Undiagnosed cases and the 'real' health burden of diabetes in Latin America and the Caribbean

Flavia Andrade
University of Wisconsin-Madison

Please address correspondence to Flavia Andrade, Center for Demography and Ecology, University of Wisconsin, 4412 Sewell Social Science Bldg., 1180 Observatory Drive, Madison, Wisconsin 53706-1393, U.S.A. (fandrade@ssc.wisc.edu).

Abstract

This paper proposes an innovative approach to estimate diabetes prevalence rates in Latin America and the Caribbean. The results indicate that diabetes prevalence estimated by self-reports is underestimated for most part. For example, the self-reported diabetes prevalence in Buenos Aires is 12.4%, while the predicted prevalence can reach 30.2%. As a consequence, the average number of years expected to be lived with diabetes are considerably higher than one would expect using self-reported measures. Finally, this paper shows that estimates of total life expectancy and disability-free life expectancy of diabetics in Mexico based on self-reports may be biased downwards. In any case, it is important to understand that diabetes reduces total life expectancy and the bulk of this reduction comes in the form of reductions in the number of years expected to be lived without disability. The new estimates shows that total life expectancy of diabetics at age 50 are about 4 years lower than total life expectancy of non-diabetics.

Introduction

Self-reported measures on diabetes are, in most cases, the only available data for countries in Latin America and the Caribbean. Self-reported measures have the advantages of being easier to collect, and less expensive. However, there is always the question of how accurate these reports are, particularly among elderly people. It is also possible that misreport varies depending on the characteristics of the respondents. Existing literature has shown that diabetes self-report is relatively accurate, even among elderly people (Bush et al. 1989, Okura et al. 2004), but most of the studies use data from developed countries, where knowledge of health status may differ considerably from developing countries. Lack of diagnosis is probably the main problem with self-reported data in Latin America and the Caribbean because health care systems in this region face large limitations of resources and coverage. For this reason, clinical diagnosis of type 2 diabetes in the region usually happens many years after its onset when symptoms (increased thirst, profuse urination, unexplained weight loss, among others) become apparent. Some individuals with type 2 diabetes can live for many years without being diagnosed, and in other cases they will never be diagnosed.

There is some evidence that undiagnosed rates are quite high in Latin America and the Caribbean. In Mexico, 42% of diabetics younger than 40 years were unaware of their condition in a national survey conducted in urban areas the early 1990s, while 74% of individuals diagnosed with diabetes aged 40 and over were aware of their diabetic status (Aguilar-Salinas et al. 2002). In Argentina, undiagnosed cases reach about half of the cases (Hernández et al. 1987) and among those diagnosed about a third do not control their diabetes by any means, while the remaining group presents poor glucose control. In Brazil, levels of undiagnosed diabetes are estimated to be around 40-50%. Malerbi and Franco (1992) found that undiagnosed diabetes

accounted for 46% of the total prevalence. In this study, glucose tolerance tests were only conducted in those positive screened individuals in the FCG test and some selected negative screened. Sakata et al. (2002) analyzed a sample of 922 of residents aged 40 and over of a small town in the South region of Brazil. Sakata and colleagues also selected only cases positive screened and found that 42.7% were undiagnosed cases. However, undiagnosed rates can be even higher among certain social groups. For instance, Brito, Lopes and Araújo (2001) find that among a selected high-risk population of obese women, undiagnosed cases reached about 70%. There are also sex differences on undiagnosed rates in Brazil. For instance, Goldenberg et al. (1996) report that almost 60% of men in São Paulo constituted undiagnosed cases, while among women the percentage was close to 41%. In Chile, undiagnosed cases reached 45% of the population aged 20 and older. However, the prevalence of diabetics aged 20-44 who had not being previously diagnosed reached 67%, and decreased as people aged reaching 37.5% in the 65 and over age group (Baechler et al. 2002). This is consistent with the fact that older people have more time to develop the disease and to present complications that may trigger medical diagnosis and treatment.

Given the fact that undiagnosed rates are relatively high in Latin America and the Caribbean, the question is whether self-reported measures can be used to estimate 'true' diabetic status. More broadly, the question is how to account for misreported cases when estimating prevalence rates. The primary goal of this paper is to estimate new prevalence rates for the SABE (Salud, Bienestar y Envejecimiento en América Latina y el Caribe Proyecto) and MHAS (Mexican Health and Aging Study) samples taking into consideration misreporting. SABE investigates the health of older people in seven major cities in Latin America and the Caribbean (Buenos Aires, Bridgetown, São Paulo, Santiago, Havana, Mexico City and Montevideo. MHAS

is a prospective two-wave panel study of a nationally representative cohort of Mexicans born prior to 1951.

The relationship between "true" probabilities of having diabetes and self-reported measures is estimated using two additional datasets which contain both self-reported measures and measures of blood glucose levels (Encuesta Nacional de Salud, ENSA, for Mexico, and Estudio de Longevidad y Envejecimiento Saludable, CRELES, for Costa Rica). Multivariate analysis is used in order to analyze the predictive value of demographic and social characteristics of the individuals and whether self-report can be used to estimate 'true' diabetic status. After, based on logistic regression procedures, the paper will then predict new estimates of the diabetes burden in the Latin America and the Caribbean using self-reported measures as the main covariate. Finally, these new estimates will be used to estimate diabetes-free life expectancy in seven urban areas of Latin America and the Caribbean using the Sullivan method and new estimates of disability-free life expectancy by diabetic status in Mexico based on multistate models. These new prevalence rates are tentative, since the intention is to have an idea about the 'maximum' burden imposed in the population.

Background

Misreport in this case not only imposes problems in adequately measuring the burden of the condition, but it also makes more difficult to make comparisons across settings. This is particularly problematic because misreport usually varies depending on the social, economic and demographic characteristics of the respondents. Often, individuals underreport prevalent chronic conditions. As a consequence, prevalence might be underestimated, but more importantly, analyses across settings and social groups may be biased if those characteristics are associated with diabetes reporting.

As shown in the literature, misreport of chronic conditions depend on the type of the disease and its severity - diseases that are less severe and more transient are more poorly reported (Bergmann, Jacobs and Boeing 2004). Other demographic and socioeconomic factors, such as gender, education and socioeconomic status also influence the quality of the self-report and its consistency over time (Wu, Li and Ke 2000, Beckett et al. 2000).

There are few studies that have evaluated the precision of assessments of chronic conditions, particularly diabetes. Moreover, most of the studies assessing diabetes self-report focus on developed countries (Krueger 1957, Tretli, Lung-Larsen and Foss 1982, Bush et al. 1989, Heliovaara et al. 1993, Kehoe et al. 1994, Kriegsman et al. 1996, Haapanen et al. 1997, Martin et al. 2000, Simpson et al. 2004, Bergmann et al. 2004), and usually their sample sizes are small (Krueger 1957, Bush et al. 1989, Martin et al. 2000). There are even fewer studies that have analyzed the accuracy of self-report of chronic conditions among the elderly population (Bush et al. 1989, Kriegsman et al. 1996, Haapanen et al. 1997, Wu et al. 2000, Goldman et al. 2002, Simpson et al. 2004). The self-report among the elderly population may be of greater concern because elderly individuals may be more likely to misreport than younger individuals (Haapanen et al. 1997). In fact, elderly individuals have more cognitive problems that may affect their self-report. In fact, some studies have shown that elderly individuals are more likely to respond 'don't know', to respond in a socially desirable manner or to inaccurately report (Sherbourne and Meredith 1992, Kriegsman et al. 1996). However, there is good evidence that self-report provided by elderly individuals is usually accurate, particularly with more severe diseases. It is also important to note that elderly individuals are more likely to visit their physicians more often, which may increase their awareness. Although, Haapanen et al. (1997)

shows that number of health services contact increased the misclassification of self-reported cardiovascular diseases in their sample.

Most of the reviewed studies focus on three measures of accuracy: Kappa statistics, sensitivity and specificity (see Table 1). Kappa is a statistical measure of agreement between two sources of information. Kappa takes into account the percentage of data values that are in agreement (diagonal of the table) given the agreement that could be expected due to chance alone. Sensitivity, also known as true positive rate, measures the probability that a person self-reported having diabetes given that the person has the disease. In other words, sensitivity indicates the percentage of patients with diabetes who are aware of their diabetic status. Underreporting is the complement of sensitivity. The proportion underreporting refers to those individuals who have diabetes, but are unaware of their diabetic status (false-negative). Specificity, also known as true negative rate, refers to the proportion of people without diabetes who self-report not having diabetes. It indicates the percentage of persons without diabetes that correctly recognize themselves as non-diabetics. The proportion underreporting is the complement of specificity. It refers to the individuals without diabetes that self-report having the condition (false-positives).

Table 1: Traditional measures of agreement

	Other method of disease assessment (clinical data, medical records, physician, etc.)		
Self-report	Diabetes present	Diabetes not present	Total
Diabetes present	a	b	a+b
Diabetes not present	c	d	c+d
Total	a+c	b+d	n
Sensitivity	a/(a+c)*100		
Specificity	d/(b+d)*100		
Self-reported prevalence	(a+b)/n*100		
Overall agreement	(a+d)/n*100		
Kappa	(2(ad - bc))/((a + b)(b + d) + (c + d)(a + c)))		

Kruger is the pioneer in the field of analyzing the accuracy of diabetes self-reports. In a paper published in 1957, Krueger analyzes a sample of residents in Baltimore in a study conducted during 1953-1955. The final sample was composed by 809 individuals that completed the diagnostic examinations and self-reported the chronic conditions. The results indicate that only 37% of the self-reported cases of diabetes matched with the diagnostic examinations.

Tretli, Lung-Larsen and Foss (1982) use panel data on over 12,000 adults aged 20-49 from Finnmark County, Norway, to analyze the reliability of self-reports of diabetes, heart disease and stroke. For positive questionnaire answers, agreement between self-report and medical records reached 66% for diabetes, 65% for stroke and 81% for myocardial infarction. The authors also report that 73% of those answering having diabetes in the first wave repeated the positive answer for the disease in the second wave. Midthjell et al. (1992) also uses a Norwegian sample from the Nord-Trondelag county to analyze the reliability of diabetes self-reports. They find that concordance was higher than found in Tretli, Lung-Larsen and Foss (1982) study. In this sample, for those self-reporting having diabetes, 96.4% had this diagnosis verified. For those who self-reported not having diabetes, 338, only one (0.3%) had diabetes.

Another important conclusion of this study was that, even though self-report of diabetes status is very accurate, self-report of diabetes duration is considerably overestimated.

Bush et al. (1989) analyzes a sample of 120 elderly (65+) volunteers residents in Dunedin (Florida) and find that there is a very high agreement (kappa of 0.93) between self-reported diabetes and medical records.

Heliovaara et al. (1993), based on data from a nationally representative sample of adults aged 30 and over in Finland, find that diabetes self-report is reasonably accurate when compared with clinical diagnostic obtained in a health examination survey (kappa 0.78, sensitivity 81.4%, specificity 99.1%).

Kehoe et al. (1994) use data from a case-control cataract study in which 1,380 residents in Boston (Massachusetts) provided self-reported medical history. They compare self-reports with information from the participants' physicians. Diabetes self-reports had the highest specificity (97%), while sensitivity reached 84%. Self-reports on use of insulin and oral hypoglycemic had sensitivity of 84% and 78%, respectively, and specificity of 99% and 98%, respectively.

Kriegsman et al. (1996) use data from Netherlands to compare self-reported diabetes with medical records. They analyze data from 2,380 community-dwelling elderly individuals aged 55-85. As in other studies, the diabetes report had the highest kappa-statistic (0.85) and the highest concordance (97.8%).

Haapanen et al. (1997) analyze self-reported data and medical records of 596 Finnish men and women aged 45-73 years. Sensitivity of diabetes self-report was 80%, specificity reached 98% when definite diagnose of medical records was used as 'gold standard'. Kappa statistics was 0.75, positive predicted value reached 75% and negative predicted value 98%.

Martin et al. (2000) analyzes a sample of 599 adults (21 years and older) residents in Colorado and subscribers of a health maintenance organization. They find that sensitivity of diabetes self-report was 73.2% and specificity was 99.3% using the medical records as the 'gold standard'. Women's self-reports presented lower sensitivity, but similar specificity of males' self-reports.

