Using Propensity Scores Subclassification to Estimate Effects of Longitudinal Treatments

An Example Using a New Diabetes Medication

Jodi B. Segal, MD, MPH, Michael Griswold, PhD,† Aristide Achy-Brou, MS,† Robert Herbert, BS,‡ Eric B. Bass, MD, MPH,*‡ Sydney M. Dy, MD,*‡ Anne E. Millman, MS,‡ Albert W. Wu, MD, MPH,*‡ and Constantine E. Frangakis, PhD†*

Background: When using observational data to compare the effectiveness of medications, it is essential to account parsimoniously for patients' longitudinal characteristics that lead to changes in treatments over time.

Objectives: We developed a method of estimating effects of longitudinal treatments that uses subclassification on a longitudinal propensity score to compare outcomes between a new drug (exenatide) and established drugs (insulin and oral medications) assuming knowledge of the variables influencing the treatment assignment.

Research Design/Subjects: We assembled a retrospective cohort of patients with diabetes mellitus from among a population of employed persons and their dependents.

Methods: The data, from i3Innovus, includes claims for utilization of medications and inpatient and outpatient services. We estimated a model for the longitudinal propensity score process of receiving a medication of interest. We used our methods to estimate the effect of the new versus established drugs on total health care charges and hospitalization.

Results: We had data from 131,714 patients with diabetes filling prescriptions from June through December 2005. Within propensity score quintiles, the explanatory covariates were well-balanced. We estimated that the total health care charges per month that would have occurred if all patients had been continually on exenatide compared with if the same patients had been on insulin were minimally higher, with a mean monthly difference of $397 [95% confidence interval (CI), $218–$1054]. The odds of hospitalization were also comparable (relative odds, 1.02; 95% CI, 0.33–1.98).

Conclusions: We used subclassification of a longitudinal propensity score for reducing the multidimensionality of observational data, including treatments changing over time. In our example, evaluating a new diabetes drug, there were no demonstrable differences in outcomes relative to existing therapies.

Key Words: longitudinal treatment, propensity score, pharmacoepidemiology

(*Med Care* 2007;45: S149–S157)

The relationships between outcomes and treatment are often unclear when studying the use of a medication outside the setting of a trial. The prescribing physician's decision to use a new medication and the patient's decision to fill and to continue to fill the prescription affect the outcomes of interest. Incomplete adherence complicates the ability to study outcomes, as outcomes are generally related to adherence as well as the factors that influence adherence. Given that many factors influence the prescription of and adherence to a new drug, all of which may impact outcomes, we aimed to develop methodology to manage these covariates, with a focus on the issues of multiple treatments and multidimensional covariate histories. Our goal was to develop methods to use with observational data, which will allow one to predict a medication's upper limit of effectiveness, in a setting outside of a trial.

We explored the usefulness of these methods in our study of a new drug for treatment of type 2 diabetes mellitus. Exenatide, manufactured as Byetta by Amylin Pharmaceuticals, (San Diego, CA), was approved by the Food and Drug Administration in April 2005 for use as an adjunctive therapy for treatment of type 2 diabetes.[1] Exenatide is a peptide that is a partial analog of glucagon-like peptide-1. It is injected twice daily to augment the "incretin effect" in which the pancreas responds with insulin secretion. It is indicated for patients who have not achieved adequate glycemic control on metformin, a sulfonylurea, or a combination of metformin and a sulfonylurea. In prelicensing trials, up to 20% of patients discontinued therapy due to nausea; however, the efficacy of the drug was good regarding the glycemic tar-

From the *Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland; and the Departments of †Biostatistics and ‡Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland.

This research was conducted by the Johns Hopkins University DEcIDE Network Center under contract to the Agency for Healthcare Research and Quality Contract 290-2005-0034-1, Rockville, MD.

The authors of this manuscript are responsible for its content. No statement may be construed as the official position of the Agency for Healthcare Research and Quality of the US Department of Health and Human Services.

Reprints: Jodi Segal, MD, MPH, 1830 E, Monument St, Room 8047, Baltimore, MD 21205. E-mail: jsegal@jhmi.edu.

ISSN: 0025-7079/07/4500-0149

gets.[2–8] We used observational data from the first 6 months after approval of this new drug to develop models to predict patient outcomes, while accounting for patient and physician factors affecting both initiation and continued use of the medication.

METHODS

Data Acquisition

This was a retrospective cohort study of employed, commercially-insured patients and their dependents whose healthcare utilization data is collected by i3Innovus, an Ingenix company (Eden Prairie, MN). No patient is uniquely identifiable, and the project received exemption from review from the Johns Hopkins Institutional Review Board. The data set is called Ingenix LabRx and contains data from UnitedHealthcare, which has beneficiaries in 50 states and the District of Columbia. LabRx currently includes 24 million insured lives. LabRx is updated monthly with information on enrollee age, sex, enrollment dates, and claims for reimbursement for billable health care services. The data includes patient diagnoses as identified by International Classification of Diseases, Ninth Revision (ICD-9) codes and medical procedures using several classification systems. Additionally, a separate, linkable file is available which includes pharmacy claims for prescription drugs, including the drug name, prescription fill date, and the number of days supply provided. Results from laboratory evaluations were available on a subset of the enrollees depending on the diagnostic testing site used.

