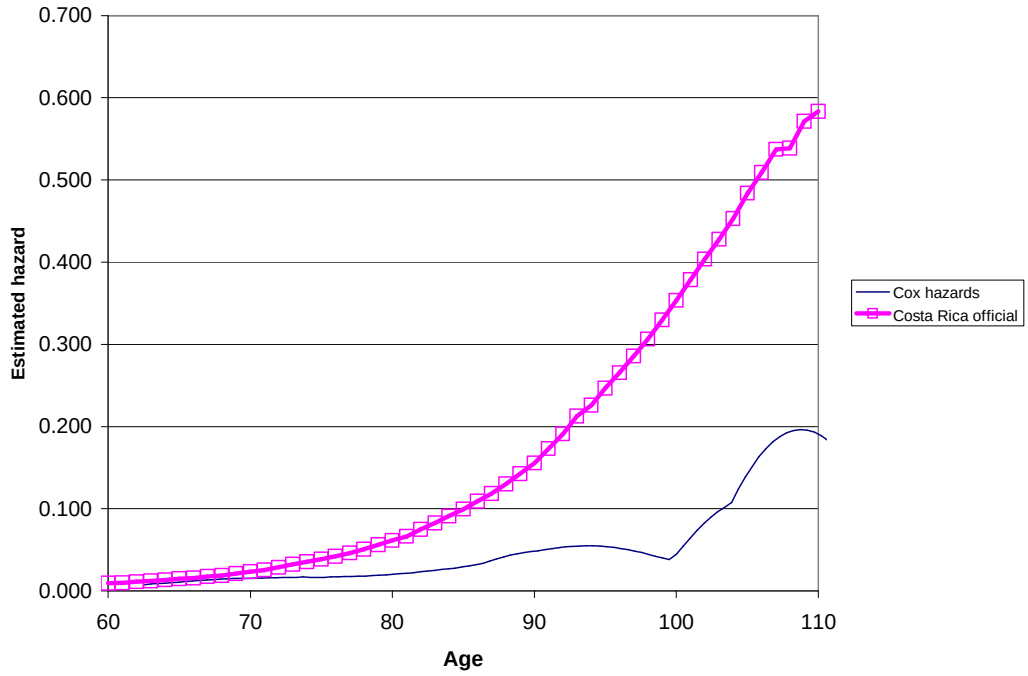


Table V. 13. Costa Rica: Elderly population in 2004-2006. Relative risks from Gompertz parametric survival regression on mortality.

	Total sample			Non-proxy sample, w/ weight change var		
Variables	RRR	(Coeff SE)		RRR	(CoeffSE)	
Diabetes	1.44	0.33		1.44	0.33	
Knee Height≤Q1 (Ref: Q2<KH≤Q3)	0.54	0.14	**	0.52	0.13	***
Q1<Knee Height≤Q2	0.71	0.18		0.70	0.18	
Knee Height>Q3	0.68	0.20		0.68	0.20	
Ln of Child Mortality Index (lnCMI)	2.73	0.87	***	2.68	0.81	***
Foreign born	25.47	28.38	***	23.39	25.24	***
Obese: BMI≥30 kg/m2	0.95	0.30		1.11	0.33	
Overweight: 25≤BMI<29.9	0.54	0.14	**	0.58	0.14	**
Missing BMI	2.58	0.73	***	2.19	0.61	***
Female	0.45	0.18	**	0.49	0.19	*
Schooling≥6 yrs.	1.21	0.25		1.30	0.27	
Living in Metropolitan Area	1.23	0.26		1.23	0.26	
Retired	1.99	0.49	***	2.00	0.49	***
Current alcoholic drinker	0.64	0.23		0.67	0.24	
Past alcoholic drinker	1.05	0.32		1.09	0.33	
Current smoker	1.12	0.44		1.10	0.44	
Past smoker	0.74	0.16		0.74	0.16	
Current physical activity	0.64	0.20		0.69	0.21	
Ever worked	1.04	0.33		1.12	0.35	
Mean couple's income<50000 col.	1.38	0.27	*	1.30	0.25	
Known DM family history	0.89	0.21		0.89	0.21	
Hospitalized during last 12 months	1.39	0.33		1.31	0.31	
Daily carb intake > 400g (Interaction with Female)	0.76	0.22		0.76	0.22	
Parity	1.13	0.12		1.10	0.12	
Scale perceived change of current weight-weigth at 25				0.96	0.08	
Missing perceived current weight				0.87	0.30	
Missing perceived weigth at 25				1.87	0.40	***
Parameter gamma	0.05	0.01	***	0.04	0.01	***
(n)	(2640)			(2640)		

Notes: *: p<.10; **:p<.05; ***: p<.01

Table V. 14. Costa Rica: Elderly population in 2004-2006. Relative risks from Gompertz parametric survival regression on mortality, using categorized CMI

	Total sample			Non-proxy sample, w/ weight change var		
Variables	RRR	(Coeff SE)		RRR	(CoeffSE)	
Diabetes	1.40	0.33		1.40	0.32	
Knee Height≤Q1						
(Ref: Q2<KH≤Q3)	0.55	0.14	**	0.54	0.14	**
Q1<Knee Height≤Q2	0.72	0.19		0.71	0.18	
Knee Height>Q3	0.68	0.20		0.68	0.20	
CMI≤18	0.57	0.24		0.55	0.24	
CMI≥31	1.37	0.25	*	1.34	0.24	
Foreign born	0.99	0.46		0.96	0.45	
Obese: BMI≥30 kg/m2	0.97	0.31		1.13	0.34	
Overweight: 25≤BMI<29.9	0.53	0.14	**	0.58	0.15	**
Missing BMI	2.53	0.72	***	2.13	0.60	***
Female	0.45	0.18	*	0.49	0.20	*
Schooling≥6 yrs.	1.21	0.25		1.30	0.27	
Living in Metropolitan Area	1.26	0.27		1.25	0.26	
Retired	2.00	0.49	***	2.01	0.49	***
Current alcoholic drinker	0.65	0.23		0.68	0.24	
Past alcoholic drinker	1.05	0.32		1.09	0.33	
Current smoker	1.13	0.45		1.11	0.44	
Past smoker	0.74	0.16		0.73	0.16	
Current physical activity	0.64	0.20		0.69	0.21	
Ever worked	1.03	0.33		1.11	0.35	
Mean couple's income<50000 col.	1.36	0.26		1.29	0.25	
Known DM family history	0.89	0.21		0.89	0.21	
Hospitalized during last 12 months	1.37	0.32		1.30	0.31	
Daily carb intake > 400g	0.74	0.22		0.74	0.22	
(Interaction with Female)						
Parity	1.13	0.12		1.11	0.12	
Scale perceived change of current weight-weigth at 25				0.96	0.08	
Missing perceived current weight				0.84	0.29	
Missing perceived weigth at 25				1.89	0.41	***
Parameter gamma	0.05	0.01	***	0.04	0.01	***
(n)	(2640)			(2640)		

Notes: *: p<.10; **:p<.05; ***: p<.01

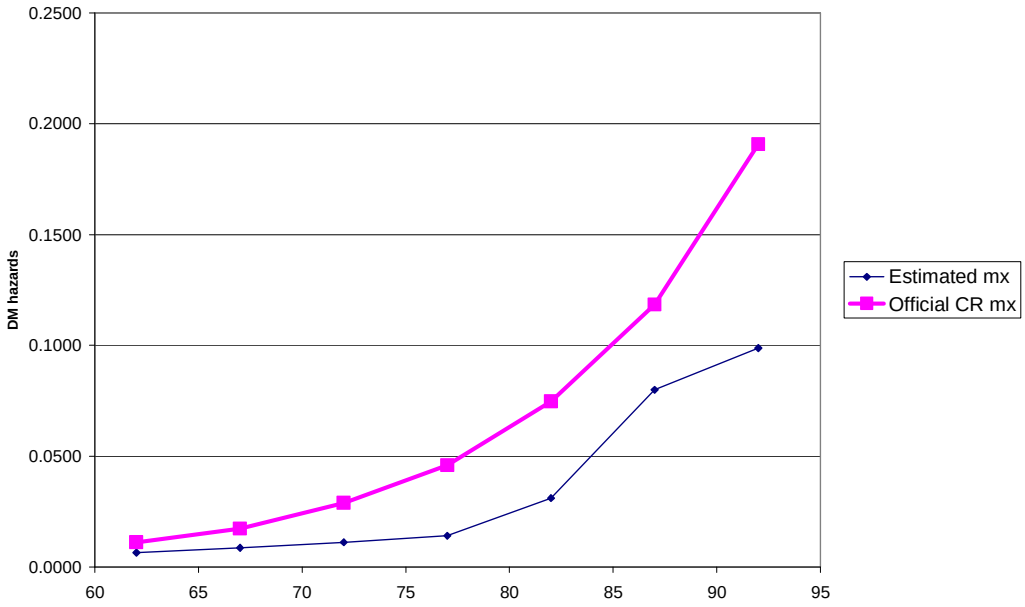
Table V. 15. Costa Rica: Elderly population in 2004-2006. Relative risks from Gompertz parametric survival regression on mortality. Models with interactions.

Variables	Total sample RRR	(Coeff SE)		Non-proxy sample, w/ weight change var RRR	(CoeffSE)	
Model with continuous CMI						
Diabetes	0.44	0.54		0.50	0.57	
Knee Height≤Q1						
(Ref: Q2<KH≤Q3)	0.56	0.15	**	0.55	0.15	**
Q1<Knee Height≤Q2	0.60	0.18	*	0.61	0.18	*
Knee Height>Q3	0.58	0.20		0.60	0.20	
(Interaction with Diabetes)						
Knee Height≤Q1						
(Ref: Q2<KH≤Q3)	0.74	0.49		0.74	0.49	
Q1<Knee Height≤Q2	1.68	1.05		1.57	0.99	
Knee Height>Q3	1.48	0.99		1.38	0.92	
Ln of Child Mortality Index (lnCMI)	2.44	0.68	***	2.41	0.65	***
Interaction Diabetes with lnCMI	1.37	0.37		1.33	0.33	
Foreign born	21.42	21.70	***	19.90	19.67	***
Model with CMI categories						
Diabetes	1.14	0.54		1.20	0.57	
Knee Height≤Q1						
(Ref: Q2<KH≤Q3)	0.56	0.15	**	0.55	0.15	**
Q1<Knee Height≤Q2	0.60	0.18	*	0.61	0.18	*
Knee Height>Q3	0.59	0.20		0.61	0.21	
(Interaction with Diabetes)						
Knee Height≤Q1						
(Ref: Q2<KH≤Q3)	0.84	0.54		0.84	0.55	
Q1<Knee Height≤Q2	1.84	1.09		1.74	1.05	
Knee Height>Q3	1.58	1.03		1.51	0.99	
CMI≤18	0.52	0.23		0.51	0.22	
CMI≥31	1.49	0.31	*	1.49	0.31	*
Interaction with Diabetes						
CMI≤18	1.40	1.30		1.38	1.28	
CMI≥31	0.70	0.30		0.65	0.28	

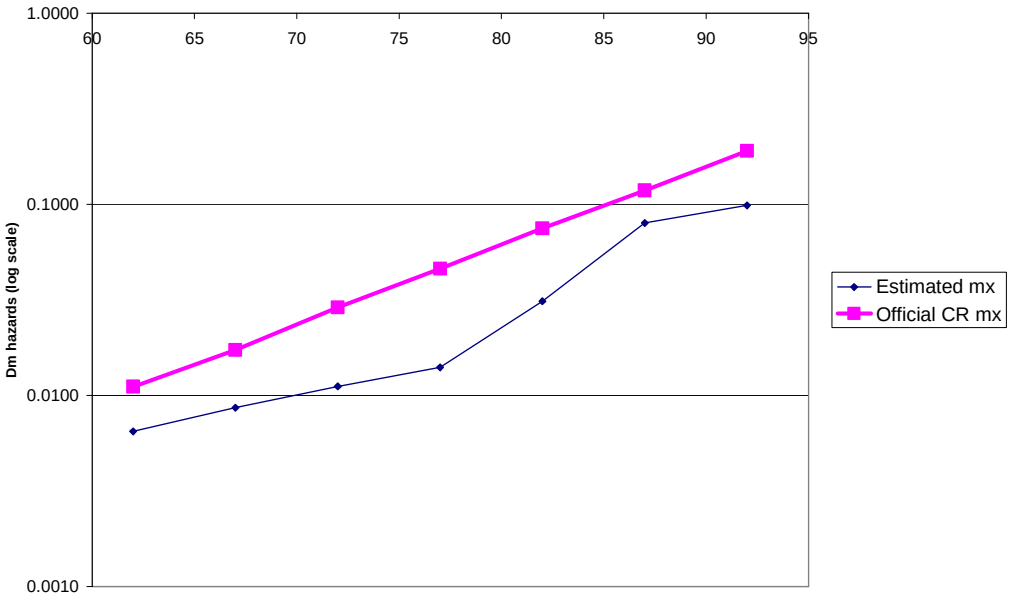
Notes: *: p<.10; **:p<.05; ***: p<.01

Figure V. 6. Comparison between death rates in official Costa Rica Life Table 1995-2005, and hazard estimates using final Gompertz survival regression. In arithmetic and logarithmic scale.

Arithmetic scale



Gompertz scale



Source: Official Costa Rican death rates from CCP-INEC (2002).

Figure V. 7. Estimated hazards of DM by KH, using Kaplan-Meier estimates.

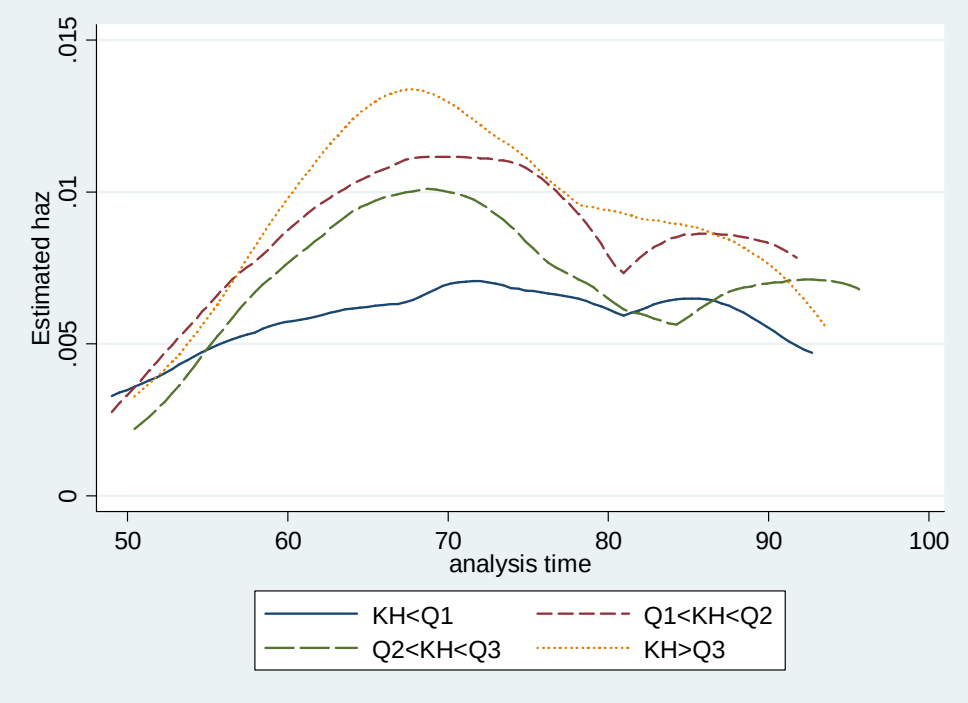


Figure V. 8. Estimated hazards of DM by CMI categories, using Kaplan-Meier estimates.

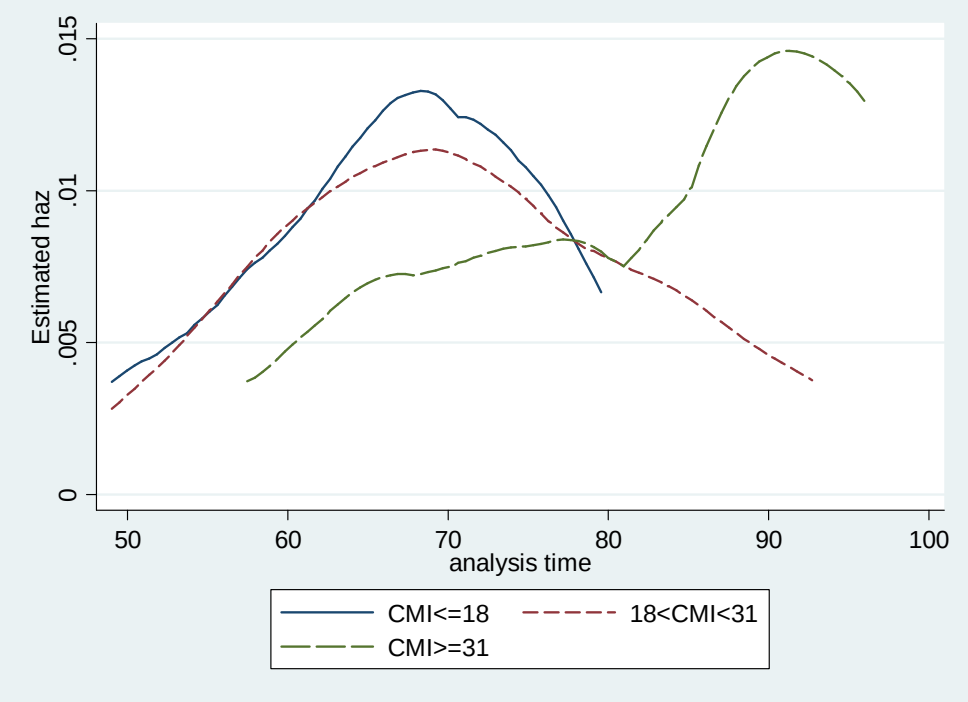


Figure V. 9. Estimated hazards of DM by KH, using Kaplan-Meier estimates, adjusted for differential survival probabilities.

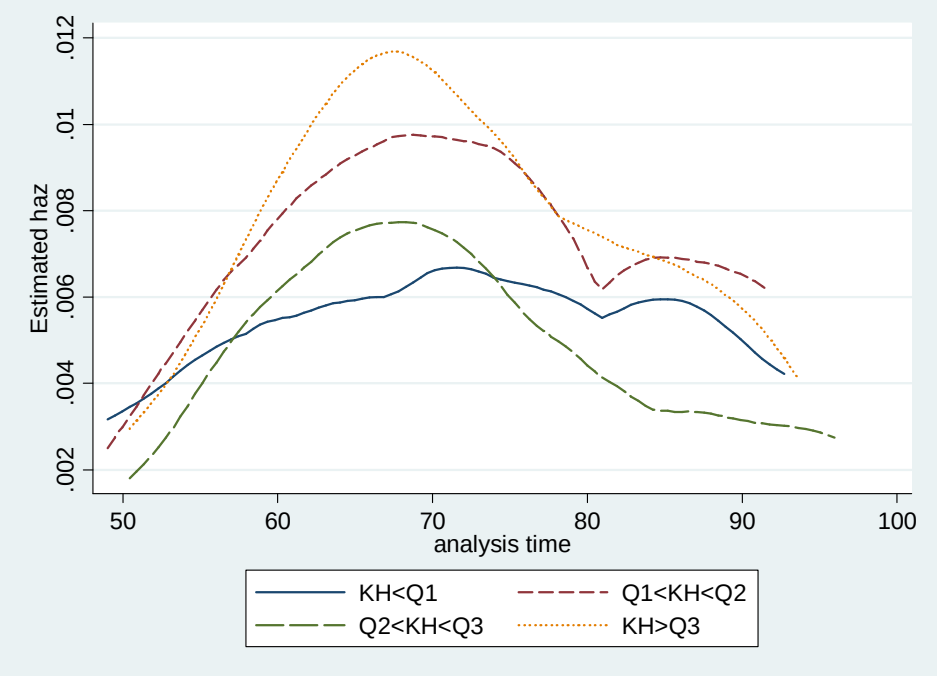


Figure V. 10. Estimated hazards of DM by CMI categories, using Kaplan-Meier estimates, adjusted for differential survival probabilities.

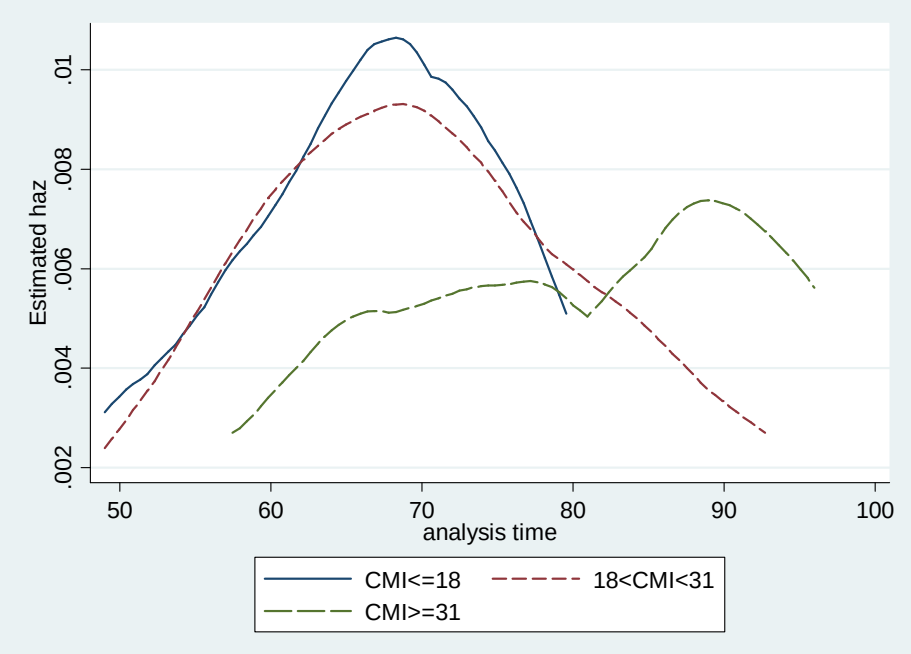
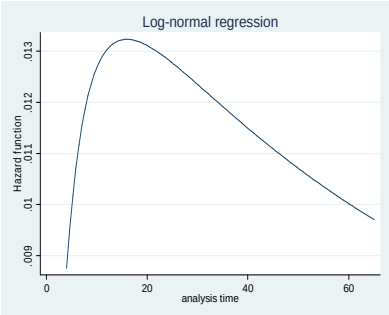


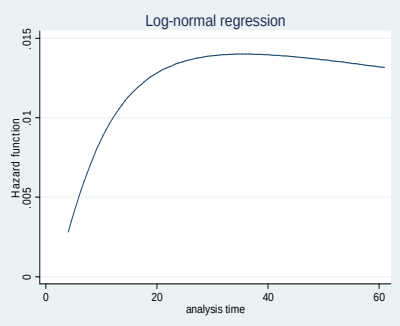
Table V. 16. Information criteria to select failure distribution for parametric survival regression models of DM hazard, using retrospective information.

Failure distribution	AIC	BIC	Log-likelihood
Weibull	1892.58	2047.70	-919.29
Exponential	1898.61	2047.98	-923.30
Log-normal	1918.44	2073.55	-932.22
Log-log	1920.62	2075.73	-933.31
Gompertz	1931.83	2086.94	-938.91

Figure V. 11. Estimated curves from log-normal event history models. Model 1 is modeled without covariates and model 2 with covariates. (Analysis time is counted from age 45 on).

Model 1





Model 2

Table V. 17. Costa Rica: Elderly population in 2004-2006. Relative risks from Weibull parametric survival regression on self-reported previous diagnoses of DM.

	Total sample		Non-proxy sample, w/ weight change var	
Variables	RRR	(CoeffSE)	RRR	(CoeffSE)
Continuous CMI				
Without mortality correction				
Knee Height≤Q1				
(Ref: Q2<KH≤Q3)	1.20	0.24	1.20	0.24
Q1<Knee Height≤Q2	1.31	0.22	1.31	0.22
Knee Height>Q3	1.16	0.20	1.16	0.20
Ln of Child Mortality Index (lnCMI)	0.83	0.17	0.83	0.16
Foreign born	0.37	0.26	0.37	0.26
With mortality correction				
Knee Height≤Q1				
(Ref: Q2<KH≤Q3)	1.13	0.22	1.12	0.22
Q1<Knee Height≤Q2	1.26	0.22	1.26	0.22
Knee Height>Q3	1.12	0.19	1.12	0.19
Ln of Child Mortality Index (lnCMI)	0.91	0.18	0.91	0.18
Foreign born	0.49	0.34	0.50	0.34
Categorical CMI				
Without mortality correction				
Knee Height≤Q1				
(Ref: Q2<KH≤Q3)	1.19	0.23	1.19	0.23
Q1<Knee Height≤Q2	1.30	0.22	1.30	0.22
Knee Height>Q3	1.17	0.20	1.17	0.20
CMI≤18	0.97	0.16	0.96	0.16
CMI≥31	0.84	0.15	0.84	0.15
Foreign born	0.63	0.20	0.63	0.20
With mortality correction				
Knee Height≤Q1				
(Ref: Q2<KH≤Q3)	1.12	0.22	1.12	0.22
Q1<Knee Height≤Q2	1.25	0.22	1.25	0.22
Knee Height>Q3	1.12	0.19	1.12	0.19
CMI≤18	0.93	0.16	0.93	0.16
CMI≥31	0.89	0.16	0.89	0.16
Foreign born	0.63	0.20	0.63	0.20
Notes:	Notes: *: p<.10; **:p<.05; ***: p<.01			

Table V. 18. Costa Rica: Elderly population in 2004-2006. Relative risks from Weibull parametric survival regression on self-reported previous diagnoses of DM, accounting for differential mortality correction.

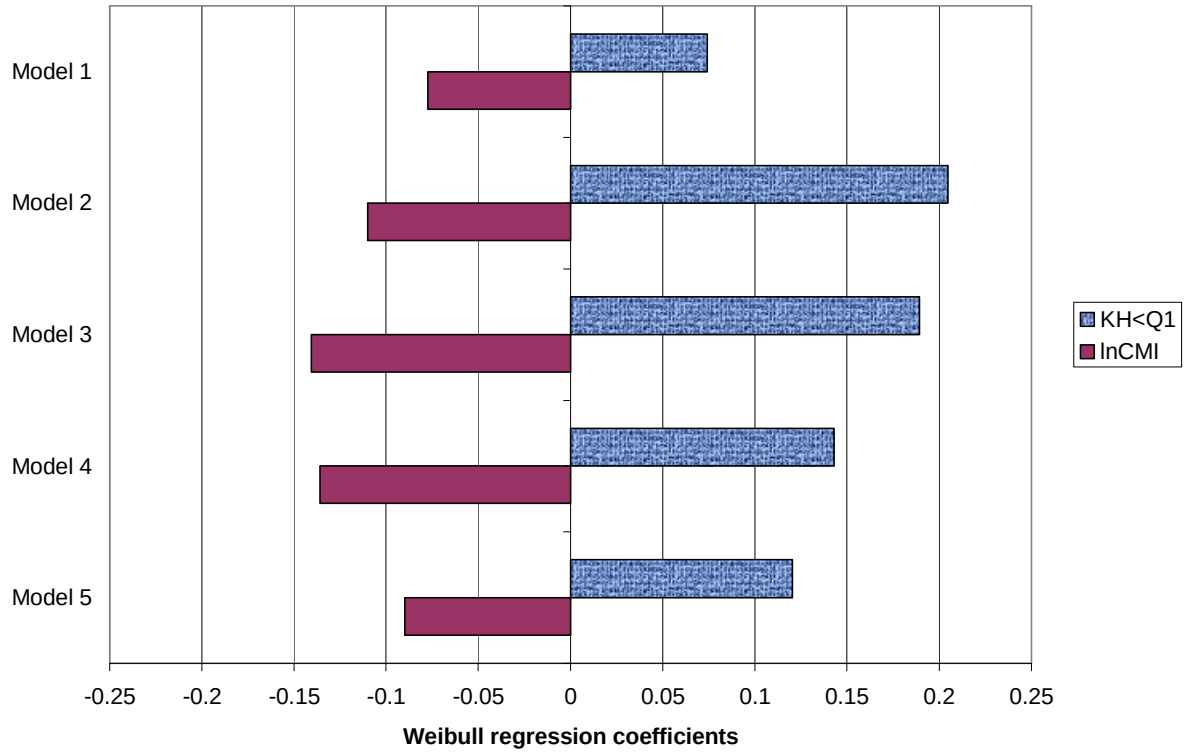
	Total sample			Non-proxy sample, w/ weight change var		
Variables	RRR	(CoeffSE)		RRR	(CoeffSE)	
Knee Height≤Q1						
(Ref: Q2<KH≤Q3)	1.13	0.22		1.12	0.22	
Q1<Knee Height≤Q2	1.26	0.22		1.26	0.22	
Knee Height>Q3	1.12	0.19		1.12	0.19	
Ln of Child Mortality Index (lnCMI)	0.91	0.18		0.91	0.18	
Foreign born	0.49	0.34		0.50	0.34	
Obese: BMI≥30 kg/m2	2.86	0.57	***	2.94	0.61	***
Overweight: 25≤BMI<29.9	1.79	0.33	***	1.82	0.34	***
Missing BMI	2.46	0.84	***	2.47	0.89	**
Female	0.82	0.20		0.82	0.20	
Schooling≥6 yrs.	1.04	0.16		1.05	0.16	
Living in Metropolitan Area	1.00	0.13		1.00	0.13	
Retired	0.89	0.11		0.89	0.11	
Current alcoholic drinker	0.96	0.17		0.96	0.17	
Past alcoholic drinker	1.22	0.21		1.22	0.21	
Current smoker	0.91	0.27		0.90	0.27	
Past smoker	0.82	0.12		0.82	0.12	
Current physical activity	0.62	0.10	***	0.62	0.10	***
Ever worked	1.20	0.22		1.21	0.22	
Mean couple's income<50000 col.	1.27	0.17	*	1.26	0.18	*
Known DM family history	2.18	0.27	***	2.18	0.27	***
Hospitalized during last 12 months	1.33	0.25		1.33	0.25	
Daily carb intake > 400g	1.10	0.20		1.10	0.20	
(Interaction with Female)						
Parity	1.11	0.08		1.11	0.08	
Scale perceived change of current weight-weigth at 25				0.99	0.04	
Missing perceived current weight				1.00	0.31	
Missing perceived weigth at 25				1.01	0.20	
Parameter /ln(p)	0.22	0.10	**	0.22	0.10	**
(n)	(2493)			(2493)		
Notes:	Notes: *: p<.10; **:p<.05; ***: p<.01					

Table V. 19. Costa Rica: Elderly population in 2004-2006. Relative risks from Weibull parametric survival regression on self-reported previous diagnoses of DM, accounting for differential mortality correction and categorized CMI.