Wu et al. (2000) using a sample of 228 elderly Taiwanese residents in three northern districts (Long-Shang, Sheng-Kang, and Shi-Ding) show that self-reported history of diabetes had the most accurate self-report among the analyzed chronic conditions (diabetes, hypertension and heart disease). However, the level of agreement between clinical and self-reported diabetes can only be categorized as moderate, with Kappa of 0.56 (Landis and Koch 1977a and 1977b). They also show that diabetes self-report reached the highest sensitivity (66. 7%) and specificity (95.2%) among the analyzed conditions. Finally, contrary to usually expected, diabetes prevalence was a little overestimated with the self-reported data. Their analysis also showed that there is a negative association between the number of self-reported health conditions and the risk of underreport diabetic status. Also, those with higher education were also less likely to over reporting diabetes. There were no age or sex differences misreporting diabetes. One important limitation of this study is that there were a very large number of non-participants (958) that differed in important aspects from participants.

Goldman et al. (2002), based on data from a nationally representative survey of adults conducted in Taiwan in 2000, compare self-reported measures with results provided by physical examinations. Their final sample is composed by near 1,000 adults aged 54 and older. They analyze the validity of self-report of hypertension and diabetes. They compare self-reported diabetes with measures of glycosylated hemoglobin. Their results indicate that sensitivity of

diabetes self-report reaches 83.9% and 48.9% for hypertension. Specificity for diabetes was 98.8%. Additional analyses also indicated that better cognitive function is associated with more accurate report of diabetes, but not age, gender or SES. This result is due, in part, because diabetes was very accurately reported. One important limitation of this study, however, is that glycosylated hemoglobin test is effective to monitor diabetes treatment, but it is not indicated for diagnostic purposes. Also, a large percentage of their initial sample did not complete the physical examinations.

Simpson et al. (2004) analyze a sample of over 1,000 disabled women 65 and older residents in Baltimore. They compare self-report of disease diagnoses and medical records. They find that self-report of diabetes seems to be valid (Kappa 0.92). Sensitivity and specificity reached 95% and 99%, respectively.

Bergmann et al. (2004) use a sample is composed by over 7,000 respondents. The sample includes women aged 35-64 and men 40-64 residents in Postdam and surroundings (Germany). They contrast the self-reported measures obtained in-person interview with a self-administered questionnaire. Their results indicate that there is a very high level of agreement between both sources on diabetes report (Kappa of 0.84). Diabetes had the second highest level of agreement subsequent to malignant tumor.

The limited literature indicates that self-reported diabetes status is fairly accurate, even among elderly individuals. In all the studies, specificity is very high, reaching more than 97% in most of the studies. Therefore, the percentage of individuals without diabetes who self-report having diabetes is generally small. However, this overreporting may have important consequences in the total self-reported prevalence because even a small proportion can make a difference when most of the individuals do not have diabetes. Sensitivity values are not as high

and they indicate that a considerable percentage of individuals who have diabetes are unaware of their diabetic status. This underreporting biases downwards the self-reported rates. However, it is important to mention that the lack of other studies in developing areas limits the scope of these conclusions. Therefore, it is important to analyze whether diabetes self-reports are useful in developing countries. Also, previous studies have also shown that there is some evidence that some demographic and socioeconomic characteristics are important determinants of the quality of report. For that reason, this paper will analyze the quality of self-reported diabetes status in two settings in Latin America and the Caribbean and how self-reports can be used to generate new estimates of diabetes prevalence that take misreport into consideration.

Table 2: Summary of literature on agreement between self-reports and other methods of assessing diabetic status

Study	Country	Age-group	Sensitivity	Specificity	Kappa
Krueger (1957)	United States	-	-	-	-
Tretli et al. (1982)	Norway	20-49	-	-	-
Bush et al. (1989)	United States	65+	-	-	0.93
Heliovaara et al. (1993)	Finland	30+	81.4%	99.1%	0.78
Kehoe et al. (1994)	United States	65.2 (mean)	84%	97%	*0.81*
Kriegsman et al. (1996)	Netherlands	55-85	*83%*	*99%*	0.85
Haapanen et al. (1997)	Finland	45-73	75%	98%	0.75
Martin et al. (2000)	United States	21+	73.2%	99.3%	-
Wu et al. (2000)	Taiwan	65+	66. 7%	95.2%	0.56
Goldman et al. (2002)	Taiwan	54+	83.9%	98.8%	-
Simpson et al. (2004)	United States	65+	95%	99%	0.92
Bergmann et al. (2004)	Germany	35-64	-	-	0.84

Note: Values in italics were estimated by the author based on information provided in the papers.

Data

Data on self-reported diabetes prevalence in Latin America and the Caribbean come from SABE (Salud, Bienestar y Envejecimiento en América Latina y el Caribe Proyecto) and MHAS (Mexican Health and Aging Study). Diagnosed diabetes was obtained by self-report in both

SABE and MHAS. Individuals who were previously told by a physician that they had diabetes were considered diabetics. In MHAS, those who reported having diabetes in the first wave were assumed to have the condition in the second wave. MHAS also provides the necessary panel data that will be used to generate multistate life tables. Additional mortality data was obtained from a variety of sources and they are described below. In addition, the paper will make use of two surveys with clinical and self-reported measures of diabetes: Encuesta Nacional de Salud (ENSA 2000), a national health survey conducted in Mexico and Costa Rica: Estudio de Longevidad y Envejecimiento Saludable (CRELES).

Disability was measured using three measures: ADL, IADL and Nagi functional limitations. Six activities were considered in the ADL measure: dressing, bathing, eating, getting in and out of a bed (transferring), toileting, and getting across the room. In MHAS, those who did not declare having Nagi limitations were assumed not to have ADL limitations. The IADL measure includes: preparing a hot meal, money management, shopping, and taking medication. The Nagi physical performance measure included: lifting or carrying objects weighted 5 Kg or over, lifting up a coin (using fingers to grasp or handle), pulling or pushing a large object such as a living room chair, stooping, kneeling or crouching, and reaching or extending arms above shoulder level. ADL, IADL and Nagi measures are in the binary form, in which those scoring '0' indicate that they do not have any limitations, while score '1' was assigned for those who have reported having difficulty performing at least one activity of each scale.

- **Salud, Bienestar y Envejecimiento en América Latina y el Caribe Proyecto (SABE)**

SABE is a multicenter survey that investigates the health and well being of older people (aged 60 and over) and, in some cases, of their surviving spouse in seven capital/major cities in countries of Latin America and the Caribbean. The cities investigated were: Buenos Aires

(Argentina); Bridgetown (Barbados); São Paulo (Brazil); Santiago (Chile); Havana (Cuba); Mexico City (Mexico); Montevideo (Uruguay) (Peláez et al., 2003). The general survey was funded and supported by the Pan American Health Organization (PAHO/WHO), Center for Demography and Ecology, University of Wisconsin-Madison and National Institute on Aging. In each country, international and national institutions contributed for the project.

The questionnaire design was intentionally geared toward the production of information that could be comparable with that retrieved in other countries. In particular, the aim was to include modules and sections modeled after the Health and Retirement Study (HRS) in the U.S. A standardized questionnaire was used to collect detailed information during face-to-face interviews. Samples were drawn using multistage clustered sample with stratification of the units at the highest levels of aggregation. Detailed information on sample selection is presented elsewhere (Palloni & Peláez, 2002).

The sample is composed by 10,602 individuals aged 60 and over. Among those, 8,782 were not previously diagnosed with diabetes, while 1,763 reported a previous diagnosis of diabetes. Other 57 individuals did not answer the question (0.54% of the sample) and were excluded from the analysis. There were no age or sex differences between those who answered or not the question regarding a previous diagnosis of diabetes. Five individuals had missing values on the sample weight variable and were excluded from the analysis. The final sample is composed by 10,540 individuals. The mean age of the sample is 70.2 years (weighted estimates) and most of the sample (59.7%) is composed by women. Among diabetics, mean age reaches 70 years and 59.3% are females. Among non-diabetics, mean age is 70.2 years and 59.8% are women.

- **Mexican Health and Aging Study (MHAS)**

MHAS is a prospective two-wave panel study of a nationally representative cohort of Mexicans born prior to 1951 (50 and older). The survey has national and urban/rural representation. Surviving spouses regardless of their age were also interviewed. The baseline interview was conducted in 2001 and the second wave during 2003. Data collection was done in collaboration with The National Institute of Statistics. The study was designed field protocol and content similar to HRS. Detailed information is presented elsewhere (Palloni, Soldo, & Wong, 2002).

In the first wave, a total of 15,144 complete interviews were obtained (response rate 94.2% at the household level). There are13,022 individuals aged 50 and over with complete information on age, sex and diabetic status. From the initial 15,144 individuals, there are 1,718 with less than 50 years of age, and they were excluded from the analysis. Other 404 individuals without diabetic status at baseline were also excluded. There were no age differences among those with complete or missing information on diabetic status, but more men lacked this information than women. Among the 13,022 cases with complete information on age, sex and diabetic status, the mean age reaches 62.7 (weighted estimate) and 53.8% are women. Among diabetics, the mean age reaches 63.4 years and 60.3% are women. Among non-diabetics, the mean age reaches 62.6 years and 52.7% are women. For the analysis of ADL limitations, the final sample is restricted to 12,050 individuals with complete information in both waves. For instrumental activities of daily living (IADL), 12,065 individuals had full information and for Nagi limitations (Nagi, 1976), 12,056 individuals were analyzed. Those with missing data on disability and mobility measures in the first wave were older ($p<0.0001$) and more likely to be men ($p=0.0001$). However, there were no differences by diabetic status. There were 546 deaths

between waves, 518 of them among those with complete information on age, sex, diabetic and disability statuses.

- **Mortality data**

Mortality data for Buenos Aires was obtained in the "Anuario Estadistico" for the years 2000 and 2001. Deaths of both years were averaged. Population estimates for Buenos Aires were obtained at INDEC based on the census data (Censo Nacional de Población, Hogares y Viviendas 2001). Data were not available for Bridgetown. Therefore, the life table produced by World Health Organization (WHO) for Barbados is being used for Bridgetown instead. The life table refers to the year 2000. The use of the Barbadian life table is reasonable because 37% of the total population lives in Bridgetown according to PAHO (Pan American Health Organization). Population of the São Paulo metropolitan area was obtained with the Brazilian Census Bureau (IBGE), mortality data was obtained in the SEADE foundation that analyzes relevant social, demographic and economic data in the São Paulo state. Santiago life table refers to the period 2001-2002 and it was published by the Chilean population bureau (Instituto Nacional de Estadísticas). The publication "Chile: Tablas Abreviadas de Mortalidad, por sexo. País y Regiones, 2001-2002" contains the life tables for disaggregated by sex and regions. Life tables for Havana were created by Esther María León Díaz from Universidad de La Habana. The life table for Mexico was obtained at WHO website, while the life table for Mexico City uses data from CONAPO. The life table for Montevideo refers to the year 2000 and it was published by the Uruguayan population bureau (Instituto Nacional de Estadísticas) and it is available in the internet (http://www.ine.gub.uy).

- **Mexican National Health Survey (ENSA) data**

Data from ENSA 2000, National Health Survey, will be used to estimate the probabilities of having diabetes taking into account self-reported diabetic status, demographic and socioeconomic characteristics. ENSA is a nationally representative survey conducted in Mexico based on a random sample of basic geographic statistical units obtained in each of the Mexican states and in the Federal District (Mexico City) from a database produced by the Instituto Nacional de Geografía y Estadística (National Institute of Geography and Statistics). A total of 47,360 households, with over 123,000 individuals, were identified. However, for the purpose of this paper, only individuals aged 50 and over will be included in the final sample. There are 13,064 individuals aged 50 and over (4,744 men and 8,320 women). Detailed information on sample selection of ENSA is presented elsewhere (Aguilar-Salinas et al. 2003, Sanchez-Castillo et al. 2003, and Rio-Navarro et al. 2004).

Demographic data and medical history were recorded using a standardized questionnaire. It included questions on age, sex, socioeconomic status (education and income), family history, and information on many chronic conditions, such as diabetes, and hypertension. For those with diabetes, additional questions assessing age at diagnosis, treatment-related issues, and symptoms were also made. Anthropometric measures such as height (Estadimeter; SECA ADEX Products, Mexico City, Mexico) and weight (Solar Scale; Tanita Corporation of America, Inc., Arlington Heights, IL) were measured to the nearest 5 mm and 0.1 kg, respectively, with the subject in light clothing without shoes. Measurement of capillary glucose concentration was requested from all participants. All of the procedures were undertaken in accordance with the ethical standards of the Helsinki Declaration of 1975, as revised in 1983. The final sample for the multivariate analysis is composed by 12,435 individuals with complete information on diabetic status,

selected symptoms and education. Those with missing data on selected variables were older than those with complete information, 64 years versus 62.9 ($p<0.05$), respectively. A higher percentage of males, 5.8%, had missing data than females, 4.2% ($p<0.001$).