We requested data on all patients with an ICD-9 code of 250.xx (diabetes mellitus) or a prescription fill for a drug used to treat diabetes between June 1, 2004 and December 31, 2005. We chose these dates for our baseline analyses so that we would have all data from the first 6 months after approval of exenatide and from a period before approval. Data elements included all available data from the medical claims file and the prescription drug file, as well as select laboratory data [specifically, hemoglobin A1c (HbA1c), fasting glucose, and lipids] on these patients from June 1, 2003 through December 31, 2005.

Defining the Cohort

For inclusion in the cohort, we required that the patient:

- Have a claim with an ICD-9 code of 250.xx at least twice[9,10]
- Have 12 months of continuous coverage before the index date, which was defined as the date at which the patient first filled exenatide or October 15, 2005 for those not filling exenatide
- Be between 18 and 64 years, inclusive
- Not be on dialysis
- Not be in a managed Medicaid health plan
- Have at least 1 visit after June 1, 2005.

Creating Variables

From the claims data, we created variables to describe our patient populations and to use for adjustment when evaluating outcomes. These variables fit into the broad categories of demographics (age, sex, census division of residence), utilization variables (hospitalizations, outpatient visits, provider specialty, total health care charges, copayments for prescriptions, medication use), and clinical variables [diabetes-associated complications, achievement of Health Employer Data Information System (HEDIS) indicators of high-quality care, side-effects, HbA1c, fasting glucose, and Johns Hopkins University ACG Case-Mix System (version 8.0 beta) for description of case-mix]. For variables that were time-varying, a unique variable was created for each month from June 2004 through December 2005. For example, total health care charges vary by month, so we created 19 variables for each patient representing total health care charges for each of the 19 months. Other variables, such as an ICD-9 code for obesity, were considered to always be present after the month in which they were first coded. The medication use indicator variables were constructed so that if a patient filled a 30-day prescription for an oral hypoglycemic medication, the month in which he or she filled it would have an indicator variable demonstrating this. If it was a 90-day supply, the patient would have an indicator signifying that he or she had the medication in the subsequent 2 months as well.

We created variables to indicate usage of exenatide and of insulin as follows: (1) we identified 2-month intervals (May–June 2005, July–August 2005, and September–October 2005); and (2) if the patient had sufficient drug-on-hand for more than 50% of the days within the 2-month interval, he or she was considered to have been on this drug for that interval. An indicator variable was made to indicate whether during the 2-month interval the patient was on (a) no injectable medications (ie, was on oral medications or no medications), (b) insulin, (c) exenatide, or (d) both. We combined patients using oral medication with patients using no medication for diabetes because these patients would most likely be similar in terms of duration of disease and complications from diabetes.

Outcome Variables

For the purpose of developing our models, we chose to explore 2 outcomes that impact drug coverage decisions: hospitalization rates and total health care charges.

Statistical Framework

We developed methodology to compare a new medication relative to established therapies, using existing, observational data. The 3 components of this approach are as follows: a framework for setting the goal, assumptions for estimability of the goal, and estimation methods for parsimoniously using the patient histories. These components are briefly explained below.

Potential Outcomes and Goal

The times where a treatment can change are denoted by $t = 1, 2, \ldots T$. At each time, let $z_t = 1, 2, \ldots K$ indicate the levels of the treatment (eg, 1 for exenatide). If patient i would have taken some longitudinal treatment of interest, $z = (z_1, z_2, \ldots z_T)$, we let $Y_i(z) = Y_i(z_1, z_2, \ldots z_T)$ be the potential outcome that will be observed at the evaluation time period.[11–13] Note that the treatment regimen of interest, z, may be differ-

ent from the treatment regimen actually observed for given patients. For example, the treatment regimen of interest may be the cumulative effect of 3 periods of treatment on a new drug, (z = 1,1,1). In the potential outcomes framework, we may ask the question, "What would we expect to see if all patients (regardless of their actual treatments) had instead received 3 periods of the new drug (z = 1,1,1)?"

For a particular longitudinal treatment, we are interested in estimating outcome quantities such as $E\{Y_i(z)\}$, which is the expectation we would observe if all patients received a particular longitudinal treatment. We wish to compare outcomes among such possible longitudinal treatments, for example, with z = 1, 1, 1 versus with z = 0, 0, 0. Note that this comparison is not the same as the comparison between the observed distribution of the 111 group versus its observed "control" (everyone other than 111), or the comparison of the observed distribution of 000 to its observed "control," or even the comparison between these 2 comparisons.

Assumption

Given that only 1 treatment assignment is actually made for each patient at each time point, and that the treatment assignment, or adherence to original treatment assignment, can change over the observation period, this contrast of interest is not directly observable. Therefore, we need 2 assumptions to make it estimable. The first assumption is that the patients are a random sample drawn from the appropriate reference population. The second assumption, already implicit in the notation $Y_i(z)$ above, is that the treatment assignment for 1 patient does not affect the outcome of a different patient (stable unit treatment values—SUTVA).[13] The third assumption is that all variables related to treatment assignment have been measured, in the sense that conditional on the observed variables up to a particular time, the assignment to the treatment at the next time is random (sequential ignorability).[13,14] Previous work has established the theoretical ability to compare treatments under these assumptions in large samples,[9,11,12] and any of several methods could be used.

