INTERNATIONAL ECONOMIC REVIEW
Vol. 57, No. 3, August 2016

SCREENING FOR A CHRONIC DISEASE: A MULTIPLE STAGE DURATION MODEL WITH PARTIAL OBSERVABILITY*

BY THOMAS A. MROZ, GABRIEL PICONE, FRANK SLOAN, AND ARSENIY P. YASHKIN[1]

Georgia State University and the Federal Reserve Bank of Atlanta, U.S.A.; University of South Florida, U.S.A.; Duke University, U.S.A.; Duke University, U.S.A.

We estimate a dynamic multistage duration model to investigate how early detection of diabetes can delay the onset of lower extremity complications and death. We allow for partial observability of the disease stage, unmeasured heterogeneity, and endogenous timing of diabetes screening. Timely diagnosis appears important. We evaluate the effectiveness of two potential policies to reduce the monetary costs of frequent screening in terms of lost longevity. Compared to the status quo, the more restrictive policy yields an implicit value for an additional year of life of about $50,000, whereas the less restrictive policy implies a value of about $120,000.

1. INTRODUCTION

According to the U.S. Centers for Disease Control (CDC), 75% of health-care expenditures and 70% of all deaths in the United States are attributable to chronic diseases, including heart conditions, cancer, stroke, and diabetes (CDC, 2009). Earlier detection of these chronic diseases can yield substantial savings and better health outcomes. To achieve these goals, the Affordable Care Act (ACA) of 2010 subsidizes not only primary preventive measures, such as improvements in diet, but also secondary preventive measures. For example, since 2014 all insurance plans must cover many screening tests without any copayment. However, the empirical evidence that increased screening will save resources or even improve health outcomes is mixed (Cutler, 2008). Knowing earlier that an individual has a chronic disease does not necessarily imply that screening can delay disease progression or increase longevity.

The "gold standard" for evaluating the benefits versus costs of alternative screening policies is the randomized controlled trial (RCT). RCTs provide a simple approach to solve the problem of unobserved heterogeneity, but their usefulness for policy evaluations can be limited in many important situations. When the outcomes monitored are relatively rare, they can be quite expensive. They are difficult to conduct for long follow-up periods. In the context of dynamic decision making, the treatment protocols specified in RCTs may not yield results generalizable to community settings. This is especially the case when there are many different outcomes occurring over long time horizons and numerous intermediate outcomes that might require additional interventions. These are all key issues when one studies diabetes. Moreover, the sample size of an RCT would have to be very large and the follow-up period very lengthy for it to have sufficient statistical power and time to measure many relevant relationships.

*Manuscript received March 2014; revised August 2014.

[1] We thank seminar participants at the University of Florida, University of Alabama Birmingham, Elon University, Twentieth European Workshop on Econometrics and Health Economics, the 2011 Annual Health Econometrics Workshop, two anonymous referees, and the editor of this journal for valuable comments. We also thank the National Institute of Aging (Grants R01-AG017473, and R01-HD047213) for their generous support and the use of the services provided by Research Computing at the University of South Florida. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

Please address correspondence to: Gabriel Picone, Department of Economics, University of S. Florida, 4202 E. Fowler Ave., CMC 207, Tampa, FL 33620. E-mail: *gpicone@usf.edu*.

A potentially important alternative is to use observational data like we do in this study. Large, longitudinal, administrative data sets are becoming increasingly available to researchers, but their effective use requires one to address directly how unobserved heterogeneity impacts both observed treatment choices and outcomes over time. Econometric solutions for dealing with such issues in analyzing observational data have been developed, including for dynamic problems like those examined here, but there is still much progress to be made. In this study, we incorporate many of these advances and add in a key component that is relevant for studying diabetes and many other diseases.

Diabetes mellitus is a complex chronic disease. It can affect eyesight, kidneys, cardiovascular systems, and nervous systems affecting the lower extremities, i.e., legs and feet. The incidence of diabetes is increasing in the United States and elsewhere, reflecting increased obesity of the population in part. Prevalence is high among the elderly (Sloan et al., 2008).

The disease progresses through several stages, each with increasingly debilitating consequences. Once an individual reaches a more advanced diabetes stage it is impossible to undo the physiological damage; diabetes complications can eventually lead to death. The early stages of the disease are typically nonsymptomatic, but without interventions, irreversible physiological damage will continue to accrue. Like many other chronic illnesses, diabetes can be much more costly to treat if detected later. Regular screening for diabetes potentially can help the patient and her physician recognize when it is appropriate to undertake behavioral modifications and start medications and other therapies to slow the disease's progression.

Screening is costly, and the optimal screening regimen depends on a comparison of the marginal benefit and the marginal cost of screening. The marginal benefit from more frequent screening depends on the probability that screening will reveal useful information and the value of this information in slowing the disease's progression. This is the primary focus of this article. This study uses a dynamic multistage discrete duration model to investigate the effectiveness of early detection of diabetes mellitus through screening in delaying the progression of complications and death.

An evaluation of screening for diabetes encounters at least four econometric issues. First, ascertainment of diagnosis in particular and care more generally is endogenous. Second is the importance of partial observability of the disease state; the person could have the disease for a long time without being diagnosed. Third, since individuals and diseases differ in aspects unobservable to the researcher, there is likely unobserved heterogeneity. Fourth, the probability of adverse outcomes increases with duration and progression of the disease.

Our estimation strategy deals with each of these four econometric problems. We address endogeneity and unmeasured heterogeneity issues by using discrete factor models (Heckman and Singer, 1984; Mroz, 1999). This approach has been used previously by Picone et al. (2003), Glewwe and Jacoby (2004), Bhattacharya (2005), Mroz and Savage (2006), and Liu et al. (2010), among others. We simultaneously account for multiple disease stages, partial observability of disease progression, endogeneity of the timing of diagnoses, and health outcomes. We control for partial observability by modeling empirically all potential exact times of disease (or stage) onset and integrating over all these potential onset times. The bounds for these integrations come from the last time period a person was known to not have the disease and the first time the individual was known to have the disease.

These time periods during which we are uncertain about precisely when the individual progressed to the next disease stage constitute a key feature of this analysis. Not only is this an econometric issue to be addressed, it is a real, substantive issue for analyzing disease progression and treatments. Many individuals will not recognize that they have progressed to diabetes or to more advanced stages if they do not see a health-care professional who can diagnose their condition. If the period of time during which the disease is present but unobserved and untreated is long, then the individual may progress much more rapidly to more severe disease stages, possibly resulting in amputation or death.

We find that earlier diagnosis of diabetes, and presumably the treatments that follow diagnosis, delays the onset of lower extremity complications (LECs) including amputation. For

TABLE 1
DISEASE PROGRESSION

Stage	Progression	Condition	ICD9-CM Code
1	Healthy		
2	Diabetes mellitus only (D_1)	Diabetes mellitus only	250.xx
3	Low severity lower extremity complication	Neuropathy	250.6 357.2 355.xx
		Paresthesia	782.xx
	(D_2)	Pain in feet	729.5
		Diabetic amyotrophy	358.1
4	Medium severity lower extremity complication	Cellulitis	681.1 682.6 682.7
	(D_3)	Charcot foot	707.1
5	High severity lower extremity complication	Osteomyelitis	730.06 730.07 730.16 730.17 730.26 730.27
	(D_4)	Gangrene	250.7 785.4

example, a one-year delay in the diagnosis of diabetes increases the probability that five years later she will have an LEC (or worse) by 11% (6.6% points out of a baseline of 59%). Transitions to high severity LECs, or worse, within five years would increase by 27%. However, because the number of individuals who transition to severe stages is relatively low, the total number of additional high severity LEC cases is correspondingly low. Our parameter estimates also allow us to conduct counterfactual analysis and policy simulations. For example, if Medicare were to cover no more than two visits per year for healthy individuals, over a 15-year span it would save Medicare $476 per beneficiary but would cost about 0.004 years of life per person. These would imply an implicit value of a year of life of about $119,000.

The rest of the article is organized as follows: Section 2 provides a background on diabetes and the decision process to screen. Section 3 describes the econometric strategy and Section 4 the data. Section 5 presents our empirical specification, including the method for accounting for endogeneity of treatments, which is followed by results in Section 6. Section 7 discusses the marginal effects and policy simulations. Finally, Section 8 presents conclusions and implications.

2. BACKGROUND

Diabetes mellitus is a complex disease, potentially adversely affecting several organ systems, including the eyes, the cardiovascular system, the kidneys, and legs and feet. It reduces life expectancy. Complications of the legs and feet, LECs, are common. They may be classified into stages, with each successive stage being increasingly more severe. In this study, we define five mutually exclusive stages (Table 1): no diabetes or "healthy" and four diabetes stages. The disease stages are: (i) diabetes only, no LEC (D_1); (ii) low LEC (D_2); (iii) medium LEC (D_3); and (iv) high LEC (D_4). Persons in D_1 may have no symptoms or at most mild symptoms, such as thirst which could reflect other factors. In D_2, the person has mild complications of the legs or feet, e.g., loss of feeling in the feet, which might be treated by prescribing special shoes to prevent foot injuries. In D_3, the person experiences possibly severe infections, bruises on a leg or swelling and/or pain in a foot. D_4 includes gangrene or infections of the bone or bone marrow, which if untreated can lead to amputation or even death.