Variables	Total sample RRR	(Coeff SE)		Non-proxy sample, w/ weight change var RRR	(CoeffSE)	
Knee Height≤Q1						
(Ref: Q2<KH≤Q3)	1.12	0.22		1.12	0.22	
Q1<Knee Height≤Q2	1.25	0.22		1.25	0.22	
Knee Height>Q3	1.12	0.19		1.12	0.19	
CMI≤18	0.93	0.16		0.93	0.16	
CMI≥31	0.89	0.16		0.89	0.16	
Foreign born	0.63	0.20		0.63	0.20	
Obese: BMI³30 kg/m2	2.86	0.57	***	2.94	0.61	***
Overweight: 25≤BMI<29.9	1.79	0.33	***	1.82	0.34	***
Missing BMI	2.46	0.84	***	2.47	0.89	**
Female	0.82	0.20		0.82	0.20	
Schooling≥6 yrs.	1.04	0.16		1.05	0.16	
Living in Metropolitan Area	0.99	0.13		0.99	0.13	
Retired	0.89	0.11		0.89	0.11	
Current alcoholic drinker	0.96	0.17		0.96	0.17	
Past alcoholic drinker	1.22	0.21		1.22	0.21	
Current smoker	0.90	0.27		0.89	0.27	
Past smoker	0.82	0.12		0.82	0.12	
Current physical activity	0.62	0.10	***	0.62	0.10	***
Ever worked	1.21	0.22		1.22	0.22	
Mean couple's income<50000 col.	1.27	0.17	*	1.27	0.18	*
Known DM family history	2.18	0.27	***	2.18	0.27	***
Hospitalized during last 12 months	1.33	0.25		1.32	0.25	
Daily carb intake > 400g	1.10	0.20		1.10	0.20	
(Interaction with Female)						
Parity	1.11	0.08		1.11	0.08	
Scale perceived change of current weight-weigth at 25				0.99	0.04	
Missing perceived current weight				1.00	0.31	
Missing perceived weigth at 25				1.01	0.20	
Parameter /ln(p)	0.22	0.10	**	0.22	0.10	**
(n)	(2493)			(2493)		

Notes: *: p<.10; **:p<.05; ***: p<.01

Figure V. 12. Coefficients for short KH (KH<Q1) and for lnCMI in models that control for different confounding variables.



Notes: Model 1: Only KH and CMI variables
Model 2: Model 1 + obesity variables
Model 3: Model 2 + physical activity
Model 4: Model 3 + family history of DM
Model 5: Model 4 + rest of covariates

Table V. 20. Costa Rica: Elderly population in 2004-2006. Coefficients from Weibull parametric survival regression on self-reported previous diagnoses of DM, accounting for differential mortality correction, and BMI and Knee Height as continuous variables.

	Total sample			Non-proxy sample, w/ weight change var		
Variables	Coeff	(CoeffSE)		RRR	(CoeffSE)	
Knee Height	0.074	0.061		0.076	0.061	
BMI	0.281	0.102	***	0.286	0.102	***
Interaction Knee Height*BMI	-0.004	0.002	**	-0.004	0.002	**
Missing BMI	2.517	0.445	***	2.574	0.466	***
Ln of Child Mortality Index (lnCMI)	-0.069	0.203		-0.072	0.202	
Foreign born	-0.633	0.702		-0.646	0.697	
Female	-0.360	0.264		-0.349	0.265	
Schooling≥6 yrs.	0.047	0.148		0.054	0.149	
Living in Metropolitan Area	-0.038	0.135		-0.042	0.135	
Retired	-0.103	0.128		-0.108	0.128	
Current alcoholic drinker	-0.108	0.179		-0.106	0.179	
Past alcoholic drinker	0.156	0.172		0.153	0.173	
Current smoker	-0.124	0.294		-0.131	0.296	
Past smoker	-0.170	0.151		-0.167	0.152	
Current physical activity	-0.440	0.157	***	-0.437	0.159	***
Ever worked	0.192	0.180		0.206	0.182	
Mean couple's income<50000 col.	0.255	0.135	*	0.250	0.137	*
Known DM family history	0.755	0.125	***	0.760	0.125	***
Hospitalized during last 12 months	0.280	0.188		0.272	0.188	
Daily carb intake > 400g	0.101	0.181		0.099	0.183	
(Interaction with Female)						
Parity	0.093	0.071		0.090	0.071	
Scale perceived change of current weight-weigth at 25				-0.016	0.037	
Missing perceived current weight				-0.109	0.322	
Missing perceived weigth at 25				0.057	0.188	
Parameter /ln(p)	0.217	0.099	**	0.211	0.102	**
(n)	(2493)			(2493)		

Notes: *: p<.10; **:p<.05; ***: p<.01

Table V. 21. Estimated hazard ratios for combinations of KH and BMI values, from Weibull event history model with continuous KH and BMI.
(Values of KH and BMI used for estimation in parentheses).

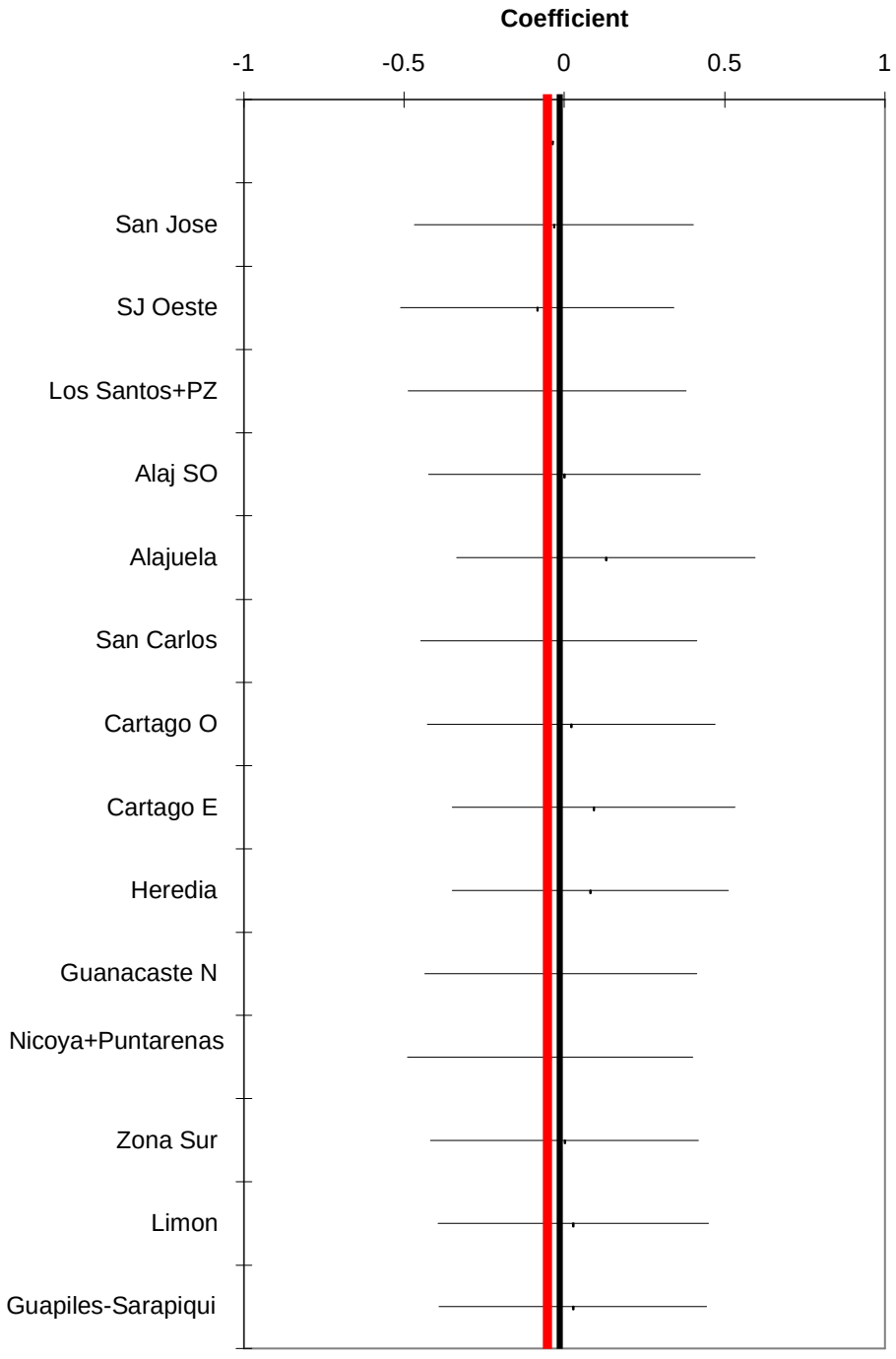
Scenarios and KH	BMI categories		
	Obese	Overweight	normal
First scenario: Observed mean values of BMI and Knee Height			
KH<Q1	**1.149** (BMI=31.6,KH=45.5)	**1.113** (BMI=26.6,KH=46.5)	**0.711** (BMI=21.6,KH=46.1)
Q1≤KH<Q2	**1.127** (BMI=32.9,KH=47.9)	**1.044** (BMI=26.7,KH=48.6)	**0.888** (BMI=21.5,KH=48.5)
Q2≤KH<Q3	**1.000** (BMI=32.6,KH=49.5)	**1.000** (BMI=26.9,KH=50.3)	**1.000** (BMI=21.8,KH=50.2)
KH≥Q4	**0.932** (BMI=33.7, KH=52)	**0.931** (BMI=27.1,KH=52.7)	**1.204** (BMI=22, KH=52.8)
Second scenario: Extreme values of BMI and Knee Height			
KH<Q1	**1.558** (BMI=40, KH=45)	**1.245** (BMI=29, KH=45)	**0.690** (BMI=18, KH=45)
Q1≤KH<Q2	**1.194** (BMI=40, KH=48)	**1.092** (BMI=29, KH=48)	**0.862** (BMI=18, KH=48)
Q2≤KH<Q3	**1.000** (BMI=40, KH=50)	**1.000** (BMI=29, KH=50)	**1.000** (BMI=18, KH=50)
KH≥Q4	**0.642** (BMI=40, KH=55)	**0.888** (BMI=29, KH=55)	**1.186** (BMI=18, KH=55)

Table V. 22. Costa Rica: Elderly population in 2004-2006. Relative risks of KH and lnCMI from Weibull parametric survival regression on self-reported previous DM diagnosis with mortality correction, testing unmeasured heterogeneity with a gamma frailty model

	Total sample			Non-proxy sample, w/ weight change var		
Variables	RRR	(Coeff SE)		RRR	(CoeffSE)	
Continuous CMI						
Observed						
Knee Height≤Q1	1.13	0.22		1.12	0.22	
Q1<Knee Height≤Q2	1.26	0.22		1.26	0.22	
Knee Height>Q3	1.12	0.19		1.12	0.19	
Ln of Child Mortality Index (lnCMI)	0.91	0.18		0.91	0.18	
Foreign born	0.49	0.34		0.50	0.34	
Gamma frailty model						
Knee Height≤Q1	1.97	1.08		2.26	1.65	
Q1<Knee Height≤Q2	1.66	0.56		1.89	0.84	
Knee Height>Q3	1.52	0.49		1.70	0.66	
Ln of Child Mortality Index (lnCMI)	0.71	0.33		0.66	0.34	
Foreign born	0.18	0.27		0.12	0.21	
Gamma frailty parameter (ln[theta]$^{-1}$)	1.29	0.41	***	1.50	0.53	***
Categorical CMI						
Observed						
Knee Height≤Q1	1.12	0.22		1.12	0.22	
Q1<Knee Height≤Q2	1.25	0.22		1.25	0.22	
Knee Height>Q3	1.12	0.19		1.12	0.19	
CMI≤18	0.93	0.16		0.93	0.16	
CMI≥31	0.89	0.16		0.89	0.16	
Foreign born	0.63	0.20		0.63	0.20	
Gamma frailty model						
Knee Height≤Q1	1.97	1.14		2.24	1.71	
Q1<Knee Height≤Q2	1.66	0.58		1.89	0.86	
Knee Height>Q3	1.52	0.50		1.68	0.66	
CMI≤18	1.11	0.41		1.13	0.48	
CMI≥31	0.77	0.27		0.73	0.28	
Foreign born	0.51	0.26		0.44	0.25	
Gamma frailty parameter (ln[theta]$^{-1}$)	1.30	0.44	***	1.50	0.55	***
Interaction BMI and KH						
Observed						
Knee Height	0.074	0.061		0.076	0.061	
BMI	0.281	0.102	***	0.286	0.102	***
Interaction KH-BMI	-0.004	0.002	**	-0.004	0.002	**
Gamma frailty model						
Knee height	0.109	0.131		0.117	0.134	
BMI	0.542	0.220	**	0.584	0.225	***
Interaction KH- BMI	-0.008	0.004	*	-0.008	0.004	*

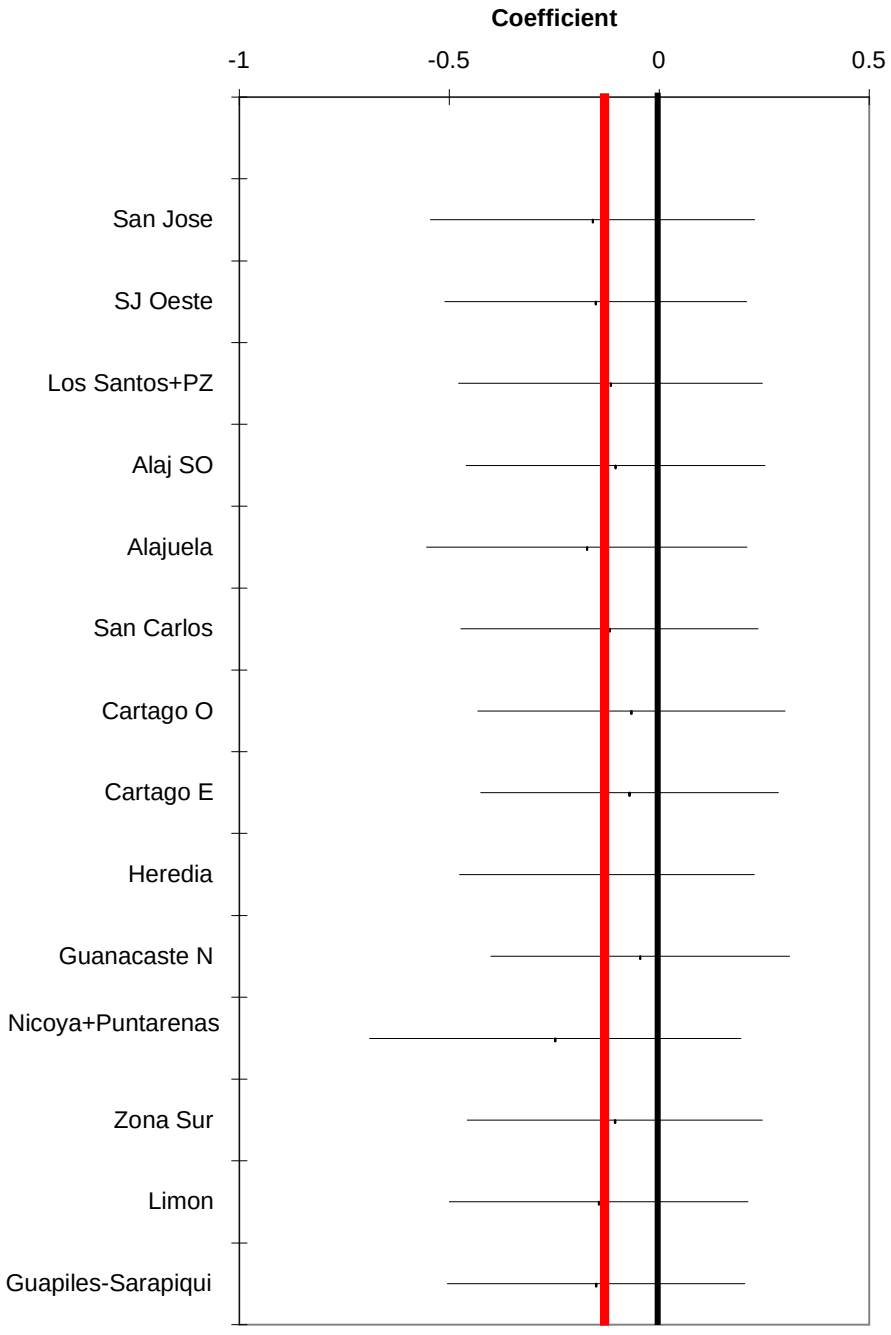
Notes: Notes: *: p<.10; **:p<.05; ***: p<.01

Figure V. 13. Point and interval estimates for coefficient of lnCMI, in logistic regression of DM hazard using first approach, excluding cases from Costa Rica's subregions one at a time.



Note: Red line denotes the coefficient estimated with all regions.

Figure V. 14. Point and interval estimates for coefficient of CMI<18, in logistic regression of DM hazard using first approach, excluding cases from Costa Rica's subregions one at a time.



Note: Red line denotes the coefficient estimated with all regions.

Figure V. 15. Estimated DM hazard rates using Weibull regression with offset, by KH groups.

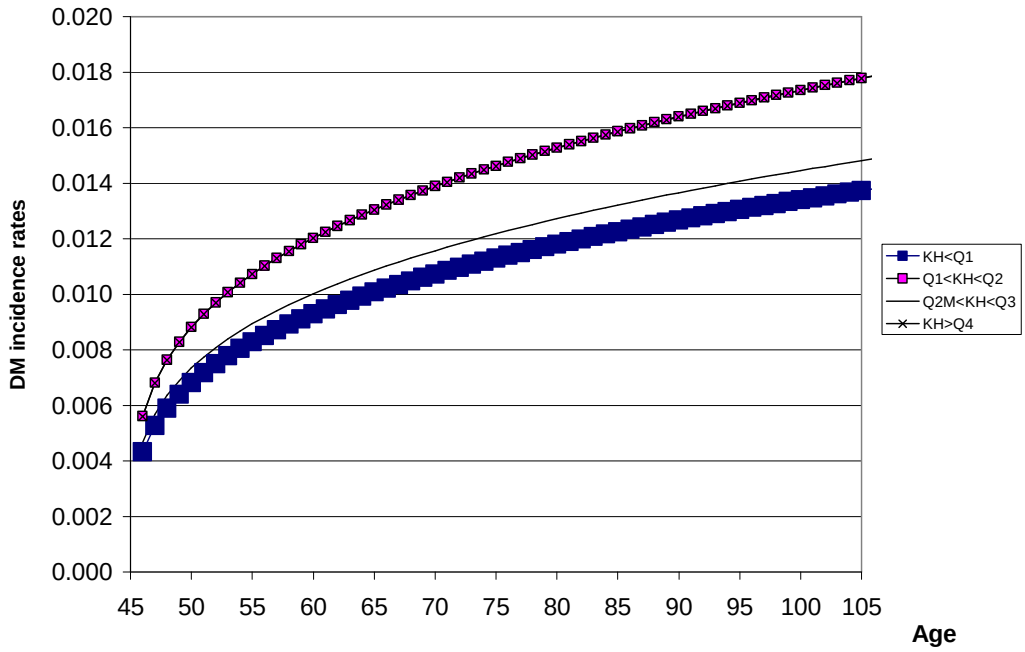


Figure V. 16. Estimated DM hazard rates using Weibull regression with offset, by CMI groups.

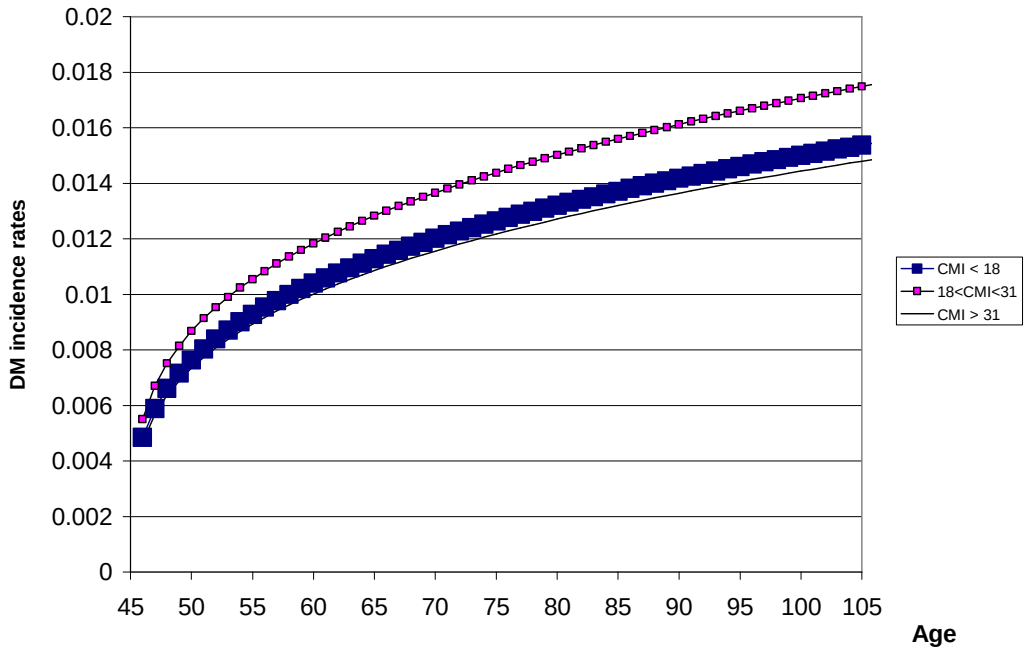


Table V. 23. Costa Rica: Elderly population in 2004-2006. Distribution by awareness of diabetes, controlling separately by knee height and CMI.

Variables	No diabetes	Unaware of diabetes	Diagnosed w/diabetes	Total	χ^2 independence test p-value
Knee height					
KH≤Q1	76	4	20	100	0.508
Q1<KH≤Q2	73	3	24	100	
Q2<KH≤Q3	77	4	18	100	
KH>Q3	74	4	22	100	
CMI					
CMI<18	75	4	21	100	0.787
18.2≤CMI<31	74	4	22	100	
CMI≥31.0	78	3	19	100	
Total	75	4	21	100	

Notes: [1/] For men's knee height, Q1=49.6 cm, Q2=51.3 cm, Q3=53.0cm. For women, Q1=45.6 cm, Q2=47.2 cm, Q3=48.7

Table V. 24. Comparison of elderly people with FSG>=200 mg/dL or HbA1C>=7% and no diagnosis of DM (outcome=1) vs people with a DM diagnosis regardless of biomarker level (outcome=0) using logistic regresion. Costa Rica, 2004-2006.

Variables	With diagnosed diabetic pop OR	(CoeffSE)	
DM symptoms			
Intense thirst	0.58	0.46	
Fatigue	0.36	0.41	**
Frequent urination	0.55	0.32	*
Obese: BMI≥30 kg/m2	0.71	0.41	
Overweight: 25≤BMI<29.9	0.60	0.39	
Female	1.25	0.65	
Age	1.04	0.02	
Schooling>6 yrs.	0.92	0.34	
Living in Metropolitan Area	0.49	0.34	**
Retired	0.75	0.34	
Current alcoholic drinker	0.53	0.51	
Past alcoholic drinker	0.52	0.45	
Current smoker	2.72	0.53	*
Past smoker	1.59	0.41	
Current physical activity	1.60	0.34	
Ever worked	0.82	0.50	
Mean couple's income<50000 col.	1.34	0.36	
Known DM family history	0.41	0.32	***
Hospitalized during last 12 months	1.34	0.55	
Daily carb intake > 400g	0.91	0.43	
Uninsured	0.59	0.89	
(Interaction with Female)			
Parity	0.67	0.20	**
(n)	(492)		

Notes: *: p<.10; **:p<.05; ***: p<.01

Table V. 25. Costa Rica: Logistic regression to estimate proportion of DM unawareness among elderly population in 2004-2006 with NO PREVIOUS DM diagnosis. (Unawareness = FSG >=200 mg/dl or HbA1C>=7% among respondents with no self-reported diagnosis).