Estimation Methods

With longitudinal treatments there is the need to control the growing dimension of history variables that need to be modeled. Such parsimony can be achieved by use of propensity scores.[15] However, the use of propensity scores for longitudinal treatments has been essentially limited to using them as weights (eg, marginal structural models).[16] As discussed in the Appendix, and as also suggested by the results of simulation studies for single time points,[17] methods that use propensity scores as stratifying variables in outcome models (possibly with the help of weights) are superior to models that do not use propensity scores as stratifying variables in outcome models. The question we address here, therefore, is how we can use propensity scores as stratifiers with longitudinal treatments. Below, we give the essential notation and steps of the methods.

For the i^{th} patient, let $X_{i,t}^{obs}$ be the vector of variables observed after the patient received a specific treatment at time $t-1$ but before taking a specific treatment at time t. Let the actual treatment received be denoted by $Z_{i,t}$, and let Z_i be the vector of these treatment assignments. Let the patient history, $H_{i,t}$, be the cumulative information observed before the patient received treatment at time t, that is:

$$\mathrm{H_{i,t}} = \{(X_{i,1}^{obs}, \mathrm{Z_{i,1}}), \ldots (X_{i,t-1}^{obs}, \mathrm{Z_{i,t-1}}), X_{i,t}^{obs}\}.$$

Let Y_i^{obs} denote the observed outcome at the end of the last period, which, based on the potential outcomes notation, is equal to $Y_i(Z_i)$. Finally, let the conditional probability for the i^{th} subject at time point t to receive treatment k, $Z_{i,t} = z$, given the history $H_{i,t}$, be denoted by:

$$\mathrm{e_{i,t,z}} = \Pr(\mathrm{Z_{i,t}} = \mathrm{z} \mid \mathrm{H_{i,t}}),$$

which is the propensity score.[18] An evaluation of $E\{Y(z)\}$ may be done by averaging the distribution of observed outcomes over a longitudinal subclassification of the patients by their propensity scores. Below we outline the estimation algorithm with application to the exenatide data.

Statistical Estimation

We had data for 3 time periods (July–August 2005, September–October 2005, and November–December 2005). Estimands of interest were specified as the expected outcomes for subjects had they received consistent treatment across the 3 time periods: (Exenatide-Exenatide-Exenatide = EEE), (Insulin-Insulin-Insulin = III), (Both-Both-Both = BBB), or (Other-Other-Other = OOO), where "both" refers to receiving both exenatide and insulin, and "other" refers to any oral diabetes medications. The 6-step algorithm for estimating contrasts in outcomes, between these treatment regimes, is as follows:

1. Estimate propensity models.

At each time point, construct a multinomial propensity score model for all treatments as a function of prior histories. For example, the multinomial generalized logit model,

$$\mathrm{logit}\{\Pr(\mathrm{Z_{i,t}} = \mathrm{k} \mid \mathrm{H_{i,t}} = \mathrm{h}) / \Pr(\mathrm{Z_{i,t}} = 1 \mid \mathrm{H_{i,t}} = \mathrm{h})\} = \mathrm{f_{,t}(k \mid h)},$$

where $f_{,t}(k\ h)$ is a function linear in the model parameters.

2. Choose a treatment group, construct a propensity score for each subject, and categorize subjects into strata.

From the models in step 1, extract the propensity score contrasting the chosen treatment (here, exenatide) against all others for each time point.

$$\mathrm{e_t\ (h)} = \Pr(\mathrm{Z_{i,t}} = 1 \mid \mathrm{H_{i,t}} = \mathrm{h}) = \Pr(\mathrm{Z_t} = \text{Exenatide} \mid \mathrm{H_{i,t}} = \mathrm{h});$$

$$1 - \mathrm{e_t\ (h)} = \Pr(\mathrm{Z_{i,t}} = 0 \mid \mathrm{H_{i,t}} = \mathrm{h}) = \Pr(\mathrm{Z_{i,t}} = \text{“Not Exenatide”} \mid \mathrm{H_{i,t}} = \mathrm{h}),$$

so that "0" treatment means "Not Exenatide." For each time point, construct K strata (usually quintiles, K = 5) of the propensity score, $e_t(H_{i,t})$ separately for patients with observed $Z_t = 1$, and $Z_t = 0$ as performed in standard subclassification propensity score techniques.[12] Remove subjects in the nonoverlaping regions of the propensity score distributions to

reduce extrapolation. Denote by $K_{it} = 1. . .K$, the stratum to which $e_t(H_{i,t})$ belongs for that subject at that time, and let $K_i = (K_{i1}. . .K_{iT})$. Check that the distributions of the covariates between the treatment groups, within quintiles of the propensity score, are similar.[12]

