New User Protocol for Diabetes, Insulin and Malignancies Study (DIMSum)
23Jan2012

1. Background

There have been several reports in the scientific literature purporting to show a relationship between the use of insulin to treat diabetes and the development of cancer. More recently, four publications appeared in the journal Diabetologia regarding the possibility that the long-acting insulin analogue glargine (Lantus) was specifically associated with an increased overall cancer risk compared with all other insulins. Overall these results have inconsistent findings across analyses, a small number of cancer cases, short periods of observation, and some methodological flaws.

2. Objective of the Study

The objective of the study is to investigate the causal effect of initiating glargine compared with human NPH insulin on the risk of cancer in people with diabetes using population-based databases.

3. Study Organization

The study is administered and coordinated by investigators at the University of North Carolina at Chapel Hill (UNC-CH). The principal investigator of the study is John Buse, MD, PhD, who is a Professor of Medicine, Chief of the Division of Endocrinology, and Executive Associate Dean for Clinical Research at the UNC School of Medicine. Dr. Buse is joined by Haibo Zhou, PhD, Professor of Biostatistics, and Investigator at the Collaborative Studies Coordinating Center, and Til Stürmer, MD, MPH, PhD, Professor of Epidemiology, head of the pharmacoepidemiology program and Director of the UNC Center of Excellence in Pharmacoepidemiology and Public Health.

The UNC-CH coordinating center supports the work of U.S. based study sites that have comprehensive, longitudinal databases of diabetic patients. UNC-CH's role on the project is to coordinate development of a common study protocol, ensure proper ascertainment of study subjects and suitability of data to conduct a meta-analysis, formulate ancillary research questions, and ensure the proper analysis and reporting of the study results.

Advisors to the study include representatives appointed by the American Cancer Society and the American Diabetes Association. The study is funded by a grant to UNC-CH from the company sanofi-aventis. UNC-CH funds collaborators from the study sites. Sanofi-aventis will have no role in formulating the final protocol, data analysis, interpretation of results, or manuscript(s) development for publication. The study is performed under a common protocol across the US study sites.

4. Data Sources

i) Ochsner Health System in Louisiana

A decade ago, the Chief Information Officer at Ochsner facilitated the development of the diabetes registry that now includes 30,000 patients. The registry has information from the institutional electronic medical record and also contains claims, laboratory, pharmacy and hospital data. Most of the diabetes patients in the registry have over five years of follow-up. Ochsner also has a cancer registry which will be linked to the diabetes registry for the analysis. Larry Blonde, MD, is PI of the Ochsner site.

ii) Partners HealthCare System in Massachusetts

The Massachusetts General Hospital Primary Care Practice-Based Research Network (PBRN) consists of a network of outpatient practices in eastern Massachusetts affiliated with Massachusetts General Hospital (MGH) and the Partners HealthCare System (PHS). The practices include three hospital-affiliated academic practices, four community health centers, and six private practice offices, together serving a wide range of communities and patient populations. PBRN practices share a common electronic health record containing all clinical and utilization data for each patient. Approximately 155,000 patients receive regular care from a primary care physician in the network. Patients with diabetes are identified using a previously validated algorithm. This diabetes cohort has about 15,700 patients (primarily type 2) who are under active care at MGH by a known primary provider. All prevalent and incident cancers of interest are identified by linkage with the Massachusetts General Hospital tumor registry. James Meigs, MD, is PI of the Partners site.

iii) MedAssurant

MedAssurant is a Maryland medical informatics company that provides analytic services to more than 200 health care organizations across the country with the goal of identifying opportunities to improve quality and outcomes of care and reduce healthcare costs. MedAssurant has built a large healthcare database, the MORE2 Registry, that contains data for over 76 million patients, 295,000 physicians, and 185,000 clinical facilities. The registry allows for tracking patients over time and across health plans (as long as data are available from all of the patient's health plans) and thus the usual follow-up time for patients in the MedAssurant database is longer than the follow-up time in databases relying on a single insurer. Katy Benjamin, PhD, is PI of the MedAssurant site.

iv) Solucia

Solucia is another medical informatics company that provides actuarial services to insurers, employers and health care providers. Their database contains data for over 215,000 diabetics with an average of two years of follow-up. They are located in Connnecticut and Tamim Ahmed is the PI of the Solucia site.

Though the sample sizes of the MedAssurant database and the Solucia database are substantially greater than that of the additional sites (Partners and Ochsner), the latter sites will provide opportunities to explore the issue of channeling of prescriptions to glargine or NPH insulin based on important potential confounders such as BMI and type of diabetes, discussed elsewhere.

