Differences between Self-Reported Diabetes and Clinical Test Results: The Usefulness of Bio-markers in Research in Developing Countries

Gilbert Brenes, *University of Wisconsin at Madison and Central American Center for Population, University of Costa Rica*

Chronic diseases are the main causes of morbidity and mortality not only in industrialized countries, but also in some developing countries. Unlike infectious diseases, they have longer durations between onset and diagnosis. This means that a certain proportion of sick people do not know what their actual health status is. The collection of biomarkers in population research that also gathers self-reported information allows studying these discrepancies. CRELES, a new survey focused on Costa Rican elderly, is one of these cases. These data were used for two purposes: to estimate diabetes prevalence in a survey with no clinical data (MHAS) using a model derived from CRELES as a standard; and to determine risk factors of discrepancies between clinical results and self-reports. Previous hospitalizations, age, hypertension, and enabling resources determine having undiagnosed diabetes, while an infarction, dizziness, absence of swollen feet, and self-perceived financial status differentiate controlled from uncontrolled diabetes.

Some developing countries are advancing with a fast pace through the epidemiological transition. Communicable diseases and malnutrition are the causes of only a small proportion of total deaths in these countries, while chronic conditions are increasingly taking a toll on mortality, morbidity, and disability. Latin America and East Asia are the regions that best exemplify this pattern, as well as a parallel process of rapid population aging (Pelaez, Palloni and Ferrer, 2000; Vallin, 1988). The importance of these phenomena has favored research projects focused on health and on aging in these countries.

In Latin America there is a growing concern about diabetes mellitus prevalence among the elderly because the survivors of these cohorts not only adopted nutritional and lifestyle behaviors of the industrialized world, but they also experienced disadvantaged health status during their younger years, which might be a risk factor for the development of diabetes (Andrade, 2005; Palloni and McEniry, 2004; Pelaez, Palloni and Ferrer, 2000; Popkin and Gordon-Larsen, 2004). Some of the surveys that have been used to study the relationship between aging and diabetes in the region have relied on self-reported answers to questions about diagnosis of the disease by a physician. There is a new project in Costa Rica, called CRELES (Costa Rica: Estudio sobre Longevidad y Envejecimiento Saludable, which means "Costa Rican Study on Longevity and Healthy Aging") that is also collecting blood and urine samples in order to produce biomarker information. CRELES has also the advantage that the questionnaire was designed in such a way to be comparable to the instruments of other aging studies in the sub-continent, like SABE -Salud Bienestar y Envejecimiento en America Latina- (Palloni and McEniry, 2004; Pelaez, Palloni and Ferrer, 2000), the Mexican Health and Aging Study MHAS (Soldo, Wong and Palloni, 2002), and the "Puerto Rican Elderly: Health Conditions" Project PREHCO (Palloni, McEniry, Guend, Davila et al, 2004).

The aim of this paper is to study how the estimate of diabetes prevalence can change when computations made with self-reported data are compared to figures produced with a model that predicts clinically-assessed diabetes prevalence. Data from CRELES are going to be used to establish such equation, which will be applied to the MHAS dataset as an example. These prediction models will also allow exploring factors that are associated with undiagnosed and controlled diabetes, in a country such as Costa Rica that has been praised for its advancements in life expectancy and health care services provision (Caldwell, 1986; Mesa-Lagos, 1992).

Literature review

Differences between self-reported diabetes and clinical exam results have been studied from two approaches: a) as a problem of error in self-reported information, and b) as a problem of undiagnosed and controlled diabetes. In the first approach, there is a series of articles warning about the accuracy of self-reported health and its utilization in causal or relational analysis. Angel and Gronfein (1988) explain that self-reported health measures are culturally and socially determined, and therefore differences between ethnic groups in subjective health status can be explained by different visions of illness and health. Hardy and Pavalko (1986) show that persons in different occupations tend to report or neglect certain conditions as limitations; they argue that a question's evaluative context affects the answer accuracy. Specifically on diabetes, several studies found a high agreement between patient's report and patient's physician records in the Netherlands, Finland and the U.S.A. (Haapanen, Miilunpalo, Pasanen, Oja and Vuori, 1997; Kriegsman, Penninx, van Eijk, Boeke and Deeg, 1996; Martin, Leff, Calonge, Garrett and Nelson, 2000). A comparison between biomarkers and self-reports among the elderly in Taiwan also found a strong agreement in diabetes, but not in hypertension (Goldman, Lin, Weinstein, and Lin, 2003). Among the factors directly related to agreement, the reviewed articles mention: age, education, cognitive ability, and recent visit to a physician.

The other approach addresses the problems of inequity of the health care system that might degenerate into cases of undiagnosed diabetes and treatment non-compliance; the approach also refers to the risks in terms of tertiary health care use and mortality that can derive from the situation of patients ignoring their real condition or from non-adherence. Prevalence of "hidden" diabetes varies across populations: around 2% in the U.S.A., but as much as almost 4% among rural Hispanic; overall, one of every three American adults with glucose intolerance do not know that they have the disease (Gregg, Cadwell, Ching, Cowie et al, 2004; Koopman, Mainous and Geesey, 2006); in Australia, this proportion mounts to 50% (Dunstan, Zimmet, Welborn, de-Courten et al, 2002); and in Denmark, to 66% (Glumer, Jorgensen and Borch-Johnsen, 2003). When compared to non-diabetics, factors directly related to "hidden" diabetes are typically risk factors of the disease: age, race (African-Americans have higher prevalence), hypertension, obesity (measured by both BMI and Waist-to-hip circumference), family history of diabetes, physical activity (Franse, Di Bari, Shorr, Resnick et al, 2001; Glumer, Borch-Johnsen and Colagiuri, 2005; Mohan, Deepa, Deepa, Somannavar and Datta, 2005; Tabaei, Engelgaut and Herman, 2005). Besides typical risk factors, some articles discuss whether there are socio-economic status (SES) differences in undiagnosed diabetes. Among middle-age Germans, women with higher occupational status and with higher income are more likely to know their diabetic condition; this pattern is not observed in men (Rathmann, Haaster, Icks, Giani et al, 2005). In the National Health and Nutrition Examination Survey (NHANES), there is no evidence of an effect of education or income among adults 20 years old or older (Wilder, Majumdar, Klarenbach and Jacobs, 2005). In terms of controlled diabetes, articles refer to compliance and adherence to treatment, and psychosocial factors are more often used to explain the behavior. Residence in poverty areas in the U.S.A. is directly linked to missing physician appointments for diabetes control (Karter, Parker, Moffet, Ahmed et al, 2004). According to a psychosocial model, education, being married, positive coping styles and not having stress are associated with better glycemic control (Peyrot, McMurray and Kruger, 1999). Lutfey and Wishner (1999) prefer the concept of “adherence” over “compliance” arguing that it is necessary to view patients as more active subjects in medical treatment decisions. As a summary, while both biomedical risk factors and psychosocial variables are taken into account for studying undiagnosed diabetes, the latter are preferred to study compliance and adherence among diabetics. One of the main problems for comparing results across these articles is the criteria established to define diabetes; some use fasting and non-fasting serum glucose concentrations; others, 2-hour Oral Glucose Tolerance Test (OGTT); while others prefer glycosylated hemoglobin.

An explanatory framework: Andersen’s Behavioral Model of Health Services Use.

The main goal of this paper is to assess the underestimation of diabetes prevalence due to using self-reported information rather than clinical examination. As it will be explained below, logistic equations reduced through stepwise procedures are used to estimate disease prevalence. Also, the original regression equations can be analyzed to study the factors associated with controlled and undiagnosed diabetes. Differences between self-reports and biomarkers can be seen as a problem of difference in health care services utilization. If a person that has been diagnosed with diabetes gets negative results from a FSG concentration test, this means that the person is either complying with physician’s prescription or improving health habits (or both). If a person does not know that he has diabetes, then this might be due to poor utilization of preventive services, especially if this person is 60 years old or older.