	Total sample			Non-proxy sample, w/ weight change var		
Variables	OR	(Coeff SE)		RRR	(CoeffSE)	
Knee Height≤Q1						
(Ref: Q2<KH≤Q3)	0.87	0.32		0.75	0.36	
Q1<Knee Height≤Q2	0.80	0.36		0.80	0.39	
Knee Height>Q3	0.91	0.33		0.84	0.36	
Ln of Child Mortality Index (lnCMI)	0.57	0.32	*	0.63	0.37	
Foreign born	0.23	1.13		0.32	1.28	
Glycemia more than 1 year away	1.60	0.28	*	1.55	0.32	
Obese: BMI≥30 kg/m2	1.48	0.34		1.52	0.44	
Overweight: 25≤BMI<29.9	0.95	0.31		1.07	0.40	
Missing BMI	1.50	0.42		-	-	
Female	1.04	0.50		1.00	0.56	
Age	1.03	0.02	**	1.03	0.02	
Schooling>6 yrs.	1.20	0.30		1.20	0.33	
Living in Metropolitan Area	0.71	0.26		0.61	0.30	*
Retired	0.95	0.29		0.98	0.34	
Current alcoholic drinker	0.47	0.41	*	0.44	0.46	*
Past alcoholic drinker	0.59	0.39		0.57	0.44	
Current smoker	1.54	0.39		1.60	0.44	
Past smoker	1.44	0.34		1.25	0.40	
Current physical activity	0.92	0.28		0.94	0.28	
Ever worked	1.33	0.37		1.07	0.44	
Mean couple's income<50000 col.	1.20	0.27		1.42	0.30	
Known DM family history	1.07	0.26		1.11	0.28	
Hospitalized during last 12 months	1.30	0.40		1.27	0.51	
Daily carb intake > 400g	1.09	0.33		0.91	0.38	
(Interaction with Female)						
Parity	0.90	0.13		0.83	0.16	
Scale perceived change of current weight-weigth at 25				0.99	0.07	
Missing perceived current weight				1.77	0.76	
Missing perceived weigth at 25				0.53	0.76	
(n)	(2110)			(1587)		

Notes: *: p<.10; **:p<.05; ***: p<.01

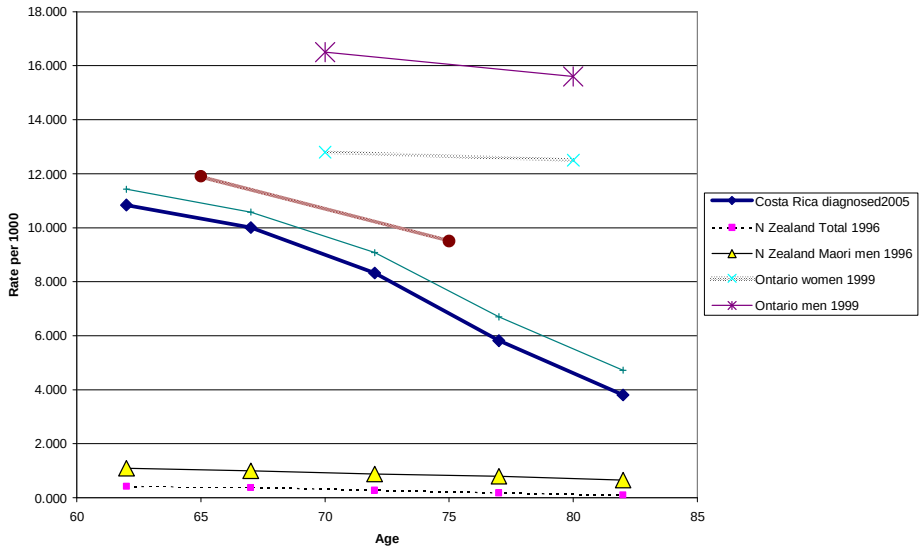
Table V. 26. Costa Rica: Logistic regression to estimate proportion of DM unawareness for elderly population in 2004-2006 with no previous DM diagnosis. (Unawareness = FSG >=200 mg/dl or HbA1C>=7% among respondents with no self-reported diagnosis). (Categorized CMI)

	Total sample		Non-proxy sample, w/ weight change var	
Variables	OR	(Coeff SE)	RRR	(CoeffSE)
Knee Height≤Q1				
(Ref: Q2<KH≤Q3)	0.85	0.32	0.74	0.36
Q1<Knee Height≤Q2	0.79	0.36	0.79	0.39
Knee Height>Q3	0.92	0.33	0.84	0.36
CMI≤18	1.05	0.35	1.08	0.37
CMI≥31	0.59	0.34	0.60	0.43
Foreign born	1.21	0.49	1.29	0.54
Glycemia more than 1 year away	1.62	0.28 *	1.57	0.32
Obese: BMI≥30 kg/m2	1.47	0.34	1.51	0.44
Overweight: 25≤BMI<29.9	0.95	0.31	1.06	0.40
Missing BMI	1.52	0.43		
Female	1.03	0.50	1.00	0.56
Age	1.03	0.02 **	1.03	0.02
Schooling≥6 yrs.	1.19	0.30	1.20	0.33
Living in Metropolitan Area	0.68	0.26	0.60	0.30 *
Retired	0.94	0.29	0.97	0.33
Current alcoholic drinker	0.46	0.41 *	0.43	0.46 *
Past alcoholic drinker	0.59	0.39	0.57	0.44
Current smoker	1.54	0.39	1.62	0.44
Past smoker	1.45	0.34	1.25	0.40
Current physical activity	0.92	0.28	0.94	0.28
Ever worked	1.36	0.38	1.08	0.45
Mean couple's income<50000 col.	1.21	0.27	1.42	0.30
Known DM family history	1.07	0.26	1.11	0.28
Hospitalized during last 12 months	1.30	0.41	1.27	0.52
Daily carb intake > 400g	1.09	0.32	0.91	0.38
(Interaction with Female)				
Parity	0.90	0.13	0.83	0.16
Scale perceived change of current weight-weigth at 25			0.99	0.07
Missing perceived current weight			1.86	0.76
Missing perceived weigth at 25			0.53	0.77
(n)	(2110)		(1587)	

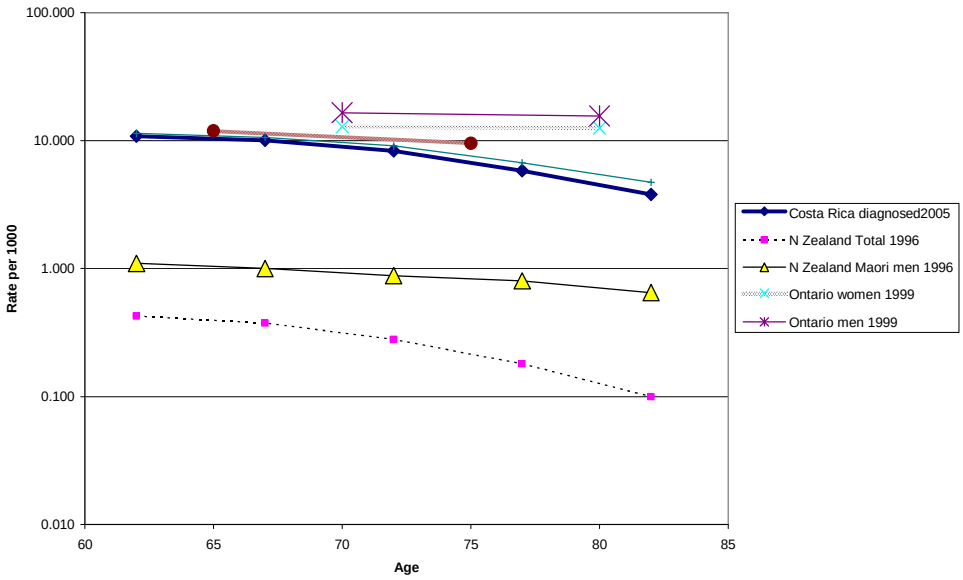
Notes: *: p<.10; **:p<.05; ***: p<.01

Figure V. 17. Estimated DM incidence curves for Costa Rica, and comparison with other countries.

(Arithmetic scale)



(Log scale)



Note: Rates were estimated using 5-year age groups.
Sources: New Zealand-Ministry of Health (2002); Hux and Tang (2002); Bonora *et al.* (2004).

Chapter VI: Diabetes mellitus projections and the future effect of early life conditions.

A. Introduction

The previous chapter aimed to test whether there is a statistically significant effect of early life conditions on DM incidence and mortality. This goal was achieved. The equations estimated to test the statistical relationship were used also to estimate DM and death hazard rates by single year of age, sex, obesity categories, KH categories, and CMI categories, controlled by other important covariates: region of residence, education, retirement condition, alcohol intake, smoking, physical activity, low income, family history of DM, being hospitalized, parity, and caloric intake. These sets of rates are used for projecting the diabetic population from 2005 to 2030. This chapter is divided in three sections. The first section will describe the procedure followed to project the population; the next section will discuss the projected figures, as well as the impact of early life conditions on the projections; the last section is short, and will explain the projection by sex and obesity categories, two covariates with distributions that vary over time given that the projection methodology allowed to incorporate covariation between at least KH (one of the main independent variables) and other covariates. Given that the projection used CMI as a continuous variable rather than as categories, it was not possible to incorporate covariation between CMI and other independent variables.

B. Projection methodology

The procedure to project the population can be understood as a variation of the cohort component method. This section contains a more detailed description of the variation of the cohort component method that was used for projecting the diabetic population.

1. The projection methodology used.

The projection methodology used is considered a variation of the cohort-component method because of the data sources and the inputs into the main projecting equations. The first variation on the method was to split the projected population into two subpopulations: the cohorts born before 1945 that are represented in the CRELES dataset, and the cohorts born after 1944 and that start to appear progressively in the projections when they become 60 years-old. Both cohorts are "extinguished" in the subsequent future by applying the sets of DM and death hazard rates, and the probabilities of having DM but being unaware of it ("unaware DM").

a) Cohorts born before 1945

The CRELES project infers to the population 60 years old and older in the period 2004-2006. This chapter assumes that the sample expanded using the sampling weights provided by the project is equivalent to the total population in 2004, born before 1945, and thus, 60 years old and older.

Using the sampling weights is important because they constitute one of the main inputs in the projection. Suppose that one of the individuals in the sample is a woman, age 65, obese

(BMI≥30 kg/m^2), and with short KH (KH<Q1), and suppose that her sampling weight, W, is equal to 32.4. In symbolic notation, this can be written as:

$${}_1W^S_{65}(2004)=32.4, \text{ where S=female and obese and KH<Q1.}$$

This means that 32.4 people in the actual Costa Rican female population age 65 and obese and with short KH are represented by this individual in the sample. Therefore, in the usual projection formulas, ${}_1W^S_{x-1}(t-1)$ is used as input in the formula representing the population size term: ${}_1N^S_{x-1}(t)$. Another important variation in the procedure is that it is not using survival ratios, $\frac{{}_1L^i_x}{{}_1L^i_{x-1}}$, but direct DM hazard rates (denoted by ${}_1\lambda_x$) and mortality hazards for diabetic patients (${}_1\mu(d)_x$) and for non-diabetic population (${}_1\mu(\bar{d})_x$).

An additional variation in the usual projection formulas is taking into account the incidence probabilities of DM unawareness. The ideal way to deal with this condition is to include it as an extra state in a multi-state model. However, as mentioned before, a multi-state hazards model was discarded because there were not enough deaths among people that already had DM, which is one of the transitions in the multi-state model. Including an additional state would have reduced the subsample size for the new transitions involved even more. These individuals who are unaware of their disease have the same mortality hazard rates as those with diagnosed DM therefore they have to be exposed to the same mortality schedule. However, they are also part of the population at risk of being diagnosed with DM; therefore, they have to be added to the "undiagnosed population" in order to estimate the DM incidence cases. In this

sense, the easiest way to deal with this group is to consider them as an additional subpopulation different from the diabetic and the non-diabetic population.

Therefore, I use the following set of expressions for the projection:

For unaware population:

$$ {}_1U_x(t+1) = \left[{}_1U_{x-1}(t)\right] * \left(1 - {}_1\mu(d)_{x-1}\right) + \left[{}_1ND_{x-1}(t)\right] * \left({}_1\delta_{x-1}\right) * \left(1 - {}_1\lambda_{x-1}\right) * \left(1 - {}_1\mu(d)_{x-1}\right) \quad (1) $$

For diagnosed diabetic population:

$$ {}_1DD_x(t+1) = \left[{}_1DD_{x-1}(t)\right] * \left(1 - {}_1\mu(d)_{x-1}\right) + \left[{}_1ND_{x-1}(t) + {}_1U_{x-1}(t)\right] * \left({}_1\delta_{x-1}\right) * \left(1 - {}_1\mu(d)_{x-1}\right) \quad (2) $$

For total diabetic population:

$$ {}_1D_x(t+1) = \left[{}_1U_x(t+1)\right] + \left[{}_1DD_x(t+1)\right] \quad (3) $$

For non-diabetic population:

$$ {}_1ND_x(t+1) = \left[{}_1ND_{x-1}(t)\right] * \left(1 - {}_1\delta_{x-1}\right) * \left(1 - {}_1\lambda_{x-1}\right) * \left(1 - {}_1\mu(\bar{d})_{x-1}\right), \quad (4) $$

Where:

${}_1U^S_x(t+1)$ is the unaware diabetic population age x, with characteristics S, at time (t+1),

${}_1DD^S_x(t+1)$ is the diagnosed diabetic population age x, with characteristics S, at time (t+1),

${}_1D^S_x(t+1)$ is the total diabetic population age x, with characteristics S, at time (t+1),

${}_1ND^S_x(t+1)$ is the non-diabetic population age x, with characteristics S, at time (t+1).

$\mu(d)$ is the death rate for diabetic population

$\mu(\bar{d})$ is the death rate for non-diabetic population,

λ is the DM incidence rate, and

δ is the incidence rate of developing DM but being unaware of it (DM unawareness).

All formulas are applied to different sub-populations defined by covariate patterns.

As mentioned before, the original 4 populations (total diabetic, diagnosed diabetic, unaware diabetic, and non-diabetic) in 2004 are represented by the sampling weights distributed according to the probability of being diabetic. Therefore, equations (1), (2), and (4), can be expressed in the following way:

For unaware diabetic population:

$$\begin{aligned} {}_1U_x(t+1) &= \left[{}_1W_{x-1}(t) * P(\text{Diabetes}) * P(\text{Unaware} \mid \text{Diabetes})\right] * \left(1 - {}_1\mu(d)_{x-1}\right) \\ &+ \left[{}_1W_{x-1}(t) * \left(1 - P(\text{Diabetes})\right)\right] * ({}_1\delta_{x-1}) * \left(1 - {}_1\lambda_{x-1}\right) * \left(1 - {}_1\mu_{x-1}\right) \end{aligned} \quad (5)$$

For diagnosed diabetic population:

$$
\begin{aligned}
{}_1DD_x(t+1) &= \left[{}_1W_{x-1}(t) * P(\text{Diabetes}) * \left(1 - P(\text{Unaware} \mid \text{Diabetes})\right)\right] * \left(1 - {}_1\mu(d)_{x-1}\right) \\
&+ \left[{}_1W_{x-1}(t) * \left(1 - P(\text{Diabetes})\right)\right] * \left({}_1\delta_{x-1}\right) * \left(1 - {}_1\mu(d)_{x-1}\right) \\
&+ \left[{}_1W_{x-1}(t) * P(\text{Diabetes}) * \left(P(\text{Unaware} \mid \text{Diabetes})\right)\right] * \left({}_1\delta_{x-1}\right) * \left(1 - {}_1\mu(d)_{x-1}\right)
\end{aligned}
\tag{6}
$$

For non-diabetic population:

$$
{}_1ND_x(t+1) = \left[{}_1W_{x-1}(t) * (1 - P(\text{Diabetes})\right] * \left(1 - {}_1\delta_{x-1}\right) * \left(1 - {}_1\lambda_{x-1}\right) * \left(1 - {}_1\mu(\bar{d})_{x-1}\right) \tag{7}
$$

Notice from equations (1), (2), and (3) that the diabetic population is composed of the survivors of DM, the non-diabetic individuals that are diagnosed with the disease and then survive, and also the individuals that develop the disease but are not unaware of having it and also survive. Meanwhile, the non-diabetic population is composed only of those who are not diagnosed with the disease, or are not unaware of having the disease, and survive death. Notice also that, because of the formula and the procedures, what the projection follows from 2004 to 2030 is the set of cohorts.

The advantage of taking this approach with the existing cohorts –that can be called "old cohorts" too–, rather than aggregating them into certain number of categories, is that DM hazard rates, mortality rates, and the probabilities of new DM unawareness are estimated with equations that include several independent variables. Therefore, taking the sampling weights as the equivalent of the population for any particular "covariate pattern"[13] allows to take into account the covariation between the main independent variables on one hand (KH and CMI), and the rest

[13] Hosmer and Lemeshow (2000) define covariate pattern as: "…a single set of values for the covariates in a model. For example, (…) if the model contains only race and sex, each coded at two levels, there are only four possible covariate patterns" (p.144).

of independent variables on the other hand. This means that, for example, people with short KH are less likely to practice physical activities; therefore, they have higher DM incidence rates, not only because of the KH effect, but also because of the physical activity effect. As it will be explained later, this accounting for covariation was not possible for the "newer" cohorts.

There are several shortcomings to this approach, too. One of them is that the single-age distribution is not very smooth, given that the weights are computed based on 5-year age groups, rather than on 1-year age groups. This causes that, when the projections are observed for single years of age in any specific year, the projected figures have a relatively high sampling error, as compared to figures for 5-year age groups. In a related topic, individuals age 60 are underrepresented because the original sample was drawn from the 2000 Population Census, and the subsequent subsample size was drawn in 2004, thus not allowing for enough people to be age 60 by the beginning of the fieldwork. The sampling weight for the group of people with exact age 60 was multiplied by 3 to have a more adequate distribution, but only for the projections, not for the statistical estimation of rates or for testing the size and statistical significance of early life effects.

An important assumption that is made, and that can be considered either as a problem or as an advantage, is that the values of the covariates remain constant over the period 2005-2030. This means that if a person age 64 in 2004 reported to practice physical activity, the estimation will consider that this same person will be still doing some physical activity, but at age 90 in 2030. It is possible to change this kind of assumptions in the procedure, but it is too cumbersome to compute the pattern of this change by age, since this will require several regression equations, one for each combination of age and another covariate.

b) Cohorts born after 1944.

The projection that includes the cohorts born in 1945 and later had to be estimated differently because they are not represented in CRELES dataset. In the cohort-component-method logic, the new cohorts are treated similarly as births (the typical entry) in a complete population projection. These incoming cohorts are equivalent to the population age 60 in each year of the period 2005-2030, and are drawn from the official Costa Rican population projections by single years of age (INEC-CCP, 2001).

The following step was to build a dataset with 1104 records, one record for each combination of 46 single years of age, 4 KH categories, 2 sex groups, and 3 obesity categories. All of these records had a zero value because there were no persons in the estimate for 2004 that were born after 1944. These dataset is the basis for creating "artificial cohorts" that represent the projected population. A weight variable was computed to be equivalent to a smoothed distribution estimated from the observed distribution by KH, sex, and obesity categories for age 60. This dataset was further expanded 81 times, in order to have 89424 records. The figure 81 corresponded to the 81 cantones (counties) that constitute the geographical subdivision of Costa Rica. Having records for each canton was important because the variable CMI (Child Mortality Index) relates to the level of child mortality in the canton where the person was born. This dataset was merged with another dataset with 81 records. This shorter dataset contains variables which record the CMI for each canton for the period 1945-1970; these are the years when the cohorts age 60 during the period 2005-2030 were born. The dataset also contains the relative distribution of canton of birth for each of these incoming cohorts. This relative distribution is

estimated from the 1984 and 2000 censuses. The relative distribution served as another weight factor similar to the weight factor computed before and based on KH, sex, and obesity.

These two weight variables are multiplied by the total number of persons age 60 in Costa Rica in each projection year, according to the official Costa Rican projections (CCP-INEC, 2001), in order to get the absolute distribution by sex, KH, obesity categories, and county of birth for each incoming cohort. Notice that this methodology implies that the process does not allow for any association between any of the first set of covariates (sex, KH, obesity) and county of birth; in other words, the process assumed that the sex-KH-obesity distribution is the same in each county of birth. Doing otherwise would have been too difficult to estimate.

In addition, it is important to clarify that each one of the final 89,424 records have the same value for the other covariates included in the model (foreigner, education, region of residence, retired, alcohol intake, smoking, physical activity, ever worked, low income, family history of DM, being hospitalized, carbohydrate intake, and parity). The assigned value is equivalent to the mean value in the whole CRELES dataset. Again, the procedure implies that there is no association between any of the main independent variables and any of the rest of the control covariates. This decision was made to simplify the estimation procedure. In this case, it is expected that this decision would not introduce a strong bias in the estimation because most of these covariates have small coefficients in the regressions for estimating DM hazard rates.

The new variable, which can be called ${}_1W^S_{60}(t)$ and is equivalent to the total cohort size multiplied by the two weights, had a similar function to the sampling weights in the procedure for the cohorts born before 1945. In their year of entry into the procedure, the cohorts age 60 are multiplied by the probability of having DM at age 60 (DM prevalence). This prevalence for each covariate pattern was estimated using the logistic regression for prevalence estimated and

presented in the previous chapter. This means that DM prevalence at age 60 is not computed as a subsequent accumulation of DM survivors estimated from DM and death hazard rates, but by prevalence estimates. The ideal procedure would have been the former (all estimates computed from hazard rates), but it is considered that the equations from the previous chapter would produce unreliable estimates of DM hazard rates for ages younger than 60, since most of the equations are estimated based on a sample of people ages 60 and over.

The following steps in the procedure are the same as in the projection of the "older cohorts", and they included the same DM hazards (${}_1\lambda_x$) and death rates (${}_1\mu(d)_x$ and ${}_1\mu(\bar{d})_x$), and probabilities for new unaware DM cases (${}_1\delta_x$). Therefore, the procedure can again be described with formulas (3), (5), (6), and (7).

For unaware diabetic population:

$$\begin{aligned} {}_1U_x(t+1) &= \left[{}_1W_{x-1}(t) * P(\text{Diabetes}) * P(\text{Unaware} \mid \text{Diabetes})\right] * \left(1 - {}_1\mu(d)_{x-1}\right) \\ &+ \left[{}_1W_{x-1}(t) * \left(1 - P(\text{Diabetes})\right)\right] * \left({}_1\delta_{x-1}\right) * \left(1 - {}_1\lambda_{x-1}\right) * \left(1 - {}_1\mu_{x-1}\right) \end{aligned} \quad (5)$$

For diagnosed diabetic population:

$$\begin{aligned} {}_1DD_x(t+1) &= \left[{}_1W_{x-1}(t) * P(\text{Diabetes}) * \left(1 - P(\text{Unaware} \mid \text{Diabetes})\right)\right] * \left(1 - {}_1\mu(d)_{x-1}\right) \\ &+ \left[{}_1W_{x-1}(t) * \left(1 - P(\text{Diabetes})\right)\right] * \left({}_1\delta_{x-1}\right) * \left(1 - {}_1\mu(d)_{x-1}\right) \\ &+ \left[{}_1W_{x-1}(t) * P(\text{Diabetes}) * \left(P(\text{Unaware} \mid \text{Diabetes})\right)\right] * \left({}_1\delta_{x-1}\right) * \left(1 - {}_1\mu(d)_{x-1}\right) \end{aligned} \quad (6)$$

For non-diabetic population:

$${}_1ND_x(t+1) = \left[{}_1W_{x-1}(t) * (1 - P(\text{Diabetes})\right] * \left(1 - {}_1\delta_{x-1}\right) * \left(1 - {}_1\lambda_{x-1}\right) * \left(1 - {}_1\mu(\bar{d})_{x-1}\right) \quad (7)$$

For total diabetic population:

$$ {}_1D_x(t+1) = \lfloor {}_1U_x(t+1) \rfloor + \lfloor {}_1DD_x(t+1) \rfloor \quad (3) $$

The final procedure is to merge both projection sets: the one for the older cohorts and the other for the new cohorts, and aggregate them according to the covariate patterns defined by the main independent variables: age, KH, sex, and obesity categories.

A summary of the projection assumptions is the following:

- DM and death hazard rates and incidence probabilities of DM unawareness remain constant over time.
- Rates and probabilities are differential across age, sex, KH, CMI, and BMI categories.
- Projections account for covariation among age, sex, KH, and BMI categories.
- CMI is assumed uncorrelated with sex, KH, and BMI categories.
- The size of the incoming cohorts in each year (people age 60) is taken from Costa Rica's official population projections.
- People with DM unawareness have the same mortality rates as persons with DM but become diabetic at the same rate as people without DM, controlling for variation by age, sex, KH, and BMI categories.
- DM burden is measured as the difference between observed prevalence and a hypothetical prevalence constructed by assuming 0 people with short KH or that CMI during the first part of the 20th century is equal to observed CMI in 1970.

C. Results.

Results for the projection are going to be presented in graphs rather than in tables because this chapter aims to describe the patterns and trends over time, rather than exact projected figures[14]. Each figure contains 2 graphs: the first one includes the people who are unaware of having DM as part of the diabetic population, while the second one excludes this group. In this way, it is possible to observe how much results vary if the information based on biomarkers is taken into account or not. Also, several figures are drawn for the same outcome variable, but changing certain assumptions of the projection.

The first graph (Figure VI.1, panel a) shows that from 2008 to 2030, prevalence grows from 27% to 35% (roughly a third of percentage point per year). The size of the diabetic population grows steadily from 60 thousand in 2005 to more than 400 thousand at the end of the period. This graph assumes that DM incidence rates are constant throughout the period, mortality rates are also constant, and the prevalence of people with short knee height (KH) at age 60 diminishes slowly to 0% in 2030. The other proportions for the rest of the KH categories are prorated to keep a relative distribution that adds up to 1. If the unaware diabetic population is excluded from the analysis, the prevalence goes from around 24% in 2005 to 30% in 2030. The size of the diabetic population reaches to just 340 thousand people. This means that the diabetic population is 18% higher if undiagnosed or unaware cases are included as diabetic persons.

[14] Besides, in population forecasting, it is difficult to argue in favor of an exact figure, because all projections are too sensitive to assumptions.

The assumption of constant mortality can be modified so as to consider the improvement in survival that is projected for the entire Costa Rican population during the following 25 years. According to official population projections, life expectancy is going to rise from 78.4 in 2005 to 80.9 in 2030 –from 76.1 to 78.5 among men, and from 80.8 to 83.5 among women– (CCP-INEC, 2002). An average rate of change in mortality rates for the elderly in Costa Rica from 2005 to 2030 is estimated from the official projected life tables. Mean changes in mortality rates among men are projected to increase linearly from -0.0089 in 2005-2006 to -0.005 in 2029-2030 among men, and from -0.0113 in 2005-2006 to -0.0064 among women. These rates of change are multiplied by the estimated mortality hazard rates to have a new set of values for each year, under the assumption of decreasing mortality throughout the period. According to Figure VI.2, modifying the assumption of mortality change does not have a strong impact on the estimates: prevalence increases in less than 0.1 percentage points, and the number of diabetic people increases in near 5 thousand persons under diminishing mortality in either of the two scenarios (with the unaware population and without them).

Figure VI.3 shows the effect of projecting the population using the DM hazard estimates from the equation where BMI and KH are operationalized as continuous variables (rather than as sets of dichotomous variables) and an interaction term is included in the analysis. The projection keeps the procedure of establishing 4 KH categories, but different mean BMI and KH values are assigned introducing more extreme variation in the estimated scenarios. Even under the assumption that all obese people have a BMI of 35 (a clearly extreme figure, given that it is the value that separates obesity from morbid obesity), the estimated DM prevalence change very little: around 0.4 of percentage point in prevalence, and less than 4,000 diabetics in the estimate of the size of the unhealthy population (scenario 2). These results prove that, even if the

equation estimated in the previous chapter has an interaction term between KH and BMI that is significantly different from zero, the size of the effect on DM burden is still small. Because there are no differences whether I use the DM hazard model with categorical KH and obesity or the DM hazard model with the continuous variables, I keep using the former model (with categorical KH and obesity).

The general objective of this dissertation is to analyze the impact of early childhood conditions on the hazard and prevalence of DM among the elderly in Costa Rica. This chapter will first focus on KH, and then on the effect of child mortality in respondents' place of birth. This analysis will keep the assumption that the proportion of people with short KH starts to decrease linearly to 0% in 2030 for the incoming cohorts. It is important to remember the results from the previous chapter: people with short KH (KH under first quartile) do not have higher chances of being diagnosed with DM (except if they are also obese) or of getting DM without being aware of it, and have lower mortality probabilities than people with KH between the second and the third quartile. Based on the hazard rates and probabilities used as inputs for the projection, DM prevalence for people with short KH keeps increasing until 2030 in the two scenarios: with and without including the unaware population (Figure VI.4, panels a and b). By the end of 2030, DM prevalence among this group is the highest: 40% in the first scenario and 34% in the second scenario. Prevalence increases for the other three groups, but not as fast as for the first group. When the assumption of constant mortality is shifted to a pattern of decreasing death rates, the observed trends by KH categories remain practically the same (Figure VI.5).

Differential effects of KH on DM incidence and mortality make the composition of the diabetic population by KH to vary across time (Figure VI.6). In the 25 year period, the proportion with the shortest KH decreases, while the proportion with the longest KH increases.

This is due mainly to the changing distribution of the new cohorts coming into the projection every year. These cohorts who are age 60 at each year have a heavier weight in the total population, given that they are the youngest cohorts. Again, changing from an assumption of constant mortality to an assumption of decreasing mortality does not modify this pattern in the projected population (Figure VI.7).

Another assumption that could have had an effect on the estimates is whether the relative composition of the elderly population by KH remains the same or changes, since there is no direct information about how this composition has shifted among cohorts born after 1944: the incoming cohorts of elderly during the following 25 years. The first published survey results about undernutrition in Costa Rica showed that malnutrition among children dropped to around 6% at the middle of the 1970s (Mata, 1980). New populations are projected under the assumption that the proportion of people with short KH (25% among the older cohorts) falls linearly to 0% among cohorts born in 1970, in order to assess the difference under an extreme assumption. The differences in the rest of the proportions are prorated to have a 100% sum in the distribution. When compared with projected figures under constant KH distribution assumption, the projected prevalence under decreasing short KH proportion decreases in 0.1 percentage points in 2030, regardless of whether mortality remains constant or not (Figure VI.8, panel a). Prevalence does not change when only diagnosed people are considered within the diabetic population (that is, excluding people that are unaware of their disease). The absolute figures decrease in around 17,000 people with DM in 2030 (Figure VI.9, panel a). This reduction amounts to less than 14,000 if unaware diabetics are excluded from the analysis (Figure VI.9, panel b). It is also worth noticing that the shapes of the trends in prevalence and absolute number of diabetic population do not vary despite the changes in the assumptions. The

only assumption modification that changes the shape is whether or not DM unawareness is included in the definition of DM.

