Introduction

Health insurance is often cited as a policy instrument with the capacity to improve equity of health care access and health outcomes. The Mexican government is currently working towards expanding social health insurance coverage to the entire population by the year 2025. Currently, more than 40% of the Mexican population is uninsured. The theoretical rationale for expanding social health insurance programs to the entire population is that insurance decreases the cost of medical care which is thought to increase health care access and utilization which subsequently increases health status for the population. Despite the increasing trend of expanding insurance programs, in developing countries there is little evidence that health insurance improves health status.

Observational studies of health insurance on health care utilization and health status suffer from the problematic selection issues surrounding insurance status. A central difficulty with studies of insurance on health care utilization and health outcomes is that the insured and uninsured are likely to differ in both observable and unobservable characteristics. However, the differences become problematic when they are unobservable and affect both the decision to obtain insurance and the decision to use health care services. There are several potential solutions for the endogeneity problem, including using either additional control variables (if available), fixed effects models or the use of two-stage least squares (a variant of instrumental variables). This analysis will attempt the use of two-stage least squares.

This proposed study investigates the effect of health insurance on the prevalence, diagnosis and management of disease in Mexico using a cross-sectional, nationally representative data set. The long-term objective is to better understand the role of economic determinants in morbidity. The contributions of this study include the attempt to address the issue of endogeneity and the use of plasma glucose concentration as well as measures of systolic and diastolic blood pressure in understanding the prevalence, detection and management of disease. The findings from this research may assist the Mexican government in the development of their national health program agenda by providing information on the pathway of disease prevalence, diagnosis and management of disease.

Conceptual Framework

A household production model, based on the health production function framework of Schultz (1984), Grossman (1972) and an example by Dow and colleagues (2003) is used to illustrate that insurance is hypothesized to lower price and increase demand for health care, which will lead to an improvement of health status.

Health care is not valued in and of itself, but rather it is an input for the health production function. Health is a function of medical inputs M_i, non-medical inputs S_i (such as sanitation and nutrition) chosen by the household, the health environment E_c in community c, and the unobserved health/biological endowment θ_i:

(1) $H_i = H(M_i, S_i, E_c, \theta_i)$

Medical input demand M_i, is chosen by the individual to maximize health with limited resources. The demand for health care services can be expressed as a function of socioeconomic characteristics W_i such as household wealth and educational attainment levels, in addition to health and unobserved preferences π_i and the local environment. Insurance I_i, which lowers the price of medical care P_c (including time costs associated with geographic access and wait times), may increase the demand for medical care:

(2) $M_i = M(I_i, W_i, P_c, H_i, \pi_i)$

Non-medical inputs S_i depend on similar factors as M, although the cross-price effect of insurance may be ignorable:

(3) $S_i = S(W_i, P_c, H_i, \pi_i)$

Insurance may be considered as endogenous to the system, depending on factors such as health status, socioeconomic status, health care access barriers, and unobserved preferences for medical care:

(4) $I_i = I(W_i, P_c, H_i, \pi_i)$

If insurance is in fact endogenous to health status, then many of the relationships in (2) work in different directions, making it difficult to sign net omitted variable biases in observational data with imperfect controls. For example, adverse selection would imply that less healthy individuals would have a higher demand for insurance, causing health care use to be positively related and health status to be negatively related to health insurance. However, positive selection of healthier and higher socioeconomic status persons into positions which provide health insurance could cause a negative relationship between health insurance and health care use and a positive relationship between health insurance and health status. Thus, it is imperative for the empirical design to account for the potential endogeneity of insurance.

Data

The Encuesta Nacional de Salud 2000 (National Survey of Health 2000) is a survey consisting of 190,214 nationally representative observations of a population of roughly 100 million. The adult health survey module, which includes biological samples and reported diagnosis of disease (n=45,273), will be used for this analysis.

Dependent Variables

Six dependent variables are available for analyses, including: individual has diabetes, undiagnosed diabetes, uncontrolled diabetes, individual has hypertension, undiagnosed hypertension and uncontrolled hypertension.

An individual is classified as diabetic if they test positive for diabetes, based on measurements from a casual plasma glucose concentration ($\geq$200mg/dl) from this survey or if they report being told by a doctor that they have diabetes or high blood sugar. In order to be classified as having undiagnosed diabetes, the individual must have a casual plasma glucose measure of $\geq$200mg/dl and report never being told by a physician or health care personnel that they have diabetes. An individual is classified as having uncontrolled diabetes if they report being diagnosed by a physician or health care personnel with diabetes or high blood sugar and also have a casual plasma glucose concentration $\geq$200mg/dl.

An individual is classified as having hypertension if they test positive for hypertension, based on two measures of systolic ($\geq$140 mmHg) and/or diastolic ($\geq$90 mmHg) blood pressure from this survey or if they report being told by a doctor that they have hypertension or high blood pressure. Undiagnosed hypertension, measures whether individuals who report never being told by a doctor or other medical personnel that they have hypertension or high blood pressure actually have two measures of systolic ($\geq$140 mmHg) and/or diastolic ($\geq$90 mmHg) blood pressure. An individual is classified as having uncontrolled hypertension if they report being diagnosed by a physician or health care personnel with hypertension or high blood pressure, yet they test positive for

hypertension with the two measures of systolic (≥140 mmHg) or diastolic (≥90 mmHg) blood pressure.

