**CNODES**

**Technical Analytical Protocol:**

**Phase IIa**

**The use of incretin-based drugs and the risk of pancreatic cancer in patients with type 2 diabetes**

**Version 1.8**

**Date Created:** September 15, 2014

**Date of Last Revision:** December 11, 2014

Table of Contents

[I. Document Control 3](#_Toc401056800)

[II. Pancreatic Cancer Case-Control Selection 4](#_Toc401056801)

[III. Exposure Definition 7](#_Toc401056802)

[IV. Confounders 9](#_Toc401056803)

[V. Statistical Analysis 10](#_Toc401056804)

[1. Primary Analyses 10](#_Toc401056805)

[2. Secondary Analyses 11](#_Toc401056806)

[3. Sensitivity Analyses 20](#_Toc401056807)

[4. Reduced Model Analyses 23](#_Toc401056808)

[5. Quality Assurance Analyses 25](#_Toc401056809)

[APPENDIX I: Site-specific protocol deviations 26](#_Toc401056810)

[APPENDIX II: Amendments 27](#_Toc401056811)

[APPENDIX III: Questions & Answers 31](#_Toc401056812)

[APPENDIX IV: Glossary of Terms 32](#_Toc401056813)

# Document Control

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Author(s)** | **Type of Change** |
| 1.0 | September 15, 2014 | MD | Initial version of Analytical Protocol Phase IIa (case-control selection, exposure definition, statistical analysis, secondary and sensitivity analyses) |
| 1.1 | October 1, 2014 | LS, LA, KBF, MD, PE, RP | Revision of initial version of Analytical Protocol Phase IIa. |
| 1.2 | October 16, 2014 | LS, LA, KBF, PE | Incorporation of project team’s comments, addition of quality assurance (simulation) analyses. |
| 1.3 | October 20, 2014 | LS, LA, KBF, PE | Clarifications for quality assurance (simulation) analyses added, definition of cumulative duration in Glossary of Terms corrected, Excel workbooks revised (cell locking). P-values for cumulative duration and time-since-initiation added to protocol and Excel workbooks. |
| 1.4 | October 24, 2014 | MD, KBF | Addition of SAS code for p-values for test of heterogeneity for duration and time-since-initiation analyses. Removal of type III tests for duration and time-since-initiation analyses. |
| 1.5 | November 13, 2014 | LS, KBF | Updated Excel Workbook Table 1. A note was added to specify that cases with no controls are excluded. Maximum caliper for duration of diabetes was increased to 365 days. |
| 1.6 | November 27, 2014 | LS, KBF, LA | The Deyo version of the Charlson Comorbidities Index has replaced the Romano version in Reduced Model #2. Notes were added to sections IV and V.4 to clarify the weighting of controls tables and to clarify how the Deyo score should be implemented into our diabetic cohorts and how sites with only 3 digits should adapt the score. The protocol deviations appendix was also revised and a protocol deviations form was added to Dropbox. |
| 1.7 | December 8, 2014 | LS, KBF, LA | A note regarding the weighing of controls in Tables 2 and 3 was corrected to specify that this is done in Table 2 (Section IV). Additional information regarding the Deyo macro was added. A note to specify that pre-cohort entry anti-diabetic drugs should not be included as covariates in the primary analysis was inserted. FAQs and corresponding answers were updated. |
| 1.8 | December 11, 2014 | KBF, LA | Previous amendment regarding pre- cohort entry anti-diabetic drugs corrected. Revised Deyo macro posted, which now examines ever history rather than only diagnostic codes in the year before cohort entry. |

**Note:** Text in **red** fontin the analytical protocol refers to Excel worksheets that should be completed. See file **“CNODES Incretins Phase IIa Results Tables (November 13, 2014).xlsx”** located in Dropbox.

**Note:** Amendments in the analytical protocol are indicated in **blue** font. See Appendix II: Amendments for specific details.