In the present analysis, those who had a history of diabetes mellitus and were under oral hypoglycemic agents and/or insulin therapy were deemed to be cases of diabetes ('true diabetes'). Individuals who meet any of the following criteria were also classified as diabetics ('true diabetes'): a) fasting plasma glucose ≥ 126mg/dL or b) casual (random) capillary glucose concentration ≥ 200 mg/dL. Blood specimens were collected through capillary puncture. The large majority of individuals (95.4%) were not fasting at the time of the blood collection. Individuals were also asked if a doctor has ever told them if they had diabetes. This information is equivalent to the self-reported measure of diabetes mellitus in SABE and MHAS. ENSA also collected information on symptoms of diabetes (polyuria, polydipsia, and unexplained weight loss). Polydipsia and weight loss are included in the analysis.

- **CRELES data**

Data from CRELES, Costa Rican Study on Longevity and Healthy Aging, refer to the population born in Costa Rica in 1945 or earlier and that was alive during the period 2004-2006. The sample in the first wave is derived from an original random sample of 8,000 individuals ages 55 and over interviewed in the 2000 Census. A subsample was selected and it is expected to produce valid information on 3,000 individuals. The fieldwork started in September 2004, and the first wave will be finished around August 2006. Therefore, data is limited to 1,832 individuals interviewed until December 31st, 2005. Detailed information on sample selection can be found in Brenes (2006).

Individuals are asked for personal consent and then interviews are carried out. Demographic data and medical history are recorded using a standardized questionnaire. It included questions on age, sex, socioeconomic status (education and income), family history, and self-reported information on many chronic conditions, such as diabetes, and hypertension. For those self-reporting having diabetes, additional questions assessing age at diagnosis, treatment-related issues, and symptoms were also made. Blood and urine samples, as well as anthropometric measures are usually collected in the next day morning. Data have been gathered by Computer Assisted Interviews, using pda's (Personal Digital Administrators) or palms. Therefore, data are immediately available and data quality assessments (check for inconsistencies) are simultaneously made (Rosero-Bixby, Hidalgo and Antich 2005 and Brenes 2006).

CRELES has information on self-reported diabetes diagnosis and clinical information based on blood samples draw from a vein (venipucture). There are two biomarkers that CRELES can use to determine diabetes: glycosylated hemoglobin levels (HbA1c $\geq$ 6.5%) and fasting serum glucose levels (FSG$\geq$ 126 mg/dL). The latter is the criterion recommended by a World Health Organization (WHO) Consultation Group (1999) for diabetes diagnosis. Glycosylated hemoglobin levels gives a good indication of how much sugar has been in a person's blood during the past months. It is usually used to monitor the effectiveness of diabetes treatment, but it is not indicated to detect diabetes cases. Therefore, it will not be included in the present analysis. There were 1,537 individuals (84%) who provided blood samples and 295 individuals to whom data is not available. Among those who provided blood samples, 1,457 (97%) were fasting for more than eight hours (the question asked if they have eaten or drunk anything since 6 pm of the previous day).

In the present analysis, those who had a history of diabetes mellitus and were under oral hypoglycemic agents and/or insulin therapy were deemed to be cases of diabetes ('true diabetes'). Individuals with FSG greater or equal than 126 mg/dL were also classified as diabetics ('true diabetes'). The FSG is the standard test performed in Costa Rica for diagnosis. Serum fasting glucose 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 informed consent was approved by the University of Costa Rica's Institutional Review Board (IRB). The final sample is composed by 1,539 individuals with complete information on diabetic status, self-reported diabetes, literacy and some diabetes-related symptoms.

Methods

The agreement between self-reported diabetes and clinical data in ENSA 2000 and CRELES will be assessed by the percentage of agreement, sensitivity, specificity, positive predictive value, negative predictive value, Kappa and Receiver Operating Characteristic (ROC) curve statistics. Positive predictive value refers to the chance that a positive self-reported diabetes result will be correct, while negative predictive value is concerned only with negative self-reports. However, it is important to mention that both positive and negative predictive values change if the prevalence of the disease changes. The ROC curve is the representation of the tradeoffs between sensitivity and specificity. The closer the area under the ROC curve is to 1.0, the better is the self-report.

Logistic regressions will also be used to analyze the association between actual diabetic status and self-reported diabetes taking into consideration demographic and social characteristics of the respondents. This analysis will be performed using the logistic regression approach. Given the predictive value of this estimation, the final equation will then be used to estimate the new

diabetes prevalence rates for SABE and MHAS settings. More details about the implemented procedure used to estimate the new diabetes prevalence rates are described below. More detailed information about the determinants of accurate self-reports and model selection is available upon request. All statistical analyses were performed using STATA 9.1 S.E. Sample weights were used to determine the predicted prevalence of diabetes. Finally, diabetes-free life expectancy will be estimated based on the Sullivan method and disability-free life expectancies by diabetic status will be estimated using IMach 0.98g.

Method for estimating new prevalence rates for SABE and MHAS

Using data from ENSA and CRELES the initial goal is to fit a predictive model of the true diabetic status using the self-reported measure of diabetes as a covariate. Logistic regression will be used to fit the predictive model since the dependent variable is binary. Basic demographic data, such as age and sex, are included in the model. As for socioeconomic characteristics, education (literacy) was also included in the model. Finally, because many individuals with undiagnosed diabetes have symptoms of the disease, excessive thirst (polydipsia) and weight loss were also used as independent variables. In the case of CRELES, fatigue was also included in the final model. The final models are:

$$P(Y=1 \mid X_{ENSA}) = \frac{e^{\alpha+\beta_i X_i(ENSA)}}{1+e^{\alpha+\beta_i X_i(ENSA)}}$$

And

$$P(Y=1 \mid X_{CRELES}) = \frac{e^{\alpha+\beta_i X_i(CRELES)}}{1+e^{\alpha+\beta_i X_i(CRELES)}}$$

Where the probability of truly having diabetes (Y=1; 'true' diabetes) is a function of the independent predictor variables (X). The selected independent predictor variables for are: age,

gender, polydipsia, weight loss and literacy. Other symptoms such as swollen feet, dizziness, polyuria and other possible predictors of diabetes, such as BMI were also analyzed. The decision to use those five independent predictors comes from the fact that they represent important demographic variables, diabetes symptoms and literacy adds a social dimension that it is necessary when dealing with different contexts. The other variables did not improve the fit of the predictive models when applied to ENSA and CRELES data. The only exception was fatigue that seemed to be a good predictor of CRELES data. Therefore, it will be included in the predictions based on CRELES. However, it is not available in ENSA. The decision to do not incorporate BMI comes from the fact that there was no available data in SABE for Buenos Aires and data from MHAS only contains BMI measures for a limited sample. Therefore, the adoption of a common model was preferred. Therefore, obesity or BMI indicators are not included here. Finally, hypertension was also analyzed but it did not improve the predictive power of the models either.

The next step is to calculate the probability of a positive outcome using SABE and MHAS data. The predictive models are:

$$P(Y=1 \mid X_{SABE}) = \frac{e^{\alpha+\beta_i X_i(SABE)}}{1+e^{\alpha+\beta_i X_i(SABE)}}$$

And

$$P(Y=1 \mid X_{MHAS}) = \frac{e^{\alpha+\beta_i X_i(MHAS)}}{1+e^{\alpha+\beta_i X_i(MHAS)}}$$

The predicted probabilities make use of the independent variables in SABE and MHAS, respectively. As it is well-known, in the logistic regression model, the predicted values for the dependent variable are always in the range 0 to 1 regardless of the regression coefficients or the magnitude of the independent variables. However, the predicted values for the dependent

variable will never be exactly 0 or 1. Therefore, in order to obtain the new estimates of diabetes prevalence, two different approaches of categorizing the predicted probabilities were adopted: a) individual predicted values of the dependent variable were averaged across groups of interest and b) those to which the predicted probability was higher than a random number were classified as having diabetes, while those with lower values than a random number were categorized as non-diabetics. This last approach was repeated 100 times; the values presented on tables are averages of these one hundred estimations. Because the results of these two approaches are almost identical, the estimates presented later on this paper are the ones obtained using the second one. Finally, for the computation of IMach, one random number out of those one hundred was selected and it was used to generate the tables with the results of total and disability-free life expectancy by diabetic status.

Results

Agreement between self-report and clinical data – ENSA 2000 and CRELES

Table 3 shows the contingency table for the self-reported diabetes versus the actual diabetic status based on ENSA 2000. The estimated diabetes prevalence using the self-reported measure reaches 16.6%, while the 'true' prevalence rate is somewhat higher – 18.0%. The contingency table also shows that there is a very high agreement (94.9%) between the two sources of information. The Kappa statistic reaches 0.82. Sensitivity reaches 81.7%, which indicates that 81.7% of the individuals with diabetes are aware (self-reported) of their diabetic status. Specificity is also very high – 97.7%, which shows that among those who do not have the disease, the vast majority, 97.7%, knows that they are not diabetic. In other words, only a small percentage (2.3%) of those who do not satisfy the medical criteria that (self) reported being diabetic. The positive predictive value reaches 88.8% and the negative predictive value 96%. The

area under the ROC curve is 0.897. Further analyses also show that levels of agreement between self-reported measures and 'true' diabetic status are very similar among males and females, 95.2% and 94.6%. Kappa values were 0.81 and 0.82 for males and females, respectively. Agreement level among individuals aged 50-75 is slightly lower (94.7%) than among older individuals aged 75+ (95.6%). Kappa is exactly the same, 0.82, for both age groups.

Table 3: Unweighted estimates of self-reported versus 'true' diabetic status, Mexico, ENSA 2000

Diabetes Status		Self-reported		
		No	Yes	Total
'True' diabetes	No	9,966	230	10,196
	Yes	410	1,829	2,239
	Total	10,376	2,059	12,435

Table 4 shows similar information presented in Table 3, but based on CRELES. Results indicate that self-reported prevalence rate reaches 18.8%, while the 'true' prevalence is much higher – 33.0%. The contingency table also shows that there is a reasonable agreement (83%) between the two sources of information. The Kappa statistic reaches 0.57, which can be considered a moderate agreement. Sensitivity reaches 52.6%, which indicates that only 52.6% of the individuals with diabetes are aware (self-reported) of their diabetic status. On the other hand, specificity is very high – 97.9%, which shows that among those who do not have the disease, the vast majority, 97.9%, knows that they are not diabetic. In other words, only a small percentage (2.1%) of those who do not satisfy the medical criteria that (self) reported being diabetic. The positive predictive value reaches 92.4% and the negative predictive value 81%. The area under the ROC curve is 0.75. Further analyses also show that levels of agreement between self-reported measures and 'true' diabetic status are very similar among males and females. Kappa values

were 0.53 and 0.59, respectively, for males and females. Agreement level among individuals aged 50-75 is slightly higher (83.8%) than among older individuals aged 75+ (82.1%). Kappa is also higher among young adults, 0.61, than among older adults, 0.52.

Table 4: Unweighted estimates of self-reported versus 'true' diabetic status, Costa Rica, CRELES

Diabetes Status		Self-reported		
		No	Yes	Total
'True' diabetes	No	1,009	22	1,031
	Yes	241	267	508
	Total	1,250	289	1,539

The results presented in this section show that estimates based on data from Costa Rica (CRELES) contrast markedly with the results from Mexico (ENSA). In Costa Rica, 'true' prevalence is 76% higher than self-reported, while in Mexico it is only 9% higher. There are several reasons that may help explain this large difference. One possibility is that these results truly reflect the knowledge of diabetes status in these two countries. In other words, it is possible that Mexicans are more aware of their diabetic status, while undiagnosed cases are more common in Costa Rica. However, it is also possible that these differences emerge due to different diagnostic procedures adopted in ENSA and CRELES. More specifically, ENSA used a random (casual) fasting glucose (cutoff of 200 mg/dL), while CRELES used a lower cutoff point (126 mg/dL) because most individuals reported being fasting before the exam. However, there is an important drawback when using this lower cutoff point at CRELES. In reality, interviewers cannot verify that respondents are really fasting when the blood sample is being drawn. Therefore, it is possible that CRELES captures more 'undiagnosed' cases if, for instance, individuals who reported fasting were not and their measured glucose levels were higher than 126 mg/dL just because they had eaten recently. If this conjecture is true, 'true' rates in CRELES

may be overestimating the actual prevalence. At the same time, it is possible that ENSA gives a 'real' rate that is still underestimated. The next section explores the factors influencing misreport and accurate report.