Each period the individual has the option of visiting a physician who diagnoses whether or not she has diabetes and the stage of the disease from D_1 to D_4. Early diagnosis of D_1 as well as later stages leads to earlier treatment. This potentially can slow down the disease progression and adverse health shocks. If these screening visits were costless, the individual would visit a physician every period and the disease would be perfectly monitored. In practice very few individuals visit a physician each period for screening, since it is costly in terms of time and out-of-pocket medical or other expenditures (e.g., for transportation). Several factors affect the individual's decision to have a screening visit including

time since last visit and whether she has disease symptoms. Writing, solving, and estimating a complete dynamic optimization model is outside the scope of our study. Instead we model the individual's decision to have a screening visit at each point in time and examine how this interacts with disease progression. This is a function of observed and unobserved disease characteristics and observed individual and location-specific characteristics as well as unobserved heterogeneity.

Screening visits serve two purposes. First, they assess whether or not the individual's diabetes' status has progressed since the previous screening visit. Second, if the disease has progressed, the patient and her physician use this information to adjust the sequence of treatments. These sequences of treatment choices affect the patient's probabilities of progressing to more advanced disease stages. The actual set of treatments chosen by the patient, however, is not observed in our data set. As a consequence, we assume that the impacts of these treatment decisions can be captured by the measures of each disease stage's durations in the hazard functions of moving to more severe, higher disease stages. Provided that the distributions of the future disease stages and exogenous variables are captured by the current period's variables (a first-order Markov assumption),[2] our formulations approximate the optimal sequence of screening visits and treatment decision rules chosen by the patient and her health-care provider.

3. A DISCRETE-TIME MODEL OF A CHRONIC DISEASE, SCREENING VISITS, AND HEALTH SHOCKS

3.1. *Setup.* The Medicare claims data used in this study record information about physician visits and diagnoses for individuals after they enroll in Medicare at age 65. In addition, before we observe the first post-age 65 screening visit we lack information about the individual's health, including presence or absence of chronic diseases such as diabetes. For this reason, we must account for left truncation of the data.

We split time into discrete time intervals. We allow for three arbitrary time periods before one becomes 65 ($t = 1, 2, 3$), and we use these time intervals to model individuals' initially observed diabetes states. As the model relies on the time spent with diabetes as a determinant of disease progression, we arbitrarily set each of these three periods to be five years in length. Once one turns 65, time periods are for a quarter ($t = 4, 5, \ldots$). During any of the first three time periods, an individual may have diabetes (D_1) and also may transition to low severity extremity complications (D_2), but it is not until the person's first post-age 65 screening visit that we as researchers can observe her diabetes status.[3]

Starting in period four and until the time of a first screening visit ($t_{FV} \geq 4$), we model potential disease progression through its initial two stages. After the first screening visit ($t > t_{FV}$), we model the quarterly progression of diabetes through all of its stages and the timing of screening visits in all subsequent time periods. Observations are right censored at time T when the person dies or leaves the sample for some other reason. We assume censoring not due to death is ignorable.

In total, there are 12 different equations in our likelihood function: two pre-age 65 hazards (for D_1 and D_2), two post-age 65 before first visit hazards (for D_1 and D_2), four post-age 65 after first doctor visit hazards (for D_1, D_2, D_3, D_4), two visit equations (first visit after age 65 and subsequent visits), and two health outcomes equations (amputation and death). Below we describe the likelihood function and our parametric assumptions.

[2] That is, in a stochastic dynamic optimization model the expected value of the future value function conditional on any set of choices made today can be perfectly forecast by current period variables. See, for example, Blundell et al. (1997) and Liu et al. (2009).

[3] We exclude from the analysis any individual who had already advanced to disease stage 3 (D_3) by the time of the first observed screening visit after becoming 65. In preliminary analyses, we used having reached stage 3 diabetes or worse by the time of the first post-age 65 visit as an endogenous selection mechanism. This had little effect on the parameter estimates for visits and disease progression after age 65, but it greatly increased the computational burden. Thus, we simplified the model by eliminating these severely ill individuals and focus on individuals whose length of time in the various diabetes stages are more precisely measured. Persons with stage 3 diabetes have severe deformities (Charcot foot and/or cellulitis) and are likely to have relatively continuous medical care.

3.2. *Functional Form Specifications.*

3.2.1. *Disease progression.* At a particular time period, an individual can be either healthy or in one of the four diabetes stages $(D_1, \ldots, D_4)$. Once a person enters a more severe disease stage, it is impossible to return to a less severe one. Let the hazard function for the progression to D_1 at time $t > t_{FV} \geq 4$ be

$$h_t^{D_1}\left(D_{1t} = 1 | X_t^{D_1}, e_k, D_{1t-1} = 0\right) = \Lambda\left(X_t^{D_1}\beta_{D_1} + \rho_{D_1}(e_k)\right), \tag{1}$$

where $\Lambda(z)$ is the logit function $e^z/(1+e^z)$, $X_t^{D_1}$ is a vector of time-invariant and time-varying explanatory variables, e_k is the unmeasured heterogeneity assumed to be discrete with K heterogeneity points, and $\rho_{D_1}(.)$ is a polynomial of degree $J \leq K-1$.

Let t_{D_j} be the time at which the individual progresses to D_j $(j = 1, \ldots, 4)$. The hazard functions for D_j $(j = 2, \ldots, 4)$ at time $t \geq t_{D_{j-1}}$ are given by

(2)

$$h_t^{D_j}\left(D_{jt} = 1 | t^{D_1}, \ldots, t^{D_{j-1}}, X_t^{D_j}, e_k, D_{jt-1} = 0\right) = \Lambda\left(X_t^{D_j}\beta_{D_j} + \sum_{l=2}^{j}\delta_{l-1,j}\left(t^{D_{l-1}}\right) + \rho_{D_j}(e_k)\right),$$

where $t^{D_{l-1}} = t - t_{D_{l-1}}$ is the length of time since the D_{l-1} onset and $\delta_{l-1,j}(.)$ is a quadratic function for $l = 2, \ldots, j$ and $j = 2, \ldots, 4$. Our specification allows for the hazard functions to depend on duration in each of the previous stages, and it allows for different sets of regressors depending on the stage. The two pre-age 65 diabetes onset hazards (D_1^{P65}, D_2^{P65}) and the two post-age 65 before first visit hazard functions (D_1^{FV}, D_2^{FV}) are

$$\Lambda\left(X_t^{D_1^{P65}}\beta_{D_1^{P65}} + \rho_{D_1^{P65}}(e_k)\right), \quad \Lambda\left(X_t^{D_2^{P65}}\beta_{D_2^{P65}} + \delta_{1,2}^{P65}\left(t^{D_1}\right) + \rho_{D_2^{P65}}(e_k)\right)$$

and

$$\Lambda\left(X_t^{D_1^{FV}}\beta_{D_1^{FV}} + \rho_{D_1^{FV}}(e_k)\right), \quad \Lambda\left(X_t^{D_2^{FV}}\beta_{D_2^{FV}} + \delta_{1,2}^{FV}\left(t^{D_1}\right) + \rho_{D_2^{FV}}(e_k)\right).$$

3.2.2. *Visits and health shocks.* The probability of a screening visit in period t $(t = t_{FV} + 1, \ldots, T)$ depends on the current stage D_j $(j = 1, \ldots, 4)$:

$$\Pr\left(V_t = 1 | t^{D_1}, \ldots, t^{D_j}, Z_t^V, e_k\right) = \Lambda\left(Z_t^V\beta_V + \sum_{l=1}^{j}\alpha_l^V\left(t^{D_l}\right) + \rho_V(e_k)\right), \tag{3}$$

where $\alpha_l^V(t^{D_l})$ is a quadratic function with an intercept. We restrict β_V and $\rho_V(e_k)$ to be the same irrespective of disease stage, but allow for different intercepts and different duration coefficients for each stage. For a healthy individual, this probability is $\Lambda(Z_t^V\beta_V + \rho_V(e_k))$, and for an individual that has only progressed to D_1, this probability is given by $\Lambda(Z_t^V\beta_V + \alpha_1^V(t^{D_1}) + \rho_V(e_k))$. Among the regressors are variables affecting the probability of a visit but otherwise not affecting health or diabetes progression (exclusion restrictions). We use distances to the nearest health providers as exclusion restrictions. The specification of the probability of the first visit is identical to Equation (3), but we do not restrict β_V and $\rho_V(e_k)$ to be the same as in Equation (3).