**Note:** For terminology clarifications, see Glossary in Appendix IV.

# Pancreatic Cancer Case-Control Selection

1. Please refer to Analytical Protocol Phase 1 Version 1.6 sections VII.i and VII.ii for a description of the cohort construction, including definitions of **study cohort entry** and **cohort exit** (defined by the end of follow-up). Construct a duration of follow-up time variable using the following definition:

Follow\_up\_dur = Study Exit date – (Study Cohort Entry Date + 365) +1

1. Cases will be defined as all patients with incident pancreatic cancer during follow-up, using the pancreatic cancer event definition described in Analytical Protocol Phase 1 Version 1.6 section VII.ii. Create a variable to identify the case series:

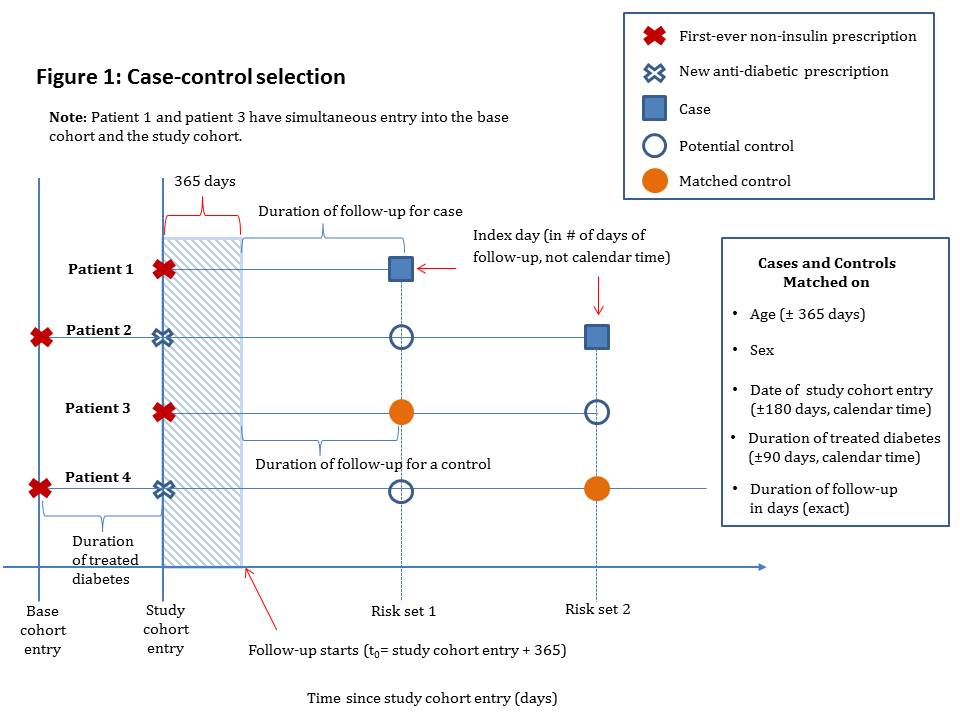
* If pancreatic cancer present: case=1;
* If pancreatic cancer absent: case=0;

1. Up to 20 controls will be randomly selected from the case’s risk set, with the risk set constructed on the follow-up time axis and the hospital admission date (on the follow-up time axis) defining the index day. (Although originally titled ‘index date’, it has been changed to ‘index day’ to reflect the follow-up time axis, rather than calendar time axis). In addition, cases will be matched to controls on age (see Table 1 below for matching calipers), sex, date of study cohort entry, duration of treated diabetes, and duration of follow-up. Controls will be assigned the index day such that they have the same duration of follow-up as their matched cases. To accomplish this matching, perform the following steps for each case of pancreatic cancer (case=1):
   1. Identify all potential controls (regardless of case status) who were at-risk of having the event in the risk set defined by the case. To be a potential member of the risk set, patients must be in the cohort at the follow-up time of the cases’ index day. Consequently, the duration of follow-up (Follow\_up\_dur) for controls must be greater than the duration of follow-up for the case. From these potentially eligible controls, identify all controls within the risk set that meet the matching criteria for age at study cohort entry, sex, date of study cohort entry, duration of treated diabetes, and duration of follow-up at the time of the risk set. The matching calipers are defined in Table 1, and this matching process is described in Figure 1.