A way of assessing the impact of KH on DM burden among Costa Rican elderly is by simulating a new projection, but assuming that the elderly that have short KH were actually taller: that is, the proportion of people with short KH is assumed to be zero in all the observed cohorts, and the rest of the proportions are prorated accordingly. When this simulation is made, the effect of short KH is close to -0.1 of percentage point of total prevalence during most of the period (Figure VI.10, panel a). In absolute numbers, the difference between the two simulations is negative, too. This result means that there would have been roughly 11,000 to 12,000 more diabetic older persons in 2030 if undernutrition (as signaled by short KH) had been eradicated since the end of the 19th century (Figure VI.11, panel a). This apparent contradiction is actually observed because people with KH under the first quartile have higher chances of surviving than taller people. Nonetheless, these are very small shifts in magnitude especially if considering that it is projected that in 2030 there will be almost 450 thousand diabetic elderly persons in the country, if DM hazard patterns remain the same throughout the period.

If only people with DM diagnosis are considered as having the disease, both the size of the diabetic population and DM prevalence would be slightly larger under the assumption of undernutrition eradication (Figures VI.10 and VI.11, panel b). Nonetheless, the main conclusion is that, under the assumption of no short KH among these cohorts, the prevalence and size of the diabetic population would remain the same and, thus, undernutrition in early life, as measured by short KH, does not seem to make a difference.

The other variable that is used to measure early childhood conditions is the CMI at the respondents' canton of birth. Here, instead of making CMI to be 0 for everybody (a level of

child mortality that has been achieved in no country in the world) to assess the impact of CMI on DM burden, the variable is set equal to the levels observed for the cohorts born in 1970, the cohorts with the lowest CMIs for the period under observation. Shifts in child mortality levels seem to have a stronger impact on DM prevalence over the period of projection than variations in KH (Figure VI.12). Regardless of the assumption of how mortality will vary in the future, if CMI had been lower than what it actually was during the first 70 years of the 20th century, then in 2015, DM prevalence would have been 5 percentage points higher throughout the period. The increase in the number of people with DM depends on the assumption of mortality behavior. If it is assumed that mortality remains constant, the size of the diabetic population is larger (in between 5 and 30 thousand people) than if mortality is assumed to decrease (Figure VI.13, panel a). If only persons with a DM diagnosis are taken into account, the difference in the prevalence increases over time, but in less than half of a percentage point. The difference in the absolute population size is also small.

Hypothetical shifts in prevalence and number of diabetic people attributed to the effect of early childhood conditions represent what Palloni *et al*. (2006) called the "tide to come" in chronic disease burden, and Finch and Crimmins (2004) called the "cohort morbidity phenotype": "Enduring effects of early environment, even if conditions improved at later periods" (1737). In the case of Costa Rica, these analyses imply that Costa Ricans, during the first part of the 20th century, had higher levels of undernutrition, as proxied by KH and CMI. Some of the survivors of those cohorts developed DM partially because of these adverse conditions. If these conditions had not existed, prevalence of DM and the total diabetic population might have been lower in present days and in the near future. The graphs discussed earlier show that the impact of early life conditions can be large or not depending on the

surrogate variable used to measure adverse early life conditions. Observing the projected trend of the difference is also useful in understanding this kind of "momentum". The difference in prevalence between the standard projection and the one with no short knee height tends to remain unchanged and negative, and the magnitude of the differences never exceeds half a percentage point (Figure VI.14, panel a). Differences due to short KH also remain negative if people that are unaware of their disease are excluded from the analysis (Figure VI.14, panel b). This means that exposition to malnutrition early in life seems not to have an important effect in current DM prevalence among Costa Rican seniors. The differences in prevalence using CMI as the measure for undernutrition are larger but still negative, and then tend to increase only among diagnosed diabetic elderly.

If the differences in the size of the diabetic population are observed rather than differences in prevalence, then the impact of early adverse conditions, as measured by CMI in the place of birth remains strong but even more negative during the period. If the construct is measured by KH, the pattern is still the same although the magnitude (in absolute values) is smaller (Figure VI.15). The negative direction in the difference is the result of lower total mortality which generates a larger population that is at similar risk of getting DM.

As a final subtopic in this chapter, there is also an analysis of the composition of the diabetic population throughout the period according to two additional risk factors that were controlled for in the projections procedure: sex and obesity. Distribution by sex remains roughly the same during the 25 year period of observation (Figure VI.16). The relative importance of men over women increases slowly from 46% in 2005 to almost 50% in 2030 (panel a and b). This trend occurs despite women having higher chances of developing DM and of surviving, and

it is partially related to the fact that males have a higher prevalence of other risk factors, particularly obesity.

Regarding this other important risk factor, the distribution of the diabetic population according to the three main categories (normal weight, overweight, and obesity) does not vary much either. Obese and overweight people represent the majority of the diabetic population throughout the period.

Finally, from a methodological point of view, controlling for other important risk factors –such as sex and weight categories– opens the possibility of simulating what might be the effect in DM burden if obesity and overweight prevalence grew.

D. Chapter summary

As a partial summary, there is an impact of early life conditions on DM burden but it is relatively small and, in absolute numbers, it has the opposite direction of what was expected. The impact is larger if CMI rather than KH is used as the surrogate for adverse early life conditions. Relaxation of the assumption of constant mortality or of using the equation with continuous variables (BMI and KH) rather than the one with categorical variables throughout the period does not make a difference in the projections.

The use of biomarkers for defining DM does make a difference, and the effect of adverse early life conditions is larger if “DM unawareness” or “undiagnosed DM” is taken into account in the analysis. If only people with diagnosed DM are included in the analysis, the impact of short KH or CMI is smaller. A hypothetical absence of adverse early life conditions increases

the burden of DM rather than decreasing it, especially when the burden is measured in absolute terms (size of the diabetic population) rather than in relative terms (DM prevalence).

E. Figures

Figure VI. 1. Costa Rica: Ages 60 and over. Projected diabetic population and projected diabetes prevalence, with constant diabetes hazard rates and constant mortality rates. 2005-2030. (Proportion of people with knee height under first quartile diminished linearly to be 6% in cohorts age 60 in 2030).

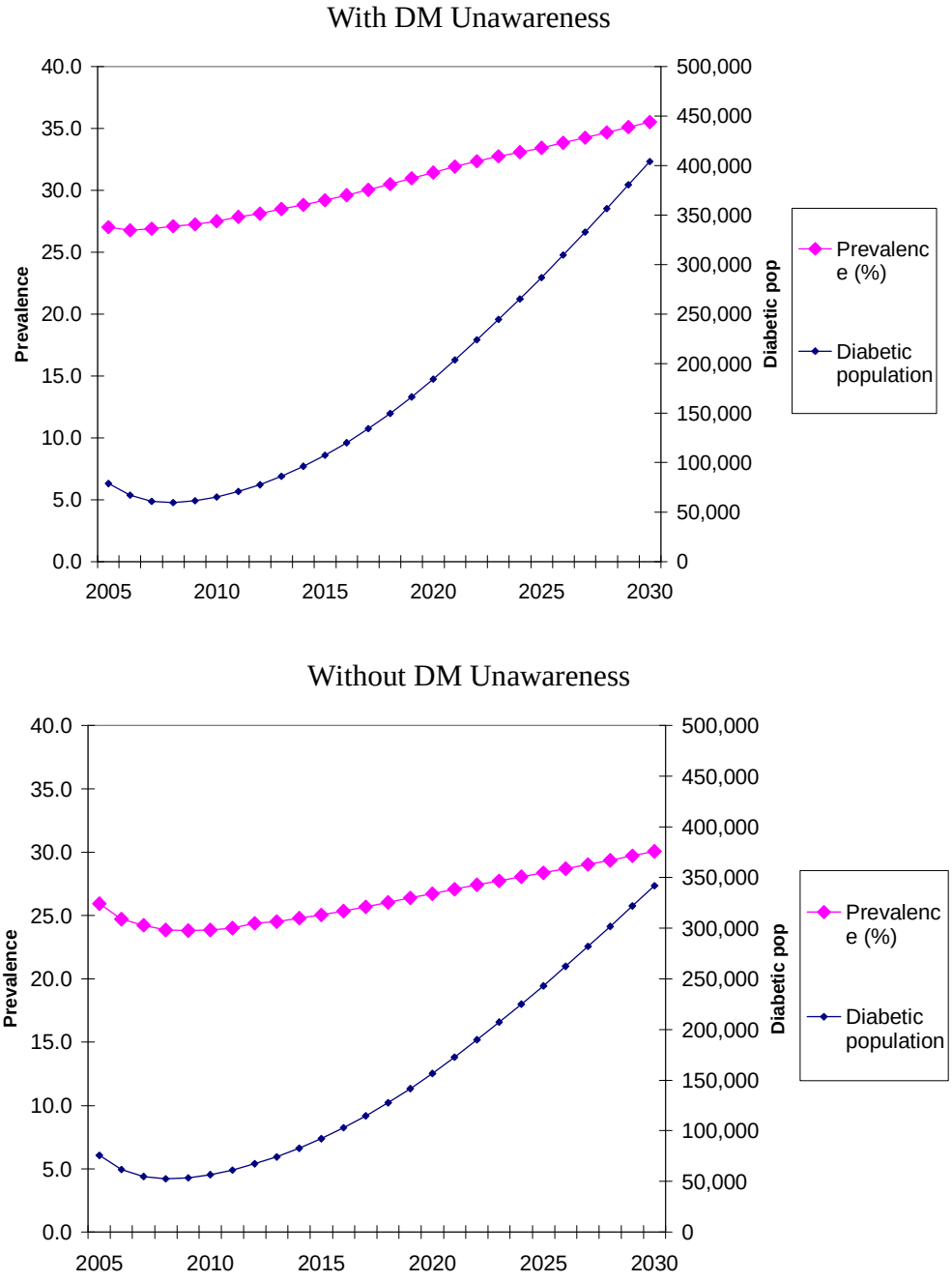


Figure VI. 2. Costa Rica: Ages 60 and over. Projected diabetic population and projected diabetes prevalence, with constant diabetes hazard rates and decreasing mortality rates according to official Costa Rican projections. 2005-2030. (Proportion of people with knee height under first quartile diminished linearly to be 6% in cohorts age 60 in 2030).

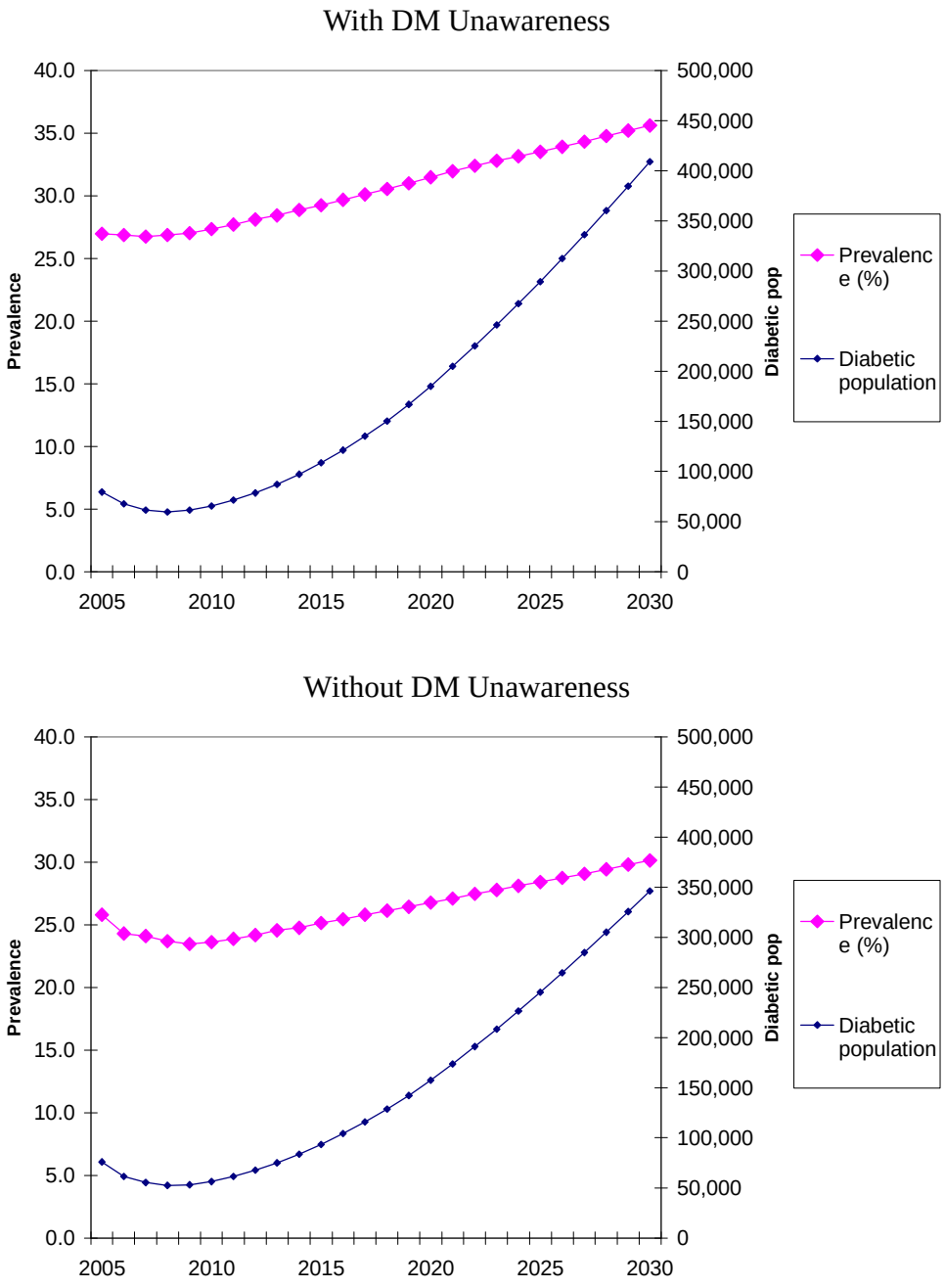


Figure VI. 3. Costa Rica: Ages 60 and over. Projected diabetic population and projected diabetes prevalence, with constant diabetes hazard rates and decreasing mortality rates in two scenarios: (a) Using hazard rates estimated from equations in which knee height is a categorical, and (b) using hazard rates estimated from equations in which knee height and BMI are continuous variables. 2005-2030. (Proportion of people with knee height under first quartile diminished linearly to be 6% in cohorts age 60 in 2030)

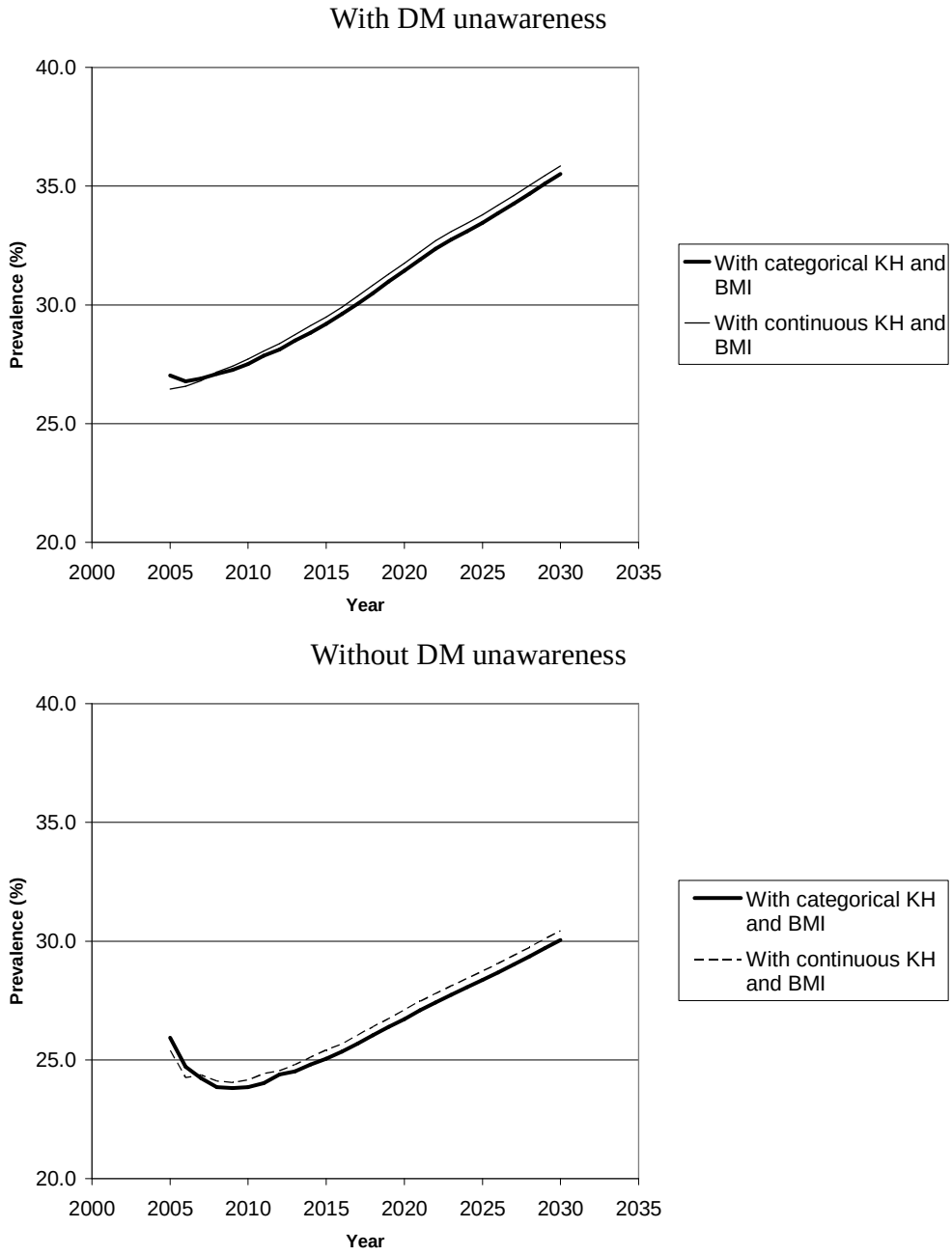


Figure VI. 4. Costa Rica: Ages 60 and over. Projected diabetes prevalence, by knee height groups, with constant diabetes hazard rates and constant mortality rates. 2005-2030. (Proportion of people with knee height under first quartile diminished linearly to be 6% in cohorts age 60 in 2030).

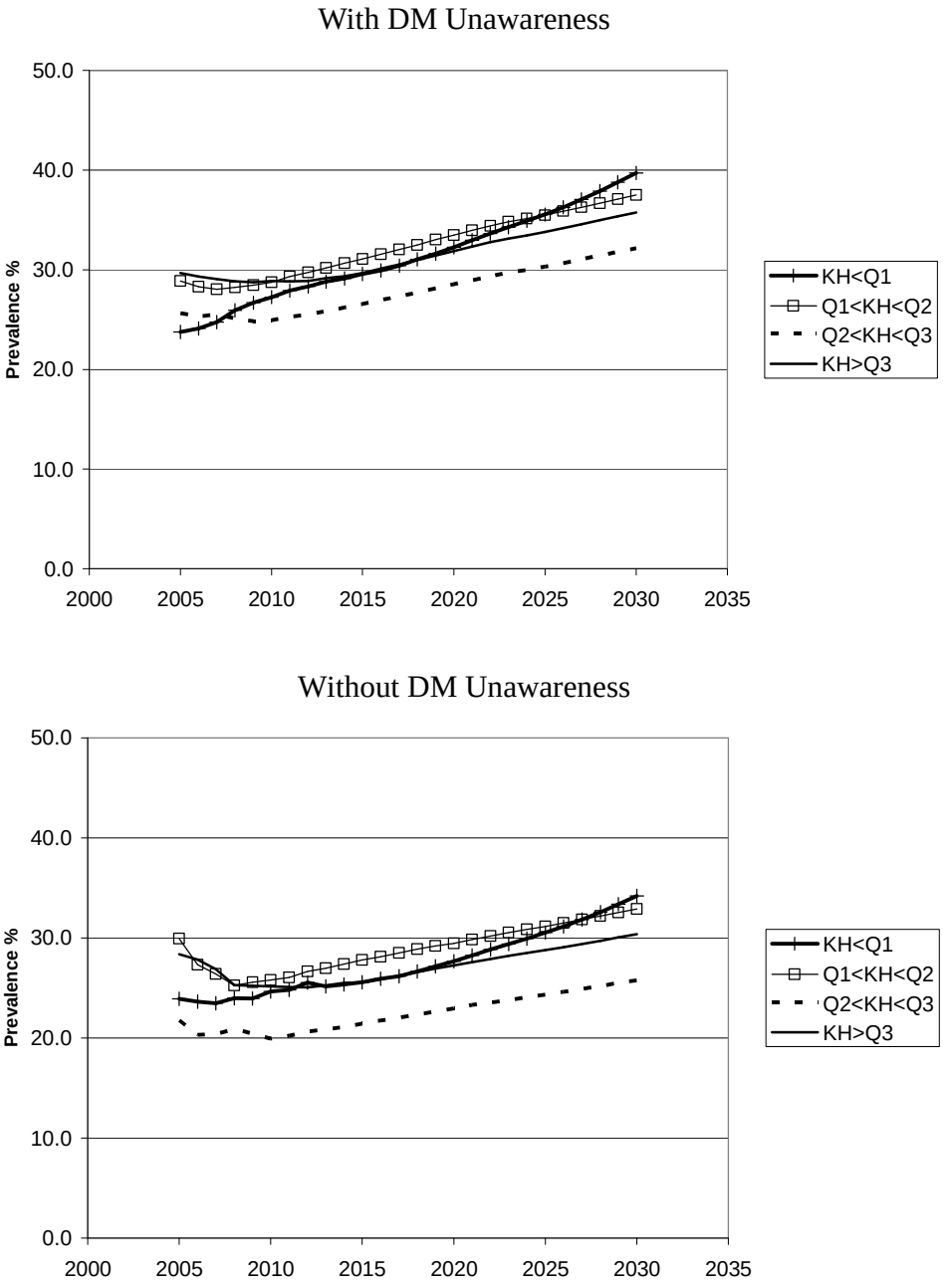
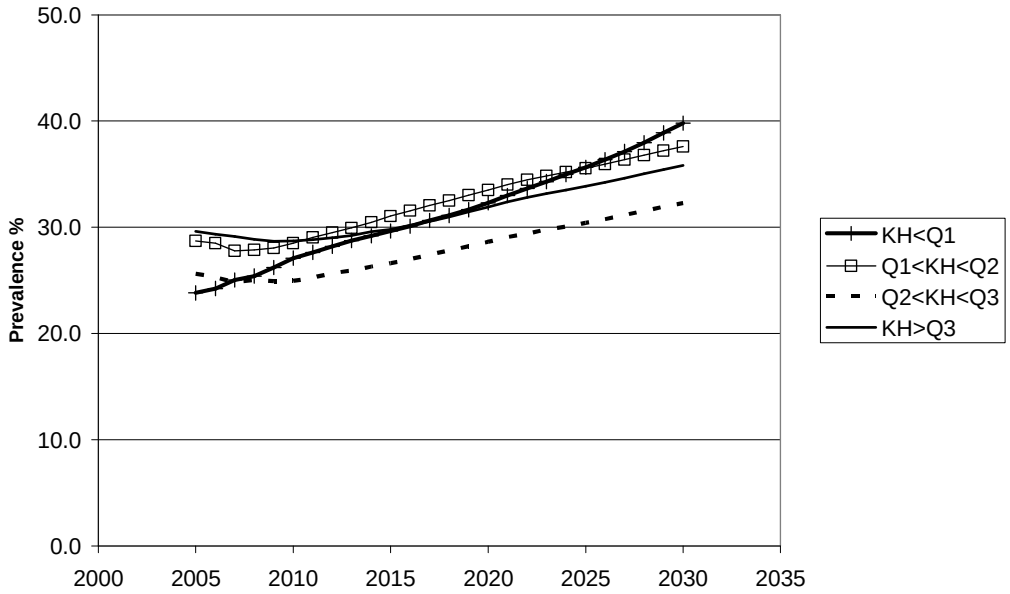


Figure VI. 5. Costa Rica: Ages 60 and over. Projected diabetes prevalence, by knee height groups, with constant diabetes hazard rates and decreasing mortality rates according to official Costa Rican projections. 2005-2030. (Proportion of people with knee height under first quartile diminished linearly to be 6% in cohorts age 60 in 2030).

With DM Unawareness



Without DM Unawareness

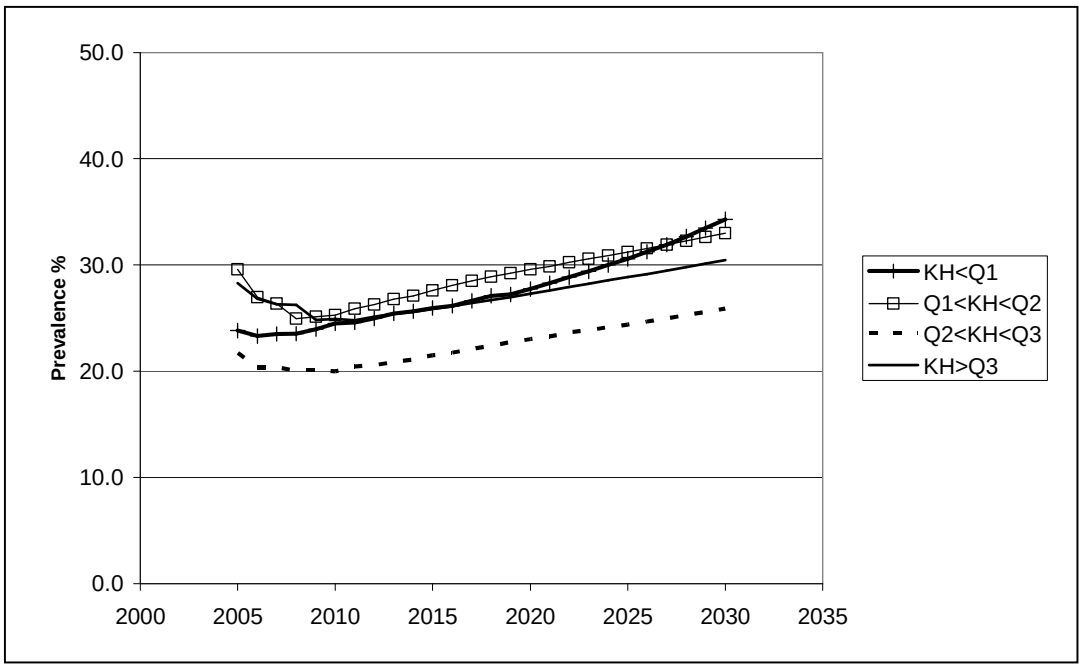


Figure VI. 6. Costa Rica: Ages 60 and over. Projected relative distribution of people with diabetes, by knee height groups, with constant diabetes hazard rates and constant mortality rates. 2005-2030. (Proportion of people with knee height under first quartile diminished linearly to be 6% in cohorts age 60 in 2030).