Explanatory Variables
The key explanatory variable, insurance status, was created from the type (if any) of social or private insurance the individual reports having. If the individual reported any type of insurance, they were coded as insured. The majority of the sample is uninsured (58.3%). Other explanatory variables include: age, martial status, educational attainment, indigenous status, household wealth, and measures of household sanitation (i.e., piped water, drainage and sanitation).

Potential Instruments
An instrument must meet two basic conditions. First, the instrument must be substantially correlated with the endogenous explanatory variable (i.e., insurance). Second, the instrument must be uncorrelated with the error of the structural equation of interest. The instrument should only be related to the dependent variable (i.e. disease status, undiagnosed disease or uncontrolled disease) via its impact on insurance. Based on the Mexican social security system, if the of a head of household receives insurance through their employment, the head and all the dependents are covered by the insurance. The two sets of potential instruments included in this analysis are household head employment type and business function.

The first potential instrument, employment type, classifies the location of household head employment in the last week into 24 categories. Examples of these categories include: in a factory, in a mechanical workshop, in the fields, etc. The employment type does not directly classify individuals as working in the formal or informal sector and many of the employment types listed may have both formal and informal components. The second set of potential instruments, function of business, classifies the primary function of the business, company or place where the household head works (i.e., cultivating, making furniture, selling clothes) into 118 categories. The rationale behind these variables as a potential instruments is that the employment function and type may be strongly predictive of whether or not a household head and the family members receive insurance (as certain types of businesses may provide insurance benefits to its employees), but may not be predictive of the health status of the individual, especially for dependents.

Methods
Knowledge of disease status is of primary importance for disease management and is critical for achieving optimal health status. I propose to estimate the effect of insurance on disease status, as well as diagnosis and management of disease for diabetes and hypertension in a national sample of Mexicans 20 years of age and older.

First, descriptive statistics will be used to describe the current distribution of insurance, prevalence, diagnosis and management of disease across socioeconomic groups. The summary statistics will describe the target population for Mexico's insurance expansion and may reveal the direction of the observed selection differences between the insured and the uninsured. I will also use t-tests to compare individuals with and without insurance and examine evidence for socioeconomic disparities.

Single Equation Methods and Results
Second, the effect of insurance on prevalence, diagnosis and management of disease will be estimated using regression analysis. Before testing and attempting to control for the potential endogeneity of insurance, I will describe and estimate the simpler single equation model. This single equation model is used as a base case and will later be compared with the models that

control for the potentially endogenous variable, insurance. Six dependent variables for prevalence, diagnosis and management will be estimated separately (resulting in six equations), including having diabetes, undiagnosed diabetes, uncontrolled diabetes, having hypertension, undiagnosed hypertension and uncontrolled hypertension.

The main equation for disease specifies disease status, diagnosis and management of disease as a function of insurance, individual-level controls and household-level controls. The estimate of this function is:

$$Y = \beta_0 + \beta_1 INSURANCE + \beta_2 INDIVIDUAL + \beta_3 HOUSEHOLD + \beta_4 HEALTH\ STATUS + \beta_5 STATES + \varepsilon \quad (1)$$

where INSURANCE is a scalar measuring insurance status, INDIVIDUAL, HOUSEHOLD, HEALTH STATUS and STATES are vectors, the β's are parameters to be estimated, and ε is an error term. The individual variables are age, male, indigenous, married and two categories of education (primary and secondary or greater – the base case is no education). The household variables include two measures of wealth status, sanitation, drainage, piped water and travel time to the nearest health care facility. The health status variables are good health (a binary variable for self-reported health status of good or better), recent health problem (a binary variable for morbidity in the last two weeks) and limitations (a binary variable for having a minimum of one limitation). The states vector consists of 32 binary state variables. Due to the survey design of this data, both population weights and design effects are incorporated in each regression.

Two-Stage Least Squares Methods and Results

The potential endogeneity of insurance will be addressed in this section of the analysis. The proposed instruments (i.e., employment type and function of business) will be tested for strength and appropriate exclusion from main equation. A test of the endogeneity of insurance will also be performed.

The 2SLS over-identified estimate of (1) is:

$$\textit{First stage}: INSURANCE = \gamma_0 + \gamma_1 EMPLOYMENT\ TYPE + \gamma_2 BUSINESS\ FUNCTION + \gamma_3 INDIVIDUAL + \gamma_4 HOUSEHOLD + \gamma_5 HEALTH\ STATUS + \gamma_6 STATES + \varepsilon \quad (2)$$

$$\textit{Second Stage}: Y = \delta_0 + \delta_1 INSURANCE^b + \delta_2 INDIVIDUAL + \delta_3 HOUSEHOLD + \delta_4 HEALTH\ STATUS + \delta_5 STATES + \varepsilon \quad (3)$$

Where δ_1 is the over-identified 2SLS estimate of β_1, $INSURANCE^b$ is the predicted value from (2), *EMPLOYMENT TYPE* and *BUSINESS FUNCTION* are each a vector of the instruments.

Discussion

Expanding health insurance programs is often cited as a way to improve health status and provide access to health services for poorer segments of the population (Trujillo, 2000). Thus, it is imperative to measure how health insurance differentially affects the prevalence, diagnosis and management of disease in the Mexican setting. Ultimately, knowledge gained from this project may be used to inform future policy debates and formulation in relation to health insurance laws in developing countries, including the relationship between health status and health insurance in Mexico.