**Table 1: Matching Criteria**

|  |  |  |
| --- | --- | --- |
| **Matching Variable** | **Caliper (Days)** | **Maximum Caliper (Days)** |
| Age | ±365 | ±1825 |
| Sex | Exact | Exact |
| Study Cohort Entry Date | ±180 | ±365 |
| Duration of treated diabetes | ±90 | ±365 |
| Duration of follow-up | Exact\* | Exact\* |

\* Although controls must have a longer duration of follow-up than cases (i.e., they must be at risk for the event on the case’s index day), they inherit the index day from their matched case.  Therefore, the duration of follow-up for the study (from start of follow-up to index day) will match exactly for the case and matched control.



* 1. Generate a random number between 0 and 1 for each patient from step 3a. For example, the RANUNI function in SAS generates random numbers between 0 and 1 from a uniform distribution. Use a positive, non-zero seed value. If you use a positive seed value, you can always replicate the stream of random numbers by using the same DATA step. In addition, create a variable for this pool of eligible controls such that control=1.
  2. Sort all patients from step 3b by ascending random number.
  3. Select as the controls for each case the first 20 non-cases in the sorted list of patients from step 3 who have the same age within ±365 days, sex, date of study cohort entry within ±180 days, duration of treated diabetes within ±90 days, and duration of follow-up at least as long as the case’s duration of follow-up.
     1. **Note:** For cases where there are fewer than 20 patients in the list to serve as controls (with the same matching factors), select all eligible controls for these cases.
     2. **Note:** For cases where there are no eligible controls that meet the matching criteria, then for those cases only, widen the age matching criterion from ±365 days to ±730 days to obtain at least one match to avoid losing the case. You may continue to increase the age caliper by ±365 day increments until a maximum of ±1825 days (5 years). If you still have cases without any controls, widen the date of study cohort entry caliper to ±365 days (re-set the age caliper to ±365 days, and increase to ±1825 days as necessary). If you still have cases without any controls, widen the duration of treated diabetes caliper to ±365 days (re-set the age caliper to ±365 days and cohort entry caliper to ±180 days, and increase them to ±1825 days and ±365 days, respectively, as necessary). Please document the number of cases and what the calipers were set to in the site-specific protocol deviations.
     3. **Note:** It is okay if a case serves as another case’s control so long as the patient was free of the outcome (i.e., in the risk set) at the time he/she was selected as a control, and only later became a case.
     4. **Note:** It is okay if a patient serves as a control for more than one case.
     5. **Note:** If you have carried forward the case variable, you will need to make sure to change case equal to 0 if a case is serving as another case’s control (only where it is acting as a control though). This can be done by saying if control=1 then case=0.

1. Assign each control within the matched set the **Index day** of their corresponding case. Thus, the index day for both the case and their corresponding matched controls will be identical and they will thus have equal person-time at risk (Index day – (Study Cohort Entry Date + 365) +1).
2. For each case, assign a matching number (**Match\_Num**) to the case and its 20 controls. For example, the first case and 20 matched controls will be assigned Match\_num=1. The second case and matched controls will be assigned Match\_Num=2.

**Note:** If you want to use a macro for matching, do not use the %match since it does not allow a patient to serve as a control for more than one case. Instead you should use and modify the template Matching\_replacement.sas to satisfy this study’s matching criteria. The template can be found in SAS macros.zip in the “Incretins Project Team\Analytic protocols\” Dropbox folder.