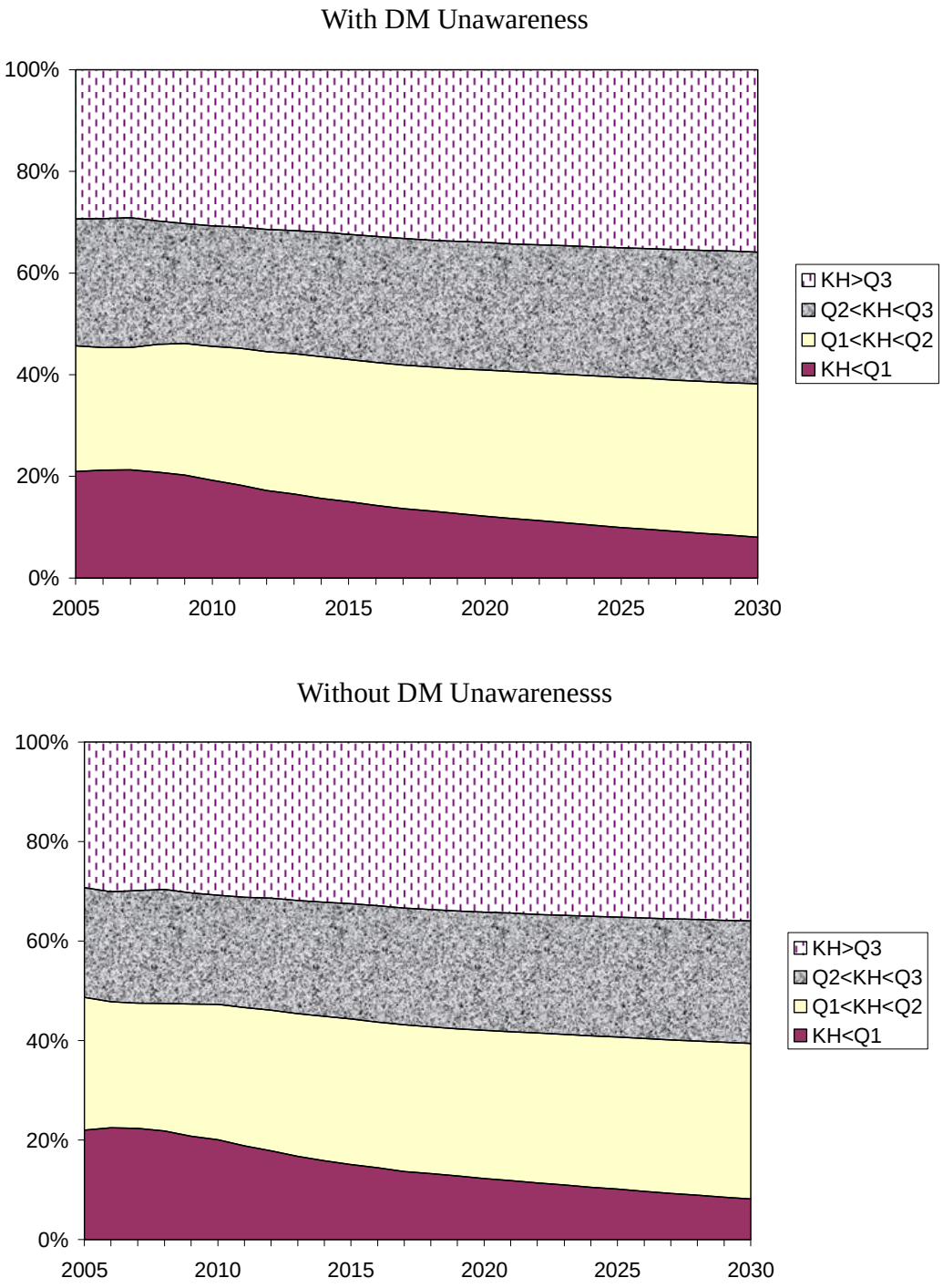


Figure VI. 7. Costa Rica: Ages 60 and over. Projected relative distribution of people with diabetes, by knee height groups, with constant diabetes hazard rates and decreasing mortality rates according to official Costa Rican projections. 2005-2030. (Proportion of people with knee height under first quartile diminished linearly to be 6% in cohorts age 60 in 2030).

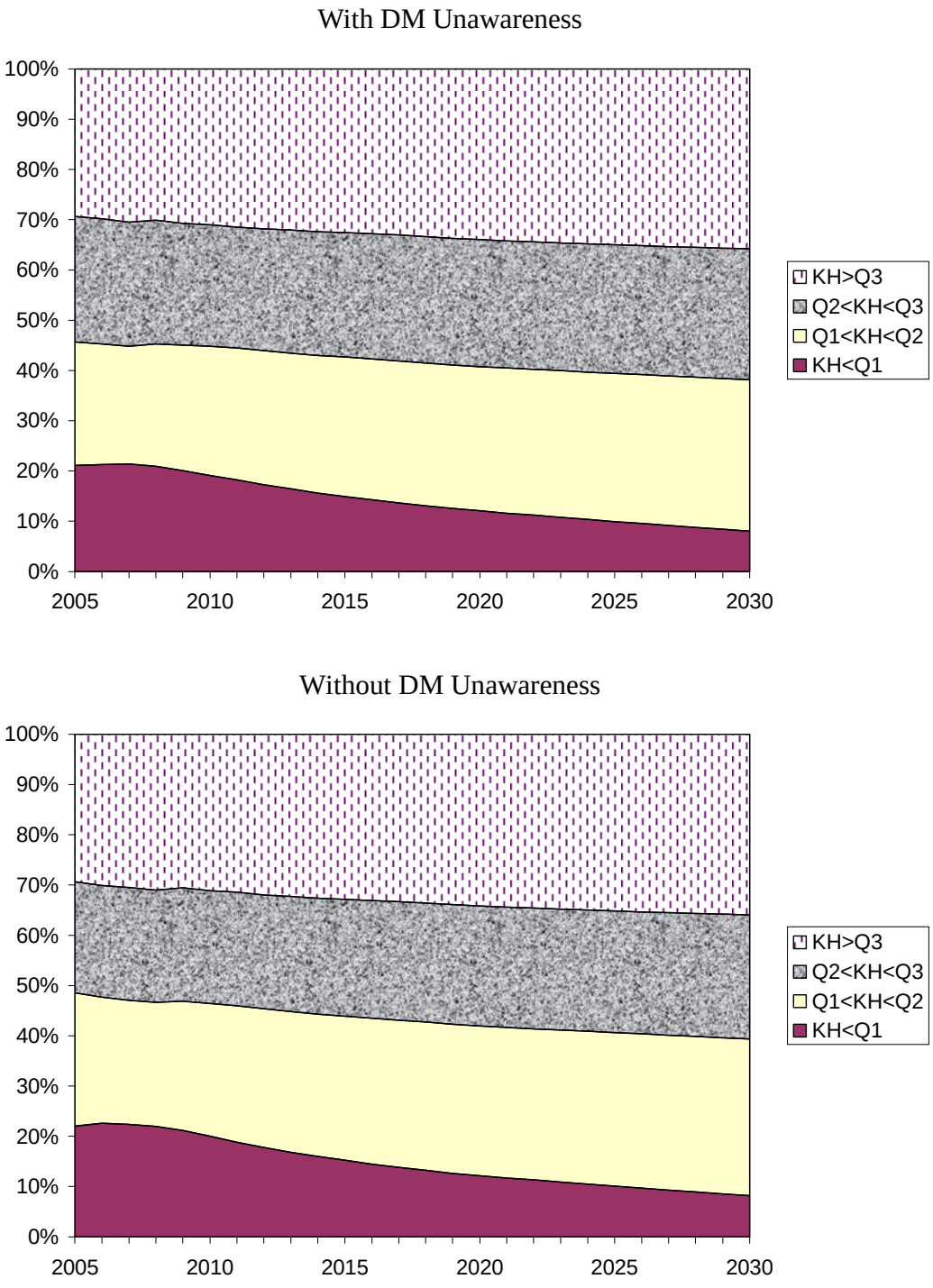


Figure VI. 8. Costa Rica: Ages 60 and over. Projected diabetes prevalence, with constant diabetes hazard rates. 2005-2030. Four scenarios: With constant knee height distribution and decreasing short knee height proportion (pro-rated proportion for the rest of knee height categories), each by constant or decreasing mortality rates.

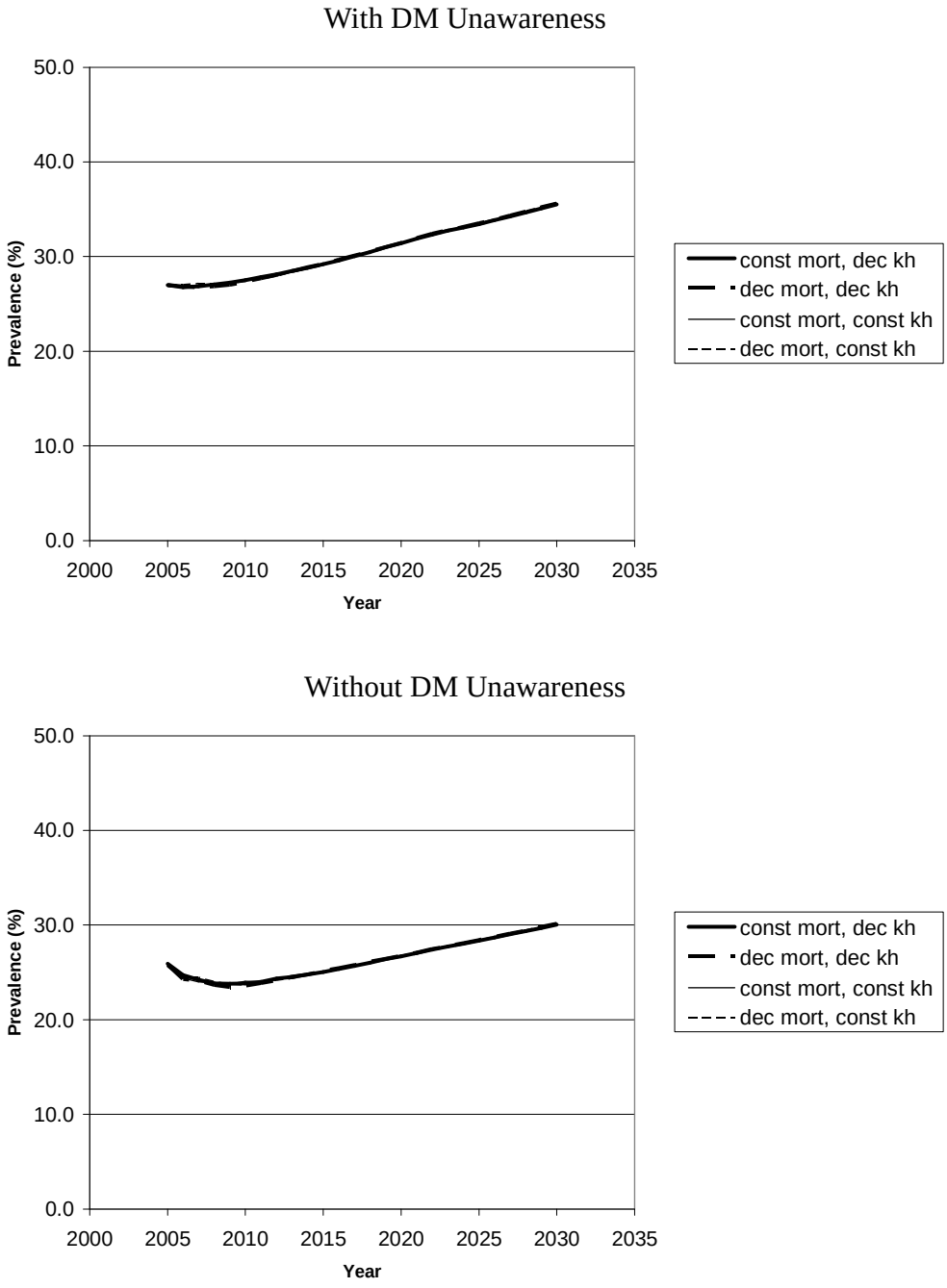


Figure VI. 9. Costa Rica: Ages 60 and over. Projected diabetic population, with constant diabetes hazard rates. 2005-2030. Four scenarios: With decreasing short knee height proportion and proportion with short knee height equal zero in all cohorts (pro-rated proportion for the rest of knee height categories), each by constant or decreasing mortality rates.

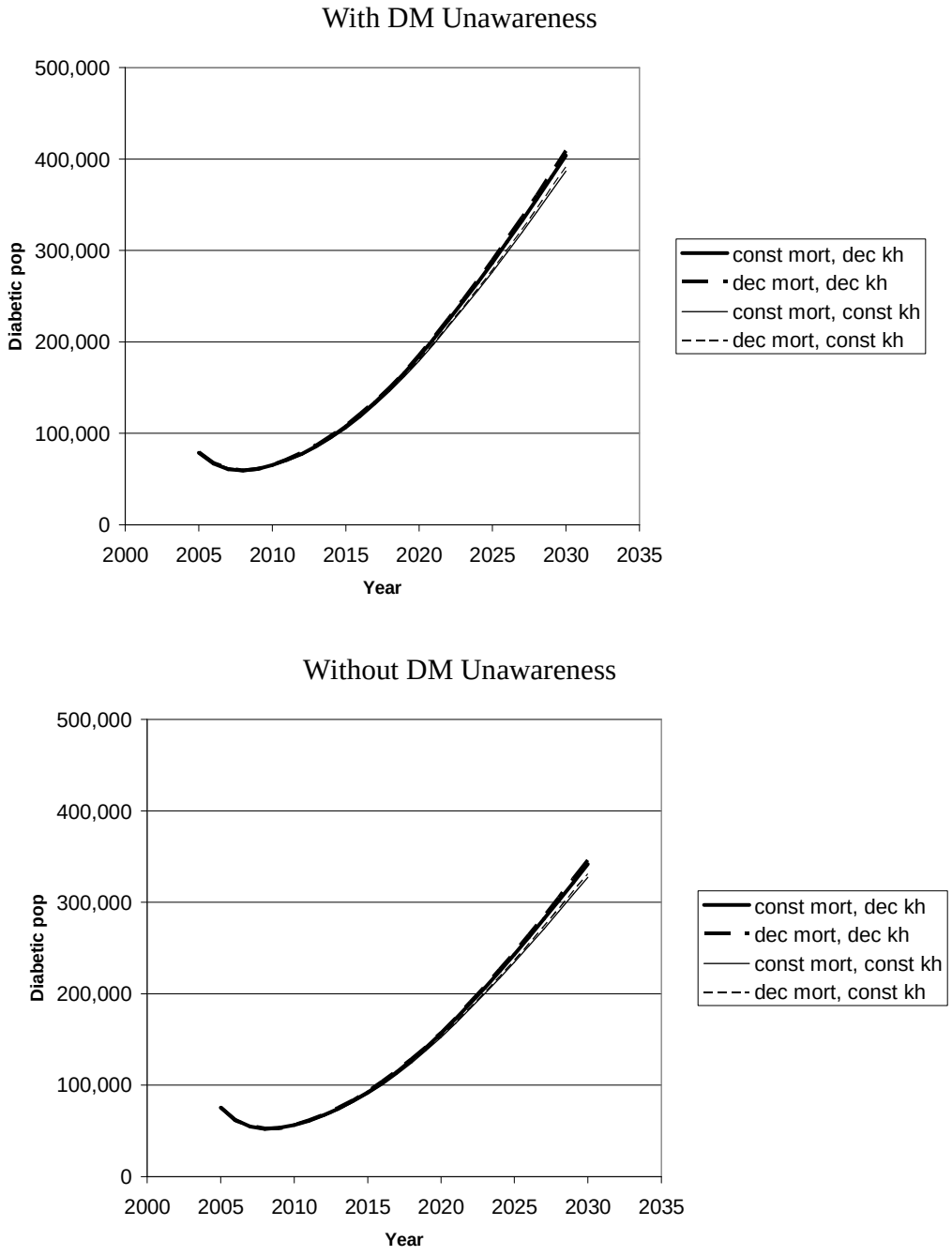


Figure VI. 10. Costa Rica: Ages 60 and over. Projected diabetes prevalence, with constant diabetes hazard rates. 2005-2030. Four scenarios: With decreasing short knee height proportion and proportion with short knee height equal zero in all cohorts (pro-rated proportion for the rest of knee height categories), each by constant or decreasing mortality rates.

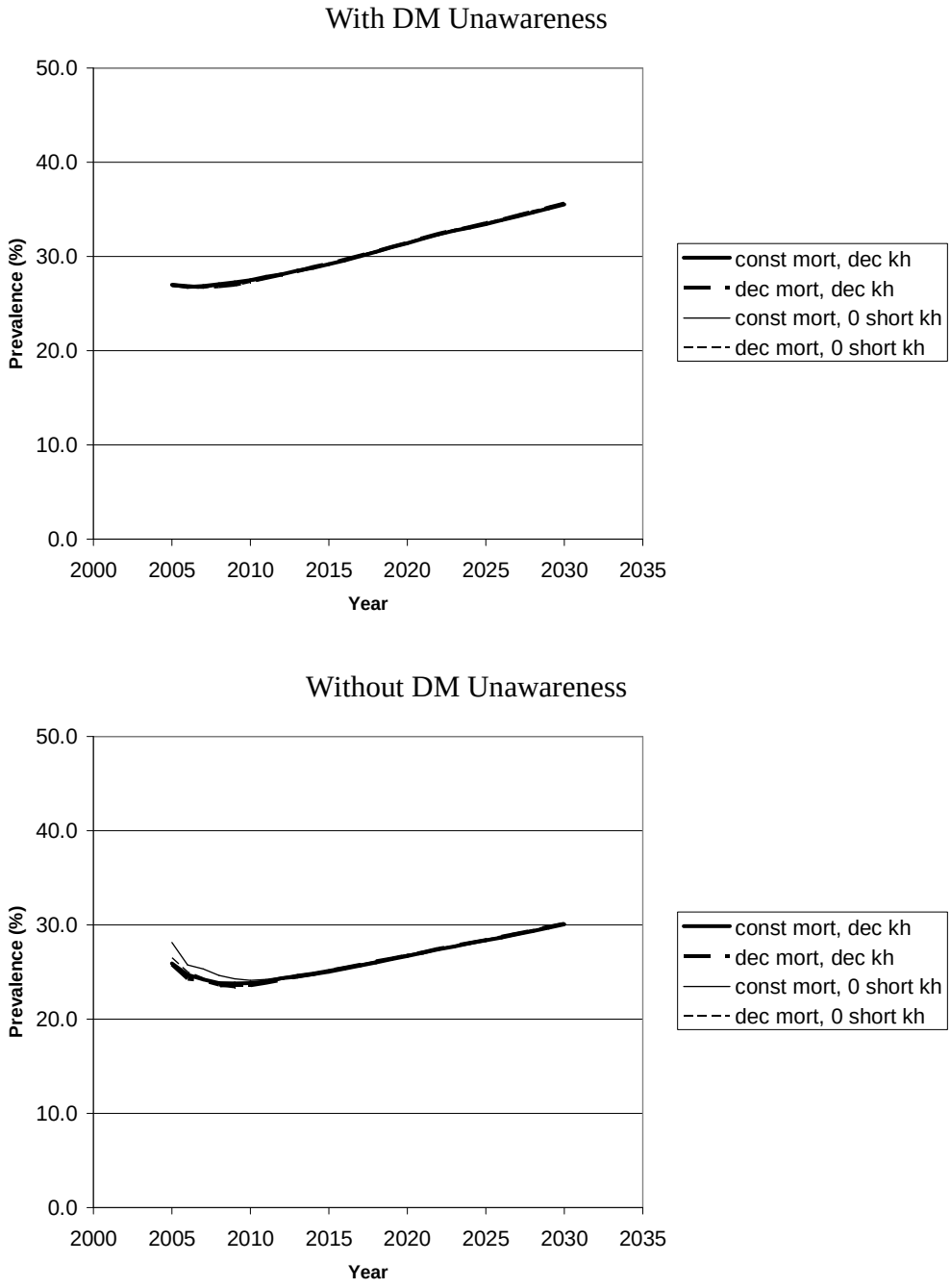
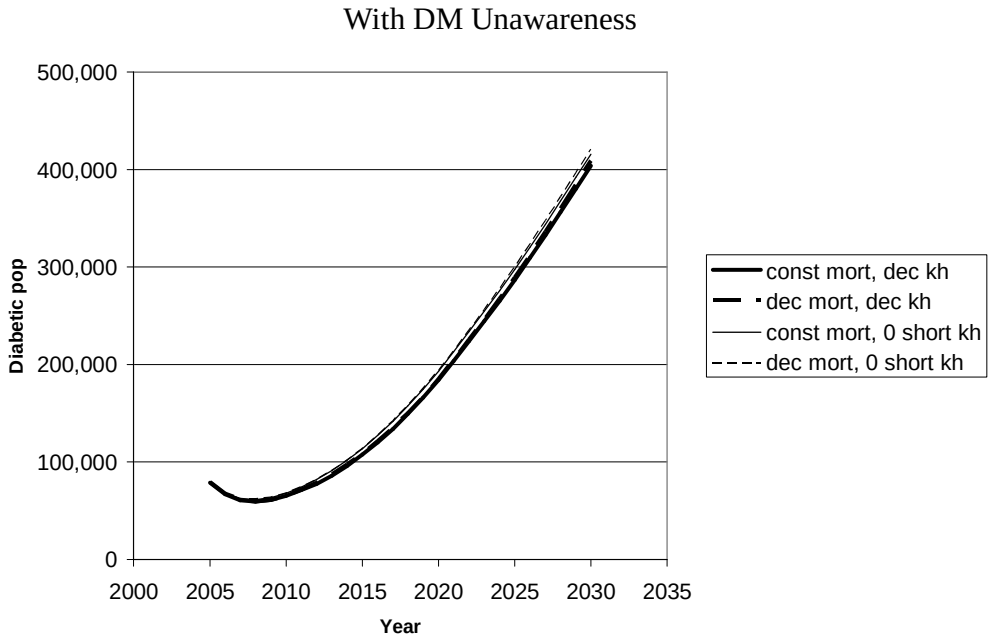


Figure VI. 11. Costa Rica: Ages 60 and over. Projected diabetic population, with constant diabetes hazard rates. 2005-2030. Four scenarios: With decreasing short knee height proportion and proportion with short knee height equal zero in all cohorts (pro-rated proportion for the rest of knee height categories), each by constant or decreasing mortality rates.



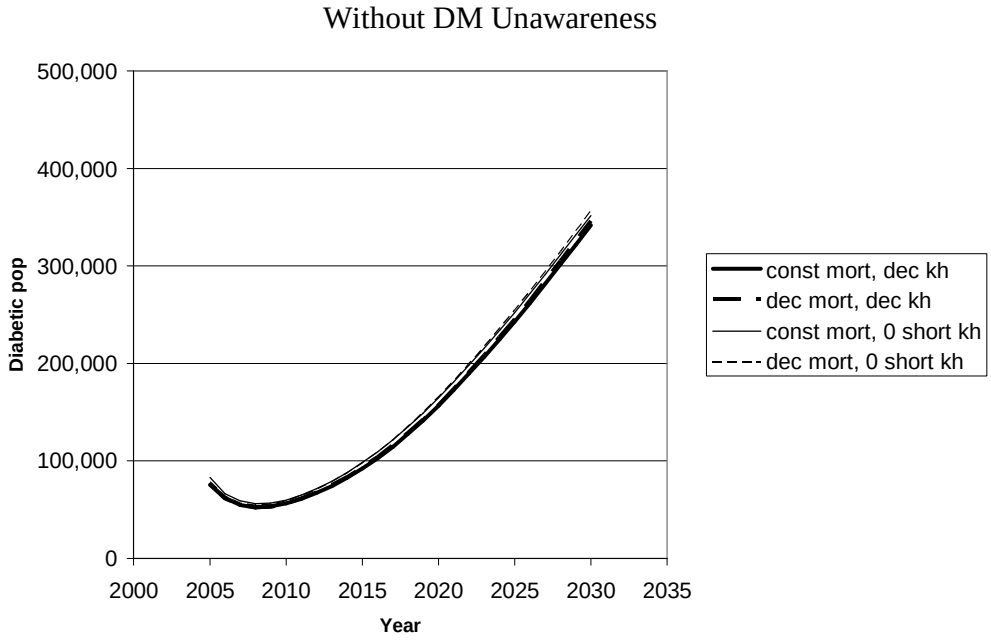
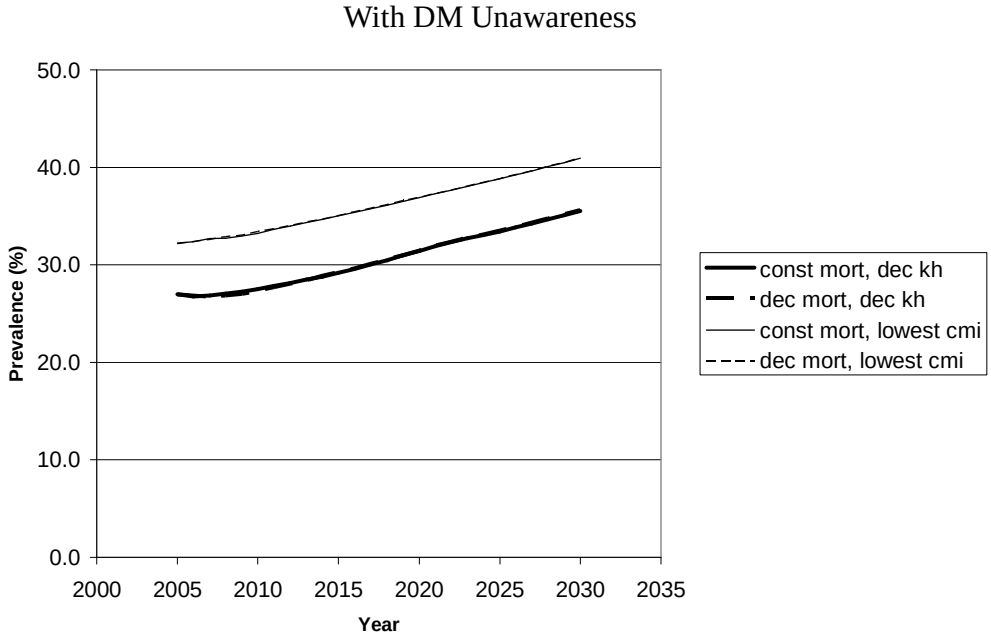


Figure VI. 12. Costa Rica: Ages 60 and over. Projected diabetes prevalence, with constant diabetes hazard rates and constant mortality rates. 2005-2030. Four scenarios: With observed Child Mortality Index CMI throughout the period, and with observed CMI of cohorts age 60 in 2030 imputed to older cohorts, each by constant or decreasing mortality rates (CMI as a continuous variable), keeping constant KH distribution.



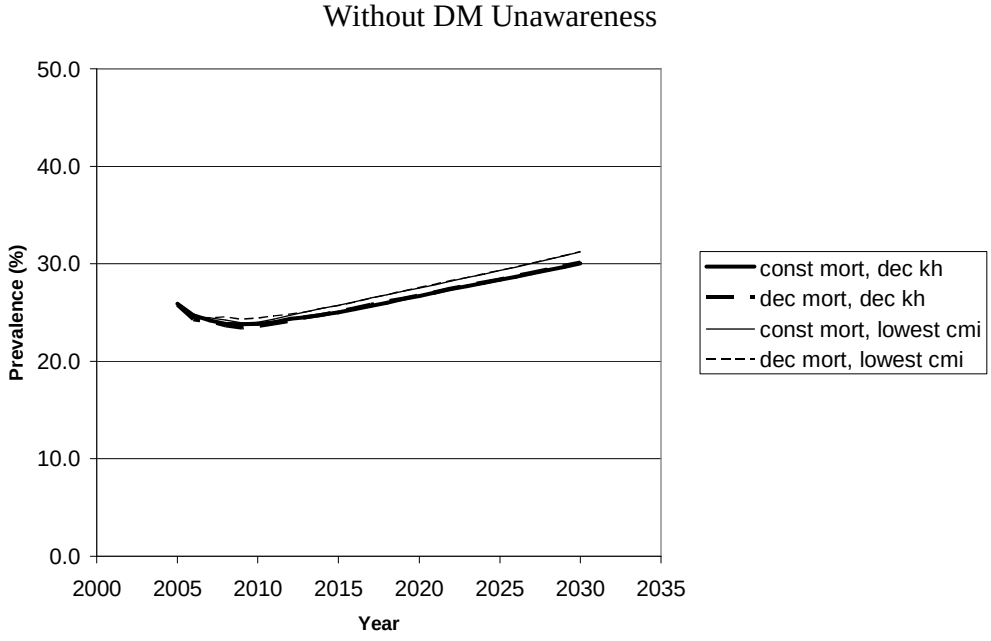


Figure VI. 13. Costa Rica: Ages 60 and over. Projected diabetic population, with constant diabetes hazard rates and constant mortality rates. 2005-2030. Four scenarios: With observed Child Mortality Index CMI throughout the period, and with observed CMI of cohorts age 60 in 2030 imputed to older cohorts, each by constant or decreasing mortality rates (CMI as a continuous variable), keeping constant KH distribution.