5. Analysis Plan

The analysis strategy corresponds to a new user cohort design in which subjects enter the cohort at time of first use of insulin (glargine or NPH; see Section 5.2 for definition of first use), and baseline information is ascertained over a fixed period prior to first use. New user designs have been increasingly embraced by pharmaco-epidemiologists since 2003, when Ray published an article which argued that the new user design can reduce the potential for many of the biases likely present in a prevalent user design. One source of bias in a prevalent user design is that prevalent users of a given treatment are not a random sample of all the patients who began using this treatment at a particular time in the past. For example, intolerance to a treatment may lead to differential dropout, which results in a survivor cohort that is more likely to do well with the therapy (i.e. the "healthy user" effect). This is likely the reason that observational studies did not demonstrate an increased cardiovascular risk associated with hormone replacement therapy, whereas randomized trials did uncover this effect. It may also be the reason that researchers were led to believe that third generation oral contraceptives were associated with a higher rate of venous thromboembolism compared with second generation oral contraceptives.

The magnitude of the risks and benefits of drugs often vary over time after the start of treatment, which introduces bias into a prevalent user design; a new user design helps to control this bias. Another difficulty in the analysis of prevalent user designs stems from the fact that disease risk factors can be affected by the treatment itself. In a new user design, this difficulty is addressed because potential confounders can be measured prior to treatment initiation. Although the benefits of the new user design are well understood and

attractive, such designs are often not employed because of the logistical complexities of identifying new users and because of the loss of sample size and thus statistical power compared with a prevalent user design.

Data from Ochsner Health System and Partners HealthCare System will be extracted and transferred to UNC-CH for analysis. The Solucia database will also be analyzed by UNC-CH and the analysis of the MedAssurant database will be performed at MedAssurant.

As a first step in the analysis, UNC-CH used data from Partners HealthCare System to determine the extent of channeling based on BMI. Based on the data available from Partners as of May 2011, channeling associated with BMI appears to be minimal; thus, it is feasible to include claims data in this project. Although BMI is not available in claims data, the inclusion of the MedAssurant claims dataset ensures sufficient power for the new user analysis.

The respective analysis groups (UNC-CH and MedAssurant) will perform the new user analysis described in the subsections that follow. The site-specific results from the larger databases will then be combined in a meta-analysis of the hazard ratios of glargine versus NPH for each of the four cancer endpoints. Estimated hazard ratios from each of the data sources will be combined in the meta-analysis by inverse variance weighting. Site-specific results will be compared to examine site-to-site differences in hazard ratios (heterogeneity).

5.1 Study Population

For the analysis of data from Partners HealthCare System and Ochsner, individuals from their diabetes registries will be eligible to enter the new user cohort on January 1, 2005 (approximately) or later at the time of their first eligible prescription of either insulin glargine or NPH insulin. Diabetes will be defined in the MedAssurant database by having at least 1 diagnostic code (ICD-9 code 250.x) for diabetes during the 18 months prior to or on the date of the first eligible prescription of glargine or NPH insulin. The earliest date when individuals in the MedAssurant database will enter the new user cohort is July 1, 2004. Individuals must have at least 19 months of continuous membership and pharmacy benefits prior to new user cohort entry; if membership is not well defined, a prescription is required in each of four 6-month periods prior to the first eligible prescription. The cohort will be restricted to adults (aged 18 years and older), since the most common cancers rarely occur in children.

Cohort exclusion criteria

Individuals will be excluded from the cohort for the following:

- Less than 18 years of age.
- Prevalent cancer. Patients with evidence of a history of any cancer (except non-melanoma skin cancer) at cohort entry will be excluded.
- Use of insulin other than 1 prescription for a short-acting insulin in the 19 months prior to the first cohort defining insulin. A period of 19 months is used to allow for an estimated one month supply from any previous insulin prescription, a 6 month grace period, and a 12 month wash-out period.

For the primary new user analysis, evidence of prior cancer or prior surgeries must occur within the 19 months prior to the first prescription for a long-acting insulin in order for the patient to be excluded from the relevant analyses.

In addition, patients may be excluded from specific analyses as described below:

- For breast cancer, women with prophylactic unilateral or bilateral mastectomy at cohort entry will be excluded from the breast cancer analysis.

- For colon cancer, men and women with evidence of complete or subtotal surgical prophylactic removal of the colon will be excluded from the colon cancer analysis.
- For prostate cancer, men with a partial or complete prostatectomy for any reason will be excluded from the prostate cancer analysis.

Subsequent Instances of New Use

Within each database, patients who satisfy the inclusion and exclusion criteria more than once (i.e., have multiple instances of initiating insulin therapy after a 19-month wash-out period) will be flagged and enumerated (Appendix A, New User Algorithm). In the primary new user analysis, only instances of new use based on the first eligible prescription of glargine or NPH will be included.

5.2 Exposures of Interest

The analysis will be based on new users of insulin glargine and new users of human NPH insulin. Patients who use premixed NPH insulin will be included with NPH insulin users in the primary analysis.

In the MedAssurant analyses, the cohorts will be defined as of the second prescription of a long-acting insulin following the first eligible prescription (see Section 5.1 Study Population). The second prescription must occur within 6 months after the first eligible prescription and patients must receive the same type of long-acting insulin (glargine or NPH) at the first and second prescription. Patients with discrepant insulin prescriptions between their 1st and 2nd prescriptions will not be included in the primary analysis but they will be identified for possible inclusion in secondary analyses. In the Ochsner and Partners HealthCare analyses, only one record indicating the start of glargine or NPH will be required to enter the cohort since the electronic medical record databases do not keep a record of every subsequent prescription.