Finally, let d_1 and d_2 be two observable health outcome shocks, amputation and death. As the disease progresses, the hazards of these shocks occurring are likely to increase. For an individual

at stage D_j, the hazard of d_{ct} during each time period is

$$\Pr\left(d_{ct}=1|t^{D_1},\ldots,t^{D_j},W_t^{cd},e_k\right)=\Lambda\left(W_t^{cd}\beta_{cd}+\sum_{l=1}^{j}\alpha_l^{cd}\left(t^{D_l}\right)+\rho_{cd}\left(e_k\right)\right) \tag{4}$$

for $c=1,2$. As with visits we restrict the coefficients (β_{cd}) to be the same irrespective of the disease stage, but allow for different intercepts and different duration coefficients for each disease stage the person entered by date t.

3.2.3. *Unmeasured heterogeneity.* Unmeasured heterogeneity, e.g., overall unmeasured health, affects the hazard of progressions to higher disease stages, the probabilities of a screening visit, amputation, and death. For each of the events, we assume a discrete heterogeneity distribution which we model as a polynomial

$$\rho_q\left(e_k\right)=\rho_{q1}\left(\frac{k-1}{K-1}\right)+\cdots+\rho_{qJ}\left(\frac{k-1}{K-1}\right)^J, \tag{5}$$

where $e_k=\frac{k-1}{K-1}$, K is the number of heterogeneity points, $k=1,\ldots,K$, and $J\leq K-1$ is the degree of the polynomial. The index q refers to the processes (equations) we describe above. In particular, there are 12 equations of interest, namely: onset of diabetes before age 65 (D_1^{P65}), onset of low LEC complications before age 65 (D_2^{P65}), onset of diabetes after age 65 but before the first doctor visit (D_1^{FV}), onset of low LEC after age 65 but before the first doctor visit (D_2^{FV}), onset of diabetes after the first doctor visit (D_1), onset of low LEC after the first doctor visit (D_2), onset of medium LEC after the first doctor visit (D_3), onset of high LEC after the first doctor visit (D_4), first doctor visit (FV), subsequent doctor visits (V), death ($1d$), and amputation ($2d$). We estimate the 12 polynomials and the probabilities of each point of support, the $\Pr(e_k)$, subject to the restrictions that each probability is nonnegative and $\sum_{k=1}^{K}\Pr(e_k)=1$. Here, we report estimates from a model with $K=8$ points of support and $J=3$.[4]

Although the eight underlying heterogeneity terms, the e_k's, are equally spaced on the unit interval, the polynomials given by Equation (5) map each e_k to a point in the real line for each equation. If each of these 12 polynomials had degree 7 ($J=K-1$), then this would correspond to an unrestricted, discrete 12-variate error structure with eight points of support.[5] Choosing a polynomial with a lower degree does impose some structure on the correlations of the multivariate distribution, but it also significantly reduces the number of parameters to estimate. We use third-degree polynomials. This choice allows for significant flexibility in the dependence of error terms across equations while reducing the number of parameters to estimate by 48.[6] This polynomial specification is an extension of the formulation proposed in Mroz (1999) to model unmeasured heterogeneity.[7] Mroz (1999) uses a single nonparametric function of k instead of a polynomial in k and it constrains the cross equation correlations to be linear functions of the unobserved nonparametric function of k. Our polynomial specification has the advantage that it allows for nonproportional relationships among the heterogeneity terms across equations.

[4] We also estimated simpler models ranging from no heterogeneity to $K=4$ and $J=3$. We report results using $K=8$ and $J=3$ because this model fits the data better than a model with four "types" (i.e., $K=4$ and $J=3$). Results from the other models are available upon request.

[5] Our specification corresponds to a simplified version of an eight "types" model. To reduce the number of parameters to estimate, we allow the intercepts terms to vary by type but assume all other coefficients are constant across types.

[6] The unrestricted polynomials have 84 parameters (12×7), but the restricted polynomials that we use have 36 parameters (12×3).

[7] Several papers have used similar specifications for the unmeasured heterogeneity, including Hoynes (1996), Blau and Gilleskie (2001), Mello et al. (2002), Bhattacharya et al. (2003), Glewwe and Jacoby (2004), and Liu et al. (2010).

3.3. *The Likelihood Function.* Based on the hazards and probabilities (Equations (1)–(4)), the likelihood function for an individual with any possible transition combination conditional on unmeasured heterogeneity e_k and the matrix of all possible explanatory variables $\mathbf{M} = (\overrightarrow{\mathbf{X}}^{D_1^{P65}}, \overrightarrow{\mathbf{X}}^{D_2^{P65}}, \overrightarrow{\mathbf{X}}^{D_1^{FV}}, \overrightarrow{\mathbf{X}}^{D_2^{FV}}, \overrightarrow{\mathbf{X}}^{D_1}, \ldots, \overrightarrow{\mathbf{X}}^{D_4}, \overrightarrow{\mathbf{Z}}^{FV}, \overrightarrow{\mathbf{Z}}^{V}, \overrightarrow{\mathbf{W}}^{1d}, \overrightarrow{\mathbf{W}}^{2d})$[8] is

$$
\begin{aligned}
L\left(t_{D_1}, \ldots, t_{D_4}, T, \overrightarrow{V}, \overrightarrow{d}_1, \overrightarrow{d}_2 | e_k, \mathbf{M}\right) = {} & L_H\left(T, \overrightarrow{V}, \overrightarrow{d}_1, \overrightarrow{d}_2 | e_k, \mathbf{M}\right)^{1\left(t_{D_1} \geq T\right)} \\
& \times L_{D_1}\left(t_{D_1}, T, \overrightarrow{V}, \overrightarrow{d}_1, \overrightarrow{d}_2 | e_k, \mathbf{M}\right)^{1\left(t_{D_2} \geq T > t_{D_1}\right)} \\
& \times L_{D_2}\left(t_{D_1}, t_{D_2}, T, \overrightarrow{V}, \overrightarrow{d}_1, \overrightarrow{d}_2 | e_k, \mathbf{M}\right)^{1\left(t_{D_3} \geq T > t_{D_2}\right)} \\
& \times L_{D_3}\left(t_{D_1}, t_{D_2}, t_{D_3}, T, \overrightarrow{V}, \overrightarrow{d}_1, \overrightarrow{d}_2 | e_k, \mathbf{M}\right)^{1\left(t_{D_4} \geq T > t_{D_3}\right)} \\
& \times L_{D_4}\left(t_{D_1}, \ldots, t_{D_4}, T, \overrightarrow{V}, \overrightarrow{d}_1, \overrightarrow{d}_2 | e_k, \mathbf{M}\right)^{1(T > t_{D4})}, \qquad (6)
\end{aligned}
$$

where $L_H(T, \overrightarrow{V}, \overrightarrow{d}_1, \overrightarrow{d}_2 | e_k, \mathbf{M})$ is the likelihood function for an individual who did not progress to D_1 by period T with a sequence of visits $\overrightarrow{V} = (V_4, \ldots, V_T)$ and health shocks $\overrightarrow{d}_c = (d_{cFV}, \ldots, d_{cT})$ for $c = 1, 2$. $L_{D_1}(.)$ is the likelihood function for an individual who entered stage D_1 at t_{D_1}, but did not progress to D_2 by terminal period T. Similarly, we define $L_{D_2}(.)$, $L_{D_3}(.)$, and $L_{D_4}(.)$. These individual likelihood functions depend on the hazard functions for disease progression, the probabilities of visits, and the probabilities of health shocks. For example $L_{D_4}(.)$ is given by

$$
\begin{aligned}
L_{D_4}\left(t_{D_1}, \ldots, t_{D_4}, T, \overrightarrow{V}, \overrightarrow{d}_1, \overrightarrow{d}_2 | e_k, \mathbf{M}\right) = {} & \left[\prod_{j=1}^{4} \Pr\left(D_j = t_{D_j} | .\right)\right] \times \Pr\left(\mathbf{V}_4^{t_{FV}} | .\right) \times \Pr\left(\mathbf{V}_{t_{FV}+1}^{T} | .\right) \\
& \times \Pr\left(\mathbf{d}_{1,t_{FV}}^{T} | .\right) \times \Pr\left(\mathbf{d}_{2,t_{FV}}^{T} | .\right),
\end{aligned}
$$

where $\Pr(D_1 = t_{D_1} | .)$ is the probability of an individual contracting D_1 at period t_{D_1} and $\Pr(D_j = t_{D_j} | .)$ for $j = 2, \ldots, 4$ is the probability of an individual contracting D_j at period t_{D_j}. $\Pr(\mathbf{V}_4^{t_{FV}} | .)$ is the probability of having the first visit at $t = t_{FV}$. $\Pr(\mathbf{V}_{t_{FV}+1}^{T} | .)$ is the probability of the sequence of visits from $t = t_{FV} + 1$ to $t = T$. $\Pr(\mathbf{d}_{c,t_{FV}}^{T} | .)$ for $c = 1, 2$ is the probability of the sequence of outcomes d_{ct} from $t = t_{FV}$ to $t = T$.[9] The online Appendix contains details on the construction of $L_H(.)$, $L_{D_1}(.)$, $L_{D_2}(.)$, $L_{D_3}(.)$, and $L_{D_4}(.)$.