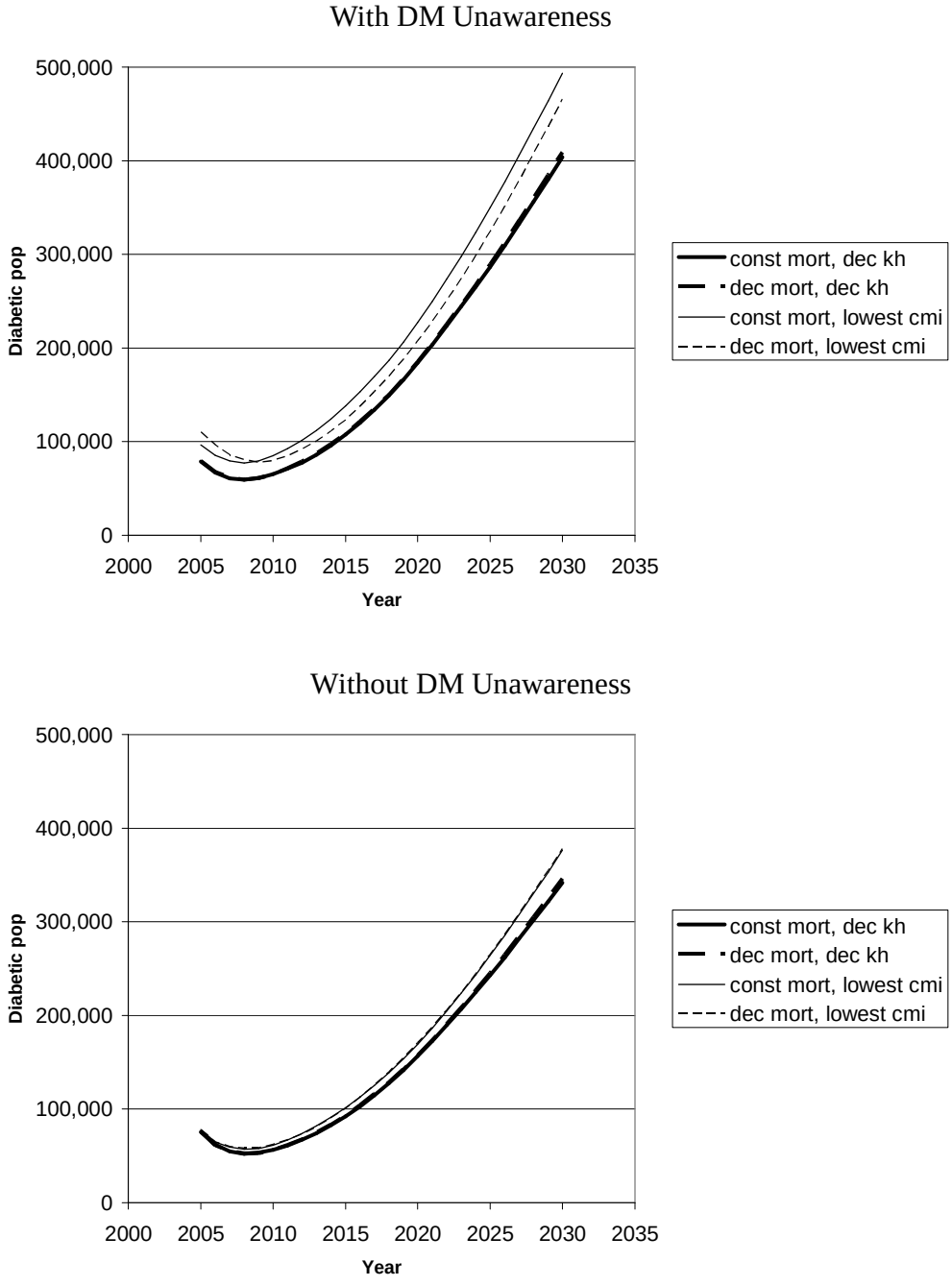


Figure VI. 14. Costa Rica: Ages 60 and over. Projected differences in prevalence due to the effect of short KH or due to the effect of high CMI, by constant or decreasing mortality.

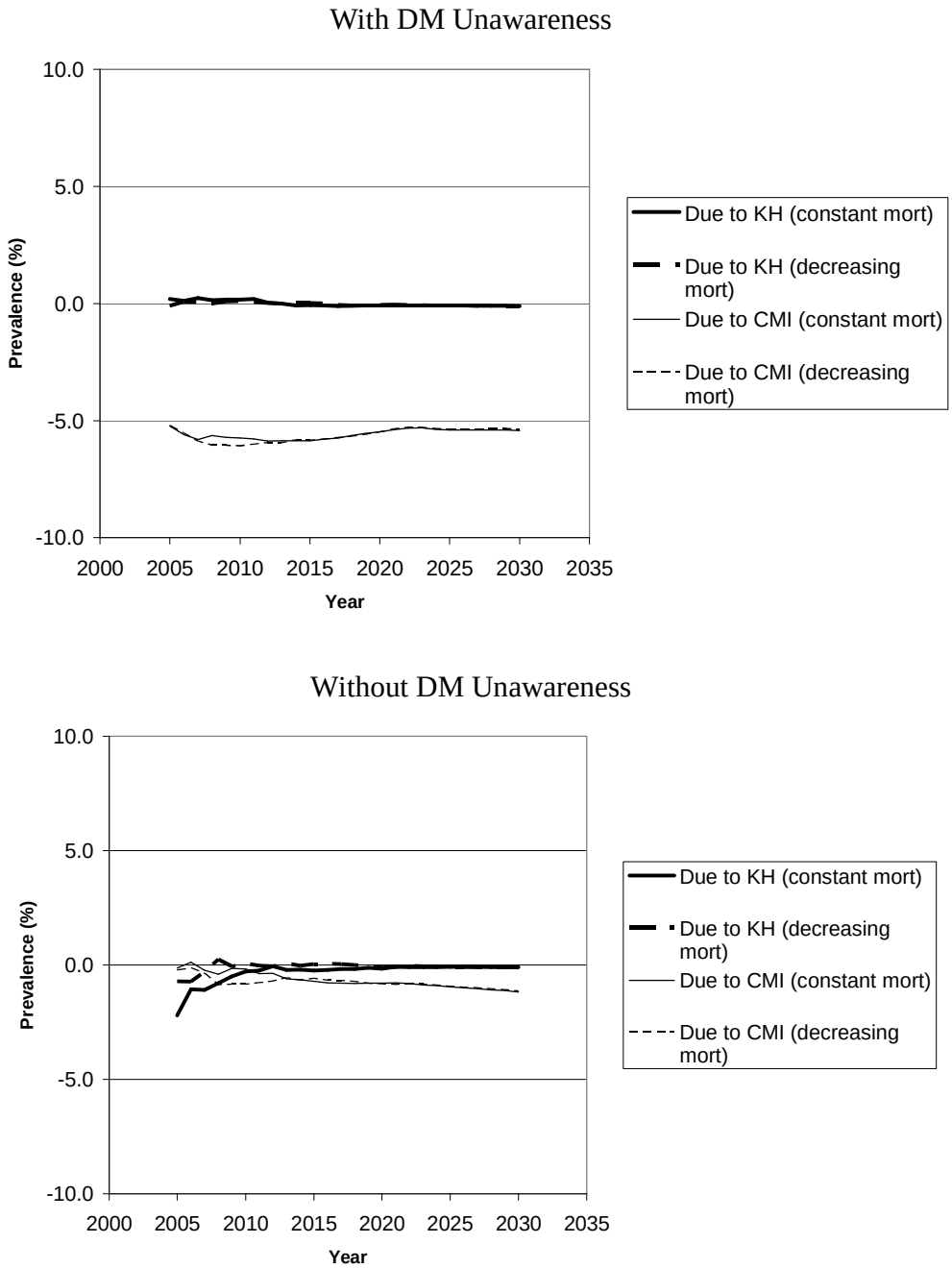


Figure VI. 15. Costa Rica: Ages 60 and over. Projected differences in size of diabetic population due to the effect of short KH or due to the effect of high CMI, by constant or decreasing mortality.

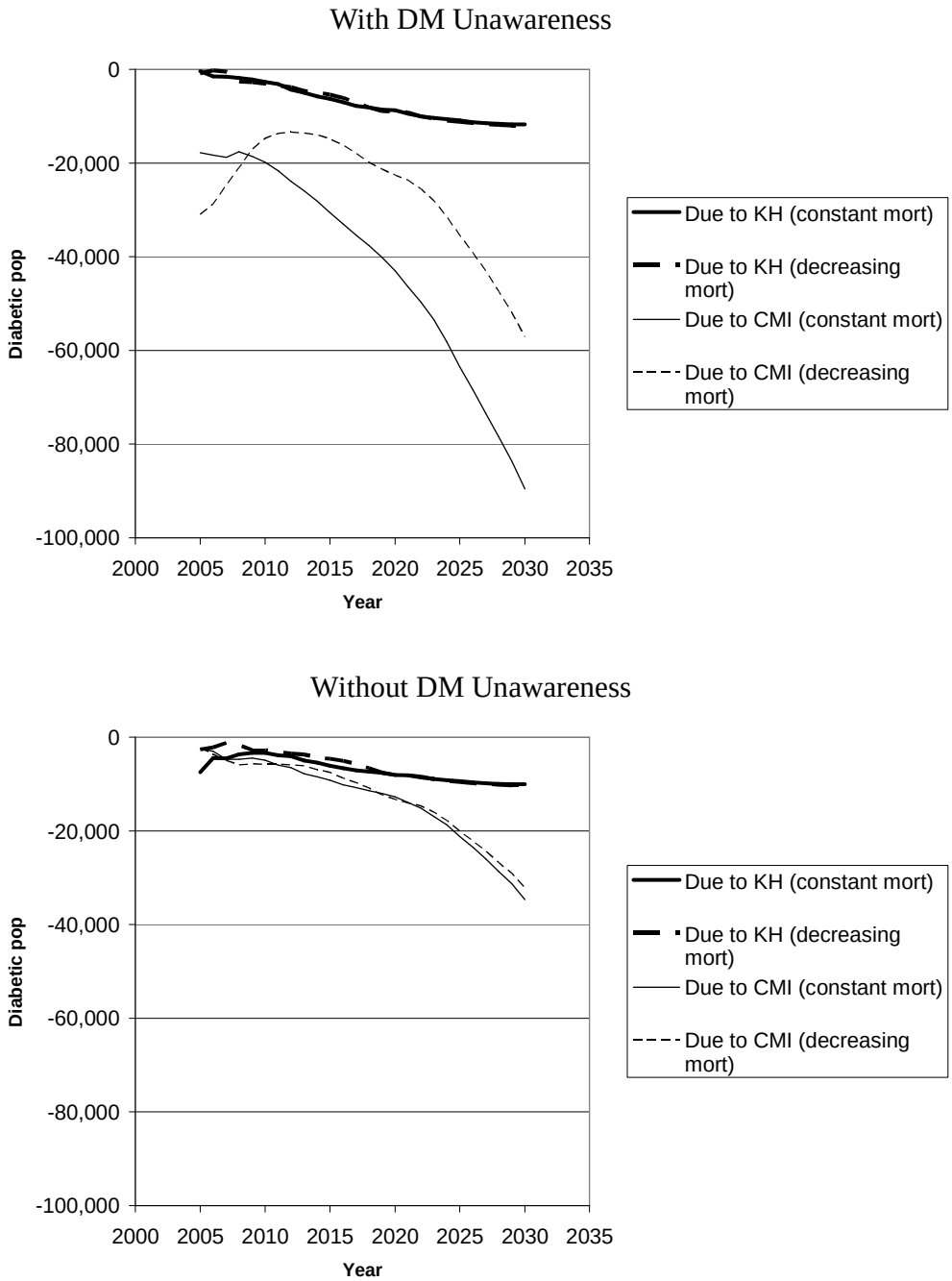


Figure VI. 16. Costa Rica: Ages 60 and over. Projected relative distribution of people with diabetes, by sex, with constant diabetes hazard rates and constant mortality rates. 2005-2030. (Proportion of people with knee height under first quartile diminished linearly to be 6% in cohorts age 60 in 2030).

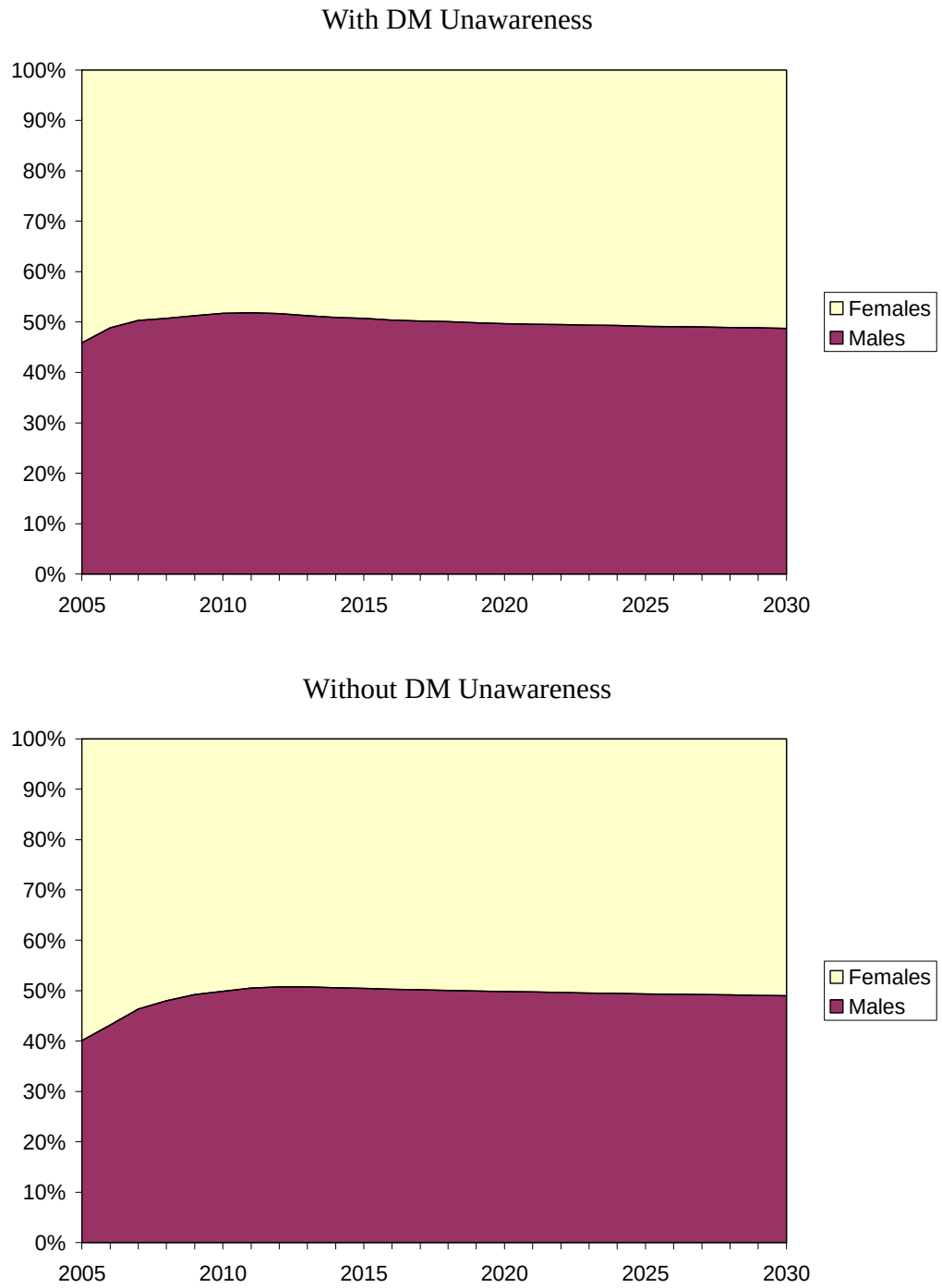
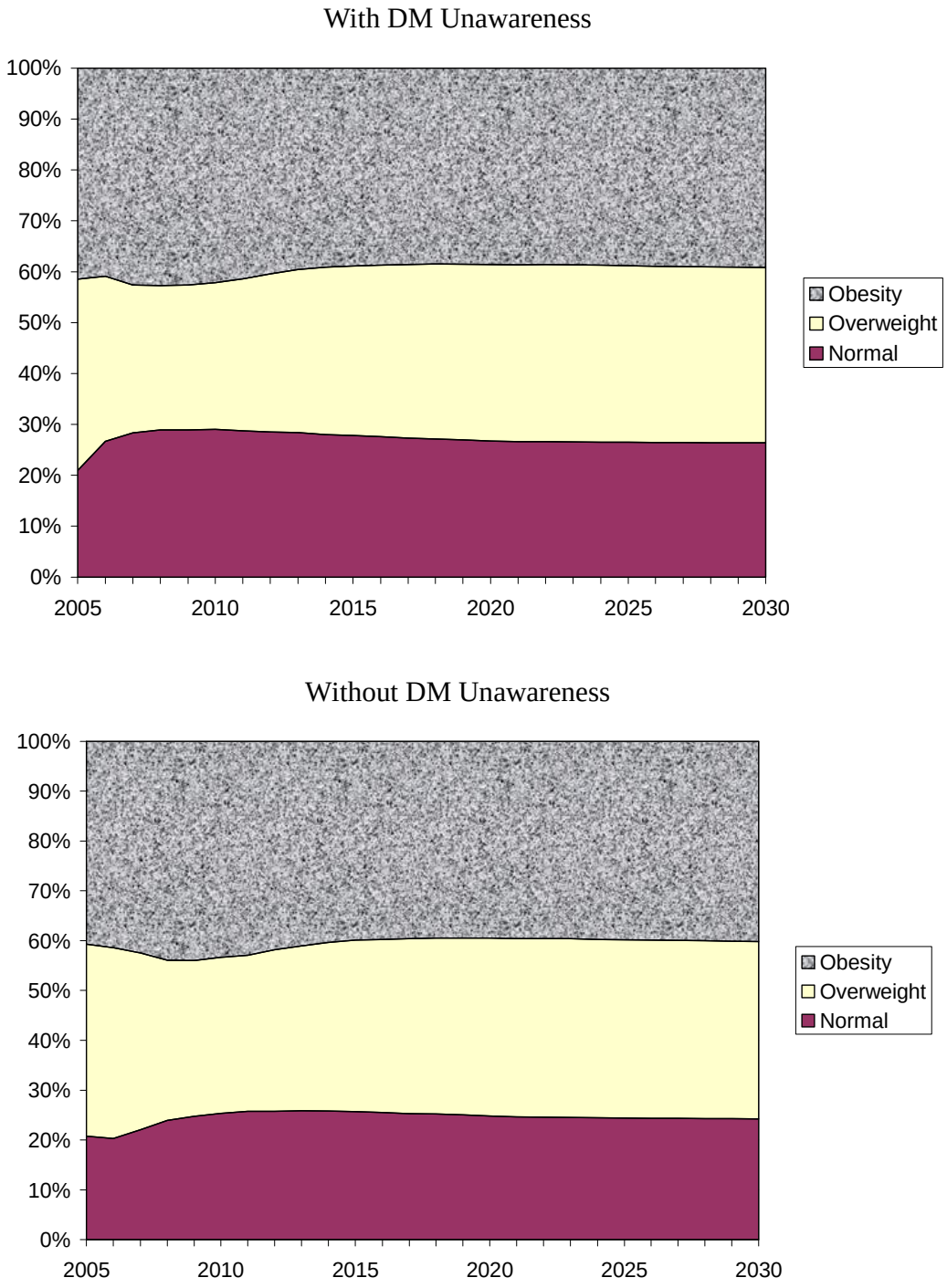


Figure VI. 17. Costa Rica: Ages 60 and over. Projected relative distribution of people with diabetes, by obesity groups (normal weight, overweight, and obesity), with constant diabetes hazard rates and constant mortality rates. 2005-2030.



Chapter VII: Discussion

The goal of this dissertation was to explore the effects of early life conditions on Diabetes Mellitus (DM) in Costa Rica. Given that there is no longitudinal study in Costa Rica that has direct measures of early life adverse conditions and morbidity later in life, I used surrogate measures to approach the concept of adverse early life conditions. In this section, I will discuss the results of this project following the outline presented below:

- I will first review some of the main theoretical arguments that motivated this dissertation's topic.
- I will describe the advantage of studying the effects of early life conditions by separately analyzing their association with incidence rates and mortality.
- I will then discuss the utility of the results related to the incidence of DM and the possible reasons for finding such a weak association.
- The next section will argue why the results of the mortality model are important.
- I will follow with an analysis of the utility of having biomarker information available for this study, and the problems of estimating an incidence rate of undiagnosed DM or DM unawareness.
- I will discuss what the implications of the dissertation's results are to the theoretical arguments on which this research is based.
- I will follow this discussion with a list of other relevant limitations faced in this research.
- I will close this section with suggestions for future research that can be drawn the results of this dissertation.

A. Review of the theoretical arguments that lead to this project

The "life course" approach in research on chronic disease etiology has been developed primarily through research projects studying populations in industrialized countries. The framework has been relevant in this region of the world because the burden of communicable diseases is small and their health care systems are dealing with increasing numbers of adult-age people who are affected by chronic diseases such as cardiovascular diseases, DM, and cancer.

However, some authors (Cubillos-Garzon *et al.*, 2004; Palloni *et al.*, 2006; Prentice and Moore, 2005) have reasoned that the processes described by some versions of the life course approach have severe implications for the health transition in developing countries. Palloni *et al.* (2006) explain that current cohorts of elderly in Latin America experienced on average very particular health trajectories because they survived their adverse childhood conditions thanks to medical innovation rather than to improved nutrition and socio-economic circumstances, and therefore –under the assumption that adverse early life conditions are risk factors for non-communicable illnesses– they have higher probabilities of being affected by certain chronic diseases such as DM and heart disease. Prentice and Moore (2005) contend that similar conditions are occurring in Asia and even in Africa, "… particularly so in countries passing through a rapid economic and nutritional transition, or in peoples from poor countries who migrate to wealthy ones" (p.430). Both groups of authors rely on Hales and Barker's "thrifty phenotype hypothesis" (Hales and Barker, 2001), according to which children who were born with low birth weight or low ponderal index, or were too thin during infancy, are more likely to develop Type 2 DM. The assumed biological mechanism relates conditions of low nutrient

intake during gestation to impaired beta cell production, which in turn leads to insulin resistance. Whereas the first versions of the thrifty phenotype hypothesis linked early undernourishment with DM, later versions stressed the importance of obesity or rapid weight gain as important mediators or interacting factors for linking gestation and infancy conditions and chronic disease.

This dissertation relies on how Prentice and Moore and Palloni *et al.* link the thrifty phenotype hypothesis with conditions in the developing world, primarily because this dissertation deals with DM. However, there is another line of research that relates adverse early life conditions with chronic disease etiology. This other line of research argues that surviving a highly infectious environment during childhood, adolescence and early adulthood leads to chronic inflammation and, subsequently, to a higher risk of cancer (Gersten and Wilmoth, 2002), and cardiovascular disease (Finch and Crimmins, 2004; Crimmins and Finch, 2006; Cubillos-Garzon *et al.*, 2004). Even some results in Palloni *et al.* (2006) provide evidence that support this hypothesis because these authors find the expected link between rheumatic fever during childhood and greater prevalence of heart disease at old ages.

B. Reasons for analyzing and estimating incidence and mortality separately

The objectives pursued in this dissertation, as well as data characteristics, resemble the analysis conducted by Palloni *et al.* (2005, 2006) with information from 6 cities in Latin America and a dataset from Puerto Rico. One of the main explanatory variables is the same: knee height as a marker of early life undernutrition. However, the outcome variable is different. These authors analyze the effect of several early life markers on the prevalence, rather than the incidence, of DM and heart disease. I started the analysis with similar models on DM prevalence

for the sake of comparability with their results. However, I also chose to analyze the effect of KH and CMI on incidence rates and on mortality to understand whether the possible association with prevalence is the result of separate processes in the incidence and mortality of the elderly in Costa Rica. Why is this relevant?

Under stationary conditions, there is the traditional epidemiological identity that states that:

$$\text{Prevalence} = \text{Incidence} * \text{Duration}$$

Given that duration is associated with the lethality of the disease and other mortality conditions, disease prevalence is a function of both the incidence and mortality schedules of the population.

According to this dissertation, among Costa Rican persons age 60 and above, KH has a negative, weak, but statistically significant association with DM hazards, and only when KH interacts with BMI levels. However, keeping everything else constant, people with short KH live longer than people with longer KH. Therefore, the weak but not significant association with DM prevalence is more strongly explained by longer duration with the disease among short KH Costa Ricans rather than by higher incidence rates. This is more directly observed in the projections, regardless of whether they are estimated with KH as a continuous variable or as a categorical variable. The prevalence of DM among people with KH under the first quartile keeps increasing over time even though the share of this group within the diabetic population decreases over time as the share within the total population also decreases. In 2030, DM prevalence among people with short KH is 30% higher among people with KH between the second and the third quartile, even though the hazards of getting DM among short KH seniors are just 20% higher

than the reference group when KH categories are used. When continuous variables are used to operationalize KH and BMI, the DM hazard among short KH persons is more than 30% higher only in extreme obesity levels (BMI≥35). This means that higher prevalence in the projected 2030 short KH population can be explained by greater durations with the disease. How much of the association found by Palloni *et al.* (2006) might be due to differential mortality rather than differential risks in getting the disease? Further research is needed with new Latin American datasets that have already made information on mortality available to researchers, as in the case of MHAS and PREHCO.

With respect to CMI, the other surrogate for adverse early life conditions, the degree of association between it and DM incidence is always very small and statistically non-significant. However, people born in cantones with high child mortality are more likely to die than people born in cantones with low child mortality. This translates into a very weak association between DM prevalence and child mortality in the respondent's place of birth.

C. The association between early life conditions and DM incidence

As mentioned above, if KH is operationalized as a set of dichotomous variables, no statistically significant effect is found for short KH on DM hazard rates. A significant effect is found when both KH and BMI are operationalized as continuous variables, and an interaction term is included in the model. Notice that obesity is not exactly mediating in the effect but interacting. This means that a higher DM hazard of mortality among people with short KH is found only if the persons are obese. Otherwise, the association might not exist or might be in the opposite direction.

The result of the interaction matches what the literature explains: the persons that are more likely to develop Type 2 DM are those who had low birth weight followed by obesity in adulthood (Forsen *et al.*, 2000; Hales and Barker, 1992, 2001). However, based on the Costa Rican data and assuming KH as a marker of impaired fetal and childhood growth, higher risk of developing DM among people with short KH only occurs strongly at very high BMI levels. If the population is classified in the 12 categories resulting from combining obesity categories (obese, overweight, normal) and KH categories (4 categories, given that they were created based on quartiles), and if mean BMI and KH are calculated for each of these 12 categories, plugging these values into the equation yields a relative risk of barely 1.15 for people with short KH as compared to people with KH between the second and the third quartile. A relative risk of 1.56 is achieved only at a BMI of 40 Kg/m2, and using the minimum and maximum KH values for the extreme categories. This relative risk is still rather small compared to relative risks found in the literature for the effect of low birth weight (without interaction and net of other effects) on DM: Some of these relative risks range from 1.37 to 3.52 (Boyko, 2000).

Were these weak results expected? In their analysis of 6 Latin American cities plus Puerto Rico, Palloni *et al.* (2005, 2006) only found statistically significant results in Santiago, Mexico City, and Puerto Rico, and the authors also conclude that the odds ratios for DM prevalence that they found represent a weak association. It is worth while to highlight the fact that, in their models, the analyses control for obesity using a dichotomous variable. However, in all three places where the effect of short KH was significant, prevalence of obesity is higher than in Costa Rica as a whole. When compared to the SABE cities, obesity prevalence in Costa Rica is in the middle, close to Barbados, higher than in Brazil and Cuba, and lower than in Mexico, Chile, and Uruguay. Would Palloni *et al.* (2005, 2006) have found more significant associations

if they had operationalized the explanatory variables as continuous and had included an interaction term?

Regardless of the operationalization, it is possible to conclude that the degree of association between KH as a surrogate of adverse early life conditions and DM incidence is weak and, therefore, does not make much difference in determining the burden of DM due to these conditions. Again, this is more evident in the projection results. The projections used the estimated DM hazard rates and the estimated mortality hazard rates as inputs. When creating the two scenarios –one that keeps the observed KH distribution, and the other that assumes that there were no people with short KH among these cohorts– the differences in the estimated DM prevalence throughout the 25 years do not even exceed 1 percentage point. These differences would have been higher if the relative risks of DM were higher.