A good theoretical framework for understanding this feature is Andersen’s Behavioral Model of Health Services Use (Aday, Begley, Lairson and Slater, 1998; Andersen, 1995). The original framework stressed on the existence of three types of factors that influence health services use: need factors, enabling resources, and pre-disposing characteristics. A summary of how these categories are understood is displayed below:

Categories	**Definition**	**Examples**
Need factors	Direct causes. Perceived or clinically determined health status that moves a patient to use the services	• Illnesses • Symptoms • Disability days • Hospitalization
Enabling resources	Financial, social, or organization resources that individuals have available for using services	• Income • Health insurance • Distance to provider • Socialized, private, or mixed health care system
Pre-disposing characteristics	Characteristics that describe the propensity to use services. They are distal factors in causality chain.	• Demographic characteristics • Social structural variables • Health beliefs • Genetic and psychological characteristics

The Behavioral Model has been widely cited for analyzing equitable access, depending on which factors are more relevant in explaining health services utilization. According to Andersen (1995), there is equitable access when demographic and need variables are the most important in accounting for service use, while there is inequitable access if enabling resources (income, health insurance), the social structure (race, ethnicity, living in poor areas), and health beliefs have an effect on differential utilization. Health beliefs are included in this group because it is argued that health beliefs can be modified through intervention. And from a policy perspective, Andersen emphasizes the concept of factor mutability: demographic and social structure variables have low mutability, while enabling resources and health beliefs are considered to be moderately or highly mutable (through policy). The Socio-Behavioral Model should be useful to understand differences between self-reported and clinically assessed diabetes, in a cultural and socio-political context that has been considered relatively equitable because of the existence of a socialized health care system.

Data

The main information is derived from the first waves of two datasets: MHAS and CRELES. The latter will be used for estimating differences between self-reports and clinical data, while with the former we will estimate underreporting of diabetes prevalence, under the assumption that the factors that determine controlled and undiagnosed diabetes have a similar behavior as they have in Costa Rica.

CRELES's sample infers to the population born in Costa Rica in 1945 or earlier and that was alive during the period 2004-2006. It is derived from an original random sample of 8,000 individuals ages 55 and over interviewed in the 2000 Census; this sample was stratified by 5-year age groups, and within each stratum, individuals were selected randomly using a systematic procedure. Their census information was linked to the Vital Registration System database in order to study their mortality patterns. A subsample of 3,300 people, which is expected to produce a total subsample of 3,000 survivors, was selected using a two stage cluster sampling design, where the clusters were the official Health Areas in which the Ministry of Health and the Social Security Institution (CCSS, for Caja Costarricense del Seguro Social) divide the country[1]. CRELES interviews only the selected person, and not the spouse. There are only two fieldwork teams, composed each by a driver, a laboratory technician, and two interviewers; this structure means that on average the project expects only 32 interviews per week and, therefore, the first wave will be concluded after two years of being started; hence, the fieldwork started in November 2004, and the first wave will be finished around October 2006. Because of this kind of schedule, we use only the first 1832 individuals interviewed until December 31st, 2005. The data is gathered by Computer Assisted Interviews (CAIs), using PDA's (Personal Digital Administrators) or palms. This technology permits to control for data inconsistencies in the field and to generate information continuously (Rosero-Bixby, Hidalgo and Antich, 2005). The usual procedure is to ask for personal consent and conduct the interview one day and, in the next morning, the laboratory technician visits the respondent for drawing the blood

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

sample by venipuncture, for collecting urine samples and -if it has not been done- for carrying out the anthropometric measurements.

CRELES measures diabetes in two ways: by asking for self-reported diabetes diagnosis and by clinical analyses of fasting blood samples. The question in the instrument is the following:

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

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 (WHO, 1999), while the former has been used to control diabetes treatment. Glyosylated hemoglobin has been proposed as an alternative to the Oral Glucose Tolerance Test (OGTT, the so-called "gold standard" for diabetes diagnosis which is also recommended by the WHO) because patients do not need to fast, to drink the glucose solution, or to wait for 2 hours before blood samples are drawn, and because it is considered to be a better biomarker for daylong blood glucose concentrations (Peters, Davidson, Schriger and Hasselblad, 1996). However, in 2003, an Expert Committee on the Diagnosis and Classification of Diabetes Mellitus suggested not to use this clinical exam for diagnosis because the lack of unified standards across laboratories (ECDCDM, 2004). In this paper, we use the criteria based on FSG because it is a standard test performed in Costa Rica for diagnosis; therefore, a person that, according to a clinical exam, has FSG greater or equal than 126 mg/dL would probably be diagnosed with "high blood sugar levels", as the question inquires. The drawbacks of using this biomarker are:

- Interviewers can not verify that respondents were really fasting when the blood sample was drawn;
- The test is less accurate for elderly populations;
- We are not considering Impaired Fasting Glycaemia (IFG) in the analysis, which is defined as having FSG concentrations greater than 110 mg/dL but lower than 126 mg/dL (WHO, 1999). The limitation consists on the situation of patients

whose physicians might have told them that they had high blood sugar levels, as a simpler way to tell them that they had IFG.

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 first laboratory used the glucose oxidase method, while the other used glucose oxidase and oxygen consumption methods. The survey was applied and blood samples were drawn after an informed consent form was read and signed by the interviewees. The informed consent was approved by the University of Costa Rica's Institutional Review Board (IRB).

The models derived in this paper have the purpose of assessing over- and underestimation of diabetes prevalence in aging studies that do not perform clinical tests. As an example, these models will be applied to MHAS's first wave dataset. CRELES Principal Investigators decided to design a questionnaire that would be comparable to the MHAS questionnaire and to the one used in the SABE project. The target population of the Mexican Health and Aging Study (MHAS) comprises Mexicans born before 1951 and their spouses and partners, and it is representative to the non-institutionalized population aged 50 and over in 2001. The sample was drawn using a multi-stage probability procedure from the sample of the National Employment Survey (Encuesta Nacional de Empleo, ENASEM). One adult is selected randomly from each household, but his or her spouse (married or in consensual union) is also interviewed. The data collection for the first wave was finished in 2001, while the second wave was ended in 2003. The total number of respondents in the first wave is 15,230 persons for an overall response rate of 92% (Palloni, Soldo and Wong, 2002; Wong and Espinoza, 2004). The anthropometric module was applied to only a 20% subsample. Since one of the most important variables for estimating diabetes prevalence is the Body Mass Index (BMI), which is better computed with direct anthropometric measures rather than with self-reports, the procedures will be applied to only this 20% sub-sample.

Methods

The article's goal is to estimate proportions of people with diabetes mellitus and without it, rather than estimating FSG levels; therefore, the estimating techniques are variations over binary logistic regression, used as alternative for classical discriminant analysis because the assumptions are less stringent (Hosmer and Lemeshow, 2000). The first strategy -which we call "two-equation procedure"- is similar to the analyses performed by Goldman, Lin, Weinstein, and Lin (2003) to study diabetes and hypertension accuracy of self-reported data in Taiwan. They fit separate models for the likelihood of positive clinical results among those that had never been diagnosed with diabetes, and for the likelihood of negative clinical results among those that had been diagnosed before. As these authors explain, the rationale behind this decision is that factors associated to each of these states might be different.

We first estimate a full model with a whole set of possible explanatory variables; then, stepwise selection procedures are carried out in order to select a reduced set of covariates; reduced equations were preferred over more complex models because the contribution of variables with non-significant coefficients to classification is very low, and over-identification may lead to distorted estimates; besides, the equations will be used to estimate prevalence in other surveys, thus parsimony makes this application easier. Backward stepwise selection was used, with a probability of exclusion from the model of 0.15; this means that all the variables with p-values lower than 0.15 were retained. The rationale of choosing such a very lax criterion is that the sample sizes are relatively small, particularly the one for the equation of controlled-uncontrolled diabetes.

An alternative and simpler procedure –which we call "one-equation" procedure"- is to estimate a logistic regression where the dependent variable is diabetes status according to clinical data; the dichotomous variable adopts the value of 1 if the FSG result is larger or equal to 126 mg/dL, and 0 otherwise. Again, a subset of predicting variables is selected through stepwise algorithms, keeping the probability of exclusion equal to 0.15. The advantage of this procedure when compared to the previous is that only one equation is required. The main disadvantage is that the subset of explanatory variables that best predict controlled/uncontrolled diabetes might not be the same as the one that might best predict "hidden" diabetes.