At the first eligible prescription following the cohort entry date, we will determine whether this is an instance of new use based on whether the patient had a prescription for insulin during the 19 months prior to the first eligible script. A 19-month evaluation period ensures that a patient has at least a 12-month insulin-free period prior to entering the new user cohort as it includes time for a 1 month supply following any prior insulin prescription plus a 6-month grace period plus a 12-month washout period. Note that a patient must have at least 19 months of medication data prior to their first prescription for glargine or NPH in order to be evaluated for the new user cohort; for patients with more than 19 months of data, only the latest 19 months of data will be used. Patients without any prescription for insulin during the 19 months prior to the first eligible script, or with only one prescription for a short-acting insulin during this time period, will be considered new users and will be included in the primary analysis.

5.3 Outcomes of Interest

Incident cases of new cancer will be identified from the respective databases. The three primary outcomes correspond to breast, prostate, and colon cancer individually. Only the primary cancer site will be used to determine whether the patient had breast, prostate or colon cancer. The secondary outcome will be all cancers combined excluding non-melanoma skin cancers. The only carcinoma in situ (CIS) cancers that will be included in the primary outcomes are cases of CIS breast cancer. For the analysis of claims data, an algorithm requiring two cancer diagnoses on different dates within a two month period will be required (Setoguchi et al, 2007).

5.4 Covariates

All covariates are defined based on available information within the 12-month period prior to and including the date of the first long-acting insulin prescription. It is important that this time is constant irrespective of the availability of additional data.

The minimum set of covariates includes age, gender, cohort entry date, co-morbidities (e.g., COPD as proxy for smoking), co-medication (including but not limited to other antidiabetic medications), health care system use (number of in- and outpatient encounters, days hospitalized), and screening behavior (especially cancer screening, see table). If available, smoking status, family history of cancer (overall and specific cancers, if available), body mass index, and duration of diabetes are also covariates of interest.

Potential confounders for specific types of cancers include use of hormone therapy and use of oral contraceptives.

Covariates by category are as follows:

Demographic and Background Covariates
Gender, age, BMI, blood pressure, race/ethnicity, smoking status at baseline, socioeconomic status, family history of cancer, duration since diabetes diagnosis, type of diabetes

Lab Results at Baseline
Glycemic control (HbA1c levels), lipid tests (triglycerides, HDL, LDL and total cholesterol), creatinine, ALT, hemoglobin, microalbumin, microalbumin/creatinine ratio

Co-morbidities at Baseline
History of pulmonary infection or COPD, congestive heart failure, renal insufficiency, diabetic retinopathy, diabetic nephropathy, diabetic neuropathy

Pharmacy data
Diabetes medication adherence, use of other medications of interest

Health care system use
Number of hospitalizations (for any reason), total number of days in the hospital, number of outpatient physician encounters (including same day procedures or surgeries), number of emergency department visits, cancer screenings (mammography, colonoscopies or sigmoidoscopies, PSA, pap smears, fecal occult blood tests), number of lipid assessments, number of ECGs

5.5 Baseline and follow-up (time at risk)

Baseline is defined as the 12-month time period up to and including the date of first prescription of glargine or NPH for new users; all covariates will be assessed based on data from this 12-month period. The start of follow-up for Ochsner and Partners HealthCare System patients will be the date of the first prescription. The start of follow-up for new users in the MedAssurant database will be defined for each cohort member as the date of the second prescription for glargine or the glargine comparator insulin. Note that the first and second prescriptions can be for different insulins as long as both insulins reside within the same group (i.e. glargine or NPH,).

The use of the second prescription allows us to remove patients who filled a prescription but were minimally or never exposed to the insulin products of interest. The potential for this kind of exposure misclassification is much smaller given the dispensing of a second prescription. The use of the second prescription as the start of follow-up results in the potential for changes in prescribed type of insulin (see above), covariates, and exclusion criteria. We will ignore changes in covariates because such changes may already be affected by the first prescription. Patients developing cancer between the first and second prescription will be flagged, enumerated, and excluded from the primary analysis.

Follow-up will generally end at: (1) diagnosis of any cancer (except non-melanoma skin cancer) for both overall cancer and specific subgroups, (2) death, (3) a gap of greater than 6 months in either membership or prescription benefits (if membership is not well defined, if no diagnostic code OR no prescription for

more than 6 months during follow-up; a final determination will be made following a review of the data), or (4) the end of the study (date to be determined with individual sites); whichever comes first. In addition, for the analyses on breast, colon, and prostate cancer risk, patient follow-up will be censored when the target organ is removed for a non-cancer related reason.

- For the breast cancer analysis, complete bilateral mastectomy;
- For the colon cancer analysis, removal of colon;
- For the prostate cancer analysis, complete prostatectomy, when done for reasons not related to prostate cancer.