Factors influencing accurate report for diabetics and non-diabetics

The following analyses divide the sample in two, diabetics and non-diabetics, following the rationale presented by Goldman et al. (2003). The first subsample is composed by individuals who have diabetes. The results based on ENSA and CRELES data are presented in Table 5. It shows that, in Mexico, older individuals who have diabetes are more likely to accurately report having the condition versus underreport. In Costa Rica, on the contrary, accuracy in the diabetes report decreases with age among those who have diabetes. Therefore, the results show that, in Mexico, individuals become more aware of their 'true' diabetic status as age increases, but this does not seem to be the case in Costa Rica. In Mexico, there are no statistical differences in the accuracy of self-report among men and women with diabetes, but in Costa Rica, women are more likely to accurately report the condition ($p<0.05$). In both countries, those who suffer from polydipsia are more likely to accurately self-report having diabetes, while those who suffer from weight loss are more likely to accurately report only in Mexico. In Costa Rica, individuals who reported fatigue are more likely to accurately report their positive diabetic status. In Mexico, those with higher education are also more likely to accurately report having diabetes ($p<0.10$).

Table 5: Estimated coefficients, confidence intervals and p-values for logistic models of the probability of accurate self-reports of diabetes for those individuals who have diabetes, Mexico (ENSA 2000) and Costa Rica (CRELES)

Variables and data source	Coefficient	p-value	95% Confidence Interval	
Mexico (ENSA)				
Age	0.019	0.004	0.006	0.032
Female	0.066	0.587	-0.172	0.303
Polydipsia	0.459	0.000	0.207	0.711
Weight loss	0.711	0.000	0.447	0.975
Literacy	0.236	0.087	-0.034	0.505
Constant	-0.266	0.566	-1.174	0.641
N=2,239				
Costa Rica (CRELES)				
Age	-0.037	0.000	-0.057	-0.018
Female	0.375	0.048	0.003	0.748
Polydipsia	0.781	0.001	0.301	1.261
Weight loss	0.042	0.888	-0.546	0.630
Fatigue	0.617	0.003	0.208	1.026
Literacy	-0.061	0.812	-0.568	0.445
Constant	2.343	0.004	0.756	3.930
N=508				

Note: Underreport is the base outcome.

The second subsample is composed by those who do not have diabetes based on clinical data. The results based on ENSA and CRELES are presented in Table 6. The baseline category is overreport. The results indicate that, in Mexico, only the symptoms (polydipsia and weight loss) are negatively associated with accurate report among those without diabetes. In Costa Rica, on the other hand, the only variable that is statistically significant is gender. The results indicate that women without diabetes in Costa Rica are less likely to accurately report their non-diabetic status. In other words, they are more likely to overreport having the condition when, in fact, they do not have diabetes.

Table 6: Estimated coefficients, confidence intervals and p-values for logistic models of the probability of accurate self-reports of diabetes for those individuals who do not have diabetes, Mexico (ENSA 2000) and Costa Rica (CRELES)

Variables and data source	Coefficient	p-value	95% Confidence Interval	
Mexico (ENSA)				
Age	-0.009	0.175	-0.022	0.175
Female	-0.060	0.668	-0.336	0.668
Polydipsia	-0.556	0.001	-0.876	0.001
Weight loss	-0.918	0.000	-1.223	0.000
Literacy	-0.296	0.076	-0.622	0.076
Constant	4.866	0.000	3.940	0.000
N=10,196				
Costa Rica (CRELES)				
Age	0.014	0.516	-0.029	0.058
Female	-1.159	0.024	-2.167	-0.152
Polydipsia	-0.403	0.409	-1.359	0.553
Weight loss	-0.060	0.925	-1.311	1.191
Fatigue	-0.612	0.176	-1.498	0.275
Literacy	-1.047	0.167	-2.533	0.439
Constant	4.785	0.014	0.982	8.588
N=1,031				

Note: Overreport is the base outcome.

The results from the multivariate analysis show that demographic and social characteristics are important factors influencing the quality of the self-report. The next section will present the models in which these characteristics function as predictors of the individual's true diabetic status in a multivariate analysis that incorporates self-report as the main covariate.

Selected predictive models based on ENSA 2000 and CRELES data

The results described in the literature and the ones from ENSA and CRELES indicate that self-report is reasonably accurate, or at least, moderately accurate in the case of CRELES. As a consequence, one can expect that self-report is a good predictor of the true diabetic status.

Therefore, the next step is to perform a multivariate analysis that includes self-report as one of the predictive variables as well other demographic, health and social characteristics. If self-reported diabetes was a perfect predictor of 'true' diabetes then all the other coefficients would be zero, however because self-report is not a perfect predictor of 'true' diabetic status, other covariates are included in the model to improve its predictive potential.

Table 7 provides the coefficients, standard errors, p-values and 95% confidence intervals of the selected models for ENSA and CRELES. The results from ENSA confirm that all selected variables are statistically significant ($p<0.10$), except for literacy. The coefficients indicate that self-reported diabetes is the most important predictor of the 'true' diabetic status. The two diabetes symptoms, polydipsia and weight loss, are also important predictors of actual diabetes. Results from ENSA also show that diabetes self-report is, by far, the most important predictor of the 'true' diabetic status, while fatigue is the only other variable that it is statistically significant. The estimates presented in Table 7 will be used in the next section to predict the diabetes prevalence using self-reported data from SABE and MHAS.

Graph 1 and Graph 2 show the estimated diabetes prevalence by age and the predicted values obtained by using the models presented in Table 7. It also shows the predicted values depending on the diabetes self-report. The results show that predicted values are closer to self-reports in the ENSA data than in the CRELES data. This finding is due to the fact that agreement between self-report and clinical data is much higher in ENSA than in CRELES. The poorer of agreement between self-reports and clinical data in CRELES comes mainly from the fact that several individuals who self-reported not being diabetic had, in fact, high glucose levels that were higher than the cutoff point for diabetes diagnostic. As a consequence of this lower agreement, Graph 2 shows that there is a considerably high underlying prevalence of diabetes

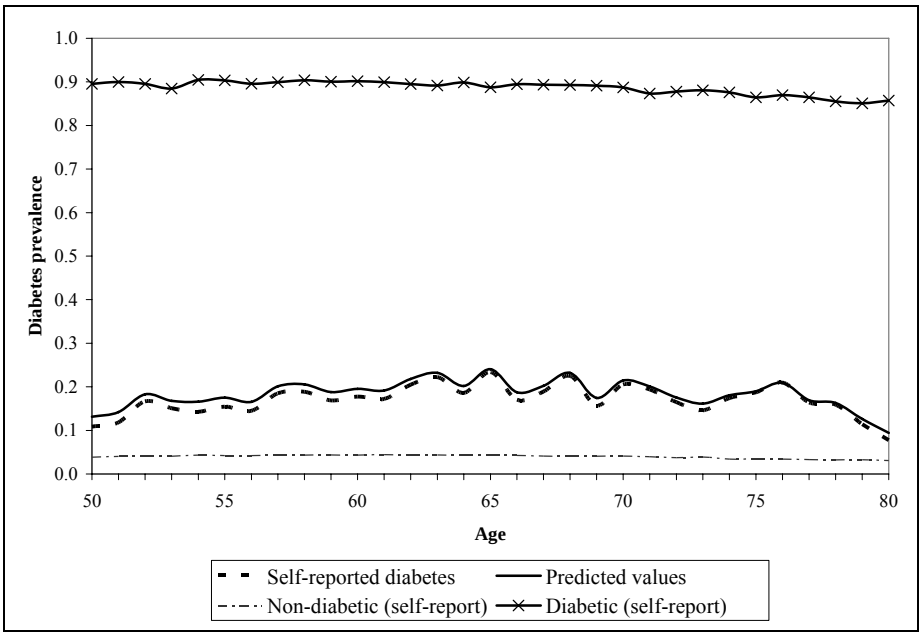
among those who self-reported being non-diabetic. In ENSA, this underlying prevalence of undiagnosed cases is considerably lower. On the other hand, both datasets indicate that, among those who self-reported being diabetic, the large majority really has it. Even though, false-positives exist and they are more prevalent in ENSA.

Table 7: Coefficients, standard errors, p-value and 95% confidence intervals of the probability of truly having diabetes, ENSA 2000 and CRELES

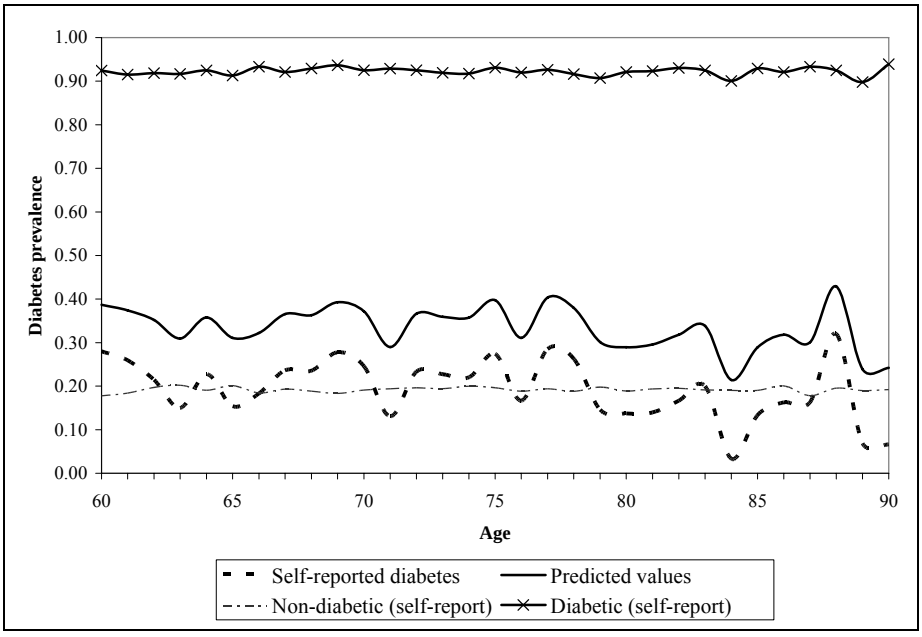
Variables	Coefficient	Robust Std. Err.	p-value	95% Confidence Interval	
Mexico (ENSA)					
Age	0.142	0.053	0.008	0.038	0.247
Age2	-0.001	0.000	0.004	-0.002	0.000
Female	0.194	0.088	0.027	0.022	0.366
Self-reported diabetes	5.117	0.088	0.000	4.945	5.289
Polydipsia	0.658	0.100	0.000	0.463	0.854
Weight loss	0.460	0.103	0.000	0.257	0.662
Literacy	0.126	0.101	0.211	-0.071	0.323
Constant	-7.822	1.738	0.000	-11.229	-4.415
N=12,435					
Costa Rica (CRELES)					
Age	0.004	0.007	0.567	-0.010	0.018
Female	-0.009	0.137	0.948	-0.278	0.260
Self-reported diabetes	4.090	0.248	0.000	3.603	4.577
Polydipsia	-0.243	0.183	0.183	-0.601	0.115
Weight loss	-0.035	0.217	0.872	-0.459	0.390
Fatigue	-0.512	0.154	0.001	-0.813	-0.210
Literacy	0.227	0.180	0.206	-0.125	0.580
Constant	-1.702	0.593	0.004	-2.865	-0.539
N=1,539					

Source: ENSA and CRELES

Graph 1: Self-reported prevalence by age and predicted probabilities by age and self-reported diabetic status, Mexico (ENSA)



Graph 2: Self-reported prevalence by age and predicted probabilities by age and self-reported diabetic status, Costa Rica (CRELES)



New estimates of diabetes prevalence data in Latin America and the Caribbean

As discussed before, diabetes self-report is usually moderately well reported, but there is evidence that self-reported prevalence rates in Latin America and the Caribbean underestimate the actual prevalence. Therefore, the purpose of this paper is to use self-reported diabetes status and other important demographic and social covariates to estimate the actual diabetes status, and ultimately obtain new prevalence rates. This selected approach was to use estimates from ENSA and CRELES, which contain both clinical and self-reported measures, to predict the probabilities of actually having diabetes based on covariates from SABE and MHAS. The underlying assumption is that the models are valid for all participating countries. This is a very strong assumption. However, there are reasons to believe that this is a reasonable one. In the case of ENSA, Mexico represents a large country from Latin America that is more closely related to countries represented in SABE and MHAS than developed countries to which this type of data would be available. Moreover, for two of the analyzed samples, Mexico City in SABE and Mexico in MHAS, we can be even more confident of its suitability. However, differences between self-reported and 'true' rates in Mexico are relatively small and this may be due to the adopted diagnostic procedure used in this survey. This brings to the advantage of using CRELES from Costa Rica. In CRELES, the diagnostic procedure produced a larger number of 'true negative' cases (undiagnosed) relatively to ENSA from Mexico. In fact, the larger number of 'newly diagnosed' cases is similar to the descriptions of some previous studies conducted in Latin America and the Caribbean. Moreover, the inclusion of these two extremes gives a better idea of the range in which the real prevalence rates may be. Finally, the use of these models constitute an alternative to the procedures used in WHO studies (King, Aubert and Herman 1998 and Wild et al. 2004) that predict prevalence rates for several Latin American and Caribbean

countries based on a single number that indicates the proportion of cases that have not being previously diagnosed. Moreover, these previous studies from WHO use prevalence data from few countries in the region that are used for many other countries. The approach adopted in this study makes use of relevant data on self-reported measures and other demographic and social characteristics of the analyzed countries to predict the 'true' diabetes prevalence.