3.3.1. *Partial observability and early detection of a disease.* To avoid the unrealistic assumption that the individual's stage is continuously monitored at each t (i.e., the individual visits a physician every period), we incorporate partial observability by the individual and her physician of the disease stage into the likelihood function. We do this by integrating over the possible time spans during which an individual is known from our data to have progressed to a higher disease stage.[10] In particular, we assume that the first date at which a claim reports a diabetes stage is

[8] $\overrightarrow{\mathbf{X}}^{D_j^{P65}} = (X_1^{D_j^{P65}}, X_2^{D_j^{P65}}, X_3^{D_j^{P65}})$ and $\overrightarrow{\mathbf{X}}^{D_j^{FV}} = (X_4^{D_j^{FV}}, \ldots, X_{FV}^{D_j^{FV}})$ for $j = 1, 2$. $\overrightarrow{\mathbf{X}}^{D_j} = (X_{FV+1}^{D_j}, \ldots, X_T^{D_j})$ for $j = 1, \ldots, 4$. $\overrightarrow{\mathbf{Z}}^{FV} = (Z_4^{FV}, \ldots, Z_{FV}^{FV})$, $\overrightarrow{\mathbf{Z}}^{V} = (Z_{FV+1}^{V}, \ldots, Z_T^{V})$, and $\overrightarrow{\mathbf{W}}^{cd} = (W_{FV}^{cd}, \ldots, W_T^{cd})$ for $c = 1, 2$.

[9] To save notation, we do not explicitly state the conditioning set at each date t. At each point in time we condition on the unobserved heterogeneity and the subset of elements in $\mathbf{M}$ that are relevant to the possible outcomes at that time period.

[10] This is similar to the strategy used by Mroz and Weir (1990) to address the partial observability of lactational amenorrhea in their life-cycle model of fertility control.

the latest period ($t_{\max j}$) a person could enter that stage. The date of the visit immediately prior to that is assumed to be the latest date that we observe the person without the diabetes stage. The period immediately following this date is the earliest date ($t_{\min j}$) a person could enter that stage. The actual (unobserved) date at which the person enters the stage is in the range

$$t_{\min j} \leq t_j \leq t_{\max j}$$

for $j = D_1, \ldots, D_4$. Conditional on e_k, the individual likelihood function for the observed series $(t_{\min D_1}, t_{\max D_1}, \ldots, t_{\min D_4}, t_{\max D_4}, T)$ is obtained by integrating over all possible starting and ending values of $t_{D_1}, t_{D_2}, t_{D_3}$, and t_{D_4}:

$$L\left(t_{\min,D_1}, \ldots, t_{\max,D_4}, T, \vec{V}, \vec{d} \,|e_k, \mathbf{M}\right)$$

$$= \sum_{t_{D_1}=t_{\min D_1}}^{\min\{t_{\max D_1}, T\}} \Bigg[\sum_{\substack{t_{D_2}=\max\{t_{D_1}, t_{\min D_2}\} \\ t_{D_2} \geq t_{D_1}}}^{\min\{t_{\max,D_2}, T\}} \Bigg[\sum_{\substack{t_{D_3}=\max\{t_{D_2}, t_{\min D_3}\} \\ t_{D_3} \geq t_{D_2}}}^{\min\{t_{\max D_3}, T\}} \Bigg[\sum_{\substack{t_{D_4}=\max\{t_{D_3}, t_{\min D_4}\} \\ t_{D_4} \geq t_{D_3}}}^{\min\{t_{\max D_4}, T\}}$$

$$L\left(t_{D_1}, \ldots, t_{D_4}, T, \vec{V}, \vec{d} \,|e_k, \mathbf{M}\right)\Bigg]\Bigg]\Bigg]. \tag{7}$$

Uncertainty about precisely when the individual progressed to the next stage is a key feature of this analysis, not just an econometric issue to be addressed. If the time during which the disease is present but unobserved and untreated is long, then the individual may progress to more severe disease stages much more rapidly.

An advantage of our model is that we can model the length of time a person spends with any stage of the disease without it having been diagnosed. We define t_{D_j} as the time at which the individual progresses to diabetes stage D_j, so $t^{D_j} = t - t_{D_j}$ is the length of time of stage D_j that the individual has been in that state since its onset. We calculate the time spent with undiagnosed stage D_j as $t^{D_j^u} = \min\left\{t^{D_j}, t_{\max D_j} - t_{D_j}\right\}$. Delay in time to diagnosis of diabetes and its complications has a potentially permanent effect on the person's health and longevity because of the consequent delay in treatment. These $t^{D_j^u}$s are our main explanatory variables. They allow us to ascertain whether or not more frequent visits, and the resulting earlier diagnosis, could reduce disease progression.

Without modeling explicitly the process of disease stage acquisition and $t^{D_j^u}$s in an environment with partial observability, it is not possible to separate the true causal effects of early diagnosis from lead time biases. These types of bias arise because more frequent screenings mechanically diagnose disease stages at shorter true duration times on average, and this leads to a spurious relationship between the frequency of screening and longer waiting times to the future adverse events. Of course, the length of time with an undiagnosed disease stage depends on the unobserved time of entering the stage, but that is captured in Equation (7).

3.3.2. *Estimation.* Finally, the unconditional log-likelihood function is

$$L(\theta) = \sum_{i=1}^{N} \ln\left(\sum_{k=1}^{K} \Pr(e_k)\, L\left(t_{\min D_{1i}}, \ldots, t_{\max D_{4i}}, T_i, \vec{V}_i, \vec{d}_i|\theta, e_k, \mathbf{M}_i\right)\right), \tag{8}$$

where θ is the vector of parameters and $\Pr(e_k)$ is the probability of the discrete heterogeneity point e_k.[11] We maximize this likelihood function using GQOPT with respect to all of the

[11] We wrote a FORTRAN program to estimate this likelihood function, which is available from the corresponding author upon request.

parameters in θ in the 12 outcome equations. We estimate the likelihood with different numbers of heterogeneity points (K), including $K = 1$ (no heterogeneity and no correlation among the equations). We report results for eight points of support ($K = 8$) and third-degree polynomials in the underlying heterogeneity factor ($J = 3$).

4. DATA

We use data from interviews conducted for the National Long-Term Care Survey (NLTCS), a longitudinal study of elderly persons. The screening process began with a random sample of persons aged 65 and older in 1982. Respondents were tracked in five-year intervals (1989, 1994, 1999, and 2004) with additional persons added to the sample in later waves. Over the five NLTCS cohorts, more than 40,000 individuals were interviewed. We only use NLTCS data from the 1994, 1999, and 2004 interviews, because Medicare claims data with the necessary information on diagnosis are only available since 1991. The NLTCS interviews provide information on the sample person's demographic characteristics, date of birth, gender, race, marital status, and years of schooling.

An advantage of NLTCS is that Medicare claims data, both for Part A (services provided by institutions such as hospitals) and Part B (services provided by physicians and other health professionals) have been merged with NLTCS. These claims data provide information on diagnoses and procedures performed by date of service. Furthermore, NLTCS respondents were merged with National Death Index (NDI) data, providing the respondent's death dates through 2004.

Using NLTCS interviews for 1994, 1999, and 2004, Medicare claims data for 1991–2004, and NDI data through 2004, we create a panel with the individual Medicare beneficiary by quarter as the observational unit. We select individuals born between 1926 and 1939. This restriction ensures that all individuals were at least age 65 when they enter the sample. We also drop individuals with less than two years of data and individuals who had progressed to medium severity LEC (D_3) or higher by the time of the first visit observed in the data. Our final sample size consists of 9,417 individuals observed over a total of 261,916 quarters (Panel A, Table 2), the time periods used in this study. A quarter may be sufficient for some individuals to notice disease symptoms and visit a doctor. Moreover, visits more frequent than once in a quarter are likely to be predominantly follow-up visits.

5. EMPIRICAL SPECIFICATION

There are 12 different equations in the likelihood function: two pre-age 65 hazards (for D_1 and D_2), two post-age 65-before first visit hazards (for D_1 and D_2), four post-age 65-after first doctor visit hazards (for D_1, D_2, D_3, D_4), two visit equations (first visit after age 65 and subsequent visits), and two health outcomes equations (amputation and death), respectively.

5.1. *Dependent Variables.* **Stages**: A person is in one or more of five mutually exclusive stages during a quarter. When a person transitions to a higher stage in a quarter, the person is assumed to have been in the higher stage throughout the quarter.