3. Estimate the transitional probabilities of the longitudinal propensity score strata.

$$\text{I. } p(k1) = \Pr(Ki,1 = k1)$$

$$\text{II. } p(k2 \mid k1, z1) = \Pr(Ki,2 = k2 \mid Ki,1 = k1\ Zi,1 = z1)$$

$$\text{III. } p(k_3 \mid k_1, z_1, k_2, z_2) = \Pr(K_{i,3} = k_3 \mid K_{i,1} = k_1\ Z_{i,1} = z_1\ K_{i,2} = k_2, Z_{i,2} = z_2)$$

For the application, we fit a series of transitional proportional odds logistic regression models to estimate the parameters of the propensity score strata transitions, an extension to logistic regression models for ordinal data with K categories. Although interaction terms may be included, we opted for 2 models with only the main effects:

$$\text{logit}(\Pr(K_2 \le k' \mid K_1, Z_1) = \alpha_{0k'} + \alpha_1 Z_1 + \sum_k [\alpha_{2k} * 1(K_1 = k)]$$

$$\text{logit}(\Pr(K_3 \le k' \mid K_1, Z_1, K_2, Z_2) = \gamma_{0k'} + \gamma_1 Z_1 + \gamma_2 Z_2 + \sum_k [\gamma_{3k} * 1(K_1 = k) + \gamma_{4k} * 1(K_2 = k)]$$

where $k' = 1, 2, \ldots 4$.

When we applied these models to our data, we calculated the following distribution of propensity score strata:

$$p(k_1, k_2, k_3; z_1, z_2) = p(k_3 \mid k_1, z_1, k_2, z_2) * p(k_2 \mid k_1, z_1) * p(k_1)$$

Note that this distribution is not the joint probability of being in the strata k_1, k_2, k_3 given treatments z_1, z_2 because in each of the 3 conditional probabilities on the right hand side of this equation, the conditioning on treatment arms (Z) differs.

4. Estimate expected outcomes in observed longitudinal treatment groups of interest.

One possibility for $E(Y^{obs}\ K, Z)$ is to use the sample mean responses within the observed strata of K and Z. This choice can lead to unstable estimates due to small sample sizes, especially for skewed data such as economic outcomes. It is better to use a series of generalized linear models to estimate the parameters of the outcome probability models, and calculate expected values, based on these parameters. For example, the resource utilization outcome (total health care charges) can be estimated with a regression model:

$$E(Y^{obs} \mid K, Z) = \beta_0 + \sum_t \{\beta_t Z_{,t} + \sum_k [\beta_{t,k} * 1(K_{,t} = k)]\}$$

Although we show here the main effects model for demonstration, interactions may also be included. When appropriate, this allows one to borrow information across

TABLE 1. Median and Interquartile Range for Continuous Variables for Entire Cohort, N = 206,345

Variable	Mean	Standard Deviation	Median	25th Percentile	75th Percentile
Number of diabetes-associated complications in preceding year (range 0–8)	1.34	1.09	1	1	2
JHH-ACG score predicting high costs in subsequent year	0.16	0.24	0.04	0.01	0.21
Mean hemoglobin A1c in 1 year preceding index date (%) (n = 20,897)	7.3	1.8	6.9	6.1	8.0
Mean fasting glucose in 1 year preceding index date (mg/dL) (n = 24,225)	145	63	127	104	165
Monthly total health care charges per person in 1 year preceding index date ($)	1299	3534	348	133	1044
Monthly prescription copays per person in 1 year preceding index date ($)	59	51	47	22	82
Monthly total number of prescriptions per person filled in 1 year preceding index date	3.7	2.9	3.2	1.6	5.3
Monthly copay for exenatide per person using exenatide ($)	35	21	33	20	50
No. outpatient visits per person, total, in year preceding index date	12.8	11.7	9	5	16
No. outpatient visits per person to internist in year preceding index data	2.2	3.4	0	0	4
No. outpatient visits per person to endocrinologist in year preceding index data	0.4	1.3	0	0	0
No. outpatient visits per person to family physician in year preceding index date	2.6	3.6	1	0	4

HEDIS indicates Health Employer Data Information System; ICD-9-CM, International Classification of Disease, Ninth Revision—Clinical Modification; JHH-ACG, Johns Hopkins Hospital ACG Predictive Model (version 8 beta).

strata and across time points to better estimate the necessary expectations for constructing the evaluation. For example, the average total health care charges for a patient continuously on exenatide who fell in the highest propensity score strata at each time point would be:

$$E\{Y^{obs} \mid K = (5, 5, 5), Z = (1, 1, 1)\} = \beta_0 + \beta_1 + \beta_{15} + \beta_2 + \beta_{25} + \beta_3 + \beta_{35}$$

5. Estimate the average, over all patients, of the potential outcome for a longitudinal treatment group of interest.

The expected outcome for a particular longitudinal treatment can be calculated as an average of the values $E(Y^{obs}\ k, z)$ over the longitudinal propensity subclasses, where the average is taken over the distribution $p(k_1, k_2, k_3; z_1, z_2)$:

$$E\{Y(z_1, z_2, z_3)\} = \sum_{k1, k2, k3} E\{Y^{obs} \mid K = (k_1, k_2, k_3), Z = (z_1, z_2, z_3)\} * p(k_1, k_2, k_3; z_1, z_2)$$