The main results are presented in Table 8 and Graph 3. Table 8 shows the self-reported prevalence and the predicted prevalence rates and 95% confidence intervals of the predicted values based on CRELES and ENSA. Important to note that sample sizes of predicted values are smaller than the ones based on self-reports due to missing values on variables used in the prediction. SABE sample is composed by 10,540 individuals, but the sample size based on ENSA is restricted to 9,976 and the one based on CRELES is limited to 9,932 individuals. MHAS sample is composed by 13,022 individuals aged 50 and over, while samples based on predicted values are restricted to 11,747 and 11,737, based on ENSA and CRELES, respectively.

Results from Table 8 indicate that diabetes self-reported prevalence is underestimated in most analyzed settings when ENSA relations are used to predict new estimates, but the self-reported prevalence is always underestimated if CRELES is used instead. Predicted rates increase by 2% in Mexico City to 19.5% in Santiago when ENSA is used and, predicted values are about 1% lower in Bridgetown than the self-reported ones. However, when CRELES is used instead, the differences between self-reported and predicted are considerably larger – varying from 64% in Bridgetown to 126% in Montevideo.

In Buenos Aires, the self-reported prevalence is 12.4%, but predicted prevalence reaches 14% when ENSA is used and 30% when CRELES is used. In Bridgetown, predicted values based on ENSA are very similar to the self-reported (21.7%), but it reaches 37% if CRELES is

used. In São Paulo, predicted prevalence based on ENSA reaches 19.4% and 33.7% when CRELES is used (compared to 18% of the self-reported rate). In Santiago, the self-reported rate is 13.3%, but it rises to 16% based on ENSA and 29% if it is based on CRELES. Predicted prevalence based on ENSA in Havana is about 10% higher than self-reported prevalence rate – 14.8% and 16.2%, respectively. However, 'true' prevalence rate in Havana would reach 32% if CRELES is used. In Mexico City, predicted rate are 2% higher than the ones estimated using self-reports, but in Mexico underreport seems to be considerably higher based on ENSA. In fact, predicted prevalence rates in Mexico based on ENSA are 13% higher than self-reported among those aged 60 and over. If CRELES is used, underreport in Mexico is also higher than in Mexico City. This finding may reflect a higher awareness in Mexico City given more access to information and to the health care system. Finally, estimated prevalence in Montevideo reaches 14%, but predicted prevalence rates based on ENSA are 13% higher (18%) and 126% higher (31%) based on CRELES.

Graph 3 shows the 95% confidence intervals of the self-reported and predicted prevalence rates of diabetes on seven large urban areas in Latin America and the Caribbean and Mexico. This graph makes clear that confidence intervals of self-reported rates and predicted rates based on ENSA overlap for most part in all cities and also in Mexico. On the other hand, predicted estimates from CRELES do not overlap at all with self-reported rates in Buenos Aires and Montevideo. In Bridgetown, São Paulo and Mexico City, predicted prevalence rates based on CRELES overlap with self-reported rates, but in Santiago and Mexico they only overlap with predicted rates based on ENSA. In Havana, predicted rates based on CRELES slightly overlap with predicted rates from ENSA.

Table 8: Self-reported and predicted diabetes prevalence rates, and 95% confidence intervals (weighted estimates), SABE and MHAS

Data sources and countries	**Self-reported prevalence**	**Predicted using ENSA**			**Predicted using CRELES**		
		Prevalence	**95% CI**		**Prevalence**	**95% CI**	
SABE							
Buenos Aires (Argentina)	*12.4*	13.9	12.1	15.6	30.2	28.6	31.8
Bridgetown (Barbados)	*21.7*	21.4	19.6	23.2	37.3	35.6	38.9
São Paulo (Brazil)	*18.0*	19.4	17.7	21.0	33.7	32.3	35.1
Santiago (Chile)	*13.3*	15.9	13.7	18.1	28.7	26.7	30.8
Havana (Cuba)	*14.8*	16.2	14.8	17.6	31.8	30.6	33.0
Mexico City (Mexico)	*21.6*	22.0	20.0	24.1	35.5	33.7	37.3
Montevideo (Uruguay)	*13.7*	15.5	13.8	17.3	31.0	29.5	32.5
MHAS							
Mexico (50+)	*15.3*	17.6	16.6	18.7	31.0	30.0	31.9
Mexico (60+)	*16.7*	18.8	17.2	20.4	31.5	30.1	33.0

Graph 3: 95% confidence intervals of self-reported (self) and predicted diabetes prevalence rates, based on ENSA and CRELES, among individuals aged 60 and over in seven large urban areas and Mexico

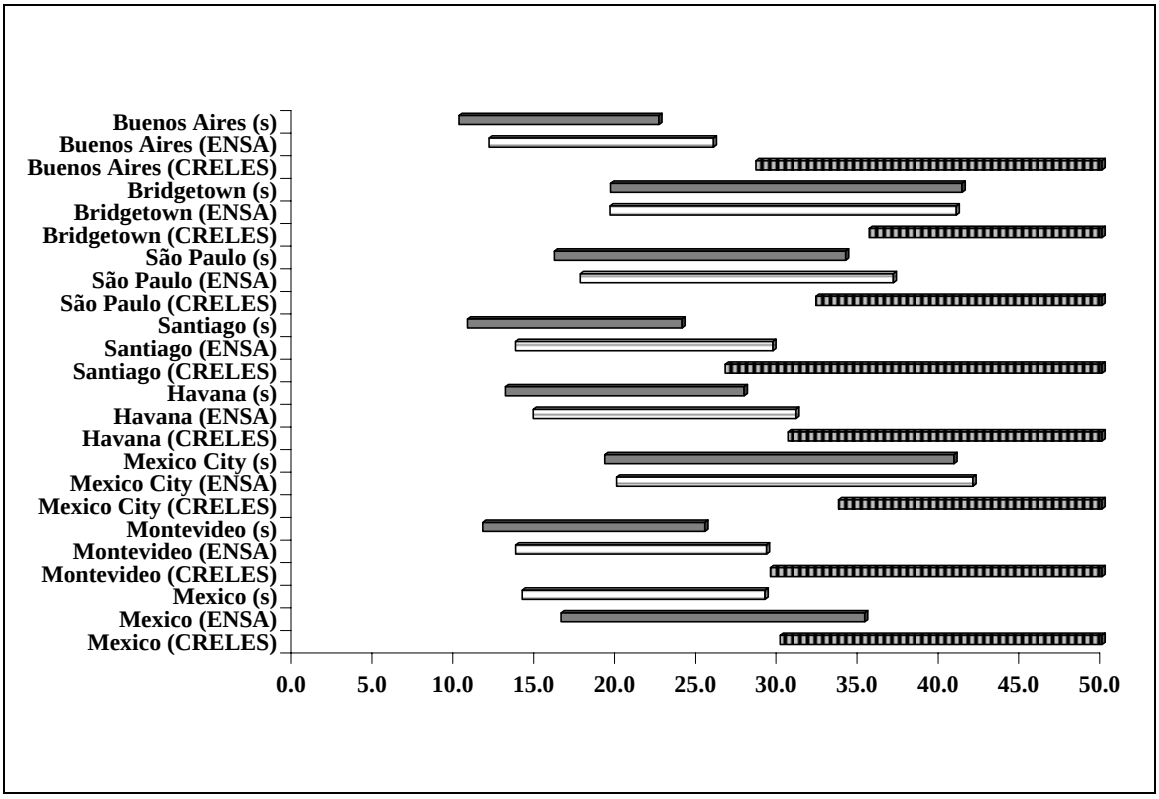


Table 9 presents the self-reported and predicted prevalence rates based on ENSA by sex. The results confirm that, for both men and women, underreporting is more common than overreporting. The only exception would be for females in Bridgetown and males in Mexico City where self-reported rates are higher than predicted rates based on ENSA. In Bridgetown, self-reported rates are very close to the predicted prevalence rates based on ENSA. In Buenos Aires, São Paulo, Santiago and Montevideo the adjustments are somewhat higher among females than among males. However, the most striking difference between self-reported and predicted rates based on ENSA is found in Havana. In this city, self-reported prevalence rates among males are much lower than women's rate. The predicted rates among men based on ENSA are still about

half of what it is found among women, but the difference between self-reported rates and predicted rates is amazing. More specifically, predicted rates based on ENSA are about 33% higher among men than self-reported rates, but only 4% higher among women. For Mexico, predicted rates based on ENSA are about 12.5% higher than self-reported ones for both sexes.

Table 10 shows similar data from Table 9, but with predicted rates based on CRELES rather than ENSA. The results indicate that, in most settings, increases in prevalence rates are higher among men than women in all settings, except in Buenos Aires and Mexico City.

Table 9: Self-reported, predicted diabetes prevalence rates based on ENSA and their relative differences (%) for elderly individuals aged 60 and over by sex, SABE and MHAS

Data sources	**Males**			**Females**		
	Self-reported	Predicted	Relative Difference	Self-reported	Predicted	Relative Difference
SABE						
Buenos Aires (Argentina)	14.0	14.5	3.6	11.4	13.5	18.7
Bridgetown (Barbados)	18.7	19.0	1.2	23.6	23.0	-2.8
São Paulo (Brazil)	17.0	18.1	6.5	18.7	20.3	8.2
Santiago (Chile)	12.0	14.4	19.2	14.1	17.0	20.9
Havana (Cuba)	7.3	9.6	32.5	20.0	20.8	4.2
Mexico City (Mexico)	22.4	22.2	-0.8	21.0	21.9	4.3
Montevideo (Uruguay)	12.4	14.0	12.5	14.5	16.4	13.4
MHAS						
Mexico	14.5	16.3	12.5	18.7	21.0	12.4

Table 10: Self-reported, predicted diabetes prevalence rates based on CRELES and their relative differences (%) for elderly individuals aged 60 and over by sex, SABE and MHAS

	Males			Females		
	Self-reported	Predicted	Difference	Self-reported	Predicted	Difference
SABE						
Buenos Aires (Argentina)	14.0	31.4	124.4	11.4	29.4	159.1
Bridgetown (Barbados)	18.7	36.0	92.3	23.6	38.0	60.9
São Paulo (Brazil)	17.0	33.6	97.8	18.7	33.8	80.2
Santiago (Chile)	12.0	27.9	132.1	14.1	29.3	107.8
Havana (Cuba)	7.3	27.3	276.2	20.0	34.9	74.8
Mexico City (Mexico)	22.4	36.5	62.9	21.0	34.8	65.9
Montevideo (Uruguay)	12.4	30.9	148.5	14.5	31.1	114.8
MHAS						
Mexico (60+)	14.5	30.0	106.8	18.7	32.9	75.9

Health status information of "newly" diagnosed diabetic cases

This section analyzes the health status of 'newly' diagnosed cases in SABE and MHAS given the predicted estimates based on CRELES. More specifically, the question is whether those 'newly diagnosed cases' face lower or higher disability than those who 'correctly' self-reported being diabetic. Predicted values based on ENSA are quite similar to self-reported ones and, for this reason, will not be explored here. Information on social and demographic information of 'newly' diagnosed cases is available upon request.

Results presented in Table 11 and Table 12 show that in SABE and MHAS those newly diagnosed cases are less likely to have ADL, IADL or Nagi limitations than those who would be classified diabetic in both methods (self-reports and predicted status based on CRELES). Individuals who are classified as non-diabetics independently of the method are less likely to have functional disabilities than those who are 'true positive' cases. These results indicate that those 'new' cases added to the truly diabetics are relatively better off in terms of functional limitations than those who would be consistently classified as diabetics.