We assume that once a person transitions to a higher stage she cannot return to a lower stage. Furthermore, a person in a particular stage has experienced all prior stages in the past, albeit at times unknown to us, unless explicitly documented in prior claims.

Visits: The role of a visit is to diagnose whether the individual has the disease (D_1) and any of its progression stages (D_2, D_3, and D_4). We assume that the following specialties screen for diabetes (with Medicare provider type codes in parentheses): general practice (01), cardiology (06), family practice (08), internal medicine (11), endocrinology (46), and clinical laboratory (69). Once a person progresses to diabetes only (D_1), we add podiatry (48) to the list of specialities. A podiatrist is a nonphysician who specializes in diseases of the lower extremities.

TABLE 2
SUMMARY STATISTICS

Panel A: Dependent Variable

	Progression		Outcomes		Screening Visits per Quarter (%)
	Individuals	Quarters	Death (%)	Amputation (%)	
Overall			7.8	1.1	54.2
Stages (%)					
Healthy	66.1	77.5	6.4	0.6	47.8
Diabetes only (D_1)	10.7	10.6	6.2	0.3	67.9
Low severity LEC (D_2)	13.9	7.2	9.9	0.5	81.3
Medium severity LEC (D_3)	5.3	2.8	14.7	1.4	85.6
High Severity LEC (D_4)	4.1	1.9	18.4	13.1	88.9
Sample size	9417	261,916			

For individuals, stage is at last quarter observed in data; for quarters, stage is at last quarter observed in data except for visits.

Panel B: Main Explanatory Variable (Undiagnosed Disease)

Disease Progression	No. of Individuals	Mean Quarter of Diagnosed	Range of Undiagnosed Disease		
			Mean	Percent of Zeros	Maximum
Diabetes only (D_1)	3187	16.24 (12.52)	3.21 (5.41)	41.5	55
Low severity LEC (D_2)	2183	22.53 (12.32)	0.89 (2.83)	72.7	42
Medium severity LEC (D_3)	878	24.74 (12.74)	0.64 (2.32)	78.9	32
High Severity LEC (D_4)	381	25.10 (12.66)	0.41 (1.72)	85.3	20

Panel C: Time-Invariant Explanatory Variables

Variable	Mean
Arthritis (%)	1.67
White (%)	87.97
Male (%)	45.16
Married (%)	57.38
High school + (%)	48.83
Year turned 65	0.67
(1991 = 0.1)	(0.36)
Mean Distance GP	0.11
(10 miles)	(0.16)
Mean Distance Podiatrist	0.55
(10 miles)	(0.84)

Diagnoses on claims to other types of health professionals are less likely to contain information on diabetes, even if this disease is present.

Outcomes: In each period, the individual may experience two types of adverse health outcomes: (i) a person's toe, foot, or leg is amputated[12] and (ii) death. Other diabetes-related health shocks, e.g., heart attacks and strokes, were not explicitly modeled because they do not affect higher LEC transitions.

[12] Since two or more amputations are rare events in our data set, we only model time to the occurrence of the first amputation.

5.2. *Explanatory Variables.* Explanatory variables fall into four categories: (i) early diagnosis (our main explanatory variable); (ii) duration dependence; (iii) demographic variables; and (iv) exclusion restrictions.

Early Diagnosis: The effects of early diagnosis are measured using the time with undiagnosed D_1, D_2, D_3, and D_4 and their squares. We defined these variables in Section 3. These variables are different from the duration of the time already spent in the four disease stages after the stage is ascertained by screening. If early diagnosis is beneficial, then one would expect undiagnosed duration to increase the probability of progression to the next stages and possibly increase mortality and amputation probabilities. Given partial observability of diabetes stages, the effects of these times with undiagnosed disease depend on the integration implicit in Equation (7).[13] We assume that a visit in which a diagnosis is made leads to treatment. Thus, earlier diagnosis leads to earlier treatment.

Duration Dependence: Duration dependence is measured by a quadratic function of the time in quarters from the period the individual enters a particular stage in Equations (2) and (3). In addition, we allow for a discrete shift in the hazard and probability arguments at entry to any stage. All duration dependence terms can vary independently for each disease stage.

Demographic and Health Characteristics: We include binary variables in all equations for gender, educational attainment, marital status, arthritis, and race.[14] We expect more highly educated and married persons to have better health outcomes. To control for age and cohort effects, we also include a year trend and its square and the year in which the individual turned age 65. Age and generational changes in diet, for example, might affect diabetes outcomes.

Exclusion Restrictions: We use distances to the nearest health providers as exclusion restrictions affecting the probabilities of visits but not directly influencing disease progression or health shocks. The NLTCS only provides area of residence information at the level of the primary sampling unit (PSU), which is a Standard Metropolitan Area for persons living in such areas and a rural area of a state for others. For each PSU, we use a random sample of Medicare beneficiaries (from the 5% Medicare sample) to calculate the average distance from the centroid of the zip code of residence in the PSU to the centroid of the zip code of nearest provider based on the specialty code. We expect an increase in the minimum distance to be negatively related to the probabilities of visits but not to affect disease progression or health shocks after controlling for visits.

6. RESULTS

Of the 9417 persons in our sample (column 1, Panel A, Table 2), 66.1% were never diagnosed with diabetes during the observational period, 10.7% had a diabetes diagnosis without an LEC (D_1), whereas 13.9%, 5.3%, and 4.1% were observed to have progressed to D_2, D_3, and D_4, respectively. Of the 261,916 quarter/person observations in our sample (column 2), 77.5% are healthy,[15] 10.6% are in state D_1, 7.2% in D_2, 2.8% in D_3, and 1.9% in D_4. Overall, 7.8% of individuals die during the observational period; we observe 27.8 quarters per person on average. Mortality is 6.4% for healthy individuals and increases to 14.7% and 18.4% for those in D_3 and D_4 states, respectively. Amputations spike upward at stage D_4, with 13.1% of those entering this stage having at least one lower extremity amputation; but amputation is very rare before this stage. The number of quarters with a visit rises by stage. Healthy individuals visit a physician in 47.8% of quarters, but once diagnosed with low severity LEC (D_2), this probability increases to over 80%.

Table 2 (Panel B) describes results for our main explanatory variable, time with undiagnosed diabetes, and its lower extremity complications. The mean quarter at which diabetes is diagnosed

[13] Mroz and Weir (1990) discuss identification of the distribution governing a partially observed process for a simpler model than that analyzed here.

[14] These variables were obtained from the NLTCS screener file using the latest available year.

[15] More precisely, 77.5% of the person/quarters take place before the first diagnosis of diabetes.

is 16.2 (slightly over four years). The mean potential range for undiagnosed diabetes is 3.2 quarters (almost 10 months), with 41.5% of our sample having a range of zero and a maximum of 55 quarters. For D_2, D_3, and D_4, the mean ranges of potential undiagnosed disease are much smaller, and the percentage of zero quarters with undiagnosed LEC is much larger (columns 2–4). This is because once a person is diagnosed with diabetes, she tends to have more frequent visits. Finally, Table 2 (Panel C) describes our time-invariant explanatory variables. Most persons are White (88%) and about half are male, married, and have at least finished high school. The mean minimum distance in miles to the nearest provider is 1.1 for visits overall and 5.5 miles for podiatrists.

Table 3 presents key parameter estimates and standard errors with eight points of support for the heterogeneity distribution. We only report transition to diabetes and LECs after the first visit, adverse health outcomes (death and amputation), and screening visits. Other transitions (two pre-age 65 hazards, two post-age 65-before first visit hazards, and first visit after 65) were estimated to control for endogenous initial conditions but are not shown (available from the corresponding author on request).

Consider the effects of the length of time with undiagnosed diabetes (D_1) on the transitions to more severe disease stages (D_1, D_3, and D_4) and death. The parameter estimates on each of these linear terms for time with undiagnosed diabetes (D_1) in row 1 of Table 3 are positive, whereas the parameter estimates on the quadratic terms are negative (row 2). These imply positive but decreasing marginal effects of time without a diagnosis on the hazards of these adverse outcomes occurring.[16] At this disease stage, the principal therapies are drugs, improved diet, and more and more regular exercise. These results imply that persons who screen less often for diabetes, and consequently have longer stretches when they have diabetes but are unaware of it, pay a long-term health penalty.

For some categories of persons with LECs, the parameter estimate on the linear term is negative, but the parameter estimate on the squared term is positive. This implies that the initial negative effect of increased time with a particular undiagnosed diabetes stage is followed by increasingly positive marginal effects. Rows 3 and 4 show that time with undiagnosed low severity LEC (D_2) has such a U-shape effect on the hazards for medium severity LEC (D_3); there is a similar pattern for high severity LEC (D_4). Both troughs, however, are at quite short durations, i.e., three and five quarters, respectively. At the low severity LEC stage, therapies include use of special shoes. Wearing such shoes often defers adverse outcomes of diabetes complications of the lower extremities. Although this is generally so, there appears to be a short period during which minor undiagnosed complications have no apparent adverse effects of not wearing specialized shoes.