The choice of covariates was driven mainly by the existence of the same variables in MHAS and other Latin American aging surveys where the prediction techniques can be applied. Table 1 has a list of the chosen covariates classified according to the three types of factors in Andersen's Behavioral Model, as explained before[2]. We also include four additional variables that are particular to CRELES, but that are of interest for a more substantive interpretation: living in the Metropolitan Area, living in urban areas, family history of diabetes (dichotomous variable), and having been visited by a primary health technician from an EBAIS (Equipos Básicos de Atención Integral en Salud: Basic Teams for Integral Health Services); these covariates are included only in the full model, but not in the equations for prediction. The use of insulin shots is included only in the model for controlled diabetes in the "two-equation" procedure (because the variable is only relevant to those that currently have a diabetes diagnosis).

After the logistic equations are defined with each of the two procedures, the estimated prevalence of diabetes is computed in two ways:

- The estimated probability is used to classify individuals in two categories: diabetes and non-diabetes, as when logistic regressions are employed like alternatives to classical discriminant analysis. The probability at which specificity and sensitivity are equal is used as a cutoff point (Hosmer and Lemeshow, 2000). This means that, if specificity and sensitivity intersect at P(X=1)=0.08 in the undiagnosed equation, all those respondents that have a probability greater than 0.08 will be classified as having undiagnosed diabetes.
- The estimated probability is used to impute zeroes (0s) and ones (1s) from a Bernoulli distribution where the estimated probability represents the distribution's "p" parameter. The imputation was performed 100 times to construct a simulated confidence interval.

The second approach incorporates uncertainty to the estimates, while the first does not.

An additional methodological comment concerns the nature of certain biomarkers used as dependent variables. FSG concentration is assessed in laboratories, and these clinical results may be exposed to measurement error. Although there are statistical-

---

[2] As discussed before, the Behavioral Model might be useful as a classification of covariates for explaining differences between controlled, uncontrolled and undiagnosed diabetes, but not as a classification for diabetes risk factors.

econometric techniques to consider the effect of categorical dependent variables measured with error (Buonaccorsi, 1996), we are not using them in this article because they are usually more closely related to interpreting models rather than in classifying individuals in categories.

Results: Prevalence estimation

As shown in Table 1, the number of covariates included initially was considerably large. Table 2 contains descriptive statistics for the whole set of covariates, for the total dataset as well as classified in the 4 main groups that can be drawn in CRELES: no diabetes (reported=0, clinical=0), undiagnosed diabetes (reported=0, clinical=1), uncontrolled diabetes (reported=1, clinical=1), and controlled (reported=1, clinical=0). Among features worth to notice in the table, are the relatively high proportion of people that have experienced hypertension, swollen feet, dizziness, and fatigue; less than a half lives in the Great Metropolitan Area (where the capital is), although more than half lives in urban areas; 63% are retired and earning a pension, and 45% have been visited during the last 12 months by a primary health technician. Among the differences across groups, consistently people with undiagnosed diabetes have experienced the list of symptoms (dizziness, fatigue, etc) and the list of risks (obesity, overweight, hypertension, etc.) more often than those without diabetes (an expected pattern), as well as those with controlled diabetes as compared to those that do not have it controlled.

For Costa Rican elderly (ages 60 and above), diabetes mellitus prevalence according to self-reported information is 21.2%. Two fifths of these people that have had a diabetes diagnosis –which represent 8.5% of the target population- have their diabetes controlled (FSG <126 mg/dL). An additional 6% of Costa Rican elderly has undiagnosed diabetes (See Figure 1). This means that roughly, one of every three seniors do not know that they have the disease; this figure is very similar to the one reported for the U.S. adult population (ages≥20), as mentioned above.

The reduced prediction equations are shown in Tables 3 and 4. Among the variables that significantly predict controlled diabetes (against uncontrolled diabetes), having had a heart attack, having swollen feet, or current smoking status are inversely related to FSG negative results, while dizziness and worse financial situation are directly related to the dependent variable. In the equation for undiagnosed diabetes, hypertension, worse financial situation, age, better education, smoking, obesity, overweight, and being retired are directly related to positive clinical results, while the coefficients for dizziness, hospitalization (during the last 12 months), and functional status (sum of ADL+IADL limitations) are associated to negative results.

As explained in the Methods section, one of the ways in which diabetes prevalence is estimated is directly classifying respondents into the two states with the logistic models. For that purpose, a cutoff point was established for each of the two equations. Figure 2 shows that the sensitivity and specificity curves intersect at a probability equal to 0.085 for the undiagnosed diabetes equation, and in 0.415 for the controlled/uncontrolled equation. The areas under the ROC curve are 0.67 for the undiagnosed equation, and 0.74 for the controlled-uncontrolled equation; these levels are moderately acceptable. For the former equation, specificity and sensitivity are higher than 70%; the negative predictive value (the probability of classifying correctly in the non-diabetes state) is excellent, but the positive predictive value is rather low; this means that, in the predictions, only around 1 of every five persons with undiagnosed diabetes will be categorized as such. This cutoff point was accepted, though, in order to provide conservative estimates of "hidden" diabetes. For the other equation, specificity and sensitivity are lower (65% and 62%) than in the previous equation; the negative predictive value is also lower (71%), but the positive predictive value is higher than 50% (55%).

The other method used to estimate diabetes prevalence is the "one-equation" procedure. Table 6 has coefficients for the logistic regression of clinically established diabetes; two equations are presented: one with the full model, the other with the reduced model through stepwise. Notice that among the variables with statistically significant coefficients, there are known risk factors and known comorbidities for diabetes, such as: hypertension, obesity, and overweight (as well as family history of diabetes). Having swollen feet increases the risk of diabetes as it will be explained later, probably because of circulation problems (although in the full model, the variable's coefficient is not significant), while feeling dizzy decreases the risk of diabetes (dizziness might be a state produced by medication). Less disability is associated with increasing diabetes; this is the only coefficient with a counterintuitive sign, but probably this variable is implicitly referring to people that receive help with their limitations in activities of daily living, and therefore, the extra help might facilitate them having their diabetes controlled. For classification, Figure 3 shows the curves for sensitivity and specificity for different cutoff points; the two curves intersect near a probability equal to 0.19. With this cutoff point, within the dataset, sensitivity nears 63% and specificity 62%; the positive predictive value is just 28%, but the negative predictive value is 88%; the proportion of the data correctly classified is almost 62.5%

Results from the different procedures and approaches to estimate diabetes among Mexican elderly are contained in table 7. According to MHAS, diabetes prevalence in people 50 years old or above in 2001 was 17%. Using the “two-equation” procedure, diabetes prevalence might be as high as 36%, which means that for each person that has been diagnosed with diabetes, there is another one that does not know his illness. Using the “one-equation” procedure, diabetes prevalence seems too high: either 42% (with random imputation) or 52% (with direct classification). Which procedure is better? As mentioned before, the “two-equation” procedure might be better because the covariates related to undiagnosed diabetes are not the same as the ones related to the controlled/uncontrolled dichotomy. In this sense, a prevalence of undiagnosed diabetes between 15% and 20% might be more feasible than the figure larger than 25%. A good source to cross-check these estimates is the information provided by Mexico’s National Health Survey ENSA-2000 (Encuesta Nacional de Salud 2000) (Olaiz, Rojas, Barquera, Shamah et al, 2003). According to the ENSA, among adults 50 years old or older, previously diagnosed diabetes prevalence is around 16% and undiagnosed diabetes prevalence is just 3%[3]. This implies that just 15% of Mexican elderly diabetics do not know that they have the disease. This last figure is low and it seems to provide possible evidence that the estimate using the Costa Rican standard is too high. However, according to ENSA researchers, their results might be underestimating undiagnosed diabetes prevalence because: they collected the blood specimens through capillary puncture instead of venipuncture (as in CRELES, and thus, results are not fully

[3] Figures were computed based on the information provided in the ENSA publication.

comparable), there was a high percentage of respondents that were not fasting (although the publication does not clarify how large was this percentage) and therefore a higher cutoff limit of 200 mg/dL was used to determine diabetes), and there was a 6% of interviewees from whom no blood was drawn (the publication does not mention the reason for this set of missing values).

Also, according to ENSA, 43% of the population ages 50 and over with previous diagnosis have their diabetes uncontrolled (Olaiz, Rojas, Barquera, Shamah et al, 2003). This figure is closer to the estimates generated with the model; they range from 48% to 67%. However, these figures are even less comparable to the ones produced with the model since, adding to the differences cited above, ENSA researchers used a cutoff point of 140 mg/dL among people with previous diabetes, in order to assess controlled diabetes. This means that if Mexicans would have used the same cutoff point as used in the Costa Rican model, the proportion of uncontrolled diabetes might have been higher than 43%; the other differences –respondents that were not fasting and capillary puncture instead of venipuncture- might be moving this figure downwards, as well.