Stopping, switching, augmenting: Because we are interested in potential harm, the primary analysis will be as treated. Therefore, in the primary analysis that compares new users of glargine with new users of NPH, patients stopping the corresponding drug (glargine or NPH), switching to another long-acting insulin (including detemir), or augmenting with another long-acting insulin will be censored at that point in time. Patients who take a premixed analog insulin during follow-up will be treated in a similar manner (i.e. censored at that point in time). Patients who do not refill their insulin prescription within a period equal to the total days of supply (plus a 6 months grace period) will be treated as if they stopped taking insulin at the end of that period (i.e. they will be censored at the last prescription date plus [the days of supply plus 6 months]). If no information on days of supply is available, an informed guess will be made based on the specific study site (i.e., median days of supply).

5.6 Statistical Analysis Methods

5.6.1. New User Analysis

The new user analysis (Ray 2003) will be based on a defined period of no insulin use prior to the first eligible prescription. To do so, all patients with a prescription of any insulin during the 19 months prior to the first eligible prescription will be excluded. The only exception will be for patients with only one prescription for a short-acting insulin. To allow for treatment titration and therefore better account for early switching, treatment augmentation, and non-chronic use, follow-up will not start at drug initiation but rather at the second prescription after initiation as outlined above. In cases where only one record is necessary to define the cohort (i.e. Ochsner and Partners HealthCare), follow-up will begin at drug initiation.

For the primary analysis, follow-up will start at the time-point defined above for all cohorts. To allow for time-varying hazard ratios (e.g., varying induction periods), we will stratify the follow-up into 0-<6 months, 6-<9 months, 9-<12 months, 12-<24 months, and 24+ months.

Augmentation with short-acting insulin during follow-up will be ignored for the primary analysis but will be described. To do so, time of a first prescription of a short-acting insulin will be ascertained as a separate variable (as well as specific formulation and dose).

The effect of cumulative insulin dose will be addressed by calculating the cumulative insulin dose as of every new prescription and categorizing patients into mutually exclusive categories of cumulative dose (e.g., >0-<=10kU, >10-<=20kU, >20-<=40kU, >40kU). We will categorize cumulative dose based on the long-acting insulin exclusively as well as on long-acting plus short-acting insulin combined. Time at risk for malignancies will start once a patient reaches the corresponding cumulative dose and end once a patient reaches the next cumulative dose level.

In the as treated new user design, follow-up time equals time on treatment since initiation and thus is an indicator of cumulative dose (while being less affected by intermediates). We will therefore also stratify follow-up time (e.g., 0-6, 6-12, 12-24, 24+ months; see above).

To adjust for potential confounding due to channeling between different long-acting insulins, we will estimate the propensity score for glargine initiation (primary analysis) and prevalent glargine use (secondary analyses) vs. glargine comparator initiation/use. In a first step, all covariates defined will be entered into the propensity score. Continuous or ordinal variables will be coded so as to be able to check

model assumptions (e.g., using dummy variables). Specific attention will be given to year of cohort entry because of the dynamics in channeling over time. Using complete cases only, we will empirically define the 10 covariates that are most strongly (Chi-square statistic) associated with channeling. We will then make a decision about whether these 10 covariates are risk factors for cancer based on substantive knowledge. We will then exclude covariates that are not risk factors for cancer. We will then re-estimate the propensity score based on the refined set of covariates. This re-estimated propensity score will be based on complete cases of the reduced set of retained covariates and thus may include more patients than the initial set of cases. Covariates in this re-estimated propensity score will be deemed confounders and we will thus exclude patients with missing values on these confounders to get an unbiased treatment effect estimate. If we lose more than 5% of our patients based on this strategy, we will consider multiple imputation methods of the covariates most strongly contributing to missingness before estimating the propensity score.

In addition, since BMI has the potential to be an important confounder given its association with the risk of breast and colon cancer, we will estimate the association between obesity (BMI >= 30 vs. < 30) and use of glargine vs. the glargine comparator. This estimate, in conjunction with information from the literature on the association between BMI and various cancer outcomes (Renehan et al., 2008) will be used to estimate the extent of potential confounding by unmeasured BMI in our main analyses (i.e. glargine vs. NPH and cancer risk) using a spreadsheet for unmeasured confounders developed by Schneeweiss et al (2005). We will use a cut point of a change in estimate of +/- 10% on the relative risk scale to define "relevant confounding by BMI" (Maldonado & Greenland, 1993).

We will implement the propensity score by stratification (deciles) and inverse probability of treatment weighting (IPTW). Balance of important covariates will be assessed within deciles of the propensity score and in the weighted pseudo population. Any imbalance will be interpreted according to the potential of the imbalanced covariate to affect the risk for the outcome. Given balanced covariates and under the assumption of no unmeasured confounding, incidence rates and survival curves are adjusted or unconfounded and thus can be directly compared. Time trends in absolute and relative hazards will be assessed by stratifying on months of follow-up time (0-6, 6-12, 12-24, 24+). The main measure of association will be the hazard ratio estimated using a Cox proportional hazards model controlling for age and sex as well as any covariates remaining imbalanced after implementation of the propensity score.