The duration dependence parameters usually imply a U-shaped effect of duration on transitions. For example, total time with diabetes, including both undiagnosed and diagnosed time (columns 2 and 3, rows 10 and 11), has a U-shaped effect on the transition probabilities to the low LEC stage (D_2) and medium LEC stage (D_3); the inflection points are around 28 quarters (seven years) and 23 quarters (six years). Similarly, time in the low LEC stage (column 3, rows 13 and 14) has a U-shape effect on the transition to the medium LEC stage (inflection point around 21 quarters), and time in the medium LEC stage has a U-shape effect on the transition probability to the high LEC stage (column 4, rows 16 and 17) with an inflection point of around 26 quarters. These somewhat lengthy estimated spells of negative duration dependence could reflect a failure to allow for disease-stage specific individual heterogeneity or other functional form specification issues.

The hazards of death and amputation increase considerably when one enters either stage D_3 or D_4 (columns 5 and 6, rows 15 and 18). Stage D_4 especially carries an extensive risk for both amputation and death. Surprisingly, we find that an increase in the time one has undiagnosed D_3 or D_4 appears to reduce, relatively, the risks of death (column 5, rows 5–8). The estimates also suggest that an increase in the total time spent with D_4 yields relatively lower (though still quite

[16] The turning points are at 20 to 55 quarters, which would be an extreme length of time with unmeasured diabetes.

TABLE 3
RESULTS WITH EIGHT POINTS HETEROGENEITY

	Disease				Outcomes		
		Lower Extremity Complication					
Explanatory Variables	D_1 (1)	D_2 (2)	D_3 (3)	D_4 (4)	Death (5)	Amputation (6)	Visits (7)
Early Diagnosis							
(1) Time with undiagnosed diabetes		0.0571*** (0.0134)	0.0067 (0.0127)	0.0440** (0.0182)	0.0690*** (0.0215)	−0.0151 (0.0295)	0.0087 (0.0177)
(2) Time with undiagnosed diabetes sq		−0.0014*** (0.0004)	−0.0002 (0.0003)	−0.0004 (0.0004)	−0.0016*** (0.0006)	0.0007 (0.0006)	−0.0009*** (0.0001)
(3) Time with undiagnosed low severity LEC			−0.0306 (0.0353)	−0.1848* (0.1014)	0.0270 (0.0494)	−0.0783 (0.2131)	−0.0008 (0.0179)
(4) Time with undiagnosed low severity LEC sq			0.0053*** (0.0018)	0.0172* (0.0092)	0.0005 (0.0035)	−0.0572 (0.0477)	0.0034*** (0.0004)
(5) Time with undiagnosed medium severity LEC				−0.1495 (0.1102)	−0.1331 (0.0930)	0.3300 (0.2362)	0.0052 (0.0177)
(6) Time with undiagnosed medium severity LEC sq				0.0007 (0.0082)	0.0032 (0.0029)	−0.0758 (0.0596)	−0.0081*** (0.0005)
(7) Time with undiagnosed high severity LEC					−0.0617** (0.0285)	0.0384 (0.0273)	0.0116 (0.0178)
(8) Time with undiagnosed high severity LEC sq					0.0004** (0.0002)	−0.0010 (0.0006)	0.0000 (0.0001)
Duration Dependence							
(9) Has diabetes					−0.4125* (0.2109)	−0.6026 (0.4776)	0.3760 (0.7075)
(10) Time with diabetes		−0.2535*** (0.0092)	−0.0194** (0.0083)	0.0215 (0.0136)	−0.0041 (0.0120)	0.0057 (0.0248)	−0.0349*** (0.0037)
(11) Time with diabetes square		0.0045*** (0.0003)	0.0004*** (0.0001)	−0.0003 (0.0002)	0.0002 (0.0002)	0.0000 (0.0002)	0.0012*** (0.0001)
(12) Has low severity LEC					1.2104*** (0.2176)	0.6192 (0.5994)	0.0807 (0.7143)
(13) Time with low severity LEC			−0.3033*** (0.0142)	−0.0202 (0.0228)	−0.0454** (0.0219)	−0.1016* (0.0537)	−0.0246*** (0.0081)
(14) Time with low severity LEC sq			0.0073*** (0.0004)	0.0008 (0.0007)	0.0005 (0.0006)	0.0028** (0.0013)	0.0003 (0.0003)
(15) Has medium severity LEC					0.2421 (0.2479)	0.2934 (0.7295)	0.1893 (0.7087)
(16) Time with medium severity LEC				−0.3640*** (0.0270)	0.0391 (0.0412)	0.1606** (0.0663)	−0.0005 (0.0121)
(17) Time with medium severity LEC sq				0.0071*** (0.0008)	−0.0018 (0.0014)	−0.0049** (0.0022)	−0.0004 (0.0004)
(18) Has high severity LEC					0.7214** (0.2893)	4.4580*** (0.5515)	0.4368 (0.7104)
(19) Time with high severity LEC					−0.0892* (0.0452)	−0.4532*** (0.0674)	−0.0682*** (0.0168)
(20) Time with high severity LEC sq					0.0040*** (0.0015)	0.0115*** (0.0022)	0.0016*** (0.0005)
Other Variables							
(21) White	−0.3636*** (0.0667)	0.1621** (0.0721)	0.0055 (0.0991)	−0.3096** (0.1523)	0.0198 (0.1019)	0.3509 (0.2560)	0.0274 (0.0521)
(22) Male	0.2863*** (0.0471)	−0.3414*** (0.0534)	0.2232*** (0.0750)	0.2865** (0.1247)	0.2194*** (0.0729)	0.6862*** (0.1807)	−0.2534*** (0.0287)

(Continued)

TABLE 3
CONTINUED

	Disease				Outcomes		
		Lower Extremity Complication					
Explanatory Variables	D_1 (1)	D_2 (2)	D_3 (3)	D_4 (4)	Death (5)	Amputation (6)	Visits (7)
(23) High school+	-0.2599^{***}	-0.1336^{*}	-0.1802^{*}	-0.0096	-1.4441^{***}	-0.4602^{**}	-0.2042^{***}
	(0.0603)	(0.0679)	(0.0914)	(0.1488)	(0.1421)	(0.2241)	(0.0396)
(24) Married	-0.2041^{***}	-0.0428	-0.1246	-0.1086	-0.1461	-0.7102^{***}	0.0524
	(0.0517)	(0.0581)	(0.0824)	(0.1357)	(0.0914)	(0.1999)	(0.0333)
(25) Arthritis	0.6789^{***}	1.1341^{***}	0.5066^{**}	0.0608	0.5434^{**}	0.5645	1.2819^{***}
	(0.1561)	(0.1772)	(0.1958)	(0.3185)	(0.2349)	(0.6194)	(0.0716)
Exclusion Restrictions							
(26) Weighted mean distance to general physician							-0.2274^{***}
							(0.0772)
(27) Weighted mean distance to podiatrist							-0.0252^{*}
							(0.0130)
Log-likelihood				−346,036.05			

NOTES: All equations include a year trend and its square and the year in which the individual turned age 65. These seven equations were estimated jointly with five others that describe the diabetes progressions' "initial conditions." Those estimates can be obtained by contacting the corresponding author. Standard errors in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

high) risks of death and amputation (columns 5 and 6, rows 19 and 20). Stages D_3 (cellulitis and charcot foot) and D_4 (osteomyelitis and gangrene) produce high levels of discomfort, which leads to visits and diagnoses. It is highly unlikely a person would spend a long time with these conditions without knowing it and without seeking medical assistance. Indeed our data show that about 80% of all observations at risk of developing either D_3 or D_4 spend no quarters with the stage undiagnosed. Our data also indicate that at most 37 individuals could have a range of undiagnosed stage D_3 diabetes for more than three quarters, and only 15 observations, at most, could have a range of undiagnosed D_4 for more than two quarters. The limited number of observations with long exposures of undiagnosed diabetes with severe symptoms also suggest that we should be careful in interpreting the duration effects associated with the higher disease stages.

Results for the demographic characteristics and arthritis on disease progression and outcomes (rows 21–25) are generally consistent with previous studies (Sloan et al., 2010). For example, Whites are less likely to develop diabetes, but once they have diabetes the effects of race on further complications, death, and amputation are not uniform and mostly not statistically significant. Being male, having low education, being not married, and having arthritis lead to worse health outcomes. For screening visits, the exclusion restrictions are statistically significant, and the parameter estimates have the anticipated signs (column 7, rows 26 and 27).