Results: Explaining the models

The full models that give origin to the prediction equations through stepwise are also useful in describing the variables associated with undiagnosed and controlled diabetes. Table 8 contains the logistic regression coefficients for the equations with all the original variables, classified in the three groups according to Andersen's Behavioral Model. Bold coefficients refer to the variables that were finally selected by stepwise procedures. Regarding the equation for controlled diabetes, the most important symptom that predicts it is dizziness. Persons that experience dizziness have odds of having their diabetes controlled 2.6 times the odds of persons that do not have this symptom. On the other hand, persons with swollen feet are more likely to have their diabetes uncontrolled (OR=0.43). Dizziness might be an effect of medication, thus a sign of compliance, whereas swollen feet might be an effect of peripheral vascular disease that diabetics face. We expected that the coefficient for having had a heart attack would have been positive, suggesting that persons with an infarction history will take care of themselves better.

However, the sign is negative, and it might be highlighting respondents with inadequate health behaviors that increase the risk of both diabetes and cardiovascular diseases. The coefficient for cough and phlegm is also positive and significant, but the fact that it was excluded by the stepwise procedure is an indication that the relationship might be spurious and produced by a high correlation between this symptom and other covariates (probably, other symptoms or other health behaviors). Besides, cough and phlegm was included because it was in questionnaire's list of symptoms, but there is no clinical evidence that can link this symptom to the disease. Results regarding enabling factors are quite remarkable. Worsening functional impairment (assessed by the sum of ADL and IADL limitations) augments the likelihood of controlling diabetes status; this suggests that people that are increasingly disabled might be complying better with their medication because they have to take care more of their health status, or they have somebody that helps them dealing with the limitation as well as with diabetes. But the most interesting result is the sign of the coefficient for perceived financial situation, measured in a scale where 1 is poor and 5 is excellent. People with worse self-rated financial situation (poorer people) are more likely to have their diabetes controlled. The usual Andersen's model classifies socio-economic status as an enabling factor because more resources imply more utilization of health services. However, Costa Rica has a public health services system, and most seniors use public hospitals and clinics. While the availability of practically free health insurance for the elderly would imply that socio-economic status should not have an effect; the negative sign is suggesting that perceived financial situation might be a marker for worse health behaviors (more frequent among the most affluent elderly in these cohorts), and thus it is more a pre-disposing factor rather than an enabling one. Finally, none of the pre-disposing group has a significant coefficient, but after the stepwise procedure, current smoking remained in the model (although $p = 0.130$). This also suggests that smoking is another proxy for inadequate health behaviors, since current smokers were less likely of having their diabetes controlled.

The equation for undiagnosed or "hidden" diabetes compare people with this status against those that have never been diagnosed with the disease and their clinical exams were negative (no disease). The best predictors of hidden diabetes (as measured by the p-values of their coefficients or by being selected in the stepwise procedure) are within the pre-disposing factors category. All of them have positive coefficients and are considered as diabetes risk factors: current smoker, obesity and overweight, being retired, and age. Better schooling and living in the Metropolitan Area are also risk factors of diabetes in Costa Rica, since among the cohorts that are old-age at the beginning of the 21st century, the ones that lived in the capital and had more education were the ones that adopted inadequate health behaviors from the developed world more frequently. The positive effect of hypertension –which was classified as a need variable- reinforces this pattern, since hypertension is also a risk factor. Among the other need variables, hospitalization is the only one with a significant coefficient (with an α-level of 0.10). Having been hospitalized decreases the risk for hidden diabetes; this was the expected direction because when people are hospitalized, there are some routine clinical exams that are performed, such as the one of FSG concentrations. Feeling dizzy was included in the prediction equation by the stepwise procedure. Dizziness protects against hidden diabetes because if dizziness is a symptom of metabolic instability, people that do not feel dizzy will more likely be people without the disease. Again, self-rated financial situation is the only significant enabling factor of undiagnosed diabetes, and its negative effect means that economically better-off elderly have a lower risk of having hidden diabetes. In the traditional Andersen's model, this would imply that there is inequity in the access to health services because poorer people seem to be less likely of receiving medical attention (Andersen, 1999). However, recall that this equation estimate the probability of hidden diabetes among the ones that had never been diagnosed with the disease, thus in this case perceived financial situation might be indicating either a risk factor for diabetes or a condition of limited resources for searching for health services.

In order to find evidence for this proposition, it would be better to estimate another logistic equation among those that have diabetes according to any of the two variables used to define it (self-report or clinical exam). The outcome variable would be equal to 1 if the condition is ignored and 0 if the condition is known. Table 9 presents these results. Three variables have significant coefficients (α=0.10), and with the expected signs: hypertension, hospitalization, and self-rated health. Having been hospitalized and having been diagnosed with high blood pressure lowers the odds of ignoring the condition; this result is expected since people with a medical diagnosis of hypertension will be warned about the consequences of comorbidities, such as diabetes. The rationale behind the effect of hospitalization was explained before; people in hospitals get routine exams, such as the one for FSG concentrations. Finally, people that feel healthy are less likely to suspect that they have a disease such as diabetes. Again, the self-rated financial situation is the only enabling predictor with significant coefficient, and the sign is negative, confirming that there might be a problem of inequity in access to exams for diabetes diagnosis. The fact that living in the Metropolitan Area (where there are more and better medical resources) might reinforce that there is a problem of access to diabetes examination. Finally, age is directly related to the likelihood of ignoring the diabetes status, and probably this might be due to more resistance to go to medical services.

Discussion

The main goal of this paper was to construct an equation that would help to estimate a corrected prevalence of diabetes among Latin American elderly, given that most of the surveys about aging in the continent gather information on self-reported diabetes status, but do not perform clinical examinations. CRELES, the Costa Rican study on aging, belongs to a small group of projects -such as the NHANES and the Taiwan Biomarkers study- that collect clinically assessed biomarkers.

Differences between self-reported diabetes diagnosis and FSG concentrations were used to build statistical models that then were applied to the MHAS dataset in order to estimate possible over or underestimation of disease prevalence among the above-50-years-old population. If the same patterns observed in Costa Rica occur in Mexico, the prevalence of diabetes among the elderly population rises from 32% to around 52%. The lowest of these figure means that for every Mexican old-person that knows to have diabetes, there is roughly another that does not know. Now, these results do not agree with the estimates generated with the ENSA, which state that undiagnosed diabetes represents only around 15% of the total diabetic population ages 50 or above. It would be expected that the real proportion should be higher in Mexico than what ENSA reported, not only due to the limitations explained by its authors, but also because the percentage of “hidden” diabetes in countries with a more comprehensive health care system (the U.S., Denmark, or Australia) range from 30% to 50%.

However, when compared with Mexico, more developed countries (as well as Costa Rica, where this proportion mounts to 30%) are in more advanced stages of the epidemiological transition, where older age structures and different life styles increase the prevalence of chronic conditions. If undiagnosed diabetes is a reflection of the disease’s latent phase, then this stage in the development of diabetes might be more frequent in countries with higher prevalence of the disease. The utilization of the model created with the Costa Rican data needs to be cross-validated with other surveys that contain both self-reported and clinical data on diabetes mellitus.

It is also possible that the model is overestimating “hidden” diabetes because there are differences between Mexico and Costa Rica. Even though the two countries share historical and cultural backgrounds, there are differences between them regarding the health care system. For example, while almost all Costa Rican elderly have health insurance (in a large proportion, provided by the Government), in Mexico almost half of the population older than 49 years does not have any right for health insurance; also, Costa Rica’s small territory has been an advantage for expanding health care services, while Mexico’s large territory has made this expansion difficult (Mesa-Lago, 1992). However, because of problems of comparability, the prediction model does not have any variable that can serve as proxy for the distance to health care services in Mexico or for regional differences that can be explained partially by the effect of Mexico’s large territory.

Nonetheless, if the estimates produced by the model are true, they have strong public health implications for Mexico, given that people that do not know that they have the disease may die earlier or may be more likely to have an increase in their disability status. Rull et al (2005) report that diabetes mellitus has been the first cause of death for women and the second for men since 2000, and the primary cause of premature retirement, kidney failure, and blindness. The estimates also imply that the Mexican Health care system will need to do a large investment in medication, health services, and preventive measures in order to control the burden of this disease, which by now is quite large. Arredondo et al (2005) calculate that the disease cost per patient is as high as

$750, and it was the highest compared to the rest of the chronic diseases studied in their analysis.