1. Fill in table describing number of controls per case **TABLE 1**.

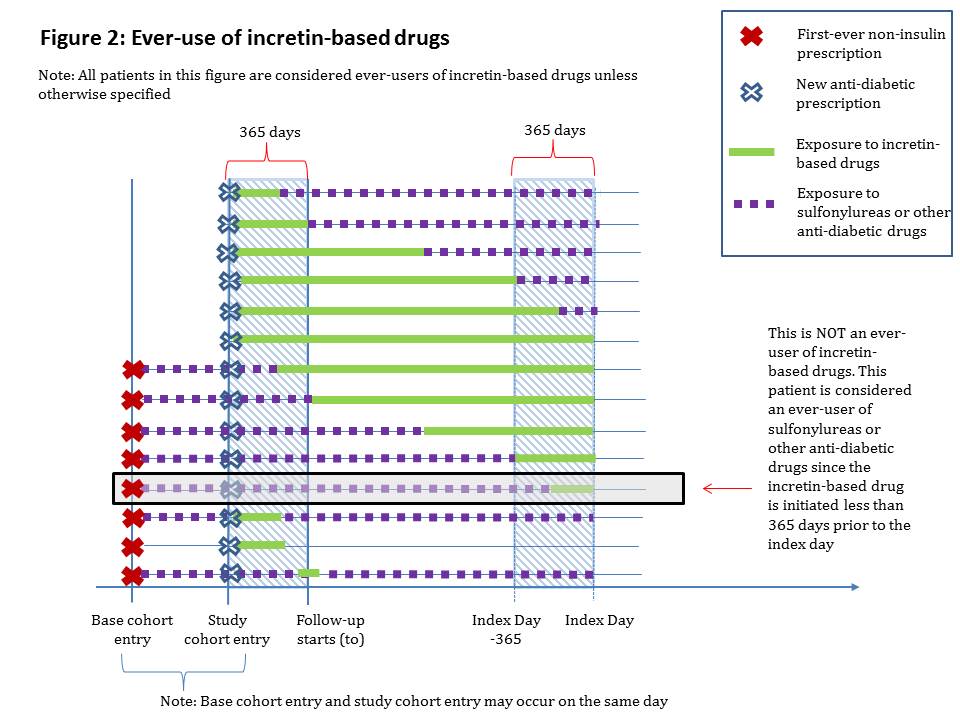
**Note**: All cases with no available controls should be excluded at this point and should not be included in subsequent analyses or data tables.