5.6.2 Prevalent User Analysis and Other Subgroup and Sensitivity Analyses

In addition to the new user analysis described above, each group (UNC-CH and MedAssurant) will conduct an analysis of prevalent users using the same dataset that was used for the new user analysis, except patients with prescriptions for insulins prior to the first eligible prescription will not be excluded. Additional details about this analysis can be found in ***Cohort study of insulin glargine and cancer risk among patients with diabetes mellitus (Kaiser Permanente, Northern and Southern California Regions)***. Results from the prevalent user analysis, including site-specific results and also combining across sites through meta-analysis, will also be reported. If the combined hazard ratios of glargine versus NPH differ substantially for the new user and prevalent user analyses for any of the four endpoints, then exploratory analyses will be carried out to explain the discrepancies.

For the new user analysis, sensitivity analyses will allow for varying latency periods of drug effects. To do so we will have additional variables censoring follow-up time 3 months, 6 months, 12 months, and 24 months after exposure defined censoring events (stopping, switching or augmenting) as described in Section 5.5. In these analyses, both events and person-time during these periods after stopping, switching, or augmenting will be counted towards the treatment preceeding these events. In addition, ITT analyses are also planned where we do not censor for augmentation, switching or stopping (i.e., counting all events until death or administrative censoring).

A complete list of planned subgroup and sensitivity analyses is included in Appendix C.

6. Power Calculations

Minimum detectable rate ratios for colon, breast, prostate and any cancer (except non-melanoma skin cancer) are presented below. The MedAssurant database contains approximately 47,227 new glargine users and 9899 new NPH users. Median follow-up time is approximately 2 years in each of the treatment groups. Age and sex-specific SEER cancer incidence rates were used to compute age and sex-adjusted cancer incidence rates that were standardized to the MedAssurant population. Additional adjustments were made to take into account cancer incidence rates in the diabetic population (Renehan, et al, The Lancet, June 2010). Using these cancer incidence rates, the following risk ratios are able to be detected with approximately 80% power at the $\alpha = 0.05$ level (two-sided test).

	Minimum RR (MedAssurant only)
Any cancer	1.2
Colon only	1.72
Breast (females)	1.53
Prostate (males)	1.58

When the MedAssurant data is pooled with the Solucia data, the power will increase. Ideally, this study would have power to detect a clinically meaningful hazard ratio of glargine versus NPH. Based on the June 2009 Diabetologia reports which examined the relationship between glargine and cancer, hazard ratios implicating glargine insulin as associated with cancer were generally greater than 1.5. The new user analysis has sufficient power to detect a hazard ratio less than 1.58 for three of the four endpoints.

7. Projected Timeline

Draft analysis results available	March 2012
Final analysis results available	June 2012
Manuscript submitted	Sept 2012

8. References

Maldonado G, Greenland S. *Simulation study of confounder-selection strategies*. American Journal of Epidemiology 1993;138:923-936.

Ray WA. *Evaluating medication effects outside of clinical trials: new-user designs.* Am. J. Epidemiol. (2003); 158 (9): 915-920.

Renehan, AG, Smith U,Kirkman MS. *Linking diabetes and cancer: a consensus on complexity*. Lancet (2010); 375: 2201-2201.

Renehan AG, Tyson M, Egger M, Heller RF, Zwalen M. *Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies.* Lancet (2008); 371: 569–578.

Schneeweiss S, Glynn RJ, Tsai EH, Avorn J, Solomon DH. *Adjusting for unmeasured confounders in pharmacoepidemiologic claims data using external information: the example of COX2 inhibitors and myocardial infarction.* Epidemiology. 2005;16:17-24.

Setoguchi S, Solomon DH, Glynn RJ, Cook EF, Levin R, Schneeweiss S. *Agreement of diagnosis and its date for hematologic malignancies and solid tumors between medicare claims and cancer registry data*. Cancer Causes Control 2007;18:561-9.

Appendix A New User Algorithm

Definitions/Conventions

Washout Period (WP) = minimum length of time that a patient must be drug-free prior to becoming eligible for the new user cohort

Grace Period (GP) = maximum length of time that a user can go without a drug before being considered discontinued from drug use

W_i = Days since start of washout period prior to 1st RX fill of i^{th} period of use for the patient

G_j = Days from last day covered by the j^{th} RX fill to the $(j+1)^{th}$ RX fill date

Cohort Eligibility

If W_1 > WP then patient's period of drug use is eligible for the new user cohort.