The format of this formula is the same as the format for the identification formula given by Robins[14] for the full multidimensional histories, but here the multidimensional covariate histories are replaced by the subclasses of the propensity scores. As an example, the expected potential outcome under the currently chosen treatment regimen from step 2 of Exenatide-Exenatide-Exenatide (Z = EEE) is:

$$E\{Y(EEE)\} = \sum_{k1, k2, k3} E\{Y^{obs} \mid K = (k_1, k_2, k_3), Z = (1, 1, 1)\} * p(k_1, k_2, k_3; 1, 1)$$

6. Iterate through steps 2–5 to estimate causal effects.

Having obtained an estimate of $E\{Y(EEE)\}$, return to step 2. Choose an alternate treatment group, say Insulin-Insulin-Insulin, and perform the same series of steps to obtain an estimate of $E\{Y(III)\}$. Construct estimates of treatment effects such as absolute $E\{Y(EEE)\} - E\{Y(III)\}$or relative $E\{Y(EEE)\}/E\{Y(III)\}$differences, and obtain uncertainty estimates (standard errors, confidence intervals, etc) using a bootstrap algorithm. Repeat this process for all longitudinal treatment regimes of interest.

RESULTS

Description of Cohort

We received data on 1,234,540 individuals meeting the initial criteria specified in our data request (patients with ICD-9 250.xx or on a medication for diabetes). After applying the exclusion criteria, 206,345 individuals remained for study. In Table 1, we present the means and medians of the relevant continuous variables for the analytic sample. The patients ranged in age from 18 to 64 years with a mean age of 51.3 years; 54% of the population was male. The population was generally healthy with a mean of only 1.3 diabetes-associated complications, out of a possible 8. The predicted probability of high cost care in the following year, from the Johns Hopkins Hospital ACG Predictive Model (JHH-ACG) scoring system, had a mean of 0.16; however, the median was only 0.04 and the 75th percentile was only 0.21, indicating that most of this population had few comorbid conditions which would be expected to be costly in the next year.

Use of Exenatide

Among these 206,345 individuals, 3225 patients filled a prescription for exenatide. The first prescription for exenatide was filled in June 2005, slightly more than 1 month after the

TABLE 2. Drug Regimens by Treatment Interval

		November–December 2005				
July–August 2005	**September–October 2005**	**Other**	**Insulin**	**Exenatide**	**Both**	**Total**
Other	Other	115,470	1424	677	5	117,576
	Insulin	800	1425	19	11	2255
	Exenatide	102	6	356	5	469
	Both	1	0	1	8	10
Insulin	Other	1082	885	26	15	2008
	Insulin	1268	7674	39	103	9084
	Exenatide	9	4	29	9	51
	Both	4	14	12	57	87
Exenatide	Other	21	3	4	0	28
	Insulin	1	3	0	0	4
	Exenatide	19	1	79	2	101
	Both	0	1	0	5	6
Both	Other	2	1	0	0	3
	Insulin	0	2	0	0	2
	Exenatide	1	0	4	2	7
	Both	2	3	4	14	23
Total		118,782	11,446	1250	236	131,714

TABLE 3. Balance of Covariates between Patients Filling Exenatide and Patients Filling Insulin after Subclassification into Propensity Score Quintiles (May–June 2005)

	Before Subclassification		After Subclassification	
	F-Statistic	***P***	**F-Statistic**	***P***
Endocrine visit June 2005	456	0.00	34	0
Charge May 2005	159	0.00	6.34	0.01
Hospital length of stay May 2005	14	0.00	3.05	0.08
Endocrine visit May 2005	326	0.00	1.97	0.16
Charge June 2005	360	0.00	1.25	0.26
Sulfonylurea May 2005	1360	0.00	0.75	0.39
Probability of high costs	5642	0.00	0.63	0.43
No. drugs May 2005	14,052	0.00	0.46	0.45
Emergency room visit June 2005	36	0.00	0.47	0.49
No. drugs June 2005	14,178	0.00	0.47	0.49
Any outpatient visit June 2005	2694	0.00	0.44	0.51
Hospital admission May 2005	22	0.00	0.41	0.52
Sulfonylurea June 2005	1384	0.00	0.39	0.53
Prescription copay May 2005	7231	0.00	0.34	0.56
Any outpatient visit May 2005	2368	0.00	0.2	0.66
Other medication May 2005	119,743	0.00	0.17	0.68
Obesity June 2005	883	0.00	0.09	0.77
Family practice visit June 2005	826	0.00	0.06	0.81
Metformin May 2005	2664	0.00	0.04	0.83
Metformin June 2005	2815	0.00	0.04	0.84
Obesity May 2005	832	0.00	0.04	0.84
Internal medicine visit June 2005	886	0.00	0.02	0.89
Other medication June 2005	155,505	0.00	0.01	0.91
Thiazolidinedione May 2005	1903	0.00	0.01	0.92
Comorbidities May 2005	10,512	0.00	0.01	0.92
Family practice visit May 2005	724	0.00	0.01	0.93
Prescription copay June 2005	6687	0.00	0.01	0.93
Emergency room visit May 2005	34	0.00	0.01	0.93
Age	171,298	0.00	0.00	0.95
Internal medicine visit May 2005	758	0.00	0.00	0.95
Thiazolidinedione June 2005	1885	0.00	0.00	0.96
Gender	41,507	0.00	0.00	0.96
No. patients on exenatide	412	0.00	0.00	0.98
Comorbidities June 2005	10,973	0.00	0.00	0.99
Side effects May 2005	31	0.00	0.00	1.00

drug was approved. The median date for first filling a prescription for this drug was mid-October 2005.