The proportion of "hidden" diabetes in Costa Rica seems also very high, especially for a country with a strong public health care system (Mesa-Lagos, 1993). Diabetes complications can take a toll on the finances of this system. Morice and Achio (2003) estimated that diabetes was the illness with the highest hospitalization total cost, and the second in outpatient costs (after hypertension) for the public hospital and clinics network. Nonetheless, the proportion of controlled diabetes among those that had a diagnosis is considerably high: almost one of every two persons with a previous diabetes diagnosis had FSG levels under 126 mg/dl. This shows that treatment compliance in Costa Rica seems to be adequate and agrees with a perception expressed by Firestone et al (2004); with a non-representative sample of diabetes patients in San Jose, they find that the levels of diabetes-specific knowledge is greater than in a sample of Spanish-speaking U.S. Latinos in Starr County, Texas, after the patients underwent an intervention program aimed to provide preventive information.

The attempt of interpreting results of the full models yielded interesting information, too. While having controlled diabetes is mainly explained by need factors (symptoms), having the disease undiagnosed is closely related to access to resources: living in the Metropolitan Area and self-rated financial situation. This result is worrisome given that almost 100% of Costa Rican elderly have health insurance offered mainly by the State. Therefore, lack of diagnosis shows a possible problem of the system as provider of preventive health care services. The fact that the visits of the ATAPs (primary health technicians) do not have an impact on the likelihood of diagnosis is another indicator of the limitations of the system in developing preventive programs against chronic conditions. Among other covariates related to the condition, undiagnosed diabetes is less likely among the elderly with hypertension and among those who have been hospitalized during the last 12 months. Also, those who rate their health as good are more likely to have "hidden" diabetes: whether this is an effect of health beliefs -people label themselves as healthy in order to avoid acknowledging their illness- or of actual absence of symptoms is unclear.

As a summary, the analyses performed in this article were useful in understanding the problem of undiagnosed diabetes in Costa Rica and in building a model that can be utilized to estimate underestimation in diabetes prevalence by other aging studies in Latin America. However, the analyses have some limitations, particularly in the predictive model building, given that the percentages of incorrectly classified cases within the original dataset (CRELES) are still considerably large: between 30% to 37%. This feature introduces bias in the classification that is difficult to control. Therefore, there is the need to externally validate the classification model with another study that contains both clinical and self-reported information on diabetes, in order to understand the limitations of the model.

Acknowledgements

"Costa Rica: estudio de longevidad y envejecimiento saludable (CRELES)" is a research project of the University of Costa Rica, conducted by the Central American Center for Population (CCP) in colaboration with Instituto de Investigaciones en Salud and the laboratories of Hospital San Juan de Dios, and funding from Wellcome Trust Foundation. Principal Investigator: Luis Rosero-Bixby. Co-investigators: Xinia Fernández and William H. Dow. Colaborating researchers: Ericka Méndez, Guido Pinto, Hannia Campos, Kenia Barrantes, Alberto Cubero, Gilbert Brenes, Fernando Morales, María de los Angeles San Román and Grettel Valverde. Technical and support staff: Daniel Antich, Aaron Ramírez, Jeisson Hidalgo, Juanita Araya and Yamileth Hernández. Fieldwork staff: José Solano, Julio Palma, Jenny Méndez, Maritza Aráuz, Mabelyn Gómez, Marcela Rodríguez, Geovanni Salas, Jorge Vindas y Roberto Patiño.

References

Aday LA, Begley CE, Lairson DR and Slater CH (1998) *Evaluating the health care system: effectiveness, efficiency, and equity*. 2nd ed. Chicago: Health Administration Press.

Andersen RM (1995) "Revisiting the Behavioral Model and access to medical care: Does it matter?" *Journal of Health and Social Behavior*, 36 (1): 1-10.

Andrade F (2005) "Obesity and central obesity in elderly people in Latin America and the Caribbean – Are we fat?", paper presented at the 2005 Meeting of the Population Association of America, Philadelphia, PA, March 31st-April 2nd

Angel R and Gronfein W (1988) "The use of subjective information in statistical models" *American Sociological Review* 53(3): 464-473.

Arredondo A and Damián T (1997). "Costos económicos en la producción de servicios de salud: del costo de los insumos al costo de manejo de caso". *Salud Pública de México* 39(2): 117-124.

Buonaccorsi JP (1996) “Measurement error in the response in the General Linear Model” *Journal of the American Statistical Association* 91(434): 633-642.

Caldwell JC (1986) “Routes to low mortality in poor countries” *Population and Development Review* 12(2): 171-220.

Dunstan DW, Zimmet PZ, Welborn TA, de-Courten M, Cameron AJ, Sicree RA, Dwyer T, Colagiuri, Jolley D, Knuiman M, Atkins R and Shaw JE (2002) “The Australian Diabetes, Obesity and Lifestyle Study” *Diabetes Care* 25: 829-834

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

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

Franse LV, Di Bari M, Shorr RI, Resnick HE, van Eijk JTM, Bauer DC, Newman AB, and Pahor Marco (2001) “Type 2 Diabetes in older well-functioning people: Who is undiagnosed?” *Diabetes Care* 24: 2065-2070.

Glumer C, Borch-Johnsen K and Colagiuri S (2005) “Can a screening programme for diabetes be applied to another population?” *Diabetic Medicine* 22: 1234-1238.

Glumer C, Jorgensen T, Borck-Johnsen K (2003) “Prevalences of diabetes and impaired glucose regulation in a Danish population: the Inter99 study” *Diabetes Care* 26: 2335-2340.

Goldman N, Lin IF, Weinstein M and Lin YH (2003) “Evaluating the quality of self-reports of hypertension and diabetes” *Journal of Clinical Epidemiology* 56: 148-154.

Gregg EW, Cadwell BL, Ching YJ, Cowie CC, Williams DE, Geiss L, Engelgau MM and Vinicor F (2004) “Trends in the prevalence and ratio of diagnosed to undiagnosed diabetes according to obesity levels in the U.S.” *Diabetes Care* 27: 2806-2812.

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” *American Journal of Epidemiology* 145(8):762-769

Hardy MA and Pavalko EK (1986) “The internal structure of self-reported health measures among older male workers and retirees” *Journal of Health and Social Behavior* 27(4): 346-357.

Hosmer DW and Lemeshow S (2000). *Applied Logistic Regression*. New York: John Wiley & Sons, Inc. 2nd ed. Wiley Series in Probability and Statistics, Texts and References Section.

Karter AJ, Parker MM, Moffet HH, Ahmed AT, Ferrara A, Liu JY, Selby JV (2004) “Missed Appointments and Poor Glycemic Control: An Opportunity to Identify High-Risk Diabetic Patients” *Medical Care* 42(2):110-115

Koopman RJ, Mainous AG III and Geesey ME (2006) “Rural residence and Hispanic ethnicity: doubly disadvantaged for diabetes?” *Journal of Rural Health* 22(1):63-68.

Kriegsman DM, Penninx BW, van Eijk JT, Boeke AJ and Deeg DJ (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" *Journal of Clinical Epidemiology* 49(12):1407-17.

Lutfey KE and Wishner WJ "Beyond 'compliance' is 'adherence'. Improving the prospect of diabetes care" *Diabetes Care* 22: 635-639.

Martin LM, Leff M, Calonge N, Garrett C, Nelson DE (2000) "Validation of self-reported chronic conditions and health services in a managed care population" *American Journal of Preventive Medicine* 18(3):215-218.

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

Mohan V, Deepa R, Deepa M, Somannavar S and Datta M (2005) "A simplified Indian diabetes risk score for screening for undiagnosed diabetic subjects" *Journal of The Association of Physicians of India* 53:759-763

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

Olaiz G, Rojas R, Barquera S, Shamah T, Aguilar C, Cravioto P, Lopez P, Hernandez M, Tapia R, Sepulveda J (2003) *Encuesta Nacional de Salud 2000. Tomo 2. La salud de los adultos.* Cuernavaca, Morelos, Mexico: Instituto Nacional de Salud Publica.