Table 4 displays the heterogeneity points of support for the different transitions and the implied probabilities for each of the eight points of support. All of the coefficients associated with these mass points and weights are statistically significant. The probabilities associated with the mass points range from 0.012 to 0.326, which indicates that there is no point with an extremely small or large weight. In addition, there are no extreme mass points in any of the specifications. The implied correlations between the heterogeneities points of the different equations (Table 5) are mostly positive and small when negative. This seems plausible in that, for example, the unmeasured heterogeneity associated with faster transitions to D_1 is positively correlated with the unmeasured heterogeneities associated with all disease progressions, amputation, death, and visits.

TABLE 4
DISTRIBUTION OF THE EFFECTS OF UNOBSERVED HETEROGENEITY BY POINT OF SUPPORT

		Disease				Outcomes		
Heterogeneity Point (e_k)	Probability	D_1	D_2	D_3	D_4	Death	Amputation	Visits
1	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.12	0.22	0.09	0.01	0.16	0.24	0.87	0.97
3	0.33	0.46	0.48	0.09	0.20	0.50	1.14	2.16
4	0.27	0.69	0.97	0.20	0.17	0.74	1.03	3.36
5	0.08	0.89	1.39	0.29	0.11	0.93	0.76	4.36
6	0.10	1.06	1.53	0.31	0.09	1.04	0.54	4.94
7	0.01	1.18	1.22	0.23	0.14	1.02	0.59	4.91
8	0.02	1.22	0.27	−0.02	0.33	0.86	1.13	4.03
Expected value		0.57	0.71	0.14	0.16	0.60	0.90	2.70
Standard deviation		0.30	0.48	0.10	0.06	0.28	0.31	1.38

TABLE 5
CORRELATION MATRIX WITH EIGHT POINTS OF SUPPORT

	D_1	D_2	D_3	D_4	Death	Amputation	Visits
Diabetes (D_1)	1.00						
Low severity LEC (D_2)	0.26	1.00					
Medium severity LEC (D_3)	0.25	0.29	1.00				
High severity LEC (D_4)	0.03	−0.05	−0.07	1.00			
Death	0.29	0.28	0.28	0.03	1.00		
Amputation	0.04	0.00	−0.01	0.28	0.06	1.00	
Visit	0.23	0.22	0.22	0.01	0.24	0.03	1.00

7. MARGINAL EFFECTS AND POLICY SIMULATIONS

We conduct two sets of simulations using the parameter estimates in Tables 3 and 4 to describe the benefits and costs of different screening trajectories for diagnosing the onset of diabetes. In the first set, we use a very mechanical rule. We compare longer-term outcomes from the immediate detection of the onset of diabetes to those same outcomes when diabetes progresses and is undetected for exactly four quarters. In the second set, we consider somewhat more policy-relevant experiments where we restrict the frequency of screening visits for diabetes to at most only once (or twice) per year; current policy does not restrict the number of times one can be screened for diabetes. We then compare longer-term outcomes from the restricted and unrestricted environments. We simulate 10,000 individuals from age 65 to 80 for each of these experiments. Demographic characteristics and a diagnosis of arthritis are based on the sample population distribution, and the unmeasured heterogeneity points of support are simulated based on results in Table 4.

In the first set of simulations, we impose that the person was healthy at the start of age 65 ($t = 4$) but becomes diabetic in the next quarter ($t = 5$). The experiment is to screen everyone at $t = 5$ for diabetes and compare those outcomes to ones obtained when there was no screening for diabetes until date $t = 9$. The former had a zero length period of unobserved diabetes whereas the latter group experiences exactly one year of undiagnosed diabetes. After the initial screening visit ($t = 5$ or $t = 9$) that detects diabetes stage D_1 (or higher for $t = 9$) with probability 1, all subsequent screening visits follow the data-generating process described by the full set of parameter estimates.

Table 6 contains estimates of this effect of a one-year delay in detecting diabetes on the progression to more severe stages, amputation, and death over a five-year time horizon. Even with detection of diabetes at time of onset, 23.8% of the simulated population progresses to low severity LEC (D_2, or worse) within one year. A one-year delay in screening would result

TABLE 6
MARGINAL EFFECT OF A ONE-YEAR DELAY IN DIAGNOSING DIABETES

	Increase in Probability of Progressing to a Higher Stage by Year				
Stage	One	Two	Three	Four	Five
Low severity LEC (D_2)	0.0391	0.0550	0.0610	0.0657	0.0660
	[0.2382]	[0.3943]	[0.4835]	[0.5436]	[0.5922]
Medium severity LEC (D_3)	0.0051	0.0174	0.0245	0.0286	0.0328
	[0.0279]	[0.1019]	[0.1614]	[0.2053]	[0.2390]
High severity LEC (D_4)	0.0004	0.0042	0.0089	0.0117	0.0155
	[0.0022]	[0.0126]	[0.0274]	[0.0420]	[0.0559]
Amputation	0.0001	0.0005	0.0018	0.0029	0.0044
	[0.0016]	[0.0042]	[0.0072]	[0.0122]	[0.0175]
Death	0.0001	0.0005	0.0018	0.0029	0.0044
	[0.0016]	[0.0040]	[0.0065]	[0.0111]	[0.0160]

NOTES: Baseline in brackets. In each of the 10,000 simulations, we specify that the person got diabetes in the second quarter at age 65 ($t = 5$). To calculate the marginal effect of early diagnosis, we compare the outcome of those who were diagnosed on the quarter of diabetes onset to those who were not diagnosed until a year later ($t = 9$).

TABLE 7
POLICY SIMULATIONS RESTRICTING THE NUMBER OF COVERED VISITS FOR HEALTH INDIVIDUALS

		Number of Quarters				
	Visits	Alive	Without Amputation	Less than High LEC	Less than Medium LEC	Less than Low LEC
Baseline	38.9393	55.9818	55.5771	54.0502	50.0424	42.3388
Reduction in Visits and Adverse Health Outcomes: 15 Years Follow-up						
Savings in quarters of screening visits						
Only one visit/year	15.6219					
Only two visits/year	9.5368					
Adverse health effects in quarters						
Only two visits/year		0.0677	0.0836	0.1172	0.0893	0.1540
Only one visit/year		0.0161	0.0284	0.0419	0.0282	0.0390

NOTES: Baseline: The predicted average of the same individuals using the parameters estimates. Only two visits/year: The predicted average of the same individuals allowing them to have at most two visits until diagnose with diabetes. Only one visit/year: The predicted average of the same individuals allowing them to have at most one visit until diagnose with diabetes.

in an additional 3.9% of the observations progressing to at least this stage. Five years after becoming diabetic, 59% of the immediately diagnosed would have progressed to low severity LEC, and this would increase by 6.6 percentage points if the diagnosis were delayed for one year for everyone. Mortality within five years of becoming diabetic would increase by 44 deaths per 10,000 because of the one-year delay in diagnosis. In general, these marginal effects are not large, but they are not trivial either.

For the second set of simulations, we examine what would happen to life-cycle trajectories from age 65 to 80 if individuals without a prior diabetes diagnosis were limited to at most only one diabetes screening visit per year. After being observed with diabetes, all subsequent visits (and disease and outcome progressions) follow the processes defined by the model estimates. We assume that the first visit per year, if there is one, is the screening visit for that year. We compare these restricted simulated outcomes to the status quo that allows one to make as many screening visits as they would like each year. We also examine a less extreme policy of allowing each person to have up to two diabetes screening visits per year before being observed with diabetes.

In the top row of Table 7, we report cumulative results for the simulations of the status quo over all quarters from age 65 to 80. In the first column, we see that with no restrictions on diabetes screening that there would be 38.9 visits on average over this 15-year period. If

we were to restrict individuals to have no more than one screening visit before they had been diagnosed with diabetes, the total number of screening visits would fall by 15.6, or about by 40% (including the unrestricted visits occurring after the first diabetes diagnosis). If instead we limited the number of screening visits for detecting D_1 to at most two per year, there would be 9.54 fewer doctor visits over the 15-year horizon compared to the status quo. These figures represent the gross cost savings in terms of the number of screening visits prevented by the policies.

The bottom panel of Table 7 describes the consequences of these policies in terms of life span and time spent in relatively healthy states (from age 65 to 80) relative to the status quo. On average, life spans would shrink by 0.0677 quarters (about six days) if there were a limit of one screening visit per year before being diagnosed with diabetes. For the less restrictive policy of no more than two screening visits per year when healthy, life spans would shrink by only 0.0161 quarters on average. The more restrictive policy change (one per year maximum) would reduce the number of quarters without any lower extremity conditions by 0.154 quarters (two weeks), and the less restrictive policy change (two visits per year) would reduce the time spent without LEC complications by 0.039 quarters.

Although these impacts seem small, they are not inconsequential. If we assume no discounting and a \$50 cost per screening visit, the more restrictive policy would save about \$780 per person entering age 65 without diabetes. In terms of the effects on one's life span, these savings come at a cost of 0.0677 quarters of life or about 0.017 fewer years of life. This policy implies an implicit value of a year of life of about \$45,900 (\$780/0.017). The less restrictive policy change would have a much smaller reduction in the number of doctor visits and considerably smaller impacts on longevity. The undiscounted cost savings from fewer visits would only be about \$476 (\$50 $\times$ 9.53), and this would only cost about 0.004 fewer years of life. These would imply an implicit marginal value of a year of life of about \$119,000 (\$476/0.004).