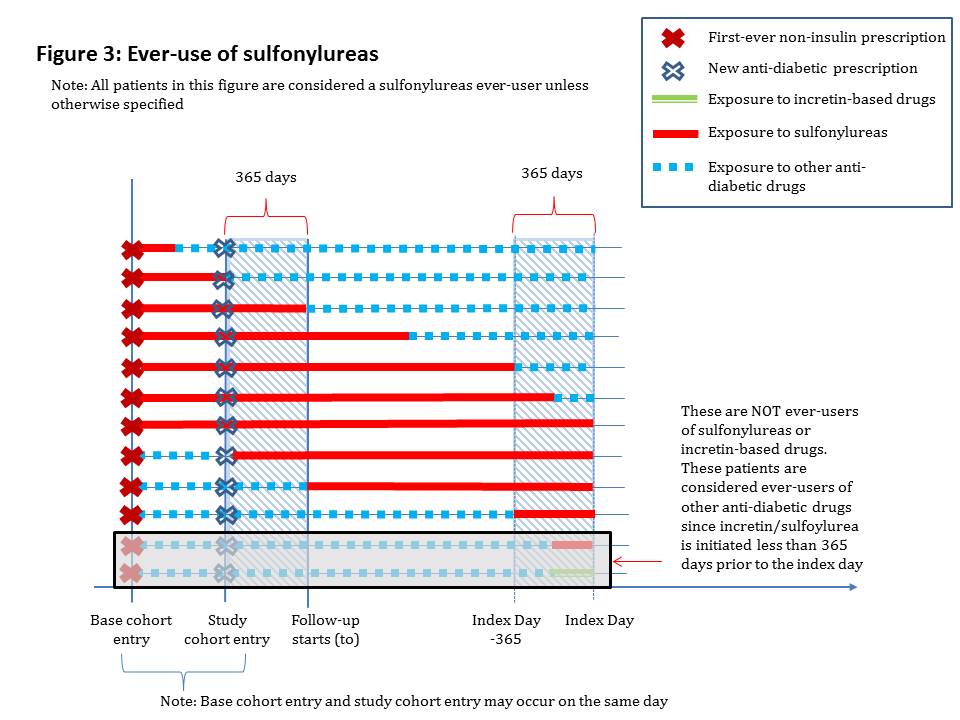
# III. Exposure Definition

The following definition of anti-diabetic agent exposure will be used for the case-control analysis. Since we are studying pancreatic cancer, exposure will be defined by **ever-use** between base cohort entry and the (index day – 365 days). We will apply a 365-day lag period to account for disease latency and potential protopathic bias by only considering prescriptions received from (and including) the Base Cohort Entry Date until (and including) the date one year prior to the Index day.

Cases and matched controls will be classified into one of the three mutually-exclusive exposure groups defined **hierarchically** (see Figures 2 and 3), and thus requires the following exposure categories to be defined in order:

1. Incretin-based drugs (**Exposure=’1’**), defined by ever-use (at least one prescription) of any of the following drugs between (and including) base cohort entry and (index day – 365 days):
   1. DPP-4 inhibitors [sitagliptin, vildagliptin, saxagliptin][ATC A10BH, A10BD07-A10BD13]; or
   2. GLP-1 analogs [exenatide, liraglutide][ATC A10BX04, A10BX07];
2. Sulfonylureas (**Exposure=’2’**) [This will be the reference category], defined as:
   1. No ever-use of incretin-based drugs (i.e., those with exposure ≠ 1); and
   2. Ever-use (at least one prescription) of any of the following drugs between (and including) base cohort entry and (index day – 365 days):ATC A10BB, A10BC;
3. Other anti-diabetic agents (**Exposure=’3’**):
   1. All patients with never-use of incretin-based drugs and never-use of sulfonylureas (i.e., those with: **Exposure≠’1’** and **Exposure≠’2’**)





# IV. Confounders

All baseline covariates were defined in Section VIII.i of the Phase I of the Analytical Protocol version 1.6.

1. Fill in cases’ and controls’ demographics (weighted and unweighted) in **TABLE 2**. Use weights for controls only, defined as the inverse number of controls per matched case.
2. Fill in ever-users of incretin-based drugs and ever-users of sulfonylureas among the controls demographics in **TABLE 3**. All demographic and clinical characteristics were defined in Analytical Protocol Version 1.6 Section VIII.

**Note:** This does not involve resampling or any patient selection. These characteristics simply need to be calculated by restricting the analysis to controls and stratifying analyses by exposure status.

**Note:** The weighting in Table 2 apply to the percentages only and not to the count data.

# V. Statistical Analysis

### Primary Analyses

For the primary analysis, conditional logistic regression will be used to estimate ORs and corresponding 95% CIs of pancreatic cancer, comparing ever-use of incretin-based drugs to ever-use of sulfonylureas, which will serve as the reference group. The models will be crude and adjusted for the potential confounders listed in Section IV (except matching variables). Please complete **TABLE 4** with the results of your primary analysis. Please provide SAS output in html format, named “Incretins – Pancreatic Cancer – Primary Analysis – Site”, replacing “Site” with your site initials.

**Note:** For continuous variables listed in Section IV, please use the categorical version of the variable in the primary analysis. The continuous versions will be used in reduced models described below.