We investigated adherence to this medication. We had data on 6753 prescription fills (for the 3225 patients); 3528 were refills of the medication. There were 1379 patients (42%) who filled 1 prescription and never filled another, although 687 (50%) of these were in December, the last month from which we have data. The majority of refills (2416 or 68%) could be considered to be "on-time," with the median time to filling a next prescription of 33 days (mean was 35 days). Among the 1112 prescriptions (31% of refills) which were filled late (defined as 7 days past the expected fill date), the median fill time was 47 days after the previous fill (mean was 52 days after). Patients who filled only 1 prescription were not included in that calculation. Twelve percent of the prescription refills were filled more than 30 days late by patients who subsequently refilled the medication.

Longitudinal Outcomes

For these analyses, we excluded patients who were not on any medications for diabetes, leaving 131,714 patients. In Table 2, we show the counts of the patients by their use of medications in the 3 time periods of interest.

Using the algorithm described in the methods, we estimated propensity scores for each treatment regimen by entering into the model 4 time-invariant covariates and 18 time-varying covariates. For each of the 4 treatment regimes, we created quintiles of the propensity scores at each of the treatment periods and populated this matrix with a probability for each patient to be in each of these 15 cells. Few patients needed to be excluded due to nonoverlapping propensity

TABLE 4. Key Components for Constructing the Estimated Outcomes

PS Strata(t)			Transitional Pr(PS Strata(t) $H(t)$)			φ	$E(Y^{obs}\ Z,\ k)$
k_1	k_2	k_3	Pr(k_1) (φ_1)	Pr(k_2 k_1, z_1) (φ_2)	Pr(k_3 k_1, z_1, k_2, z_2) (φ_3)	$\varphi = \varphi_1\varphi_2\varphi_3$	$E\{Y(Z = 111, k)\}$
1	1	1	0.2	2.63 E-06	3.52 E-07	1.85 E-13	$2005
1	1	2	0.2	2.63 E-06	8.00 E-07	4.22 E-13	$1242
1	1	3	0.2	2.63 E-06	2.04 E-06	1.08 E-12	$1194
1	1	4	0.2	2.63 E-06	7.56 E-06	3.99 E-12	$1190
1	1	5	0.2	2.63 E-06	9.99 E-01	5.27 E-07	$1330
1	2	1	0.2	5.49 E-06	1.90 E-07	2.09 E-13	$1765
1	2	2	0.2	5.49 E-06	4.32 E-07	4.75 E-13	$1001
1	2	3	0.2	5.49 E-06	1.10 E-06	1.21 E-12	$ 954
1	2	4	0.2	5.49 E-06	4.08 E-06	4.49 E-12	$ 950
1	2	5	0.2	5.49 E-06	9.99 E-01	1.09 E-06	$1090
:	:	:	:	:	:	:	:
5	5	5	0.2	0.999	0.999	1.99 E-01	$1720

The first 3 columns indicate 1 of the 5 propensity score quintiles at each of the 3 time periods (k_1, k_2, k_3). For example, the first row is patients who have the lowest propensity, based on their covariates, to receive exenatide at each of the 3 time periods. The next 3 columns are the transitional probabilities, ie, the probabilities for patients to transition from one quintile to another quintile at the next time period. As described in step 3), the transitional probabilities in columns 5 and 6 were estimated by proportional odds logistic regression using data from the entire cohort. Overall, the probability of belonging to the listed strata (row) is the product of the probability of being in the listed quintile at the first time period times the transitional probabilities to be in the listed quintile at the next 2 time periods. This is φ in column 7. As described in step 4), the expected outcome (in this case, charges) for each strata was estimated by linear regression using data from the entire cohort (in column 8).

PS indicates propensity score; Transitional PR, probability of transitioning between quintiles conditional on previous quintiles and previous treatments; φ, product of transitional probabilities; $E(Y^{obs}\ Z,\ k)$ = expected outcomes from generalized linear models.

TABLE 5. Average Monthly Total Health Care Charges and Absolute Differences in Charges in a Longitudinal Causal Framework

Medication in Each Time Period	Average Monthly Total Health Care Charges ($)	95% Confidence Interval	Difference	95% Confidence Interval
Insulin-Insulin-Insulin	1091	1035–1158	Reference	Reference
Other-Other-Other	978	963–993	−113	−184 to −49
Exenatide-Exenatide-Exenatide	1488	880–2127	397	−219 to 1054

TABLE 6. Average Monthly Hospitalization Frequency and Relative Risk of Hospitalization a Longitudinal Causal Framework

Medication in Each Time Period	Average Monthly Hospitalization Frequency	95% Confidence Interval	Effect	95% Confidence Interval
Insulin-Insulin-Insulin	0.012	0.011–0.014	1	Reference
Other-Other-Other	0.010	0.010–0.011	0.82	0.75–0.91
Exenatide-Exenatide-Exenatide	0.013	0.004–0.024	1.02	0.33–1.98

scores across treatment groups. As hoped, the explanatory covariates were well balanced between treatment groups after adjusting for quintile, as shown in Table 3.