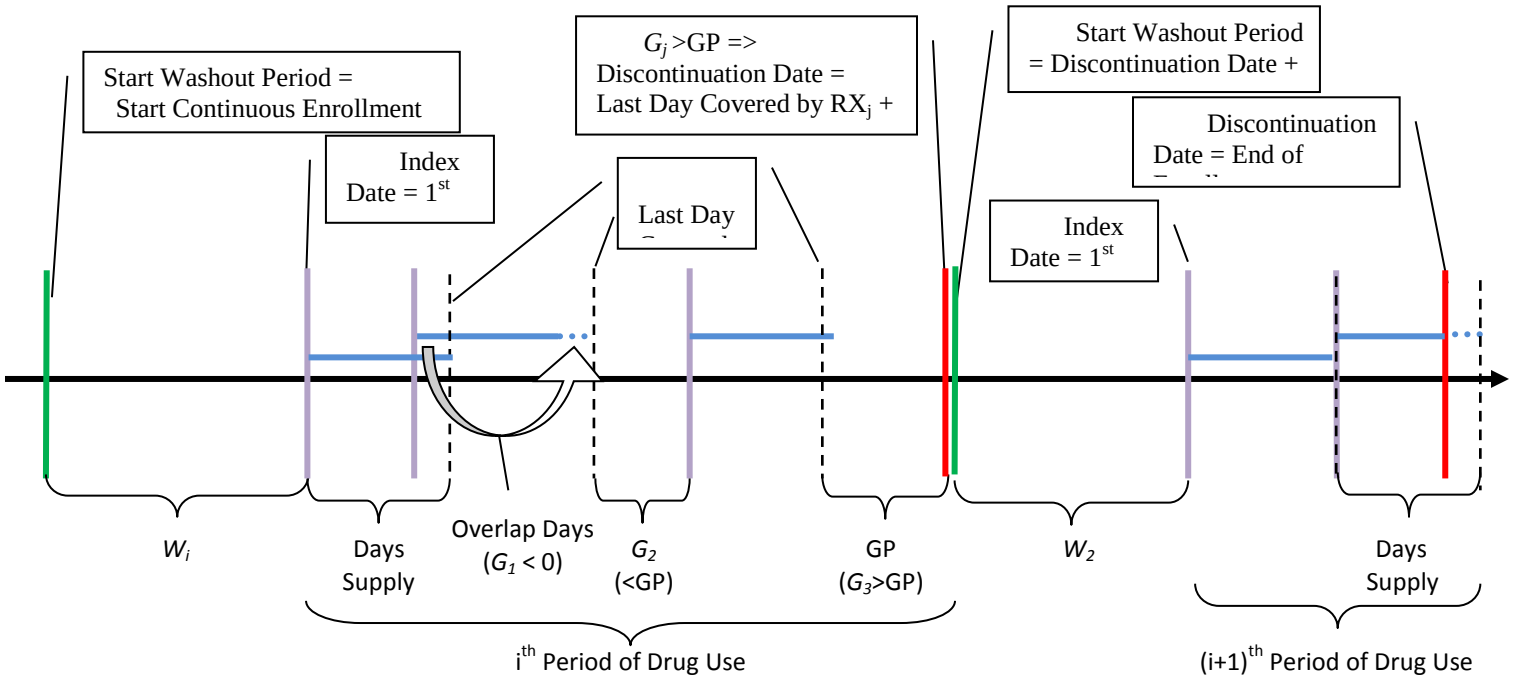
If W_i > WP and i>1 then patient's period of drug use is eligible for new user cohort **IFF** the analysis allows for previous users to become new users.

If i=1 then patient's period of drug use is eligible for the prevalent user cohort regardless of the value of W_1

Drug Discontinuation/Censor Date

If G_j > GP then the patient is considered discontinued from drug use on the last day covered by the j^{th} RX fill + GP

If (End of Enrollment) – GP < (Last Day Covered by an RX Fill) then the patient is censored at End of Enrollment



Algorithm

1. Set (Start of Washout Period) = (Start of Continuous Enrollment).
2. Set (Index Date) = (1st RX Fill Date following Start of Washout Period). Let *W* = (Index Date) – (Start of Washout Period). Flag the period of drug use for cohort eligibility based on the following criteria:
   a. If this is the patient's first RX of record, then flag as the index date for the prevalent user cohort.
   b. If this is the patient's first RX of record and *W* > (Washout Period), then flag as the index date for the primary new user cohort.
   c. If *W* > (Washout Period), then flag as the index date for the secondary new user cohort.
3. Set (Last Day Covered) = (Index Date) + (Days Supply).
4. Let *G* = (RX Fill Date) – (Previous Last Day Covered). Sequentially cycle through the subsequent prescription claims for the patient, applying the appropriate step below, until (Discontinuation Date) is set:
   a. If *G* > (Grace Period) then set (Discontinuation Date) = max(Previous Last Day Covered, Date of RX Fill) + (Days Supply) + (Grace Period).
   b. If *G* <= (Grace Period), set (Last Day Covered) = max(Previous Last Day Covered, Date of RX Fill) + (Days Supply). If (Last Day Covered) + (Grace Period) > (End of Continuous Enrollment) and the patient has no additional RX claims with (RX Fill Date) <= (End of Continuous Enrollment), then set (Discontinuation Date) = (End of Continuous Enrollment). Otherwise, repeat Step 3 for the next prescription.
5. Output a record containing Index Date, Discontinuation Date, and the cohort inclusion flags set in Step 2.
6. Set (Start of Washout Period) for the next period of use:
   a. If the patient is continuously enrolled from (Discontinuation Date) to the next (RX Fill Date), set (Start of Washout Period) = (Discontinuation Date) + 1.
   b. If the patient has a gap in enrollment between (Discontinuation Date) and the next (RX Fill Date), set (Start of Washout Period) = (Start of Next Period of Continuous Enrollment).
7. Repeat Steps 2-7 for the patient's remaining RX fills.