**Note:** Matching variables need not be included as covariates in the models.

Model to Produce Crude Odds Ratios:

PROC LOGISTIC DATA=<datasetname> descending;

CLASS EXPOSURE (REF=’2’) / PARAM=REF;

MODEL PANCREATIC\_CANCER = EXPOSURE;

STRATA MATCH\_NUM;

CONTRAST ‘INCRETIN VS SULFONYLUREA’ EXPOSURE 1 0 / ESTIMATE=BOTH;

CONTRAST ‘ANY OTHER VS SULFONYLUREA’ EXPOSURE 0 1 / ESTIMATE=BOTH;

RUN;

Model to Produce Fully Adjusted Odds Ratios:

PROC LOGISTIC DATA=<datasetname> descending;

CLASS EXPOSURE (REF=’2’) <*all\_categorical\_variables>* / PARAM=REF;

MODEL PANCREATIC\_CANCER = EXPOSURE <*all\_confounder\_variables>*;

STRATA MATCH\_NUM;

CONTRAST ‘INCRETIN VS SULFONYLUREA’ EXPOSURE 1 0 / ESTIMATE=BOTH;

CONTRAST ‘ANY OTHER VS SULFONYLUREA’ EXPOSURE 0 1 / ESTIMATE=BOTH;

RUN;

**Note:** Some sites will have a small number of cases in their case-control analyses. For this reason, we will repeat our primary analysis using two reduced models (described in Section V.4 below).

**Note:** Sites for which models do not converge (either full or reduced models) must note this information in the site-specific protocol deviations (Appendix I).

### Secondary Analyses

Secondary analyses will focus on repeating the primary analysis with sub-categories of incretin-based drugs, with these sub-categories used to examine the relationship between the risk of pancreatic cancer and 1) cumulative duration of incretin-based drug use and 2) time since initiation of incretin-based drug use. These analyses should be conducted among all patients included in the primary analysis. There will be 5 secondary analyses to be completed.

#### Analysis by type of incretin-based drug

Sub-classify ever-use of incretin-based drugs as DPP-4 inhibitors (sitagliptin, vildagliptin, linagliptin) and GLP-1 analogs (exenatide, liraglutide) using the following **hierarchical** exposure definitions:

1. DPP-4 inhibitors (**Type\_Exposure=’1’**), defined by ever-use (at least one prescription) of any of the following drugs between (and including) study cohort entry and (index day – 365 days): sitagliptin, vildagliptin, or saxagliptin [ATC A10BH, A10BD07-A10BD13];

**Note:** These users may have been exposed to GLP-1 analogs;

1. GLP-1 analogs (**Type\_Exposure=’2’)** , defined by :
   1. No ever-use of DPP-4 inhibitors (i.e., **Type\_Exposure≠’1’**); and
   2. Ever-use (at least one prescription) of any of the following drugs between (and including) study cohort entry and (index day – 365 days): exenatide or liraglutide [ATC A10BX04, A10BX07];
2. Sulfonylureas (**Type\_Exposure=’3’**) [This will be the reference category], defined as:
   1. No ever-use of DPP-4 inhibitors or GLP-1 analogs (those with **Type\_Exposure≠’1’** and **Type\_Exposure≠’2’**); and
   2. Ever-use (at least one prescription) of any of the following drugs between (and including) base cohort entry and (index day – 365 days): ATC A10BB, A10BC;
3. Other anti-diabetic agents (**Exposure=’4’**):
   1. All patients not meeting one of the other 3 type\_exposure categories, i.e., no ever-use of DPP-4 inhibitors, GLP-1 analogs, or sulfonylureas (those with **Type\_Exposure≠’1’**, **Type\_Exposure≠’2’**,and **Type\_Exposure≠’3’**).

Conditional logistic regression will be used to estimate ORs and corresponding 95% CIs of pancreatic cancer, comparing ever-use of DPP-4 inhibitors and GLP-1 analogs to ever-use of sulfonylureas, which will serve as the reference group. The models will be crude and adjusted for the potential confounders listed in Section IV. Please complete **TABLE 5** with the results of this secondary analysis, using the crude and adjusted models provided for the primary analysis but with:

1. Type\_Exposure as the exposure variable;
2. The reference category set to Type\_Exposure=’3’;
3. Using the following contrast statements:

* CONTRAST ‘DPP-4 VS SULFONYLUREA’ TYPE\_EXPOSURE 1 0 0 / ESTIMATE = BOTH;
* CONTRAST ‘GLP-1 VS SULFONYLUREA’ TYPE\_EXPOSURE 0 1 0 / ESTIMATE = BOTH;
* CONTRAST ‘ANY OTHER VS SULFONYLUREA’ TYPE\_EXPOSURE 0 0 1 / ESTIMATE = BOTH;

Please provide SAS output in html format, named “Incretins – Pancreatic Cancer – Type Analysis – Site”, replacing “Site” with your site initials.