In terms of the effect of CMI on DM hazard rates, its associated parametric Weibull regression coefficient is not significantly different to zero. It is even negative instead of positive, which would have been the expected direction in the association. The lack of association is contradictory because, in an analysis of whether CMI is a surrogate of early undernutrition (Barker's thrifty phenothype hypothesis) or of highly infectious environments (Finch and Crimmins's "cohort morbidity phenotype" hypothesis), CMI was directly associated with SG levels, although not with HbA_{1C} levels. Given that the size of the linear regression coefficient did not vary much after controlling for C-reactive protein (CRP), I concluded that the results more closely resemble Barker's framework more than Finch and Crimmins's. However, the size of the coefficient in the SG equation is also small. Passing from a CMI value of 31 to a CMI value of 18 (roughly the first and fourth quintiles) causes the average SG level to increase by

nearly 7 mg/dl, a relatively small figure, considering that the cutoff point that I am using in this dissertation to classify a respondent as diabetic is 200 mg/dl.

Assuming that CMI and KH are good markers of adverse early life conditions, the results of the event history analysis for DM hazard rates show that the link is weak. Whether CMI and KH are good markers and whether there are other methodological reasons to argue against the results are topics that will be analyzed in a further subsection of this chapter.

Aside from the statistical relationships between early life conditions and DM hazard rates, these models confirmed the strong role of known DM risk factors: body fat, exercise, and family history of the disease. According to the Weibull regression equation with categorical covariates, obese Costa Rican seniors are almost 3 times as likely to develop DM as people with normal weight. The relative risk for overweight people is 1.8 when compared to normal BMI people. The hazard ratio for elderly with a diabetic family member is also very high: Their risk of getting the disease is more than twice that of persons who do not know a family member with DM. Finally, Costa Rican seniors who do not report regular physical activity have a 61% higher risk of developing DM than regularly active elderly. Even if these hazard ratios are derived from cross-sectional data, these variables refer to well-known and clearly established DM risk factors. Given the size of the coefficients, it is safe to state that variations in the population distribution of these risk factors can have a greater impact on future trends of DM prevalence than what the prevalence of early childhood conditions can have.

D. The association between early life conditions and mortality.

Given the weak associations described in the hazard models for DM incidence, the statistical analysis of mortality provided additional insights into the relationship between early childhood conditions and health, with death as the ultimate indicator of poor health status. According to the Gompertz model, Costa Rican seniors born in cantones with high child mortality are more likely to die than people born in cantones with lower child mortality. The positive association between CMI and mortality risk matches findings shown in articles exploring the relationship using both aggregate and microdata (Crimmins and Finch, 2006; Leon and Smith, 2000; Dorling *et al.*, 2000). Even though it is not clear whether CMI is a good marker of early undernutrition, it can be understood as a marker of unhealthy environment during infancy. As in most developing countries, infant and child mortality in Costa Rica during the first part of the 20th century was caused mostly by communicable illnesses. Therefore, people born in cantones with high CMI survived an unhealthy environment characterized by the death of members of their same cohorts.

If the association between CMI and mortality was in the expected direction, this is not the case with KH. According to the model, people with short KH are more likely to survive than people with KH between the second and the third quartile. The coefficients for the three dummy variables jointly describe an inverse-V shape, according to which Costa Rican seniors with average KH are the ones more likely to die. As explained before, this finding can have profound implications in the analysis of prevalence data because it can imply that a positive correlation between having short KH and DM prevalence might be explained not by higher risks of DM, but by longer lives among short KH persons and, hence, longer disease duration. The interaction terms for DM (as a dichotomous variable) and KH were not significantly different from zero, but the size of the coefficient for the interaction between DM and short KH suggests that, even

within the diabetic population, short KH persons live longer than persons with longer KH. The direction of the association was considered unexpected because, in industrialized countries, mortality is inversely related with either height, leg length, or both (Gunnell *et al.*, 1998; Smith *et al.*, 2000).

Finally, other associations found in the mortality model have been described by researchers in other countries. Overweight elderly are less likely to die than elderly with normal or low weight. This relationship has been observed in other elderly populations, especially among those in institutions (Weiss *et al.*, 2007). The positive association between retirement and mortality, after controlling for other risk factors, has been described by the literature, too (Morris, Cook, and Shaper, 1994).

The negative association between income (operationalized as a dummy variable) and mortality was statistically significant, but at a significance level of 0.10. The death hazard ratio for low-income elderly in Costa Rica is 1.38. Even if this ratio is significantly different from one (at a 0.10 level), its size suggests a weak association between one of the typical SES variables and mortality in Costa Rica, a finding that has been confirmed by other researchers who have studied this country (Rosero-Bixby, 1996; Rosero-Bixby, Dow, and Lacle, 2005). The weak SES gradient in mortality and health in Costa Rica remains a fruitful field of research that remains scantily explored.

E. The availability of biomarkers in this study

Utilization of clinical information derived from laboratory analysis of biomarkers is common in health sciences research. However, this is not the case in health-themed research

conducted by demographers and other social scientists. There is limited availability of biomarkers in studies that are representative of country-wide populations. The CRELES project is very special because it is one of the few population-based studies in a Latin American country that collects blood specimens.

This study shows that, if information on SG and HbA_{1C} is taken into account, DM prevalence among Costa Ricans age 60 and older is at least 4 percentage points higher than if only self-reported information is used for computing prevalence. This figure was calculated using relatively high cutoff points for defining DM: SG≥200 mg/dl or HbA_{1C}≥7%. However, even if using lower cutoff points (like SG≥126 mg/dl), the proportion of possibly undiagnosed DM among Costa Rican seniors is not greater than 25%. The information is quite useful for understanding current DM burden in this population.

However, the use of biomarkers produced problems in determining DM incidence rates. Hazard rates of diagnosed DM were estimated using retrospective information, but the availability of biomarkers only permit computing prevalence of high levels of SG or HbA_{1C}, not incidence rates. Nonetheless, I tried to approach the estimation of incidence rates of what was called DM unawareness or undiagnosed DM with the information that was available. These figures are clearly overestimated and, therefore, projections based on these figures are highly unreliable.

Biomarker data are becoming popular in social science research (Butz and Torrey, 2006). Knowing the limitations and requirements of biomarker data is important for social scientists looking to collect them. In this dissertation, the most important conclusion that can be derived from the biomarker information is that actual DM prevalence is underestimated if the figure is entirely based on self-reported information of previous DM diagnosis. However, some of the

statistical methods that were applied to analyze retrospective information traditional in demographic research were not fit for analyzing biomarker data.

F. A comprehensive view of the results and their link to the theoretical framework.

As discussed above, the link between adverse early life conditions and DM incidence inspected in this research project was weak. The effect of KH and CMI on DM burden, as measured by current and projected prevalence, was not only weak but practically non-existent. Do these results mean that the presaged tide of chronic diseases and, in particular, of DM in Latin America is not going to happen? An adequate answer to this question needs to consider different factors that are very likely affecting or not affecting current and future characteristics of the elderly population in Latin America:

- If incidence rates of DM remain the same as the ones estimated in this dissertation, there will be an enormous increase in the size of the Costa Rican diabetic population for the next 25 years. Even if the rates of undiagnosed DM are excluded from the analysis, the projected diabetic population age 60 and over is going to more than triple, from around 100,000 in 2005 to nearly 350,000 in 2030. This increase is mainly due to the rapid population aging process that Costa Rica is currently experiencing. This is the case in most Latin American countries that are well advanced in their demographic and epidemiological transition, from Mexico and Colombia to the Southern Cone countries. Among the transition laggards (most Central American countries, plus Bolivia, Peru, and

Ecuador), this increase in the absolute number of people with chronic diseases may begin soon.

- A further increase in DM burden might occur if the prevalence of obesity and sedentary lifestyles rises in Costa Rica and in the region in general, because physical activity and body fat (as well as genetic factors) remain the most relevant risk factors in the development of DM. I am reluctant to believe that the obesity epidemic in the developing world is going to resemble the one observed in the United States (Olshansky *et al.*, 2005), as Popkin and Gordon-Larsen (2004) suggest. Nonetheless, it is impossible to deny that there is an upward trend in the prevalence of obesity in the developing world and the process of urbanization and economic modernization is bringing more sedentary lifestyles to the country and the region, and if the trend in these processes persist, it is very likely that the burden of DM is going to increase.

- Based on CRELES data, it is not possible to assert that the increase in the burden of DM among the Costa Rican elderly during the next 25 years can be explained by the prevalence of undernutrition or other adverse early life conditions. For this to happen, the joint prevalence of obesity and short KH should increase. This is not likely to happen because the share of the elderly population represented by the cohorts that might have experienced harsh childhood conditions is starting to decline, and obesity among Costa Rican seniors is not as serious a problem as in other Latin American settings, such as Santiago (Chile), Mexico City (Mexico), or Puerto Rico (Palloni *et al.*, 2005, 2006).

- Moreover, the positive association between CMI and mortality and the null association between the same early life marker and DM has the opposite effect from what was expected in the projections. Under the hypothetical scenario of low child mortality at the beginning of the 20th century, the population age 60 and over would have been greater than what is currently observed. This, in turn, makes the hypothetical diabetic population larger too, given that the distribution of the rest of the risk factors remains constant. Therefore, if the effects of early life conditions on DM burden are measured as the difference between the observed and hypothetical scenarios, the results must be interpreted as meaning that undernutrition during the first part of the 20th century (as measured by CMI) "contributes" in having a smaller diabetic population because the people that would have had survived would get DM.

If arguments in favor of the hypothesis that a sizable burden of DM is accounted for early life conditions were so compelling, what prevented Costa Rica from following this scenario? I will first try to present possible substantive (rather than methodological) reasons that can explain Costa Rica's divergence:

- It was unlikely that Costa Rica would have a large proportion of elderly people who were at the same time obese and with adverse childhood experiences because obesity and sedentary lifestyles linked to it are associated with the process of urbanization, and Costa Rica has one of the highest proportions of elderly living in rural areas among Latin American countries (Del Popolo, 2001). The adoption of "industrialized" or "developed" lifestyles might have been slower among these cohorts. So, the proportion of Costa

Rican seniors that experienced adverse childhood conditions might be high and the proportion of Costa Rican seniors that are obese might be increasing, but the proportion of these cohorts that experience both events seems to be small. This might also be an explanation of why Palloni *et al.* (2006) encountered a significant association between KH and DM in only two of the five SABE cities: Santiago and Mexico City. The population living in these two cities experienced a rapid health transition (unlike Montevideo and Havana, where the process was slower, and took more time because it started earlier) and a very rapid urbanization process led by sizable rural-urban migration flows (unlike Barbados, where the urbanization process is much more recent, and most of the elderly cohorts worked in agriculture before the service sector started to flourish late in the 20th century). One rebuttal to this argument is that Sao Paulo (the Brazilian SABE city) has similar characteristics to Santiago and Mexico City, but obesity prevalence was smaller.

- Even if the biological mechanism is taking place among Costa Rican elderly, the proportion of DM prevalence due to early life conditions is small in general. After an extensive literature review, Boyko (2000) concludes that the fraction of DM cases attributed to impaired fetal growth (PAF=Population Attributable Fraction), which is the main explanatory variable in most of Barker's analyses, is not greater than 25% and most of the estimates are even smaller than 10%. Boyko analyzes primary information from populations in the United States (including the Pima Indians) and Sweden, but even if this author is not analyzing results from developing countries, his results suggest that it is

difficult to find higher PAFs due to impaired fetal growth, especially if obesity (BMI≥30) has a PAF of between 25% and 43%.

- Further selection diminished a cohort that was selected anyway. From a theoretical point of view, the main argument of why there should be a tide of diabetic population in Latin America is based on a premise of selection. Frail children during the first part of the 20th century in Costa Rica survived undernutrition and communicable diseases thanks to the introduction of medical technology and public health measures, and thus the argument assumes that these frail children grew up to become old adults during the first part of the 21st century. So, they appear to be a "selected" group of adults that might be frailer than if adverse early conditions never happened. However, it is possible that the frailer members of these cohorts did survive their childhood, but died during early or middle adulthood (between 20 to 59 years old), and therefore, it is not possible to observe them within their cohorts at the beginning of the 21st century. This hypothesis is already suggested by Palloni *et al*. (2006) and is supported by observing the trends in life expectancy during the last 30 years. Costa Rica experienced a process of slowdown or even reversal in mortality decline between 1985 and 1995 (CCP, 2006). Cardiovascular mortality (which is associated with DM) was already the main cause of death among Costa Ricans during that period. This mortality reversal has not been studied yet, but it is interesting to note that it was not particular to Costa Rica: several Caribbean islands (including Bahamas, Puerto Rico, and Trinidad & Tobago) experienced such a process (CELADE, 2004). Bahamas and Trinidad & Tobago have demographic and epidemiological transitions that are very similar to the ones in Costa Rica. Is this

mortality reversal (which lasted less than 10 years) related to the death of frail people who lived through adverse early life conditions? Future research in historical demography is needed to understand whether the thrifty phenotype hypothesis can explain these processes.

- The link between early childhood conditions and DM risk may be mediated by widespread preventive health behaviors and access to preventive health services throughout the life course. As described in the introduction, the Costa Rican Public Health Care System has been praised in the past for its effectiveness (Caldwell, 1986; Mesa-Lago, 1992). The Public Health Care Institution (the CCSS) covers more than 90% of the population age 60 and over and, according to CRELES, a majority of the elderly used its services rather than private services. Articles on which I am currently working show that DM control and preventive services utilization, especially services aimed towards vaccination and cardiovascular disease prevention, is very high among Costa Rican elderly, even when compared to levels in industrialized countries. Costa Ricans are particularly prone to health prevention behaviors, not only because the CCSS provides free services, but because Government's promotion of preventive services has been common throughout the 20th century. Palmer (2003) documents that Costa Rica was, in 1914, the first Latin American Government to implement the Rockefeller Foundation anti-hookworm campaign, and it was one of the few countries in which the campaign's goals were achieved. According to Palmer, the success of this campaign can be explained by the active involvement of Government employees and rural teachers, and the willingness of the Costa Rican population to trust in these public workers. With

regard to this dissertation's statistical analysis, the model could have controlled for preventive services utilization, but the effect of controlling by these variables might be limited since a better variable to measure this proneness to health prevention should consider information from different points through the life course.

- It is possible that this framework of the effect of early life conditions on chronic disease prevalence is still relevant but for explaining other diseases rather than DM. Strong evidence links infections during childhood and youth with certain types of cancer, like stomach, liver, and cervical cancer. As mentioned several times before, this is what Gersten and Wilmoth (2002) have named the "Cancer Transition". In Costa Rica, stomach cancer is still one of the most important causes of morbidity and death among the adult population, producing high costs to the Public Health Care System. CRELES data is not well fit for studying this links, given that the prevalence of such cancer illnesses is not as high as DM or hypertension. A larger sample size is needed for exploring the effect of early life conditions on the prevalence of these diseases in Costa Rica and Latin America.

G. Implications of this dissertation's results on the relationship between economic development and the demographic and epidemiological transition.

Early theories about economic development highlighted how social and economic modernization would be interlinked with better health, lower mortality, lower fertility, and smaller out-migration flows (Kaufman, 1990). One of the most recent variants of modernization

theory suggests that fertility decline in a society opens a window period during which people in productive ages represent a high share of the total population, and if the wealth produced by these cohorts can be invested in developing productive and social infrastructure, this society will have enough resources to support its aging process. This investment is the so-called "demographic bonus" or "demographic dividend" (Bloom, Canning, and Sevilla, 2003).

The basic theoretical premise that leads this dissertation –that the survivors of mortality decline will increase the burden of chronic diseases in the near future— implies that the economic development process that is intertwined with mortality decline can lead to a "negative demographic bonus". This is relevant from an economic perspective because the "stickiness" of early life conditions (Palloni *et al.*, 2006) leads necessarily to increased costs in the provision of health care services, and therefore countries in early stages of economic and demographic transition would need to plan far ahead so that the process will not affect the advantages that economic development and the demographic transition can bring.

This dissertation's results imply that the size of this "negative demographic bonus" is practically non-existent, at least in Costa Rica. However, Palloni *et al.* (2006) are finding weak but significant links between early life conditions and DM in Santiago, Mexico City, and Puerto Rico. How much planning and policy-making should the Governments of these places conduct in order to avoid the effects of this "negative demographic bonus" is an important topic to investigate.

H. Methodological limitations in the analysis.

A previous section discussed theoretical reasons for why I found a weak association between early life conditions and DM burden. However, results could have been affected by methodological limitations that the analysis had:

- The use of retrospective information from cross-sectional data is always a limitation for computing incidence rates. The problems are greater if the population under study is composed of elderly people because the effect of mortality selection is stronger. CRELES is a longitudinal research project, but the data from the second wave are not yet available. I tried to correct for the selection problem that arises from using retrospective information on year or age at diagnosis, using a variation of a method proposed by Keiding (Keiding, 1991, 2006; Keiding, Holst and Green, 1989). However, this correction can still be insufficient for computing unbiased estimates of incidence rates.

- As mentioned before, using retrospective information is inconsistent with the availability of biomarker information because it is not possible or easy to assign a date of diagnosis for respondents without self-reported DM diagnosis but that have SG≥200 mg/dl or HbA_{1C}≥7%. Given that the equation for estimating "undiagnosed DM" or "DM unawareness" might be very unreliable, a way of making the two types of information compatible with each other would have improved the estimation and may have even yielded different results regarding the relationship between early childhood events and the risk of DM.

- The limitations of KH as a marker of undernutrition during gestation and infancy might be introducing too much measurement error into the analysis. Given the lack of direct information on early childhood conditions, the rationale for using KH instead of just height or other anthropometric measures is understandable: when compared to total height, KH is the anthropometric measure less affected by bone damage among the elderly (Palloni *et al.*, 2006) and there is a statistically significant association between nutrient intake during the first 4 years of life and leg length and height (Wadsworth *et al.*, 2002). However, even if Gunnell (2002) does not seem to reject the use of leg length as a marker for adverse exposures during early life stages, the literature shows that there are other factors aside from early nutrition that account for a larger share of the variability in leg length (or KH). The genetic factor (Wadsworth *et al.*, 2002) is very important, and not accounting for the variability in height due to genetics might have introduced measurement error in the statistical models and estimates from them. Moreover, even if researchers have demonstrated a statistical association between leg length and nutrient intake before age four, this does not mean that findings from British populations can be applied to Latin American populations. In a cross-country analysis of the linear relationship between height in one side, and GDP per capita and infant mortality in the other, Deaton (2007) shows that Latin Americans are on average too short for their levels of GDP per capita and child mortality, while Africans are on average too tall for their levels in these two indicators. Deaton also reminds the readers that there is more within-country variation than between-country variation in height, and therefore the interpretation of height or KH as markers for nutrition and well-being should be done cautiously.

- Besides genetic information (Horikawa *et al.*, 2000), there might also be other covariates that were missing from the analysis and their absence might be introducing omitted variable bias in the estimates. Information about preventive behaviors over the life course and the role of the health care system are two of these constructs that would have been useful for avoiding such a bias. Even the information about carbohydrate intake that is in CRELES might be limited because it is measured from a food-tracer questionnaire, and refers mainly to the time before the survey. How much variation in food intake patterns has this population experienced and how this variation is related to DM among the elderly in Costa Rica are two questions that cannot be answered with the data.

I. Suggestions for future research that can be drawn from this dissertation's results.

I have mentioned several possible research topics that can be studied given the results of this dissertation. I will address them again in this section, adding others that have not been commented:

- It is important to study the temporal mortality reversal in some countries in the Caribbean basin. Can the decline in life expectancies at birth observed in Costa Rica, Puerto Rico, Bahamas, and Trinidad & Tobago be explained by the same factors? Is this decline due to early life conditions? Can it be studied as a cohort effect? There is very little

comparative demographic research about Caribbean countries, and this research might produce fruitful results given that most of these small countries have higher life expectancies than continental countries in the Latin American region.

- There should be more anthropometric research about how good are height, leg length, and KH as markers of nutrition during childhood, and about whether it is possible to account for the measurement error in models that use these anthropometric measures as surrogates for adverse early childhood conditions. The subjects of longitudinal cohort studies in the developing world are reaching their adult ages (the Cebu Study, the Pelotas Cohort, the INCAP study) and the information from these studies can allow studying these relationships. However, unlike social scientists, researchers from health sciences are reluctant to make these datasets public so that other researchers can use them.

- The role of the Public Health Care System in health outcomes in Costa Rica is a topic that has been producing interesting scientific research during the last few decades, but this research is still scant. If the Costa Rican Health Care System is as effective as certain authors suggest (Caldwell, 1986; Mesa-Lago, 1992; Rosero-Bixby, 1996, 2004; Rosero-Bixby, Dow, and Lacle, 2005), it could become an interesting role model for other developing countries.

- Analyses with biomarkers in this dissertation did not yield much evidence for Finch and Crimmins's hypothesis that infection early in life results in chronic inflammation and, subsequently, heart disease. However, there are in Latin America some infectious

diseases that were very prevalent in the past and that might degenerate into an epidemic of chronic diseases. Chagas disease and malaria are two examples. It would be interesting for future research to perform analyses similar to what was pursued in this dissertation, but focusing on the link between communicable morbidity during childhood and early adulthood, and chronic disease prevalence in later life.

Chapter VIII: References

ADA American Diabetes Association (2007) “Standards of Medical Care in Diabetes-2007”. *Diabetes Care* 30(Suppl 1): S1-S41.

Alba F and Potter JE (1986). Population and development in Mexico since 1940: An interpretation. *Population and Development Review* 12(1):47-75.

Anastasiou E, Alevizaki M, Grigorakis SJ, Philippou G, Kyprianou M and Souvatzoglou A (1998). “Decreased stature in gestational diabetes mellitus”. *Diabetologia* 41(9): 997-1001.

Arauz AG, Sánchez G, Padilla G, Fernández M, Roselló M, and Guzmán S (2001). “Community diabetes educational intervention at the primary care level”. *Revista Panamericana de Salud Pública* 9(3): 145-153.

Aschner P (2002). “Diabetes trends in Latin America”. *Diabetes/Metabolism Research and Reviews* 18:S27-S31.

Ballesteros MN, Cabrera RM, Saucedo MS, Aggarwal D, Shachter NS, and Fernández ML (2005). “High intake of saturated fat and early occurrence of specific biomarkers may explain the prevalence of chronic disease in northern Mexico”. *Journal of Nutrition* 135(1): 70-73.

Barceló A, Aedo C, Rajpathak S, and Robles S (2003). “The cost of diabetes in Latin America and the Caribbean”. *Bulletin of the World Health Organization* 81(1):19-27.

Barker DJ, Osmond C, Simmonds SJ, and Wield GA. (1993). “The relation of small head circumference and thinness at birth to death from cardiovascular disease in adult life”. *British Medical Journal* 306:422-426.

Barker DJ, Winter PD, Osmond C, Margetts B, Simmonds SJ (1989). “Weight in infancy and death from ischaemic heart disease”. *Lancet* 2(8663):577-580.

Barker DJP (1998). “In utero programming of chronic disease”. *Clinical Science* 95:115-128.

Barker DJP and Osmond C (1986). “Infant mortality, childhood nutrition and ischaemic heart disease in England and Wales”. *Lancet* 1:1077–1081.

Barker DJP, Eriksson JG, Forsén T and Osmond C (2002). “Fetal origins of adult disease: strength of effects and biological basis”. *International Journal of Epidemiology* 31:1235-1239.

Barker DJP, Osmond C, Winter PD, Margetts B and Simmonds SJ (1989). “Weight in infancy and death from ischaemic heart disease”. *Lancet* 2:577-580

Barker DJP, Eriksson JG, Forsén T, and Osmond C (2002). "Fetal origins of adult disease: strength of effects and biological basis". *International Journal of Epidemiology* 31:1235-1239.

Barzilay JI, Abraham L, Heckbert SR, Cushman M, Kuller LH, Resnick HE, and Tracy RP (2001). "The relation of markers of inflammation to the development of glucose disorders in the elderly: the Cardiovascular Health Study". *Diabetes* 50: 2384-2389.

Baylin A and Campos H (2004). "Arachidonic acid in adipose tissue is associated with nonfatal acute myocardial infarction in the central valley of Costa Rica" *Journal of Nutrition* 134(11):3095-3099.

Baylin A, Kabagambe EK, Ascherio A, Spiegelman D, and Campos H (2003). "High 18:2 trans-fatty acids in adipose tissue are associated with increased risk of nonfatal acute myocardial infarction in Costa Rican adults". *Journal of Nutrition* 133(4):1186-1191.

Benítez-Zenteno R. (2000). "Los cambios de la población y la situación, perspectivas y consecuencias del envejecimiento en México". En: *Encuentro latinoamericano y caribeño sobre las personas de edad*. Memorias de Encuentro realizado en Santiago de Chile, de 8 al 10 de septiembre de 1999. Pp.333-369.

Ben-Shlomo Y and Kuh D (2002). "A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives". *International Journal of Epidemiology* 31:285-293

Bloom D, Canning D, and Sevilla J (2003). *The Demographic Dividend: A new perspective on the economic consequences of population change*. Santa Monica, CA: RAND.

Bogin B and Keep R (1999). "Eight thousand years of economic and political history in Latin America revealed by anthropometry". *Annals of Human Biology*, 26(4): 333-351.

Bonora E, Kiechl S, Willeit J, Oberhollenzer F, Egger G, Meigs JB, Bonadonna RC, and Muggeo M (2004). "Population-based incidence rates and risk factors for Type 2 Diabetes in white individuals. The Bruneck Study". *Diabetes* 53: 1782-1789.

Boyko EJ (2000). "Proportion of type 2 diabetes cases resulting from impaired fetal growth". *Diabetes Care* 23:1260-1264.

Branchtien L, Schmidt MI, Matos MCG, Yamashita T, Pousada JMDC, and Duncan BB (2000). "Short stature and gestational diabetes in Brazil". *Diabetologia* 43(7): 848-851.

Brown DC, Byrne CD, Clark PMS, Cox BD, Day NE, Hales CN, Shackleton JR, Wang TWM, and Williams DRR (1991). "Height and glucose tolerance in adult subjects". *Diabetologia* 34(7): 531-534.

Butz WP and Torrey BB (2006). "Some frontiers in social science". *Science* 312: 1898-1900.

Caldwell JC (1976). “Toward a restatement of demographic transition theory”. *Population and Development Review* 2(3-4): 321-366

Caldwell JC (1986). "Routes to Low Mortality in Poor Countries". *Population and Development Review* 12:179-220.

Cardoso E and Fishlow A (1992). “Latin American Development: 1950–1980”. *Journal of Latin American Studies* 24: 197–218.

CCP Centro Centroamericano de Población (2005). Consulta a censos y grandes bases de datos estadísticas. In: http://censos.ccp.ucr.ac.cr/.

CCP Centro Centroamericano de Población CCP (2006). “Tablas de vida abreviadas anuales. Costa Rica, 1970-2006”. At: http://ccp.ucr.ac.cr/observa/CRindicadores/tv70_03.htm.

CCP-INEC Centro Centroamericano de Población e Instituto Nacional de Estadística y Censos (2002). *Costa Rica: estimaciones y proyecciones de población 1970-2100 actualizadas al año 2000 y evaluación del Censo 2000 y otras fuentes de información: informe metodológico*. San Jose, C.R.: CCP, 2002.

CCSS Caja Costarricense del Seguro Social (2005). “Egresos hospitalarios”. In: http://censos.ccp.ucr.ac.cr/.

CCSS Caja Costarricense del Seguro Social (2006). *Informe anual: Evaluación Compromisos de gestión 2000*. Dirección de Compra de Servicios de Salud, CCSS.

CELADE Centro Latinoamericano y del Caribe de Demografía (2004). *América Latina: Tablas de Mortalidad, 1950-2025*. Boletín Demográfico No.74.

Chackiel J (2004). “La dinámica demográfica en América Latina”. Serie Población y Desarrollo No. 52. Santiago de Chile: CELADE.

Coale AJ (1973). “The demographic transition reconsidered”. In: International Population Conference. Liege: IUSSP 1:53-72.

Coale AJ and Hoover EM (1958). *Population Growth and Economic Development in Low-Income Countries*. Princeton, N.J.: Princeton University Press.

Collet D (1994). *Modeling survival data in medical research*. Chapman & Hall.

Congdon P (1993). “Statistical graduation in local demographic analysis and projection”. *Journal of the Royal Statistical Society*, Series A 156(2): 237-270.