Appendix B
Comparison of Analysis Definitions by Data Source

	MGH	Ochsner	Claims
Identifying patients with diabetes	Validated algorithm	Registry	One diagnosis code of 250.x within the one year period prior to the start of NPH/glargine
Exposure definition and Inclusion/exclusion criteria	One record indicating glargine or the glargine comparator was started. Must be >=18, no history of cancer (except non-melanoma skin cancer); patients who have had unilateral or bilateral mastectomies, colectomies or prostatectomies (for reasons other than cancer) will be excluded from the analyses of breast, colon and prostate cancer, respectively	One record indicating glargine or the glargine comparator was started. Must be >=18, no history of cancer (except non-melanoma skin cancer); patients who have had unilateral or bilateral mastectomies, colectomies or prostatectomies (for reasons other than cancer) will be excluded from the analyses of breast, colon and prostate cancer, respectively	Two prescriptions for the same medication within 6 months are required. Must be >=18, no history of cancer (except non-melanoma skin cancer); patients who have had unilateral or bilateral mastectomies, colectomies or prostatectomies (for reasons other than cancer) will be excluded from the analyses of breast, colon and prostate cancer, respectively. Evidence of a history of cancer, mastectomy, colectomy or prostatectomy must be present in the 18 month baseline period.

		MGH	Ochsner	Claims
Cancer outcomes		Cancer registry	Cancer registry	Validated algorithm according to Setoguchi et al
Follow-up/censoring primary analysis		Follow-up starts at first record indicated start of glargine/NPH. Censoring at stopping, switching, augmenting (except short acting insulin), any cancer (except non-melanoma skin cancer), death, end of database, end of eligibility	Follow-up starts at first record indicated start of glargine/NPH. Censoring at stopping, switching, augmenting (except short acting insulin), any cancer (except non-melanoma skin cancer), death, end of database, end of eligibility	Start follow-up at second script after exclusion of incident cancer between first and second script. Censoring at stopping, switching, augmenting (except short acting insulin), any cancer (except non-melanoma skin cancer), end of database, end of eligibility
Follow-up/censoring secondary analyses		Various lag periods after stopping, switching, augmenting up to ITT analysis Stratification by time since treatment initiation	Various lag periods after stopping, switching, augmenting up to ITT analysis Stratification by time since treatment initiation	Various lag periods after stopping, switching, augmenting up to ITT analysis Stratification by time since treatment initiation
Covariates		To be defined based on data within 12 months prior and including the date of initiation of glargine/NPH	To be defined based on data within 12 months prior to and including the date of initiation of glargine/NPH	To be defined based on data within 12 months prior to and including the date of initiation of glargine/NPH
	Demographic/Baseline	Age Gender BMI Race/ethnicity Smoking status Diabetes type Duration of diabetes	Age Gender BMI Race/ethnicity	Age Gender
	Other medication use	Metformin	Metformin	Metformin

		MGH	Ochsner	Claims
		TZDs Sulfonylureas Other hypoglycemic Estrogen Progestins Oral contraceptives Testosterone Statins Bile acid sequestrants Fibrates Niacin Other cholesterol Anticholinergic Beta-2 agonist Beta blocker Theophylline Corticosteroid Cardiac glycoside ACE/ARB Diuretic (loop) Diuretic (non-loop) Antidepressant CCB	TZDs Sulfonylureas Other hypoglycemic Estrogen Progestins Oral contraceptives Testosterone Statins Bile acid sequestrants Fibrates Niacin Other cholesterol Anticholinergic Beta-2 agonist Beta blocker Theophylline Corticosteroid Cardiac glycoside ACE/ARB Diuretic (loop) Diuretic (non-loop) Antidepressant CCB	TZDs Sulfonylureas Other hypoglycemic Estrogen Progestins Oral contraceptives Testosterone Statins Bile acid sequestrants Fibrates Niacin Other cholesterol Anticholinergic Beta-2 agonist Beta blocker Theophylline Corticosteroid Cardiac glycoside ACE/ARB Diuretic (loop) Diuretic (non-loop) Antidepressant CCB
	Lab Results	HbA1c Lipids, Creatinine, ALT, hemoglobin, microalbumin, microalbumin/creatinine ratio	HbA1c Lipids, Creatinine, ALT, hemoglobin, microalbumin, microalbumin/creatinine ratio	Not available
	Co-morbidities	COPD/pulmonary infection CHF Retinopathy Nephropathy Neuropathy	COPD/pulmonary infection CHF Retinopathy Nephropathy Neuropathy	COPD/pulmonary infection CHF Retinopathy Nephropathy Neuropathy