We estimated the transition probabilities, that is, the probability of transitioning to a different propensity score quintile between treatment periods. Patients whose covariate profile at the second time period suggested that they were unlikely to fill a second prescription for exenatide (eg, they had side effects, had a high copay for this drug, or did not see an endocrinologist) would have a high probability of transitioning from exenatide to another group.

In Table 4, we show select components for calculating expected outcomes, for the outcome of total health care charges. As described in the algorithm, the expected outcome (charges) for each strata was estimated by linear regression using data from the entire cohort (as shown in the last column). Using the first formula of step 5 where the weights (ie, the probabilities of being in each strata) are defined as in step 3, the expected total health care charges per month for patients consistently taking exenatide across the 3 time periods was calculated as a weighted average of the expected charges across all strata. This is $1488 with a 95% confidence

interval from −$219 to $1054. A negative value can be interpreted as the patient receiving credit for past charges.

In Table 5, we show the expected total health care charges per month for the comparison groups of interest; patients consistently filling exenatide, patients consistently filling insulin, and patients consistently filling only oral medications for diabetes. The patients taking oral medications had significantly lower total health care charges than patients consistently taking insulin and patients consistently taking exenatide. The total charges for exenatide users and insulin users were similar. These results suggest that if patients consistently took oral medications for 6 months, exenatide for 6 months, or insulin for 6 months, the monthly total health care charges of those in the oral medication group would be lowest by a small amount. Similarly, in Table 6, we show the average rates of hospitalization and the relative risk of hospitalization projected for patients consistently on the listed medications, which were again similar. The odds for a hospitalization in any month would be a nonsignificant 2% higher for patients taking exenatide throughout the observation period relative to patients taking insulin throughout the period.

DISCUSSION

Our goal was to explore new methodology for the analysis of observational data to estimate treatment effectiveness when a patient's treatment changes over time. This methodology advances the management of: (1) treatments which change over time due to varying adherence or treatment choices; (2) covariates, associated with medication assignment, that change over time; and (3) limited follow-up time or a small treatment group. We demonstrated the use of this methodology with our analyses of the effect of a new drug for treating diabetes, exenatide.

This methodology has particular value for the study of new drugs. Because the drug was only recently approved, we had limited follow-up time. However, with this methodology, we were able to "borrow power" from the other observed individuals to make predictions as to the expected outcomes from its use. Additionally, when studying a new drug, it is even less clear what are the predictors of use of the drug and predictors of adherence, compared with more established medications. With this methodology, use of the drug and adherence are modeled using all available potentially influential covariates, without the need to know definitively before hand which the strongest predictors are. In this instance, the outcomes of total monthly health care charges and hospitalizations differed little among users of exenatide and users of other medications for diabetes control.

When evaluating the efficacy of a medication, a randomized controlled trial is considered the gold standard, as covariates are balanced between groups and differences can be attributed to the medication assignment. When clinical trials are not practical, or precluded by the need for a very large sample to detect an outcome, observational data must be used to make estimates about a drug's efficacy or effectiveness, with the inherent difficulties posed by lack of random assignment to therapy.

In this article, we advance the methods for analyzing observational data by extending use of subclassification using the propensity scores in longitudinal treatments. With longitudinal treatments, there is the need to control the growing dimension of history variables that need to be modeled in the outcome distribution.[10] The longer the observation period, the greater an issue this becomes. We demonstrate that when the history contained in the covariates is reduced to the history of the propensity scores, these propensity scores can be used as stratifying variables to estimate longitudinal treatment effects.

Our analyses have limitations. We did not include laboratory data in the propensity models because laboratory data were only available on a small subset of the population, due to the source of this data. We applied our methods to this subset having laboratory data, and included Hba1c and glucose in the propensity score models. Very few people with laboratory data were using exenatide, so the analyses were limited to patients using insulin or other medications. The outcome estimates were qualitatively similar to the results presented in the article. In future work we will further investigate whether the inclusion of laboratory data improves the fit of the propensity score models and/or changes the estimated outcomes. If we demonstrate definitively that the laboratory data is not essential when studying drugs for treatment of diabetes, this will have broad implications as most observational datasets do not have laboratory data. Another limitation is the constraint that we predict outcomes for patients whom we assume to be invariantly on the assigned medication across the treatment period. These models do not allow us to predict what outcomes would be for patients who start and stop therapy. However, estimates from our methodology should approximate the upper limit of effectiveness of a medication, in which there is complete adherence to the assigned treatment.

We anticipate that these methods will increase the usefulness of large observational databases in future studies of the comparative effectiveness of new drugs and other treatments. Our methodology advances the use of a causal inference framework by generating potential outcomes using propensity scores generated from the covariates at multiple time points, with the predicted outcomes generated by regression models using the whole cohort of diabetic subjects. In this way, we have avoided an important bias associated with observational data, confounding of outcomes by observed covariates associated with treatment assignment. These methods have also let us make stable inferences about outcomes for a small treatment group; the actual number of patients on exenatide across all 3 time periods was small, but with our ability to "borrow power" from other subjects, we made predictions about their outcomes.