Conlisk AJ, Barnhart HX, Martorell R, Grajeda R, and Stein AD (2004) “Maternal and child nutritional supplementation are inversely associated with fasting plasma glucose concentration in young Guatemala adults”. *Journal of Nutrition* 134: 890-897.

Crimmins EM and Finch CE (2006) “Infection, inflammation, height, and longevity”. *Proceedings of the National Academy of Science of the United States of America* 103(2): 498-503.

Cubillos-Garzón LA, Casas JP, Morillo CA, and Bautista LE (2004). “Congestive heart failure in Latin America: The next epidemic”. *American Heart Journal* 147:412-417

Deaton A (2007). “Height, health, and development”. Proceedings *of the National Academy of Science* 104: 13232-13237.

Del Popolo F (2001). “Características sociodemográficas y socioeconómicas de las personas de edad en América Latina”. Santiago de Chile: CELADE. Serie Población y Desarrollo No. 19.

Doblhammer G. (2004). *The late life legacy of very early life”*. Berlin; New York: Springer.

Dunne F. (2005). “Type 2 diabetes and pregnancy”. *Seminars in fetal & neonatal medicine*. 10(4):333-339.

Durán-Varela BR, Rivera-Chavira B, and Franco-Gallegos F (2001). “Apego al tratamiento farmacológico en pacientes con diagnóstico de diabetes mellitas tipo 2”. *Salud Pública de México* 43(3): 233-236.

ECDCDM Expert Committee on the Diagnosis and Classification of Diabetes Mellitus (2004) “Follow-up Report on the Diagnosis of Diabetes Mellitus” *Clinical Diabetes* 22(2):71-79. Reprinted from original publication in *Diabetes Care* 26:3160-3167.

Elo IT and Preston SH (1992). "Effects of Early-Life Conditions on Adult Mortality: A Review." *Population Index* 58: 186-212.

Festa A, D’Agostino R Jr, Tracy RP, Haffner SM, Insulin Resistance Atherosclerosis Study (2002). “Elevated levels of actue-phase proteins and plasminogen activator inhibitor-1 predict the development of type 2 diabetes: the insulin resistance atherosclerosis study”. *Diabetes* 51:1131-1137.

Finch CE and Crimmins EM (2004). “Inflammatory exposure and historical changes in human life-spans”. *Science* 305(5691):1736-1739

Firestone DN, Jiménez-Briceño L, Reimann JO, Talavera GA, Polnsky WH, and Edelman SV (2004). “Predictors of diabetes-specific knowledge and treatment satisfaction among Costa Ricans”. *The Diabetes Educator* 30(2): 281-292.

Forsen T, Eriksson J, Tuomilehto J, Reunanen A, Osmond C, and Barker DJP (2000). “The fetal and childhood growth of persons who develop Type 2 Diabetes”. *Annals of Internal Medicine* 133:176-182.

Gersten O and Wilmoth JR(2002). "The Cancer Transition in Japan since 1951". *Demographic Research* 7(5): 271-306.

Goldhaber-Fiebert JD, Goldhaber-Fiebert SN, Tristán ML, and Nathan DM (2003). "Randomized controlled community-based nutrition and exercise intervention improves glycemia and cardiovascular risk factors in type 2 diabetic patients in rural Costa Rica". *Diabetes Care* 26(1): 24-29.

González-Barranco J and Ríos-Torres JM (2004). "Early malnutrition and metabolic abnormalities later in life". *Nutrition Reviews* 62(7 Pt 2):S134-S139.

Gordis L (2004) *Epidemiology*. Philadelphia, PA: Elsevier Inc., 3rd ed.

Greene WH (2003). *Econometric Analysis*. New Jersey: Prencie Hall. 5th ed.

Grilli E (2005). "Political economy and economic development in Latin America in the second half of the 20th century" *Journal of Policy Modeling* 27(1):1-31.

Guerrero-Igea FJ, Lepe-Jimenez JA, Garrido-Serrano A, and Palomo-Gil S (2001). "Association among hyperinsulinemia, family history of diabetes, and diminutive stature in normoglycemic premenopausal women" *Diabetes Care* 24(3): 602-603

Gunnell D (2002). "Commentary: Can adult anthropometry be used as a 'biomarker' for prenatal and childhood exposures?" *International Journal of Epidemiology* 31: 390-394

Gunnell D, Berney L, Holland P, Maynard M, Blane D, Frankel S and Smith GD (2000). "How accurately are height, weight and leg length reported by the elderly, and how closely are they related to measurements recorded in childhood?" International *Journal of Epidemiology* 29(3): 456-464.

Gunnell D, Oliver SE, Donovan JL, Peters TJ, Gillatt D, Persad R, Hamdy FC, Neal DE, and Holly JMP (2004). "Do height-related variations in insulin-like growth factors underlie the associations of stature with adult chronic disease?" *The Journal of Clinical Endocrinology and Metabolism* 89(1): 213-218.

Gunnell D, Whitley E, Upton MN, McConachie A, Smith GD, and Watt GCM (2003). "Associations of height, leg length, and lung function with cardiovascular risk factors in the Midspan Family Study". *Journal of Epidemiology and Community Health* 57: 141-146

Guzmán JM (2002). "Envejecimiento y desarrollo en América Latina y el Caribe". Serie Población y Desarrollo No. 28. Santiago de Chile: CELADE.

Haggard S (1992). *Pathwasy from the periphery. The politics of growth in the newly industrialized countries*. Ithaca and London: Cornell University Press.

Hales CN and Barker DJP (2001). "The thrifty phenotype hypothesis". *British Medical Bulletin* 60: 5-20.

Hales CN and DJP Barker (1992). "Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis". *Diabetologia* 35: 595-601.

Hales CN, Barker DJP, Clark PMS, Cox LJ, Fall C, Osmond C, and Winter PD (1991). "Fetal and infant growth and impaired glucose tolerance at age 64". *British Medical Journal* 303(6809): 1019-22.

Hanson RL, Nelson RG, McCance DR, Beart JA, Charles MA, Pettitt KJ, and Knowler WC (1993). "Comparison of screening tests for non-insulin-dependent diabetes mellitus". *Archives of Internal Medicine* 153: 2133-2140.

Holst-Schumacher I, Monge-Rojas R, Cambronero-Gutiérrez P, and Brenes G (2005). "Genetic, dietary, and other lifestyle determinants of serum homocysteine levels in young adults in Costa Rica". *Revista Panamericana de Salud Publica* 17(4):263-270.

Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner TH, Schwarz PEH, del Bosque-Plata L, Horikawa Y, Oda Y, Yoshiuchi I, Colilla S, Polonsky KS, Wei S, Concannon P, Iwasaki N, Schulze J, Baier LJ, Bogardus C, Groop L, Boerwinkle E, Hanis CL, and Bell GI (2000). Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nature Genetics* 26, 163 – 175.

Horiuchi S (1999). "Epidemiological transition in human history". In: Chamie J and Cliquet RL (eds.). *Health and mortality issues of global concern*. Proceedings of the Symposium on Health and Mortality, Brussels, 19-22. November 1997. New York: United Nations and CBGS, 1999: Pp. 54-71

Horvitz DG and Thompson DJ (1952). "A generalization of sampling without replacement from a finite universe". *Journal of the American Statistical Association* 47: 663-685

Hux JE and Tang M (2003). "Patterns of prevalence and incidence of diabetes". In: *Diabete in Ontario. An ISES Practice Atlas.* Canadian Diabetes Association. Institute for Clinical Evaluative Services, pp. 1.1-1.18

INEC-CCP Instituto Nacional de Estadística y Censos and Centro Centroamericano de Población (2002). *Costa Rica: Estimaciones y proyecciones de población actualizadas al año 2000, período 1970-2050*. San José, Costa Rica: Instituto Nacional de Estadística y Censos (INEC).

Jang HC, Min HK, Lee HK, Cho NH, and Metzger BE (1998). "Short stature in Korean women: a contribution to the multifactorial predisposition to gestational diabetes mellitus". *Diabetologia* 41(7): 778-783.

Jeffreys M, Lawlor DA, Galobardes B, McCarron P, Kinra S, Ebrahim S, and Smith GD. (2006). "Lifecourse weight patterns and adult-onset diabetes: the Glasgow Alumni and British Women's Heart and Health studies". *International Journal of Obesity* 30(3):507-512.

Kabagambe EK, Baylin A, Ruiz-Narvaez E, Rimm EB, and Campos H (2005a) "Alcohol intake, drinking patterns, and risk of nonfatal acute myocardial infarction in Costa Rica". *American Journal of Clinical Nutrition* 82(6):1336-1345

Kabagambe EK, Baylin A, Ruiz-Narvarez E, Siles X, and Campos H (2005b). "Decreased consumption of dried mature beans is positively associated with urbanization and nonfatal acute myocardial infarction". *Journal of Nutrition* 135(7): 1770-1775.

Kabagambe EK, Baylin A, Siles X, and Campos H (2003). "Individual saturated fatty acids and nofatal acute myocardial infarction in Costa Rica". *European Journal of Clinical Nutrition* 57(11): 1447-1457.

Kabagambe EK, Furtado J, Baylin A, and Campos H (2005c). "Some dietary and adipose tissue carotenoids are associated with the risk of nonfatal acute myocardial infarction in Costa Rica". *Journal of Nutrition* 135(7): 1763-1769.

Kaufman RR (1990). "How societies change developmental models or keep them: Reflections on the Latin American experience in the 1930s and the Postwar World". In: Gereffi G and DL Wyman (1990) *Manufacturing Miracles: Paths of industrialization in Latin America and East Asia*. Princeton, NJ: Princeton University Press. Pp. 110-137.

Keiding N (1991). "Age-specific incidence and prevalence: a statistical perspective (with discussion)". *Journal of the Royal Statistical Society Series A* 154: 371-412

Keiding N (2006). "Event history analysis and the cross-section". *Statistics in Medicine* 25:2343-2364.

Keiding N, Holst C, and Green A (1989). "Retrospective estimation of diabetes incidence from information in a prevalent population and historical mortality". *American Journal of Epidemiology* 130:588-600.

Kousta E, Lawrence NJ, Penny A, Millauer A, Robinson S, Johnston DG and McCarthy MI (2000). "Women with a history of gestational diabetes of European and South Asian origin are shorter than women with normal glucose tolerance". *Diabetic Medicine* 17: 792-797.

Koutsky LA, Galloway DA, and Holmes KK (1988). "Epidemiology of genital human papillomavirus infection". *Epidemiological Review* 10:122-63. Review.

Kuh D and Ben-Shlomo Y (1997). "Introduction: a life course approach to the aetiology of adult chronic disease". In: Kuh D and Ben-Shlomo Y (1997). *A life course approach to chronic disease epidemiology*. Oxford: Oxford University Press: 3-14

Kuh D and Smith GD (1997). "The life course and adult chronic disease: an historical perspective with particular reference to coronary heart disease". In: Kuh D and Ben-Shlomo Y (1997). *A life course approach to chronic disease epidemiology*. Oxford: Oxford University Press, 1997: Pp.15-41.

Lacle A and Peralta C (2006) "Tamizaje de Diabetes Mellitus Tipo 2 en atención primaria". *Acta Médica Costarricense* 48:17-23

Lara-Esqueda A, Aguilar-Salinas CA, Velazquez-Monroy O, Gomez-Perez FJ, Rosas-Peralta M, Mehta R and Tapia-Conyer R (2004). "The body mass index is a less-sensitive tool for detecting cases with obesity-associated co-morbidities in short stature subjects". *International Journal of Obesity* 28: 1443-1450.

Leitch I (1951). "Growth and health". *British Journal of Nutrition* **5:**142–151.

Leon DA and Smith GD (2000). "Infant mortality, stomach cancer, stroke, and coronary heart disease: ecological analysis". *British Medical Journal* 320:1705-1706.

Luros P (1942). *Aspectos biodemográficos de la población de Costa Rica: informe correspondiente al año 1940*. San José: Secretaría de Salubridad Pública y Protección Social.

Mackay J and G Mensah (2004). *The Atlas of Heart Disease and Stroke*. Geneva: World Health Organization and Center for Disease Control and Prevention.

Martínez-Ortiz JA, Fung TT, Baylin A, Hu FB, and Campos H (2006). "Dietary patterns and risk of nonfatal acute myocardial infarction in Costa Rican adults" *European Journal of Clinical Nutrition* 60(6):770-777

Martorell R, Ramakrishnan U, Schroeder DG, Melgar P, and Neufeld L (1998). "Intrauterine growth retardation, body size, body composition and physical performance in adolescence" *European Journal of Clinical Nutrition* 52(S1): S43-S53.

Martorell R, Ramakrishnan U, Schroeder DG, Melgar P, and Neufeld L (1998) "Intrauterine growth retardation, body size, body composition and physical performance in adolescence". *European Journal of Clinical Nutrition* 52(S1): S43-S53.

Martorell R, Stein AD, and Schroeder DG (2001). "Early nutrition and later adiposity". *Journal of Nutrition* 131:874A-880S

Martyn CN (1991). "Childhood infection and adult disease". In: Bock GR and Whelan J (org.). *The childhood environment and adult disease*. Chichester: John Wiley & Sons Ltd.; 1991; pp. 93-108.

Mata LJ (1980). "Child malnutrition and deprivation--observations in Guatemala and Costa Rica". *Food Nutrition* 6(2):7-14.

McKeigue P (1997). "Diabetes and insulin action". In: Kuh D and Ben-Shlomo Y (1997). *A life course approach to chronic disease epidemiology*. Oxford: Oxford University Press, 1997: Pp.78-100

McKeown T (1976). *The modern rise of population*. New York: Academic Press.

McQuestion MJ (2000). *Social effects, health behaviors and the Latin American mortality transition*. Dissertation submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in Sociology at the University of Wisconsin-Madison.

Méndez Chacón E, Rosero-Bixby L, Fernández Rojas X, and Barrantes Jiménez K (2007). "Comparación de los resultados de pruebas de laboratorio seleccionadas de un estudio poblacional de adultos mayores de Costa Rica". *Población y Salud en Mesoamérica. Revista Electrónica* 5(1):4.

Mesa-Lago C (1992). *Health Care for the poor in Latin America and the Caribbean.* Washington, D.C.: Pan American Health Organization and Inter-American Foundation.

Meza E, Barraza L, Martinez G, Fernandez V, Ramos-Jarquez E, Cano-Vargas C, Valdez-Torres A and Izaguirre R. (1995). "Gestational diabetes in a Mexican-U.S. border population: prevalence and epidemiology". *Revista de Investigaciones Clinicas* 47(6): 433-438

Ministerio de Salud-CR (2005). *Encuesta Multinacional de Diabetes Mellitus, Hipertensi[on Arterial y factores de riesgo asociados. Área Metropolitana de San José 2004*. San Jose, CR: Ministerio de Salud, Organización Panamericana de la Salud.

Ministry of Health-New Zealand (2002). *Modelling Diabetes: A Summary*. Wellington: Ministry of Health. Public Health Intelligence Occasional Bulletin No. 1.

Mitchell HS (1962). "Nutrition in relation to stature". *Journal of the American Dietetict Association* 40:521-524

Modan M, Halkin H, Karasik A, and Lusky A. (1984). "Effectiveness of glycosylated hemoglobin, fasting plasma glucose and a single post load plasma glucose level in population screening for glucose intolerance". *American Journal of Epidemiology* 119:431-444.

Morice A and Achío M (2003). "Tendencias, costos y desafíos para la atención de las enfermedades crónicas en Costa Rica". *Revistas de Ciencias Administrativas y Financieras de la Seguridad Social* 11(1): 18-34.

Morris JK, Cook DG, and Shaper AG (1994). "Loss of employment and mortality". *British Medical Journal* 308: 1135-1139.

Moses RG and Mackay MT (2004). "Gestational diabetes. Is there a relationship between leg length and glucose tolerance?" *Diabetes Care* 27: 1033-1035.

Murray CJ and Lopez AD (1997). "Mortality by cause for eight regions of the world: Global Burden of Disease Study". *Lancet* 349:1269-1276.

Nakanishi N, Shiraishi T, and Wada M. (2005). "Association between C-reactive protein and insulin resistance in a Japanese population: the Minoh Study". *Internal Medicine* 44: 542-547.

Njolstad I, Arnesen E, and Lund-Larsen PG (1998). "Sex differences in risk factors for clinical diabetes mellitus in a general population: a 12 year follow-up of the Finmark Study". *American Journal of Epidemiology* 147: 49-58

Notestein FW (1945). *Population: The long view*. In Schultz, T.W. (ed.) Food for the World. Chicago: University of Chicago Press. Pp. 36-57.

Núñez-Rivas HP, Monge-Rojas R, León H, and Roselló M (2003). "Prevalence of overweight and obesity among Costa Rican elementary school children". *Pan American Journal of Public Health* 13(1):24-32.

Olatunbosum ST and Bella AF (2000). "Relationship between height, glucose intolerance, and hypertension in an urban African black adult population: a case for the 'thrifty phenotype' hypothesis". *Journal of the National Medical Association* 92(6): 265-268.

Olshansky SJ, Passaro DJ, Hershow RC, Layden J, Carnes BA, Brody J, Hayflick L, Butler RN, Allison DB, and Ludwig DS. (2005). "A potential decline in life expectancy in the United States in the 21st century". *New England Journal of Medicine* 352: 1138-1145.

Olshansky SJ and Ault AB (1986). "The fourth stage of the epidemiologic transition: the age of delayed degenerative diseases". *Milbank Quarterly* 64(3): 355-391

Omran A (1971). "The epidemiologic transition: A theory of the epidemiology of population change". *Milbank Memorial Fund Quarterly* 49: 509-538.

Pagán JA and Puig A (2005). "Differences in access to health care services between insured and uninsured adults with diabetes in Mexico". *Diabetes Care* 28(2): 425-426.

Palloni A (1981). "Mortality in Latin America: Emerging patterns". *Population and Development Review* 7: 625-650

Palloni A (2002). "Increment-Decrement Life Tables". In: Preston, S.H., P. Heuveline and M. Guillot. *Demography. Measuring and modeling population processes*. Oxford: Blackwell Publisher Ltd. First edition. First reprint.

Palloni A (2004). "Estimating current status models with non-homogeneous competing risks". Paper presented at the Conference REVES 2004, in Brugge, Belgium, 17-19 May.

Palloni A, McEniry M, Wong R, and Pelaez M (2006). "The tide to come. Elderly health in Latin America and the Caribbean". *Journal of Aging and Health* 18(2):180-206.

Palloni A and Wyrick R (1981). "Mortality decline in Latin America: Changes in the structures and causes of death 1950-1975". *Social Biology* 28(3-4):187-216.

Palloni A, Hill K, and Pinto-Aguirre G (1996). "Economic swings and demographic changes in the history of Latin America". *Population Studies* 50: 105-132

Palloni A, McEniry M, Dávila AL, and García-Gurucharri A (2005). "Health among elderly Puerto Ricans: comparative analysis of a new data set". Paper presented at the IUSSP XXV International Population Conference, Tours, France, 2005.

Palloni A, Soldo BJ, and Wong R (2002). "Health status and functional limitations in a national sample of elderly Mexicans", paper presented at the Population Association of America Conference, Atlanta, May 2002.

Palmer S (2003). *From Popular Medicine to Medical Populism: Doctors, Healers, and the Public Power in Costa Rica, 1800-1940.* Durham: Duke Univ. Press, 2003.

Pan WH, Yeh WT, Hwu CM, and Ho LT (2001). "Undiagnosed diabetes mellitus in Taiwanese subjects with impaired fasting glycemia: impact of female sex, central obesity, and short stature". *Chinese Journal of Physiology* 44(1): 44-51.

Peláez M, Palloni A, and Ferrer M (2000). "Perspectivas para un envejecimiento saludable en América Latina y el Caribe". En: *Encuentro latinoamericano y caribeño sobre las personas de edad*. Memorias de Encuentro realizado en Santiago de Chile, de 8 al 10 de septiembre de 1999. Pp.171-194.

Peters AL, MB Davidson, DL Schriger, and V Hasselblad 1996 "A clinical approach for the diagnosis of diabetes mellitus: an analysis using glycosylated hemoglobin levels. Meta-analysis Research Group on the Diagnosis of Diabetes Using Glycated Hemoglobin Levels" *Journal of the American Medical Association* 276(15):1246-1252. Erratum in: *Journal of the American Medical Association* 1997, 277(14):1125.

Phillips DIW, Barker DJP, Hales CN, Hirst S, and Osmond C (1994). "Thinness at birth and insulin resistance in adult life". *Diabetologia* 37:592-596.

Popkin BM and Gordon-Larsen P (2004). "The nutrition transition: worldwide obesity dynamics and their determinants". *International Journal of Obesity* 28:52-59.

Prentice AM and Moore SE (2005). "Early programming of adult diseases in resource poor countries". *Archives of Disease in Childhood* 90:429-432.

Preston SH (1980). "Causes and consequences of mortality declines in less developed countries during the twentieth century". In: Easterlin R (ed) *Population and economic change in developing countries*. University of Chicago Press, pp 289-360.

Preston SH and Haines M (1991). *Fatal years: Child mortality in late-nineteenth century America*. Princeton, NJ: Princeton University Press.

Preston SH, Heuveline P, and Guillot M (2002). *Demography: measuring and modeling population processes*. Oxford, UK: Blackwell Publishers Ltd.

Ramakrishan U, Barnhard H, Schroeder DG, Stein AD, and Martorell R (1999). "Early childhood nutrition, education and fertility milestones in Guatemala" Journal *of Nutrition* 129(12): 2196-2202

Remmer KL (1991). "The Political Impact of Economic Crisis in Latin America in the 1980s". *American Political Science Review* 85: 777-800

Riste L, Khan F, and Cruickshank K (2001). "High prevalence of Type 2 Diabetes in all ethnic groups, including Europeans, in a British inner city". *Diabetes Care* 24: 1377-1383.

Rivera JA, Barquera S, González-Cossío T, Olaiz G, and Sepúlveda J (2004). "Nutrition transition in Mexico and in other Latin American countries". *Nutrition Reviews* 62(7): S149-S157.

Roselló-Araya M. (2003a). "Prevalencia de diabetes tipo 2, intolerancia a la glucosa y diabetes provisional en El Guarco, Cartago". *Revista Costarricense de Ciencias Médicas* 24(1-2): 15-24.

Roselló-Araya M. (2003b). "Factores de riesgo asociados a glicemia elevada en ayunas en pacientes de la clínica de salud, de El Guarco". *Revista Costarricense de Ciencias Médicas* 24(1-2): 25-32.

Rosero-Bixby L, Dow WH, and Laclé A (2005). "Insurance and other socioeconomic determinants of elderly longevity in a Costa Rican panel". Journal of Biosocial Science 37(6):705-720.

Rosero-Bixby L (1996). "Adult mortality decline in Costa Rica". In: Timaeus, IM, J Chackiel and L Ruzicka (1996). *Adult mortality in Latin America*. Oxford, UK: Clarendon Press.

Rosero-Bixby L (2004). "Impact evaluation of the Health Sector Reform in Costa Rica". In: Kyriopoulos, J (2004). *Health Systems in the World. From Evidence to Policy*. Athens, Greece: Papazisis Publishers. Proceedings of the Fifth FICOSSER General Conference.

Rosero-Bixby L (2005). "Costa Rican nonagenarians: are they the longest living male humans?" Paper presented at the IUSSP XXV International Population Conference, Tours, France, 2005.

Rosero-Bixby L, Brenes-Camacho G, and Collado-Chaves A (2004) "Tablas de vida para cálculo actuarial de rentas vitalicias y retiro programado. Costa Rica circa 2000" *Población y Salud en Mesoamérica Revista Electrónica* 1:4.

Sánchez-Castillo CP, Velásquez-Monroy O, Lara-Esqueda A, Berber A, Sepúlveda J, Tapia-Conyer R, and James WP (2005). "Diabetes and hypertension increases in a society with

abdominal obesity: results of the Mexican National Health Survey 2000". *Public Health and Nutrition* 8(1): 53-60

Schmidt MI, Duncan BB, Bang Heejung, Pankow JS, Ballantyne CM, Golden SH, Folsom AR and Chambless LE (2005). "Identifying individuals at high risk for diabetes" *Diabetes Care* 28: 2013-2018

Schofield R and Reher D (1991) "The Decline of Mortality in Europe." In Schofield, R, D Reher, and A Bideau (eds). *The Decline of Mortality in Europe*. Oxford: The Clarendon Press. pp. 1-17.

Smith DG, Greenwood R, Gunnell D, Sweetnam P, Yamell J and Elwood P (2001). "Leg length, insulin resistance, and coronary heart disease risk: The Caerphilly Study". *Journal of Epidemiology and Community Health* 55: 867-872

Smith JP (2007). "Nature and causes of trends in male diabetes prevalence, undiagnosed diabetes, and socioeconomic status health gradient". Proceedings *of the National Academy of Science* 104: 13225-13231.

Soldo BJ, Wong R and Palloni A (2002) "Why Should We Care About Aging in Mexico?", paper presented at the Gerontological Society of America Conference, Boston, November 2002.

Spittel M (1999). "Testing network theory through an analysis of migration from Mexico to the United States". *CDE Working Paper* 99-01.

Stanner SA, Bulmer K, Andres C, Lantseva OE, Borodina V, Poteen VV and Yudkin JS (1998). "Does malnutrition in utero determine diabetes and coronary heart disease in adulthood? Results from the Leningrad siege study, a cross-sectional study". *British Medical Journal* 317(7152): 210.

Tabak AGY, Kerenyi ZS, Nagy E, Bosnyak ZS, Madarasz E and Tamas GY. (2002). "Letters. Height and gestational diabetes mellitus". *Diabetic Medicine* 19: 344-348.

Taylor DJ, Thompson CH, Kemp GJ, Barnes PR, Sanderson AL, Radda GK, and Phillips DI (1995). "A relationship between impaired fetal growth and reduced muscle glycolysis revealed by 31P magnetic resonance spectroscopy". *Diabetologia* 38(10):1205-1212.

Thorand B, Löwel H, Schneider A, Kolb H, Meisinger C, Fröhlich M, Koenig W. (2003). "C-reactive protein as a predictor for incident diabetes mellitus among middle-aged men: results from MONICA Augsburg cohort study, 1984-1998". *Archives of Internal Medicine* 163: 93-99.

Trussell J and Richards T (1985). "Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer Procedure". *Sociological Methodology* 15: 242-276.

United Nations Organization (1956) *Manual III: Methods for preparing population projectioins by sex and age.* New York, NY: UNO. Series: Demographic studies No. 25.

Velázquez-Monroy O, Rosas-Peralta M, Lara-Esqueda A, Pastelón-Hernández G, Sánchez-Castillo C, Attie F, and Tapia-Conyer R (2003). "Prevalence and interrelations of noncommunicable chronic diseases and cardiovascular risk factors in Mexico. Final outcomes from the National Health Survey 2000. *Archivos de Cardiología de México* 73(1):62-77

Villarreal, R. (1990). "The Latin American strategy of import substitution: Failure or paradigm of region". In: Gereffi G and DL Wyman (1990) *Manufacturing Miracles: Paths of industrialization in Latin America and East Asia*. Princeton, NJ: Princeton University Press. Pp. 292-320

Wadsworth MEJ, Hardy RJ, Paul AA, Marshall SF, and Cole TJ (2002) "Leg and trunk length at 43 years in relation to childhood health, diet and family circumstances; evidence from the 1946 national birth cohort". *International Journal of Epidemiology* 31:383-390

Wang Z and Hoy WE (2007). "C-reactive protein and the risk of developing type 2 diabetes in Aboriginal Australians". *Diabetes Research and Clinical Practice* 76: 37-43.

Weiss A, Beloosesky Y, Boaz M, Yalov A, Kornowski R, Grossman E. (2007). "Body Mass Index is inversely related to mortality in elderly subjects". *Journal of General Internal Medicine* (in press).

WHO World Health Organization (1999). *Definition, diagnosis and classification of diabetes mellitus and its complications: report of a WHO Expert Consultation. Part 1: Diagnosis and classification of diabetes mellitus*. Geneva: WHO Department of Noncommunicable Disease Surveillance.

Wiener K and Roberts NB (1998). "The relative merits of haemoglobin A1C and fasting plasma glucose as first line diagnostic tests for diabetes mellitus in non-pregnant subjects". *Diabetic Medicine* 15: 558-563.

Zavala de Cosío ME (1992). "La transición demográfica en América Latina y en Europa". *Notas de Población* 20(56): 11-32.