Palloni A and McEniry (2004) "Aging and Health Status of Elderly in Latin America and the Caribbean". CDE Working Paper 2004-09

Palloni A, McEniry M, Guend H, Davila AL, Garcia A, Mattei H and Sanchez M (2004) "Health among elderly Puerto Ricans: analysis of a new data set" CDE Working Paper 2004-11.

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

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

Peyrot M, McMurray Jr. JF, Kruger DF (1999) "A biospsychosocial model of glycemic control in diabetes: Stress, coping and regimen adherence" *Journal of Health and Social Behavior* 40(2): 141-158.

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

Rathmann W, Haastert B, Icks A, Giani G, Holle R, Meisinger C, Mielck A (2005) "Sex differences in the associations of socioeconomic status with undiagnosed diabetes mellitus and impaired glucose tolerance in the elderly population: the KORA Survey 2000" *European Journal of Public Health* 15(6):627-33

Rosero-Bixby L, Hidalgo J, and Antich D (2005) "Improving the Quality and Lowering Costs of Household Survey Data Using Personal Digital Assistants (PDAs): An Application for Costa Rica". Paper presented at the 2005 meeting of the Population Association of America, Philadelphia, PA, March 31st-April 2nd

Rull, JA, Aguilar-Salinas CA, Rojas R, Ríos-Torres JM, Gómez-Pérez FJ and Olaiz G (2005). "Epidemiology of type 2 diabetes in Mexico". *Archives of Medical Research* 36(3): 188-196.

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

Tabaei BP, Engelgau MM and Herman WH (2005) "A multivariate logistic regression equation to screen for dysglycaemia: development and validation" *Diabetic Medicine* 22: 599-605.

Vallin J (1988) "L'évolution de la mortalité en Asie méridionale et orientale" *Population* 43(6): 1157-1160.

WHO World Health Organization (1999). *Definition, diagnosis and classification of diabetes mellitus and its complications"* Report of a WHO Consultation.

Wilder RP, Majumdar SR, Klarenbach SW and Jacobs (2005) "Socio-economic status and undiagnosed diabetes" *Diabetes Research and Clinical Practice* 70: 26-30.

Wong, R and M Espinoza (2004). "Response rates in the Mexican Health and Aging Study (MHAS/ENASEM) 2003". In: http://www.mhas.pop.upenn.edu/espanol/documentos/TasasdeRespuesta/Response-rates-2003-v1.pdf.

Table 1. Variables selected and not-selected for estimation models, classified according to Anderson's Health Behavior Model

Needs		**Enabling factors**	**Pre-disposing factors**
	Selected for estimation models		
• Hospitalization • Self-rated health (in a scale, and compared to others) • Hypertension • Heart attack • Other heart disease • Stroke	• Swollen feet • Dizziness • Intense thirst • Fatigue • Cough and phlegms • Burning when urinating • Insulin * • Interact insulin-swollen feet *	• Lives alone • Household size • Number of alive children • Functional status (Activities of Daily Living ADL's and Instrumental Activities of Daily Living (IADL's) • Self-rated financial situation • Currently working	• Age • Completed elementary school • Sex • Smoking status • Obesity and overweight • Retired • Cognitive impairment
	Not selected for estimation models		
		• Visit from primary health technician	• Living in Metropolitan Area • Living in urban areas • Family history of diabetes

Notes: *: Variables that were included in the equation for controlled diabetes, but not for hidden diabetes.

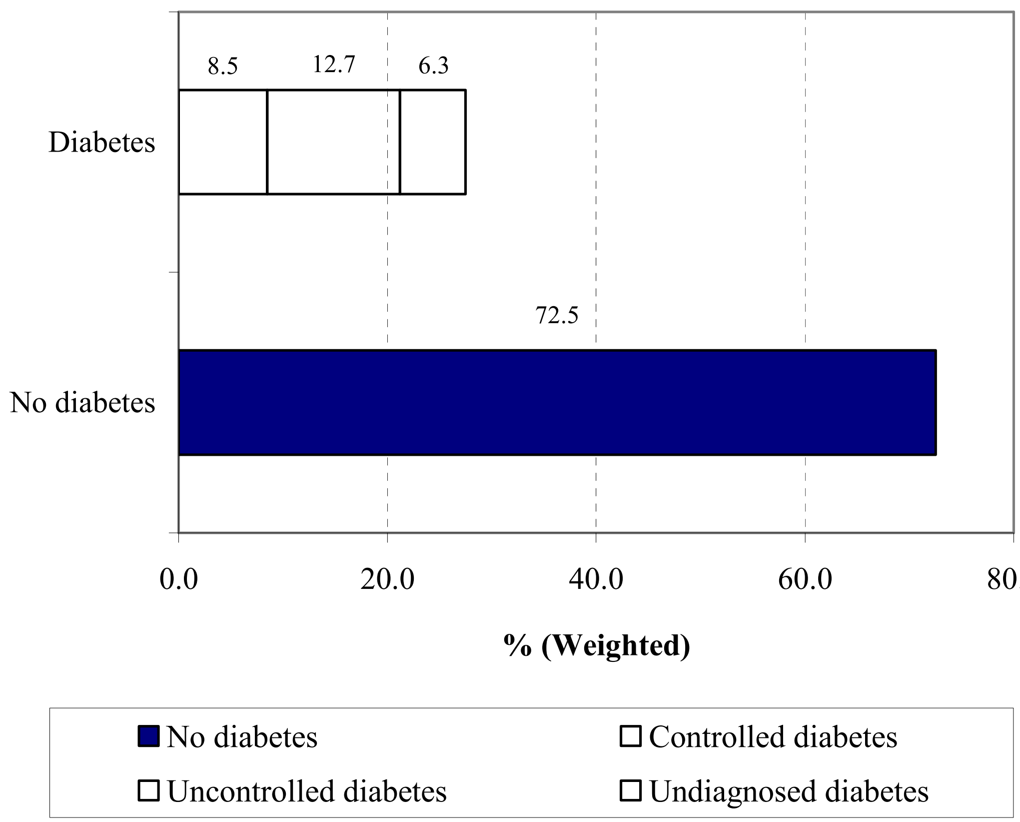
Table 2. Descriptives of all explanatory variables considered in the paper, by diabetes status (weighted samples).

Explanatory variables	**Total**	No diabetes	Undiagn diabetes	Uncon-trolled diabetes	Controll-ed diabetes
n	*1485*	*1113*	*92*	*159*	*121*
Weighted % dist	*100.0*	*72.5*	*6.3*	*12.7*	*8.5*
Needs					
Proportions					
-Hospitalized	**0.12**	0.13	0.13	0.03	0.18
-Hypertension	**0.47**	0.41	0.69	0.52	0.72
-Heart attack	**0.05**	0.05	0.09	0.03	0.04
-Other heart disease	**0.13**	0.12	0.19	0.10	0.12
-Stroke	**0.06**	0.06	0.06	0.05	0.08
-Swollen feet	**0.31**	0.30	0.43	0.26	0.35
-Dizziness	**0.41**	0.41	0.40	0.27	0.52
-Intense thirst	**0.23**	0.22	0.33	0.15	0.31
-Fatigue	**0.40**	0.39	0.49	0.25	0.46
-Cough and phlegms	**0.20**	0.20	0.19	0.15	0.26
-Burning when urinating	**0.15**	0.14	0.16	0.09	0.19
-Use of insulin shots [1]	**0.06**	-	0.32	-	0.27
Means					
-Self-rated health (1=Poor, 5=Exc)	**2.6**	2.6	2.4	2.0	2.3
Enabling factor					
Proportions					
-Live alone	**0.11**	0.12	0.08	0.15	0.12
-Currently working	**0.21**	0.22	0.19	0.20	0.17
-Visited by primary health tech.	**0.45**	0.46	0.43	0.38	0.47
Means					
-Household size	**3.4**	3.4	3.5	3.2	3.4
-Number of alive children	**6.0**	6.0	6.2	5.7	5.8
-Sum of ADL's and IADL's	**1.3**	1.4	1.0	0.8	1.5
-Self-rated financial situation (1=Poor, 5=Exc)	**2.3**	2.3	2.4	2.3	2.1
Pre-disposing factors					
Proportions					
-Males	**0.47**	0.49	0.36	0.46	0.37
-Completed elementary school	**0.30**	0.29	0.33	0.43	0.26
-Currently smoking	**0.09**	0.08	0.08	0.15	0.07
-Obese	**0.22**	0.17	0.35	0.29	0.37
-Overweight	**0.40**	0.39	0.49	0.39	0.42
-Retired	**0.63**	0.64	0.58	0.70	0.61
-Living in Metropolitan Area	**0.42**	0.39	0.50	0.62	0.42
-Urban	**0.58**	0.56	0.66	0.74	0.60
Means					
-Age	**76.2**	76.9	72.8	76.1	75.0
-Cognitive impairment scale (Number of correct items from a total of 15)	**11.8**	11.8	12.0	12.0	11.7

Source: CRELES

Notes: [1.] Use of insulin shots is asked only to those with a previous diabetes diagnosis

Figure 1. Prevalence of controlled, uncontrolled and undiagnosed diabetes mellitus among Costa Ricans ages 60 and above, according to CRELES.