Table 11: Coefficients and 95% confidence intervals of the odds of having at least one ADL, IADL and Nagi limitation in SABE depending on the diabetes status classification based on CRELES

	Coefficient	p-value	95% CI	
ADL (N=9,878)				
Age	0.067	0.000	0.061	0.073
Female	0.425	0.000	0.315	0.535
Newly diagnosed *	-0.473	0.000	-0.649	-0.296
True negative	-0.388	0.000	-0.524	-0.252
Overreport	0.244	0.235	-0.159	0.647
Constant	-6.249	0.000	-6.719	-5.778
IADL (N=9,709)				
Age	0.107	0.000	0.100	0.113
Female	0.691	0.000	0.580	0.801
Newly diagnosed *	-0.567	0.000	-0.740	-0.393
True negative	-0.525	0.000	-0.660	-0.390
Overreport	0.212	0.305	-0.193	0.618
Constant	-8.992	0.000	-9.488	-8.496
Nagi (N=9,904)				
Age	0.057	0.000	0.051	0.062
Female	0.949	0.000	0.863	1.035
Newly diagnosed *	-0.540	0.000	-0.693	-0.388
True negative	-0.467	0.000	-0.592	-0.343
Overreport	-0.057	0.774	-0.447	0.333
Constant	-3.773	0.000	-4.184	-3.363

Note: * "True positive" (individuals who self-reported being diabetic that would also be classified as diabetics based on CRELES) is the base outcome.

Table 12: Coefficients and 95% confidence intervals of the odds of having at least one ADL, IADL and Nagi limitation in Mexico (MHAS) depending on the diabetes status classification based on CRELES

	Coefficient	p-value	95% CI	
ADL (N=10,707)				
Age	0.064	0.000	0.058	0.070
Female	0.449	0.000	0.318	0.580
Newly diagnosed *	-0.818	0.000	-1.042	-0.595
True negative	-0.538	0.000	-0.695	-0.380
Overreport	0.067	0.806	-0.466	0.599
Constant	-6.222	0.000	-6.677	-5.767
IADL (N=10,706)				
Age	0.089	0.000	0.082	0.096
Female	0.815	0.000	0.660	0.970
Newly diagnosed *	-0.883	0.000	-1.130	-0.635
True negative	-0.710	0.000	-0.885	-0.534
Overreport	0.223	0.433	-0.335	0.781
Constant	-8.358	0.000	-8.891	-7.825
Nagi (N=10,708)				
Age	0.055	0.000	0.051	0.060
Female	0.752	0.000	0.670	0.834
Newly diagnosed *	-0.705	0.000	-0.846	-0.563
True negative	-0.580	0.000	-0.691	-0.468
Overreport	-0.050	0.803	-0.445	0.344
Constant	-3.843	0.000	-4.155	-3.530

Note: * "True positive" (individuals who self-reported being diabetic that would also be classified as diabetics based on CRELES) is the base outcome.

The next section will explore the impact of having additional cases of diabetes on diabetes-free life expectancy. Finally, the paper will evaluate the impact of the diabetes reclassification on total and disability-free life expectancy by diabetic status.

New estimates of diabetes-free life expectancy taking into account misreport of diabetes status

Table 13 shows the diabetes-free life expectancy and the average number of years that individuals aged 60 and over are expected to live with diabetes when self-reported data is

adjusted by misreport using ENSA. In all settings, women are expected to live longer, but also with a larger number of years with diabetes. Except for Buenos Aires and Mexico City, in all other settings women are also expected to live a larger proportion of their lives with diabetes.

Results presented in Table 13 based on ENSA show that males aged 60 years in Buenos Aires are expected to live, on average, 2.5 years with diabetes, while their female counterparts are expected to live 2.9 years. In Bridgetown, diabetes life expectancy reaches 3.5 years among men aged 60, while women at the same age are expected to live 5.2 years, on average, with diabetes. Men aged 60 in São Paulo are expected to live, on average, 3 years with diabetes while women of the same age are expected to live 4.3 years with diabetes. In Santiago, the diabetes life expectancy at age 60 reaches 2.7 years among men and 3.8 years among women. In Havana, the difference in diabetes life expectancy between males and females is remarkable – men aged 60 are expected to live, on average, 1.7 years with diabetes, while their female counterparts are expected to live 4.6 years. In Mexico City, diabetes life expectancy at age 60 reaches 4.4 years among men and 4.7 years among women. However, in Mexico the difference in diabetes life expectancy between males and females is greater than in Mexico City. In Mexico, diabetes life expectancy at age 60 reaches 3.3 years for males and 4.4 years for females. Finally, results from Montevideo show that men aged 60 are expected to live, on average, 2.4 years with diabetes and their female counterparts, 3.5 years.

The results presented in Table 13 based on ENSA are very similar to those using self-reported rates. This happens because adjustments in the predicted rates are somewhat small when ENSA is used. Therefore, the absolute variation in the diabetes life expectancies is consequently small. In fact, for females in Bridgetown, the new estimate of diabetes life expectancy is 0.2 years smaller than it would be if self-reported rates were being used. On the other extreme,

women in Santiago have their diabetes life expectancy increased in 0.6 years when predicted rates based on ENSA are used instead of self-reported rates. In relative terms, diabetes life expectancy among males is increased by 32% in Havana, 19% in Santiago, 13% in Montevideo, 11% in Mexico, 5% in São Paulo, 4% in Buenos Aires, and 1% in Bridgetown. In Mexico City, predicted rates are smaller than self-reported ones, therefore expected number of years to be lived with diabetes are reduced by 3% among males. Among females, the largest increases are found among women in Buenos Aires (18%), Santiago (18%), and Montevideo (14%). Among older women in Mexico and São Paulo, changes in the expected number of years to be lived with diabetes are in-between, 8% and 6%, respectively. In Havana and Mexico City, adjustments are smaller – 3% and 2%, respectively. In Bridgetown, the predicted diabetes life expectancy taking misreport into account is 4% smaller than the one obtained with self-report.

Table 14, on the other hand, shows significant changes in diabetes-free life expectancy and diabetes life expectancy. Table 14 shows the results when CRELES is used to predict the prevalence rates of SABE and MHAS. Diabetes life expectancies are considerably higher than self-reported rates in all settings. The smaller increase is found among males in Mexico City, 2.8 years, while the larger increase is found among females in Buenos Aires, 4 years.

Results from Table 14 confirm that women live longer lives, but with a larger number of years with diabetes. Women in Bridgetown, Santiago, Havana and Mexico are also expected to live a larger percentage of their lives with diabetes. Results based on CRELES show that males in Mexico City are expected to live 7.4 years with diabetes and 13 without after they reach age 60. Their female counterparts are expected to live 7.7 years with diabetes and 14.5 without based on predicted rates using CRELES. In Bridgetown, diabetes life expectancy reaches almost 9 years among females and is about 6 years among males. The largest differential in diabetes life

expectancy between males and females is found in Havana. In this urban area, men aged 60 are expected to live, on average 5 years with diabetes and their female counterparts 7.7 years. The smallest difference is found in Mexico City – diabetes life expectancy only differs in 0.3 years.

In relative terms, the increases in diabetes life expectancies when CRELES is used are also remarkable. In Buenos Aires, diabetes life expectancy based on predicted prevalence rates based on CRELES are 130% and 162% higher than self-reported ones for males and females, respectively. In Bridgetown, predicted rates based on CRELES are 88% and 61% higher than self-reported ones, for males and females, respectively. In São Paulo, increases are in the order of 100% and 82%, respectively, for males and females. Males in Santiago would have their diabetes life expectancy increased by 140% and females in 111% if rates estimated using CRELES are the real ones. In Havana, male diabetes life expectancy is 280% higher when predicted rates based on CRELES are used rather than self-reported ones. Among females in Havana, this increase would be around 74%. In Mexico City, adjustments would be relatively smaller than in other cities – around 63% to 66%. In Montevideo, male diabetes life expectancy would be 157% higher than when self-reported rates are used and, among females, the increase would be near 127%. Finally, in Mexico diabetes life expectancy more than double among males and among females the diabetes life expectancy is 77% higher when prevalence rates are predicted using CRELES than when self-reported prevalence rates are used.

Table 13: Total life expectancy, diabetes-free life expectancy and diabetes life expectancy at age 60, by sex, based on predicted diabetes prevalence rates obtained using ENSA

Sex and region	Total life expectancy at age 60 (e_{60})	Diabetes life expectancy (DLE_{60}) based on ENSA	Diabetes-free life expectancy ($DFLE_{60}$) based on ENSA	s.e. (DFLE)
Males				
Buenos Aires (Argentina)	17.4	2.5	14.9	0.01
Bridgetown (Barbados)	17.6	3.5	14.2	0.29
São Paulo (Brazil)	17.2	3.0	14.2	0.01
Santiago (Chile)	19.9	2.7	17.1	0.24
Havana (Cuba)	18.4	1.7	16.7	0.01
Mexico City (Mexico)	20.3	4.4	16.0	0.01
Montevideo (Uruguay)	17.6	2.4	15.2	0.02
Mexico	20.4	3.3	17.2	0.17
Females				
Buenos Aires (Argentina)	22.1	2.9	19.1	0.01
Bridgetown (Barbados)	22.8	5.2	17.6	0.31
São Paulo (Brazil)	21.9	4.3	17.6	0.01
Santiago (Chile)	24.0	3.8	20.1	0.25
Havana (Cuba)	22.0	4.6	17.4	0.02
Mexico City (Mexico)	22.1	4.7	17.5	0.01
Montevideo (Uruguay)	22.9	3.5	19.4	0.02
Mexico	22.3	4.4	17.9	0.01

Table 14: Total life expectancy, diabetes-free life expectancy and diabetes life expectancy at age 60, by sex, based on predicted diabetes prevalence rates obtained using CRELES

Sex and region	Total life expectancy at age 60 (e_{60})	Diabetes life expectancy (DLE_{60}) based on CRELES	Diabetes-free life expectancy ($DFLE_{60}$) based on CRELES	s.e. (DFLE)
Males				
Buenos Aires (Argentina)	17.4	5.4	12.0	0.01
Bridgetown (Barbados)	17.6	6.4	11.2	0.35
São Paulo (Brazil)	17.2	5.8	11.5	0.01
Santiago (Chile)	19.9	5.5	14.4	0.32
Havana (Cuba)	18.4	5.0	13.4	0.02
Mexico City (Mexico)	20.3	7.4	13.0	0.01
Montevideo (Uruguay)	17.6	5.4	12.2	0.02
Mexico	20.4	6.1	14.3	0.30
Females				
Buenos Aires (Argentina)	22.1	6.5	15.6	0.01
Bridgetown (Barbados)	22.8	8.6	14.1	0.36
São Paulo (Brazil)	21.9	7.3	14.6	0.02
Santiago (Chile)	24.0	6.9	17.1	0.31
Havana (Cuba)	22.0	7.7	14.3	0.02
Mexico City (Mexico)	22.1	7.7	14.5	0.01
Montevideo (Uruguay)	22.9	7.0	15.9	0.02
Mexico	22.3	7.3	15.1	0.01

The results from this section indicate that diabetes imposes a considerable burden on these populations. Individuals aged 60 and over in these countries are expected to live between 10% and 23% of their remaining lives with diabetes if misreport is adjusted using ENSA, but these percentages can be even higher – 27% to 38% if misreport is estimated using CRELES. In any case, because diabetes is associated with increases in the prevalence and incidence of disability, it imposes important economic and social costs on these populations. Also, the direct costs of diabetes (drugs, consultations and hospitalizations) are substantial (Barceló et al. 2004), particularly for developing countries.

New estimates of total and health life expectancy of diabetics and non-diabetics taking misreport into account

This final section uses data from Mexico that was corrected for misreport using CRELES. The goal is to present new estimates of total and health life expectancy of diabetics and non-diabetics taking misreport into account. Estimates based on ENSA are not presented here because they are very similar to the ones obtained with self-reports. Table 15 to Table 17 show that sample sizes are smaller when using predicted values because of the missing data on variables selected to predict the new values.