REFERENCES

1. Gabry K. Byetta™ Injection (exenatide) NDA No. 21-773 FDA Advisory Committee Briefing Document. Medical Review. http://www.fda.gov/cder/foi/nda/2005/021773_Byetta_medr.PDF.Accessed April 20, 2005.
2. Buse JB, Henry RR, Han J, et al. Effects of exenatide (exendin-4) on glycemic control over 30 weeks in sulfonylurea-treated patients with type 2 diabetes. *Diabetes Care*. 2004;27:2628–2635.
3. DeFronzo RA, Ratner RE, Han J, et al. Effects of exenatide (exendin-4) on glycemic control and weight over 30 weeks in metformin-treated patients with type 2 diabetes. *Diabetes Care*. 2005;28:1092–1100.

4. Fineman MS, Bicsak TA, Shen LZ, et al. Effect on glycemic control of exenatide (synthetic exendin-4) additive to existing metformin and/or sulfonylurea treatment in patients with type 2 diabetes. *Diabetes Care*. 2003;26:2370–2377.
5. Fineman MS, Shen LZ, Taylor K, et al. Effectiveness of progressive dose-escalation of exenatide (exendin-4) in reducing dose-limiting side effects in subjects with type 2 diabetes. *Diabetes Metab Res Rev*. 2004;20:411–417.
6. Heine RJ, Van Gaal LF, Johns D, et al. Exenatide versus insulin glargine in patients with suboptimally controlled type 2 diabetes: a randomized trial. *Ann Intern Med*. 2005;143:559–569.
7. Kendall DM, Riddle MC, Rosenstock J, et al. Effects of exenatide (exendin-4) on glycemic control over 30 weeks in patients with type 2 diabetes treated with metformin and a sulfonylurea. *Diabetes Care*. 2005;28:1083–1091.
8. Poon T, Nelson P, Shen L, et al. Exenatide improves glycemic control and reduces body weight in subjects with type 2 diabetes: a dose-ranging study. *Diabetes Technol Ther*. 2005;7:467–477.
9. Hebert PL, Geiss LS, Tierney EF, et al. Identifying persons with diabetes using Medicare claims data. *Am J Med Qual*. 1999;14:270–277.
10. Miller DR, Safford MM, Pogach LM. Who has diabetes? Best estimates of diabetes prevalence in the Department of Veterans Affairs based on computerized patient data. *Diabetes Care*. 2004;27(Suppl 2):B10–B21.
11. Neyman J. On the application of probability theory to agricultural experiments. Essay on principles. (1928) Section 9, Translated. *Stat Sci*. 1990;5:465–480.
12. Rubin DB. Estimating causal effects of treatments in randomized and non-randomized studies. *J Educ Psychol*. 1974;66:688–701.
13. Rubin DB. Bayesian inference for causal effects: the role of randomization. *Ann Stat*. 1978;6:34–58.
14. Robins JM. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *J Chronic Dis*. 1987;40(Suppl):139s–161s.
15. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41–55.
16. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11:550–560.
17. Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Stat Med*. 2004;23:2937–2960.
18. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Statist Ass*. 1984;79:516–524.
19. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression-coefficients when some regressors are not always observed. *J Am Stat Assoc*. 1994;89:846–866.

APPENDIX

We summarize here the theoretical justification for the use of the propensity score as a stratifier, depending on whether the model for outcomes is assumed to be correctly specified or not.

If the outcome model given the propensity score is *correctly* specified, then, it is a result of the Complete Class Theorem of decision theory (Fergusson 1967) that an estimator that depends on the propensity score, in ways other than through the model's sufficient statistics (eg, as does an estimator that uses the score as a weight), is inadmissible. It is always inferior to some estimator that depends on the score only through the sufficient statistics.

If the outcome model, given the propensity score, is allowed to be incorrectly specified, then theoretical arguments[19] and simulation evidence[17] suggest that in order for an estimator to be admissible, it should still use, at least in part, a possibly incorrect model of the outcome on treatment and the covariates. In this case it is important to note that the estimator's precision (not bias) generally deteriorates with increasing distance of the incorrect outcome model from the truth. Because this distance is expected to be large when trying to model too many covariates through an incorrect model, it becomes important that, even in this case, one makes use (eg, in the outcome component of the doubly robust estimator) of an outcome model given *the propensity score* as the single covariate per time point, as opposed to an outcome model *given all covariates*, when the latter can be mis-specified.

Of course, estimators relevant to the second case, that combine the propensity scores as weights together with outcome models that use covariates (eg, doubly robust), are different than usual ones. But the point is that if one wishes to use such combinations, it is important to use, for the second component of such combination, the outcome model given the propensity scores—not given the original covariates—as subclasses, particularly when the latter can be mis-specified. To do this, it is essential to know how to construct such a model, when, as in this article, interest lies in the effects of longitudinal treatments.