We use a similar approach to calculate the costs in terms of "quality-adjusted" years, i.e., years without an LEC. The one visit per year restriction implies a willingness to pay of about \$20,300 (\$780/[0.1540/4]) for an additional year of quality lifetime. The less restrictive policy implies a value of a year of quality life of more than \$48,000 (\$476/[0.039/4]). Many current estimates place the value of a year of life at about \$100,000 (Viscusi and Aldy, 2003), so the more restrictive policy change to only one visit per year might undervalue its implicit costs relative to the cash savings from funding fewer screening visits; the less severe restriction on screening visits would slightly undervalue the savings relative to the benefits from the extended life spans.

8. DISCUSSION, CONCLUSION, AND EXTENSIONS

This article develops a model of screening for a chronic disease that allows for analysis of the effects of early diagnosis of the disease on disease progression and other health outcomes. The model incorporates partial observability of the disease as a key component. This allows us to assess the benefits and costs of different screening scenarios. We apply the model to diabetes mellitus.

According to the U.S. Preventive Services Task Force (USPSTF), currently there is insufficient evidence on the benefits of routine screening for type 2 diabetes in asymptomatic adults unless they have high blood pressure (USPSTF, 2008). The American Diabetes Association recommends testing at three-year intervals, but there are no conclusive RCTs that have documented health benefits of more frequent screening of diabetes (USPSTF, 2008). Our results indicate there are small individual health benefits for early diagnosis of diabetes. These are consistent with limited findings from the RCTs (USPSTF, 2008).

Although many individual features of our model are not new, they have not been combined in a single application. Relevant previous studies include multiple-state duration models with endogenous treatment (Eberwein et al., 1997, 2002; Heckman and Navarro, 2007; Abbring and

Heckman, 2008; Richardson and van den Berg, 2008); competing risk (Honore and Lleras-Muney, 2006); unobserved heterogeneity (Heckman and Singer, 1984); partial observability (Mroz and Weir, 1990); and discrete factors (Heckman and Singer, 1984; Mroz 1999). Due to improvements in computation, combining these features is feasible now when it would have been infeasible a decade ago.

The study benefits from recent availability of administrative data. Such data allow researchers to study dynamic processes for many individuals over a long time period. Administrative data, of course, have important limitations, but they are relatively inexpensive and are likely to become more detailed and even less expensive to collect and use. They will constitute key components of future research because of the inherent limitations of RCTs for assessing longer-term outcomes. Developing methods for analyzing data from surveys linked with administrative records, as in this study, and dealing with their important deficiencies should receive a high priority.

Perhaps the most important feature of our study is in accounting for partial observability. By modeling the earliest and latest possible dates of onset of the stage, we infer the length of being in a stage without knowledge of the exact time of entering the stage. We study how this measure affects subsequent transitions in diabetic stages and outcomes. As in past research, accounting for unobserved heterogeneity has proven important here. However, we do not allow for the emergence of chronic health shocks, and these are likely to be important for modeling many health transitions.

Further research should incorporate these extensions. First, rather than defining a general measure of visits, there should be more explicit measurement of specific treatments. Data on the Medicare prescription drug program is just becoming available, and it is not yet possible to incorporate long-term effects of drug treatment for this reason. Second, in our model the determinants of doctor visits do not depend on the disease stage, a restriction that could be relaxed. Third, our unmeasured heterogeneity is time invariant within each equation. A potential extension would be to introduce short- and long-term unobserved health shocks evolving as the individual ages.

In conclusion, this study provides a promising approach for evaluating how the duration of unobserved events affects outcomes of interest using administrative data.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Online Appendix.

REFERENCES

Abbring, J., and J. Heckman, "Dynamic Policy Evaluation," in L. Matyas and P. Sevestre, eds., *The Econometrics of Panel Data*, 3rd edition (Dordrecht: Kluwer Academic Publishers, 2008), 795–863.

Bhattacharya, J., "Specialty Selection and Lifetime Returns to Specialization within Medicine," *Journal of Human Resources* 40(1) (2005), 115–41.

———, D. Goldman, and N. Sood, "The Link between Public and Private Insurance and HIV-Related Mortality," *Journal of Health Economics* 22(6) (2003), 1105–22.

Blau, D., and D. Gilleskie, "Retiree Health Insurance and the Labor Force Behavior of Older Men in the 1990s," *Review of Economics and Statistics* 83(1) (2001), 64–80.

Blundell, R., T. Magnac, and C. Meghir, "Savings and Labor Market Transitions," *Journal of Business and Economic Statistics* 15(2) (1997), 153–64.

Centers for Disease Control and Prevention, "Chronic Diseases: The Power to Prevent, The Call to Control," 2009, available at http://www.cdc.gov/chronicdisease/resources/publications/AAG/pdf/chronic.pdf, accessed June 2012.

Cutler, D., "Are We Finally Winning the War on Cancer?" *Journal of Economic Perspectives* 22(4) (2008), 3–26.

EBERWEIN, C., J. HAM, AND R. LALONDE, "The Impact of Being Offered and Receiving Classroom Training on Employment Histories of Disadvantage Women: Evidence from Experimental Data," *Review of Economic Studies* 64 (1997), 655–82.

———, ———, AND ———, "Alternative Methods of Estimating Program Effects in Event History Models," *Labour Economics* 9 (2002), 249–78.

GLEWWE, P., AND H. JACOBY, "Economic Growth and the Demand for Education: Is There a Wealth Effect?" *Journal of Development Economics* 74(1) (2004), 33–51.

HECKMAN, J., AND S. NAVARRO, "Dynamic Discrete Choice and Dynamic Treatment Effects," *Journal of Econometrics* 136 (2007), 341–96.

———, AND B. SINGER, "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica* 52(2) (1984), 271–320.

HONORE, B., AND A. LLERAS-MUNEY, "Bounds in Competing Risks Models and the War on Cancer," *Econometrica* 74(6) (2006), 1675–98.

HOYNES, H., "Welfare Transfers in Two-Parent Families," *Econometrica* 64(2) (1996), 295–332.

LIU, H., T. A. MROZ, AND L. ADAIR, "Parental Compensatory Behaviors and Early Child Health Outcomes in Cebu, Philippines," *Journal of Development Economics*, 90(2) (2009), 209–30.

———, ———, AND W. VAN DER KLAAUV, "Maternal Employment, Migration, and Child Development," *Journal of Econometrics* 156(1) (2010), 212–28.

MELLO, M., S. STEARNS, AND E. NORTON, "Do Medicare HMOs Still Reduce Health Services Use After Controlling for Selection Bias?" *Health Economics* 11 (2002), 323–40.

MROZ, T. A., "Discrete Factor Approximations in Simultaneous Equation Models: Estimating the Impact of a Dummy Endogenous Variable on a Continuous Outcome," *Journal of Econometrics* 92 (1999), 233–74.

———, AND T. SAVAGE, "The Long-Term Effects of Youth Unemployment," *Journal of Human Resources* 41(2) (2006), 259–93.

———, AND D. WEIR, "Structural Change in Life Cycle Fertility During the Fertility Transition: France before and after the Revolution of 1789," *Population Studies* 44 (1990), 61–87.

PICONE, G., F. SLOAN, D. TAYLOR, AND S. Y. CHOU, "Does High Hospital Cost Imply Higher Quality of Care?" *Review of Economics and Statistics* 85(1) (2003), 51–62.

RICHARDSON, K., AND G. VAN DEN BERG, "Duration Dependence versus Unobserved Heterogeneity in Treatment Effects: Swedish Labor Market Training and the Transition Rate to Employment," Working Paper, Institute for Labor Market Policy Evaluation, 2008.

SLOAN, F. A., A. BETHEL, D. RUIZ JR., A. M. SHEA, AND M. N. FEINGLOS, "The Growing Burden of Diabetes Mellitus in the US Elderly Population," *Archives of Internal Medicine* 168(2) (2008), 192–9.

———, M. FEINGLOS, AND D. GROSSMAN, "Receipt of Care and Reduction of Lower Extremity Amputations in a Nationally Representative Sample of U.S. Elderly," *Health Services Research* 45(6) (2010), 1740–62.

U.S. PREVENTIVE SERVICES TASK FORCE, "Screening for Type 2 Diabetes Mellitus: Update of 2003 Systematic Evidence Review for the U.S. Preventive Task Force," 2008, available at http://www.uspreventiveservicestaskforce.org/uspstf/uspsdiab.htm, accessed October 2013.

VISCUSI, W. K., AND J. E. ALDY, "The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World," *Journal of Risk Uncertainty* 27(1) (2003), 5–76.