Table 15 show the total life expectancy, disability-free life expectancy and disabled life expectancy in Mexico based on ADL limitations. It compares information provided using self-reports and the ones based on the predicted values using CRELES. The results indicate that total life expectancy and disability-free life expectancy of diabetics is higher when diabetes information is corrected for misreport. In other words, estimates of total life expectancy and disability-free life expectancy of diabetics based on self-reports may be biased downwards. However, the estimates of disabled life expectancy of diabetics are quite similar disregarding the type of method used to classify individuals. At the same time, estimates of total life expectancy and disability-free life expectancy of non-diabetics are slightly lower when diabetes status is corrected by misreport. Therefore, there is some evidence that traditional methods of estimating total life expectancy and disability-free life expectancy of diabetics may be biased downwards. As discussed previously, individuals that are 'newly diagnosed' are relatively better off than those who self-report being diabetic – fact that it is consistent with the fact that undiagnosed individuals are usually at earlier stages of the disease progression. This conclusion is, however, not definitive because sample sizes are different between self-reports and predicted values. In

any case, it is important to understand that diabetes reduces total life expectancy and the bulk of this reduction comes in the form of reductions in the number of years expected to be lived without disability. In the restricted sample with predicted values, total life expectancy of diabetics at age 50 is about 4 years lower than total life expectancy of non-diabetics. Moreover, the bulk of this difference is due to reduction in years to be lived without disability. Similar conclusion is found when disability-free life expectancy is calculated based on IADL and Nagi limitations (Table 16 and Table 17).

Table 15: Total life expectancy, disability-free life expectancy and disabled healthy expectancy based on ADL limitations by age and diabetic status, Mexico, MHAS

Impairment and sample size	**Age**							
	50	(s.d)	**60**	(s.d)	**70**	(s.d)	**80**	(s.d)
Self-reports (N=12,050)								
Non-diabetics								
TLE	32.8	(0.606)	24.2	(0.585)	16.6	(0.567)	10.4	(0.531)
DFLE	28.8	(0.520)	20.5	(0.498)	13.4	(0.478)	7.8	(0.437)
DLE	3.9	(0.224)	3.7	(0.230)	3.2	(0.239)	2.6	(0.247)
Diabetics								
TLE	24.7	(0.927)	17.2	(0.782)	11.1	(0.656)	6.6	(0.537)
DFLE	20.5	(0.739)	13.4	(0.599)	7.9	(0.470)	4.0	(0.345)
DLE	4.2	(0.381)	3.8	(0.360)	3.3	(0.345)	2.6	(0.336)
Predicted values based on CRELES (N=10,043)								
Non-diabetics								
TLE	32.2	(0.663)	23.6	(0.627)	16.1	(0.591)	10.0	(0.539)
DFLE	28.5	(0.578)	20.2	(0.544)	13.1	(0.509)	7.6	(0.455)
DLE	3.7	(0.226)	3.4	(0.228)	3.0	(0.232)	2.4	(0.236)
Diabetics								
TLE	28.2	(0.823)	20.3	(0.739)	13.5	(0.666)	8.3	(0.575)
DFLE	24.0	(0.670)	16.4	(0.584)	10.2	(0.504)	5.6	(0.410)
DLE	4.1	(0.327)	3.8	(0.323)	3.4	(0.323)	2.7	(0.323)

Table 16: Total life expectancy, disability-free life expectancy and disabled healthy expectancy based on IADL information by age and diabetic status, Mexico, MHAS

Impairment	**Age**							
and sample size	**50**	(s.d)	**60**	(s.d)	**70**	(s.d)	**80**	(s.d)
Self-reports (N=12,065)								
Non-diabetics								
TLE	32.6	(0.594)	24.0	(0.568)	16.4	(0.549)	10.2	(0.518)
DFLE	29.0	(0.485)	20.5	(0.454)	13.1	(0.429)	7.2	(0.384)
DLE	3.6	(0.239)	3.5	(0.248)	3.4	(0.264)	3.0	(0.288)
Diabetics								
TLE	24.8	(0.901)	17.2	(0.757)	10.9	(0.636)	6.4	(0.536)
DFLE	21.3	(0.731)	13.8	(0.589)	7.8	(0.455)	3.7	(0.321)
DLE	3.5	(0.355)	3.4	(0.350)	3.1	(0.353)	2.7	(0.371)
Predicted values based on CRELES (N=10,043)								
Non-diabetics								
TLE	31.9	(0.652)	23.3	(0.615)	15.8	(0.583)	9.8	(0.538)
DFLE	28.5	(0.539)	20.0	(0.498)	12.7	(0.460)	6.9	(0.403)
DLE	3.4	(0.249)	3.3	(0.256)	3.2	(0.270)	2.8	(0.292)
Diabetics								
TLE	28.3	(0.794)	20.2	(0.707)	13.3	(0.627)	8.0	(0.542)
DFLE	24.9	(0.654)	16.9	(0.564)	10.1	(0.476)	5.2	(0.374)
DLE	3.4	(0.303)	3.3	(0.306)	3.1	(0.315)	2.7	(0.332)

Table 17: Total life expectancy, disability-free life expectancy and disabled healthy expectancy based on Nagi limitations by age and diabetic status, Mexico, MHAS

Impairment and sample size	**Age**							
	50	(s.d)	**60**	(s.d)	**70**	(s.d)	**80**	(s.d)
Self-reports (N=12,056)								
Non-diabetics								
TLE	33.0	(0.604)	24.4	(0.583)	16.7	(0.564)	10.5	(0.530)
DFLE	19.1	(0.357)	12.4	(0.336)	7.3	(0.308)	3.8	(0.252)
DLE	14.0	(0.429)	11.9	(0.424)	9.4	(0.420)	6.7	(0.408)
Diabetics								
TLE	24.9	(0.923)	17.4	(0.769)	11.2	(0.640)	6.6	(0.518)
DFLE	11.6	(0.531)	6.8	(0.379)	3.5	(0.255)	1.6	(0.159)
DLE	13.3	(0.710)	10.6	(0.601)	7.7	(0.517)	5.0	(0.442)
Predicted values based on CRELES (N=10,043)								
Non-diabetics								
TLE	32.4	(0.660)	23.8	(0.623)	16.2	(0.591)	10.1	(0.548)
DFLE	18.7	(0.399)	12.1	(0.363)	7.1	(0.323)	3.6	(0.258)
DLE	13.7	(0.464)	11.7	(0.448)	9.2	(0.436)	6.4	(0.421)
Diabetics								
TLE	28.7	(0.819)	20.7	(0.726)	13.7	(0.649)	8.3	(0.555)
DFLE	15.6	(0.485)	9.8	(0.390)	5.4	(0.307)	2.7	(0.220)
DLE	13.1	(0.579)	10.9	(0.529)	8.3	(0.489)	5.6	(0.440)

Discussion

Most studies in Latin America and the Caribbean rely on self-reported measures as a way to assess the prevalence and incidence of chronic conditions. Self-reported measures are less expensive to be collected, but their validity is usually questioned.

The review of the literature shows that there are very few studies that try to identify which variables are associated with diabetes misreport (Kehoe et al. 1994, Kriegsman et al. 1996), even less using data from less developed countries (Wu, Li and Ke 2000, Goldman et al. 2003). The present paper contributes to this literature and it analyzes the demographic and social characteristics associated with diabetes misreport in two settings in Latin America and the

Caribbean. Therefore, the estimated parameters are of great use for other countries lacking data on clinical diagnosis, but with available data on self-reports. Moreover, the potential usefulness of the selected models is substantial because most data sources have the necessary demographic and socioeconomic data to use them as predictive models.

Another advantage of this work is that contrary to most of the previous studies, this work uses of very large samples to analyze the accuracy on diabetes self-reports. Moreover, previous studies were usually limited to particular geographic regions. Some of these studies have used medical records to validate their self-reported measures, but the availability of medical records is usually limited to a selective population. Others have used diagnostic examinations, medical records or physician's records as 'gold standard', but they are also usually limited to a small population. This study validates self-reports of diabetes data from nationally representative data of older adults from Mexico and Costa Rica that were also screened for diabetes.

Another important advantage of this study is that it is based on nationally representative data. Therefore, the estimates of sensitivity and specificity are unbiased estimators. Also, the large sample size allows the examination of how demographic and social characteristics of the individuals, such as age, gender, education and disease related symptoms, are associated with the accuracy of their report.

Another finding from this study comes from the comparison between self-reports and clinically diagnosed diabetes. The results clearly show that the definition of 'clinically diagnosed diabetes' is crucial. For instance, ENSA results indicate that diabetes self-report is reasonably accurate and yields satisfactory prevalence estimates in Mexico. Several studies, particularly from developed countries, are in agreement with this finding that diabetes self-reports is quite accurate. However, there is a good reason to believe that this result from ENSA is, at least in

part, due to the diagnostic criteria adopted by ENSA. In ENSA, individuals who meet any of the following criteria were also classified as diabetics ('true diabetes'): a) fasting plasma glucose ≥ 126mg/dL or b) casual (random) capillary glucose concentration ≥ 200 mg/dL. Most of the individuals (95.4%) were not fasting at the time of the blood collection; therefore, the second criterion was used more often. Moreover, blood specimens were collected through capillary puncture. In CRELES conducted in Costa Rica, the blood samples draw from a vein (venipucture) rather than capillary. Also, the large majority (97%) reported that they were fasting for more than eight hours. In this survey, those with fasting serum glucose levels (FSG) greater or equal than 126 mg/dL were classified as diabetics ('true diabetes'). In both surveys, individuals who self-reported being previously diagnosed with diabetes who were under treatment with oral agents and/or insulin were also classified as 'true diabetics'. The results from CRELES, differently from the ones from ENSA, show that diabetes self-reports are moderately accurate, but several individuals are undiagnosed. There is a good reason to believe that this difference between ENSA and CRELES is not only a difference across settings (Mexico and Costa Rica). The adoption of a lower cutoff point at CRELES has an important drawback. In truth, interviewers could not verify that respondents were really fasting when the blood sample is being drawn. Therefore, it is possible that CRELES captures more 'undiagnosed' cases if, for instance, individuals who reported fasting were not and their measured glucose levels were higher than 126 mg/dL just because they had eaten recently. If this conjecture is true, 'true' rates in CRELES may be overestimating the actual prevalence. At the same time, it is possible that ENSA gives a 'real' rate that is still underestimated.

Results presented in this paper based on data from ENSA and CRELES suggest that self-reported diabetes is the stronger predictor of 'true' diabetic status. Other diabetes related

symptoms are also very strong predictors of actual diabetic status. Demographic variables are also important predictors of the actual diabetic status. Education was not statistically significant in Mexico or in Costa Rica, but the decision to incorporate it to the models comes from the fact that literacy is a reasonable socioeconomic indicator that is available in most data sources in Latin American and the Caribbean that is being used to take into account some of the social differences across countries.

The selected models based on self-reported measures and other important predictive demographic and social characteristics were used to predict new estimates of the prevalence of diabetes in selected Latin American and Caribbean countries. These new estimates provide the necessary additional boundaries for the self-reported measures. The results indicate that diabetes prevalence estimated by self-reports is underestimated for most part. In Buenos Aires, the self-reported diabetes prevalence is 12.4%, while the predicted prevalence based on ENSA is 12% higher (13.9%). However, the new prevalence estimates based on CRELES are more than two times higher (30.2%) than the self-reported. A previous study from Hernández et al. (1987) reported that undiagnosed cases reached about half of the diabetes cases in Argentina, which means that prevalence rates would double with undiagnosed cases would be taken into account.

In São Paulo, the predicted prevalence rate based on ENSA is 8% higher than the self-reported rate, while new estimates based on CRELES indicate that 34% of all individuals aged 60 and over have diabetes (88% increase compared with the self-reported rate). Previous studies from Malerbi and Franco (1992) and Sakata et al. (2002) based on different samples and regions in Brazil find that undiagnosed diabetes accounted for about 40% of the total prevalence. This estimate is closer to the one provided by the CRELES estimation in which 'undiagnosed' cases represents about 47% of the total prevalence.

Santiago is the only city in which self-reported prevalence rate is outside the 95% confidence interval of the predicted rates based on ENSA. In this city, predicted prevalence rates based on ENSA are almost 20% higher than self-reported ones. New prevalence estimates based on CRELES raise the prevalence rate from 13.3% (self-reported) to 28.7%. Previous analysis in the Seventh region of Chile (Baechler et al. 2002) has shown that undiagnosed cases reached 45% of the population aged 20 and older. However, the prevalence of diabetics aged 20-44 who had not being previously diagnosed reached 67%, and decreased as people aged reaching 37.5% in the 65 and over age group (Baechler et al. 2002).