		MGH	Ochsner	Claims
	Health care system use	Number of hospitalizations Number of days in hospital Number of ED visits Number of physician visits (including same day procedures or surgeries) PSA Mammography Endoscopy Pap smear Lipid assessments ECG	Number of hospitalizations Number of days in hospital Number of ED visits Number of physician visits (including same day procedures or surgeries) PSA Mammography Endoscopy Pap smear Lipid assessments ECGs	Number of hospitalizations Number of days in hospital Number of ED visits Number of physician visits (including same day procedures or surgeries) PSA Mammography Endoscopy Pap smear Lipid assessments ECGs

Appendix C
Table of Planned Secondary and Sensitivity Analyses

Rank	**Primary Analysis**	**Secondary Analyses**	**Rationale**
Secondary Analyses			
1	Incident users	Perform analysis of prevalent users	Increase number of patients in study; confirm that results are similar to the new user analysis results.
2	All patients with diabetes	Exclude patients less than 40 years old or who do not have at least two prescriptions for oral hypoglycemic medications not more than 6 months apart during the baseline period.	Patients who meet these criteria will nearly all be Type II diabetics.
Sensitivity analyses			
1	Entire available follow-up	Stratify follow-up into 0-6 months, 6-12 months, 12-24 months and >24 months after baseline	Allow for changing hazard ratios with time after initiation
2	Baseline (date of second prescription)	Begin follow-up 6 months, 9 months, 12 months and 24 months after baseline	Allow for varying induction periods Note that this analysis is very similar but not exactly the same as the above

Rank	Primary Analysis	Secondary Analyses	Rationale
3	Censor immediately at stopping (plus grace period), switching, augmenting	Censor follow-up time 3 months, 6 months, 12 months and 24 months after a censoring event* (stopping, switching or augmenting). Both events and person-time during these periods after stopping, switching, augmenting will be counted towards the treatment preceding these events. * Note that this applies only to censoring due to stopping, switching, augmenting; all other censoring (e.g., eligibility, calendar year) remain unchanged	Allow for varying latency periods for treatment effects
4	As treated	Do not censor follow-up time after a censoring event* (ITT analysis) * Note that this applies only to censoring due to stopping, switching, augmenting; all other censoring (e.g., eligibility, calendar year) remain unchanged	Minimize the potential for selection bias
5	Exclude only non-overlapping propensity score regions	Exclude 1%, 2.5%. and 5% of those in the tails of the overlapping propensity score distribution, i.e., asymmetric trimming of those treated contrary to prediction	Most likely to have unmeasured confounders

Table of Additional Secondary and Sensitivity Analyses
(To be Performed If Resources Allow)

Rank	Primary Analysis	Secondary Analyses	Rationale
Secondary Analyses			
1	NPH insulin	Compare new glargine users with new users of other long-acting human or analog insulins	Increase the size of the glargine comparator group
2	All ages	Censor patients when they reach age 85	Older patients may not receive the same diagnostic testing as younger patients
3	In-situ included	Exclude all cases of carcinoma in situ in the outcome of all cancers	Confirm that results are similar if these cases are excluded
4	Colon cancer	Outcome of colorectal cancer	Compare with the results of the Northern European study
5	Rate ratios and hazard ratios	Estimate absolute rates and rate differences (rather than ratios) for all outcomes	Allow for discussion of absolute rather than relative rates
Sensitivity analyses			
1	6 months grace period for stopping	No new prescription within days supply plus 3 month and 1 month grace periods before patient is classified as having stopped and censored	Shorter grace periods will reduce both misclassification of exposure and median follow-up time
2	Limit to first occurrence (episode) of new use in database	Examine instances where a patient satisfies the inclusion and exclusion criteria more than once (i.e., have multiple instances of initiating insulin therapy after a 19-month period). If a sufficient number of such cases exist,	Confirm that results are similar if these cases are included

		a sensitivity analysis which includes all these cases may be performed. Note that this may also apply to initially prevalent users	
3	Include only patients with concurrent prescriptions within 6 month (note: group concurrence sufficient for comparator cohort)	Include patients with discrepant insulin prescriptions between their 1st and 2nd prescriptions. If a sufficient number of such cases exist, a sensitivity analysis which includes all these cases may be performed.	Confirm that results are similar if these cases are included
4	Start follow-up (person-time) at time of second prescription	Define follow-up as beginning at the date of first prescription for glargine or NPH rather than the second prescription	Confirm that results are similar when follow-up begins at first prescription
5	Estimate propensity score for glargine initiation over all calendar years	Estimate calendar-year specific rather than marginal propensity score models and run propensity score adjusted analyses stratified by calendar year	Allow for changes in treatment decisions over time other than increasing use of glargine
6	Censor patients at first occurrence of a nonspecific cancer code even if they have subsequent records for a specific cancer of interest	Ignore nonspecific cancer codes when defining incident colon, prostate or breast cancer cases	Confirm that results are similar