Source: CRELES

Table 3. Equation for estimating controlled diabetes, among those that were diagnosed with diabetes, reduced by step-wise procedure (in log-odds scale).

Explanatory variable	**Coeff**	**(S.E.)**	**p-value**
Heart attack	-1.477	(0.668)	0.027
Swollen feet	-0.738	(0.297)	0.013
Dizziness	0.893	(0.298)	0.003
Self-rated financial situation	-0.281	(0.161)	0.081
Currently smoking	-0.916	(0.604)	0.130
Constant	-1.419	(0.600)	0.018

Table 4. Equation for estimating undiagnosed diabetes, among those that have not been diagnosed with diabetes, reduced by stepwise procedure. (in log-odds scale)

Explanatory variable	**Coeff**	**(S.E.)**	**p-value**
Hypertension	0.494	(0.283)	0.081
Dizziness	-0.535	(0.305)	0.080
Hospitalized	-1.939	(0.964)	0.044
ADL's+IADL's	-0.208	(0.093)	0.025
Self-rated financial situation	-0.280	(0.126)	0.026
Age	0.034	(0.018)	0.064
Completed elementary school	0.576	(0.275)	0.036
Currently smoking	1.179	(0.371)	0.001
Obese	1.276	(0.336)	0.000
Overweight	0.976	(0.298)	0.001
Retired	0.592	(0.305)	0.052
Constant	-7.293	(1.284)	0.000

Table 5. Sensitivity and specificity of equations with CRELES sample.

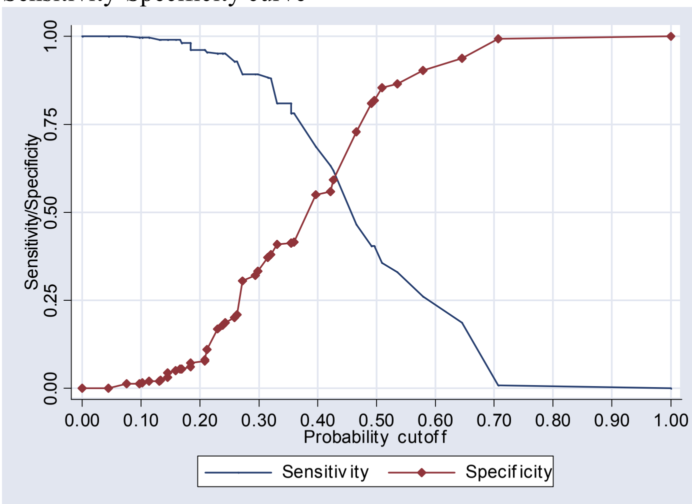
Indicators	Undiagnosed	Controlled
Optimal cutoff point *	0.085	0.415
Sensitivity	0.728	0.622
Specificity	0.700	0.646
Positive predictive power	0.176	0.551
Negative predictive power	0.967	0.710
Proportion correctly classified	0.703	0.636

Notes: * Determined by intersection of specificity and sensitivity curves with different cut-off points.

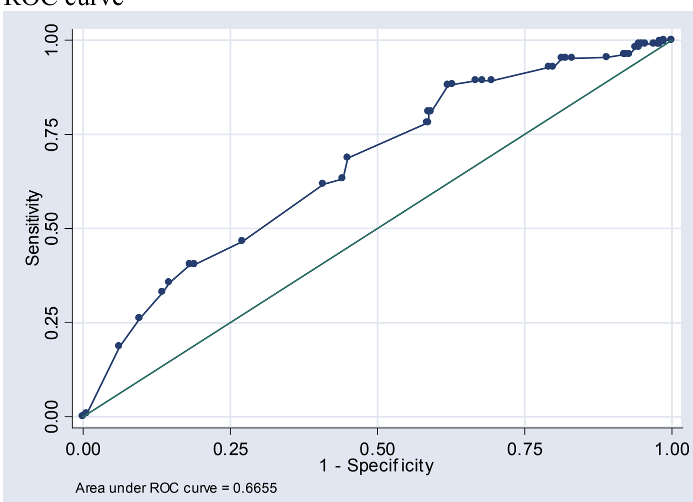
Figure 2.ROC curves and "Sensitivity and Specificity" curves for the two logistic equations.

Controlled diabetes

Sensitivity-Specificity curve

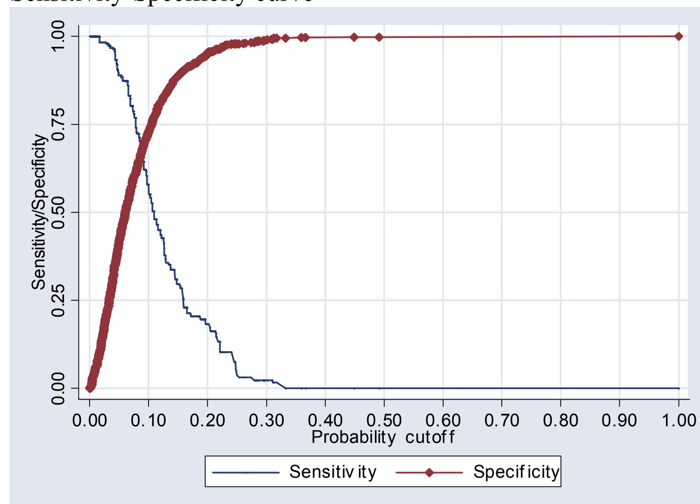


ROC curve



Undiagnosed diabetes

Sensitivity-Specificity curve



ROC curve

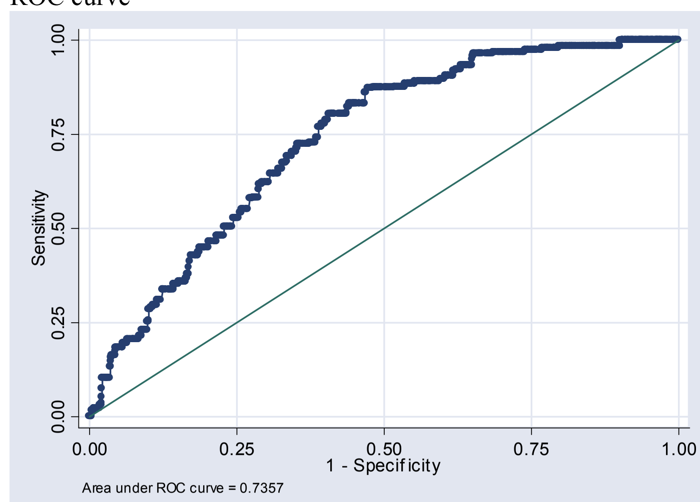


Table 6. Coefficients for logistic regression of diabetes status according to biomarker. Full and reduced model (in log-odds scale).