Predicted prevalence in Havana based on ENSA was about 10% higher than self-reported prevalence rate, while predicted prevalence rates based on CRELES are 115% higher. In Havana, self-reported prevalence rate reaches 15% of the population aged 60 and over, but it could reach 32% depending on the adopted diagnostic criterion. In Montevideo prevalence rates based on ENSA were about 13% higher than self-reported measures, but if estimates are based on CRELES then predicted rates would be 126% higher, and they would indicate that near a third of the elderly population has diabetes. In Mexico City, predicted rates based on ENSA were 2.4% higher than the ones estimated using self-reports, but in Mexico underreport seems to be considerably higher- about 15% higher than self-reported ones for those aged 50 and over and 12.7% higher among those aged 60 and over. The lower percentage of undiagnosed cases in Mexico City versus Mexico is consistent with the fact that individuals have better access to health care in urban areas. This finding is also found when CRELES is used to estimate new prevalence rates in Mexico and Mexico City. Also the finding that undiagnosed rates decrease with age was also reported by a previous study from Aguilar-Salinas and colleagues (2002). They found that in urban areas in Mexico, 42% of diabetics aged 20-39 years were unaware of

their condition in the early 1990s, while 74% of individuals diagnosed with diabetes ages 40 to 69 were aware of their diabetic status (Aguilar-Salinas et al. 2002).

Diabetes-free life expectancies corrected by misreport based on ENSA do not diverge significantly from those obtained with self-reported prevalence rates. However, when CRELES is used to predict the prevalence rates of SABE and MHAS, diabetes-free life expectancies are considerably higher than the ones obtained with self-reported prevalence rates. The smaller increase is found among males in Mexico City, 2.8 years, while the larger increase is found among females in Buenos Aires, 4 years. Results based on predicted estimates confirm that women live longer lives, but with a larger number of years with diabetes. Results based on CRELES show that males in Mexico City are expected to live 7.4 years with diabetes and 13 without after they reach age 60. Their female counterparts are expected to live 7.7 years with diabetes and 14.5 without based on predicted rates using CRELES. In Bridgetown, diabetes life expectancy reaches almost 9 years among females and is about 6 years among males. The largest differential in diabetes life expectancy between males and females is found in Havana. In this urban area, men aged 60 are expected to live, on average 5 years with diabetes and their female counterparts 7.7 years. The smallest difference is found in Mexico City – diabetes life expectancy only differs in 0.3 years.

This paper also provides new estimates of total and disability-free life expectancies of Mexicans (MHAS) based on new estimates that have been corrected by misreport using CRELES. The results indicate that total life expectancy and disability-free life expectancy of diabetics is higher when diabetes information is corrected for misreport. In other words, estimates of total life expectancy and disability-free life expectancy of diabetics based on self-reports may be biased downwards. However, the estimates of disabled life expectancy of

diabetics are quite similar disregarding the type of method used to classify individuals. At the same time, estimates of total life expectancy and disability-free life expectancy of non-diabetics are slightly lower when diabetes status is corrected by misreport. Therefore, there is some evidence that traditional methods of estimating total life expectancy and disability-free life expectancy of diabetics based on self-reports may be biased downwards. As discussed in this paper, individuals that are 'newly diagnosed' are relatively better off than those who self-report being diabetic – fact that it is consistent with the fact that undiagnosed individuals are usually at earlier stages of the disease progression. In any case, it is important to understand that diabetes reduces total life expectancy and the bulk of this reduction comes in the form of reductions in the number of years expected to be lived without disability. The new estimates shows that total life expectancy of diabetics at age 50 is about 4 years lower than total life expectancy of non-diabetics. Moreover, the bulk of this difference is due to reduction in years to be lived without disability. In sum, this paper confirms a significant reduction in total and disability-free life expectancy for diabetics. These reductions in length of life and quality of life make clear that this chronic condition imposes considerable economic, social and individual costs. Given the estimated increase in the prevalence of diabetes in Latin America and the Caribbean, the associated burden is expected to increase in the next decades unless preventive measures are taken. Recent studies have indicated that changes in lifestyle, particularly on diet and exercise, and some medications can delay the onset of diabetes. Therefore, these societies should promote campaigns that emphasize that healthy eating and exercising can be translated in longer and more active lives.

Finally, this paper proposes an innovative approach to estimate diabetes prevalence rates, which uses predicted prevalence rates based on data from two nationally representative studies

conducted in Mexico and Costa Rica. This approach is more comprehensive than simply adjusting underreporting using previous estimates of undiagnosed cases. The main advantage of using such a procedure comes from the fact that demographic, social and medical characteristics can be used to predict new prevalence rates. Moreover, it adjusts for misreport and not only for underreporting. However, it assumes that the associations, between demographic, social and medical conditions and diabetes prevalence, which prevail in Mexico or Costa Rica, are the same in the other settings. However, this can be a very strong assumption. Therefore, in order to have a range of possible prevalence values, this paper made use of two data sources (ENSA and CRELES) that use different diagnostic criteria. Also, the use of this method is less dependent on the assumptions of likeness than the ones used in other studies that predict prevalence rates for several Latin American and Caribbean countries (King, Aubert and Herman 1998, Wild et al. 2004) because this method uses relevant data on self-reported measures and other demographic and social characteristics of the analyzed countries to predict the actual diabetes prevalence. This represents a substantial advance in the use of self-reported data that it is usually available. However, additional efforts are necessary to validate these models with other surveys that contain both self-reported and clinical data with standard procedures to categorize the diabetes status.

References

Aguilar-Salinas, C. A., Velazquez Monroy, O., Gomez-Perez, F. J., Gonzalez Chavez, A., Esqueda, A. L., Molina Cuevas, V., Rull-Rodrigo, J. A., & Tapia Conyer, R. (2003). Characteristics of patients with type 2 diabetes in Mexico: Results from a large population-based nationwide survey. Diabetes Care, 26(7), 2021-6.

Baechler, R., Mujica, V., Aqueveque, X., Ramos, L., & Soto, A. (2002). Prevalencia de diabetes mellitus en la VII Región de Chile [Prevalence of diabetes mellitus in the Seventh Region of Chile]. Rev Med Chil, 130(11), 1257-64.

Beckett, M., Weinstein, M., Goldman, N., & Yu-Hsuan, L. (2000). Do health interview surveys yield reliable data on chronic illness among older respondents? Am J Epidemiol, 151(3), 315-23.

Bergmann, M. M., Jacobs, E. J., Hoffmann, K., & Boeing, H. (2004). Agreement of self-reported medical history: comparison of an in-person interview with a self-administered questionnaire. Eur J Epidemiol, 19(5), 411-6.

Brenes, G. (2006). Differences between Self-Reported Diabetes and Clinical Test Results: The Usefulness of Bio-markers in Research in Developing Countrie . Paper presented at the 2006 Population Association of America Annual Meeting.

Brito, I. C., Lopes, A. A., & Araujo, L. M. B. (2001). Associação da Cor da Pele com Diabetes Mellitus Tipo 2 e Intolerância a Glicose em Mulheres Obesas de Salvador, Bahia. Arq Bras Endocrinol Metab, 45(5), 475-80.

Bush, T. L., Miller, S. R., Golden, A. L., & Hale, W. E. (1989). Self-report and medical record report agreement of selected medical conditions in the elderly. Am J Public Health, 79(11), 1554-6.

Goldenberg P , Franco LJ, Pagliaro H, Silva Rd, Santos CA. (1996). Diabetes mellitus auto-referido no Município de São Paulo: prevalëncia e desigualdade [Self-reported diabetes mellitus in the city of São Paulo: prevalence and inequality]. Cad Saúde Pública 12(1):37-45

Goldman, N., Lin, I., Weinstein, M., & Lin, Y. (2002). Evaluating the Quality of Self-Reports of Hypertension and Diabetes . Working Paper No. 2002-3.

Haapanen, N., Miilunpalo, S., Pasanen, M., Oja, P., & Vuori, I. (1997). Agreement between questionnaire data and medical records of chronic diseases in middle-aged and elderly Finnish men and women. Am J Epidemiol, 145(8), 762-9.

Heliovaara, M., Aromaa, A., Klaukka, T., Knekt, P., Joukamaa, M., & Impivaara, O. (1993). Reliability and validity of interview data on chronic diseases. The Mini-Finland Health Survey. J Clin Epidemiol, 46(2), 181-91.

Hernández, R. E., Cardonnet, L. J., Libman, C., & Gagliardino, J. J. (1987). Prevalence of diabetes and obesity in an urban population of Argentina. Diabetes Res Clin Pract, 3(5), 277-83.

Kehoe, R., Wu, S. Y., Leske, M. C., & Chylack, L. T. J. (1994). Comparing self-reported and physician-reported medical history. Am J Epidemiol, 139(8), 813-8.

King H, Aubert RE, Herman WH. 1998. Global burden of diabetes, 1995-2025: prevalence, numerical estimates, and projections. Diabetes Care 21(9):1414-31.

Kriegsman, D. M., Penninx, B. W., van Eijk, J. T., Boeke, A. J., & Deeg, D. J. (1996). Self-reports and general practitioner information on the presence of chronic diseases in community dwelling elderly. A study on the accuracy of patients' self-reports and on determinants of inaccuracy. J Clin Epidemiol, 49(12), 1407-17.

Krueger, D. E. (1957). Measurement of prevalence of chronic disease by household interviews and clinical evaluations. Am J Public Health, 47(8), 953-60.

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics, 33(2), 363-74.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159-74.

Malerbi, D. A., & Franco, L. J. (1992). Multicenter study of the prevalence of diabetes mellitus and impaired glucose tolerance in the urban Brazilian population aged 30-69 yr. The Brazilian Cooperative Group on the Study of Diabetes Prevalence. Diabetes Care, 15(11), 1509-16.

Martin, L. M., Leff, M., Calonge, N., Garrett, C., & Nelson, D. E. (2000). Validation of self-reported chronic conditions and health services in a managed care population. Am J Prev Med, 18(3), 215-8.

Midthjell, K., Holmen, J., Bjorndal, A., & Lund-Larsen, G. (1992). Is questionnaire information valid in the study of a chronic disease such as diabetes? The Nord-Trondelag diabetes study. J Epidemiol Community Health, 46(5), 537-42.

Okura, Y., Urban, L. H., Mahoney, D. W., Jacobsen, S. J., & Rodeheffer, R. J. (2004). Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. J Clin Epidemiol, 57(10), 1096-103.

Palloni, A., Soldo, B.J., & Wong, R. (2002). Health Status and Functional Limitations in a National Sample of Elderly Mexicans, presented at the Population Association of America Conference, Atlanta.

Palloni, A., & Pelaez, M. (2002). Survey on Health and Well-Being of Elders: preliminary report. PAHO.

Peláez M, Palloni A, Alba C, Alfonso JC, Ham-Chande R, Hennis R, Lebrăo ML, León-Díaz E, Pantelides A, Pratts O. (2003). Encuesta Salud, Bienestar y Envejecimiento, 2000. Organización Panamericana de la Salud (OPS/OMS)

Rio-Navarro, B. E. d., Velazquez-Monroy, O., Sanchez-Castillo, C. P., Lara-Esqueda, A., Berber, A. , Fanghanel, G., Violante, R., Tapia-Conyer, R., & James, W. P. (2004). The high prevalence of overweight and obesity in Mexican children. Obes Res, 12(2), 215-23.

Sakata, K., Bigolin, S., Bryk Junior, A., Komatsu, M. C. G., Sakata, L., Vanzo, L. R. C., & Ruthes, H. I. (2002). Estudo dos conhecimentos de pacientes com hipertensão, diabetes ou glaucoma sobre suas doenças. Arq Bras Oftalmol, 65, 467-469.

Sanchez-Castillo, C. P., Velazquez-Monroy, O., Berber, A., Lara-Esqueda, A., Tapia-Conyer, R., &

James, W. P. (2003). Anthropometric cutoff points for predicting chronic diseases in the Mexican National Health Survey 2000. Obes Res, 11(3), 442-51.

Sherbourne, C. D., & Meredith, L. S. (1992). Quality of self-report data: a comparison of older and younger chronically ill patients. J Gerontol, 47(4), S204-11.

Simpson, C. F., Boyd, C. M., Carlson, M. C., Griswold, M. E., Guralnik, J. M., & Fried, L. P. (2004). Agreement between self-report of disease diagnoses and medical record validation in disabled older women: factors that modify agreement. J Am Geriatr Soc, 52(1), 123-7.

Tretli, S., Lund-Larsen, P. G., & Foss, O. P. (1982). Reliability of questionnaire information on cardiovascular disease and diabetes: cardiovascular disease study in Finnmark county. J Epidemiol Community Health, 36(4), 269-73.

Wild, S., Roglic, G., Green, A., Sicree, R., & King, H. (2004). Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care, 27(5), 1047-53.

Wu, S. C., Li, C. Y., & Ke, D. S. (2000). The agreement between self-reporting and clinical diagnosis for selected medical conditions among the elderly in Taiwan. Public Health, 114(2), 137-42.