Explanatory variables	**Full model**			**Reduced model**		
	Coeff	**(SE)**		**Coeff**	**(SE)**	
-Hospitalized	-0.028	(0.310)				
-Hypertension	0.729	(0.201)	***	0.792	(0.182)	***
-Heart attack	0.745	(0.389)	*			
-Other heart disease	0.157	(0.293)				
-Stroke	-0.173	(0.464)				
-Swollen feet	0.385	(0.221)	*	0.354	(0.193)	*
-Dizziness	-0.487	(0.224)	**	-0.349	(0.186)	*
-Intense thirst	0.145	(0.226)				
-Fatigue	0.189	(0.213)				
-Cough and phlegm	-0.257	(0.240)				
-Burning when urinating	-0.088	(0.269)				
-Self-rated health	-0.079	(0.121)				
-Live alone	-0.113	(0.329)				
-Currently working	-0.204	(0.273)				
-Visited by primary health tech.	-0.051	(0.189)				
-Household size	-0.026	(0.049)				
-Number of alive children	0.040	(0.030)				
-Sum of ADL's and IADL's	-0.210	(0.092)	**	-0.093	(0.057)	*
-Self-rated financial situation	0.035	(0.114)				
-Males	-0.195	(0.237)		-0.285	(0.184)	
-Completed elementary school	0.042	(0.232)				
-Currently smoking	1.057	(0.309)	**	0.937	(0.280)	***
-Obese	0.878	(0.285)	**	0.936	(0.253)	***
-Overweight	0.621	(0.254)	**	0.697	(0.230)	***
-Retired	0.126	(0.207)				
-Living in Metropolitan Area	0.421	(0.219)	*			
-Urban	0.375	(0.227)	*			
-Age	-0.018	(0.015)				
-Cognitive impairment scale	-0.075	(0.051)				
-Family history of diabetes	0.917	(0.192)	***			
Constant	-0.634	(1.436)		-2.402	(0.243)	***
-Log-Likelihood		*507.04*			*615.24*	
n		*1150*			*1362*	

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

Figure 3. Sensitivity and specificity curves.

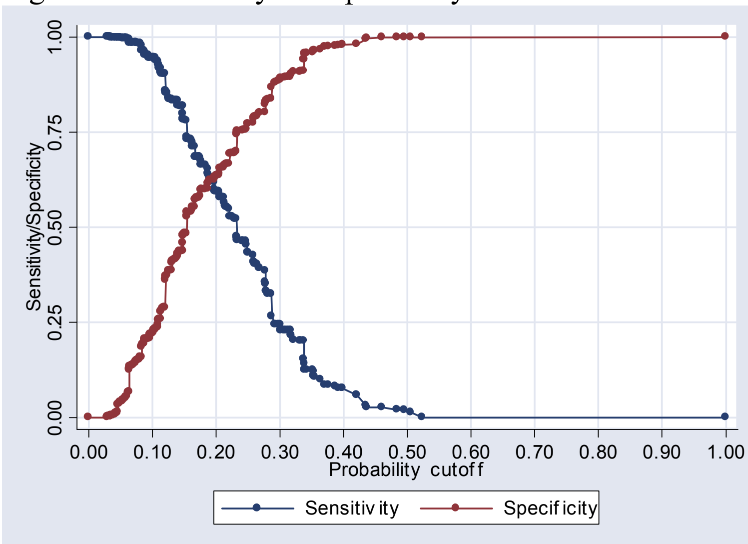


Table 7. Estimates of diabetes status in MHAS using CRELES models, according to the methods used.

Categories	Observed	Direct classification	Random imputation Med	(p5,	p95)
"Two-equation" approach					
Total	**100.0**	**100.0**	**100.0**		
% no diabetes	**82.8**	**63.8**	**67.5**		
% diabetes	**17.2**	**36.2**	**32.5**		
% Controlled (over total)		9.0	7.5	(7.0,	8.1)
% Uncontrolled (over total)		8.2	9.3	(8.7,	9.8)
% Undiagnosed (over total)		19.0	15.7	(15.1,	16.2)
"One-equation" approach					
Total	**100.0**	**100.0**	**100.0**		
% no diabetes	**82.8**	**47.6**	**57.8**		
% diabetes	**17.2**	**52.4**	**42.2**		
% Controlled (over total)		8.3	5.5	(5.1,	5.9)
% Uncontrolled (over total)		9.8	11.3	(10.8,	11.6)
% Undiagnosed (over total)		34.3	25.4	(24.4,	26.4)

Table 8. Coefficients for logistic regressions of undiagnosed and controlled diabetes in CRELES.

Explanatory variables	**Controlled diabetes**			**Undiagnosed diabetes**		
	Coeff	**(SE)**		**Coeff**	**(SE)**	
Needs						
-Hospitalized	-0.062	(0.448)		**-1.671**	**(0.950)**	*
-Hypertension	0.112	(0.382)		0.419	(0.317)	
-Heart attack	**-1.781**	**(0.845)**	**	-0.629	(1.129)	
-Other heart disease	-0.644	(0.523)		-0.296	(0.553)	
-Stroke	1.076	(0.800)		1.080	(0.859)	
-Swollen feet	**-0.845**	**(0.447)**	*	-0.237	(0.359)	
-Dizziness	**0.945**	**(0.367)**	**	-0.331	(0.359)	
-Intense thirst	-0.444	(0.402)		-0.203	(0.447)	
-Fatigue	-0.299	(0.400)		-0.209	(0.352)	
-Cough and phlegm	1.076	(0.393)	***	-0.117	(0.390)	
-Burning when urinating	0.215	(0.412)		-0.371	(0.476)	
-Use of insulin shots	-0.149	(0.552)				
-Insulin shots + swollen feet	-0.897	(0.769)				
-Self-rated health	-0.267	(0.240)		0.180	(0.195)	
Enabling factor						
-Live alone	0.687	(0.732)		0.681	(0.456)	
-Currently working	0.141	(0.464)		-0.087	(0.415)	
-Visited by primary health tech.	-0.044	(0.361)		-0.157	(0.325)	
-Household size	0.020	(0.097)		0.031	(0.071)	
-Number of alive children	-0.021	(0.060)		0.031	(0.050)	
-Sum of ADL's and IADL's	0.354	(0.204)	*	-0.155	(0.159)	
-Self-rated financial situation	**-0.412**	**(0.218)**	*	-0.466	(0.169)	***
Pre-disposing factors						
-Males	-0.161	(0.379)		-0.110	(0.377)	
-Completed elementary school	0.343	(0.387)		0.400	(0.352)	
-Currently smoking	**-0.778**	**(0.651)**		**1.341**	**(0.384)**	***
-Obese	0.017	(0.566)		**1.250**	**(0.389)**	***
-Overweight	-0.061	(0.475)		**1.029**	**(0.328)**	***
-Retired	0.428	(0.389)		0.504	(0.328)	
-Living in Metropolitan Area	-0.089	(0.386)		**0.799**	**(0.373)**	**
-Urban	-0.132	(0.398)		0.375	(0.365)	
-Age	0.009	(0.026)		0.032	(0.020)	
-Cognitive impairment scale	0.042	(0.090)		-0.004	(0.095)	
-Family history of diabetes	-0.472	(0.378)		0.327	(0.282)	
Constant	0.277	(2.472)		-6.282	(2.133)	***
-Log-Likelihood	*129.61*			*223.03*		
n	*227*			*923*		

Notes: In bold letter, coefficients of the variables that were selected by stepwise procedure for prediction equation.
*: p<.10, **: p<.05, ***: p<.01

Table 9. Coefficients for logistic regression of unknown diabetes condition (against report of diagnosis) among elderly classified as diabetics by either self-report or bio-markers.

Explanatory variables	**Coeff**	**(SE)**	
Needs			
-Hospitalized	-1.994	(0.989)	**
-Hypertension	-0.700	(0.368)	*
-Heart attack	-0.973	(1.446)	
-Other heart disease	-0.486	(0.538)	
-Stroke	0.565	(0.994)	
-Swollen feet	-0.523	(0.434)	
-Dizziness	-0.280	(0.407)	
-Intense thirst	-0.487	(0.480)	
-Fatigue	-0.201	(0.424)	
-Cough and phlegm	0.069	(0.412)	
-Burning when urinating	-0.128	(0.590)	
-Self-rated health	0.562	(0.206)	***
Enabling factor			
-Live alone	0.100	(0.566)	
-Currently working	0.086	(0.519)	
-Visited by primary health tech.	0.137	(0.369)	
-Household size	-0.005	(0.096)	
-Number of alive children	-0.025	(0.060)	
-Sum of ADL's and IADL's	-0.056	(0.296)	
-Self-rated financial situation	-0.807	(0.242)	***
Pre-disposing factors			
-Males	-0.024	(0.439)	
-Completed elementary school	0.603	(0.406)	
-Currently smoking	0.710	(0.492)	
-Obese	-0.116	(0.530)	
-Overweight	0.081	(0.440)	
-Retired	0.400	(0.390)	
-Living in Metropolitan Area	0.866	(0.402)	**
-Urban	0.281	(0.448)	
-Age	0.064	(0.028)	**
-Cognitive impairment scale	0.170	(0.109)	
Constant	-7.574	(2.831)	***
-Log-Likelihood		*123.85*	
n		*302*	

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