**Note**: For sites that do not have both GLP-1 analogs and DPP-4 inhibitors, there is no need to run this analysis.

#### Analysis by cumulative duration of incretin-based drugs

The second will assess whether the risk of pancreatic cancer varies with cumulative duration of incretin-based drug use (i.e., reclassifying ever-use of incretin-based drugs in primary exposure definition). For this analysis, cumulative duration of use among ever-users will be calculated by summing the days’ supply from all incretin-based drug prescriptions from, and including study cohort entry date until, and including, index day (Figure 4).

**Note:** The cumulative duration of use analysis will be conducted among patients deemed to be ever-users of incretin-based drugs in the primary analysis.

**Note:** If a patient receives more than one incretin-based drug prescription on the same day or another prescription before the end of the previous prescription’s duration, the duration of both prescriptions should be considered when determining the cumulative duration.

**Note:** All incretin-based drug prescriptions dispensed (or written) between study cohort entry and index day will be used for the calculation of cumulative duration. This will include prescriptions dispensed during the one-year lag period, provided that the first-ever prescription was dispensed (or written) before the lag period.

**Note:** When calculating duration of prescriptions, please be sure to not count prescription duration that extends beyond the index day; prescriptions must be truncated at the index day. For example, if a patient receives their first prescription (a 30-day prescription) for incretin-based drugs 10 days before the index day, the duration of continuous current exposure should be 10 days rather than 30 days.

Cumulative duration of use will then be classified into one of the five mutually-exclusive categories defined below:

1. Incretin-based drug cumulative duration of < 365 days (**Cumul\_dur=’1’**), defined by a cumulative duration of incretin-based drugs of less than one year between (and including) study cohort entry and index day.
2. Incretin-based drug cumulative duration of 365-729 days (**Cumul\_dur=’2’**), defined by a cumulative duration of incretin-based drugs of 1 to 2 years between (and including) study cohort entry and index day.
3. Incretin-based drug cumulative duration of ≥730 days (**Cumul\_dur=’3’**), defined by a cumulative duration of incretin-based drugs of two or more years between (and including) study cohort entry and index day.
4. Sulfonylureas (**Cumul\_dur=’4’**) [This will be the reference category], defined as:
   1. No ever-use of incretin-based drugs (i.e., those with **Cumul\_dur≠’1’, Cumul\_dur≠’2’, and Cumul\_dur≠’3’**); and
   2. Ever-use of any of the following drugs between (and including) base cohort entry and (index day – 365 days): ATC A10BB, A10BC.
5. Other anti-diabetic agents (**Cumul\_dur=’5’**):
6. All patients not meeting one of the other 4 Cumul\_dur categories, i.e., no ever-use of DPP-4 inhibitors, GLP-1 analogs, or sulfonylureas (those with **Cumul\_dur≠’1’**, **Cumul\_dur≠’2’**, **Cumul\_dur≠’3’,** and **Cumul\_dur≠’4’**).

Conditional logistic regression will be used to estimate ORs and corresponding 95% CIs of pancreatic cancer, comparing cumulative duration of use of incretin-based drugs (< 365 days, 365-729 days, and ≥730 days) to ever-use of sulfonylureas, which will serve as the reference group. The models will be crude and adjusted for the potential confounders listed in Section IV. Please complete **TABLE 6** with the results of this secondary analysis, using the crude and adjusted models provided for the primary analysis but with:

1. Cumul\_dur as the exposure variable;
2. The reference category set to Cumul\_dur =’4’;
3. Using the following contrast statements:
   * CONTRAST ‘INCRETIN < 365 DAY DUR VS SULFONYLUREA’ CUMUL\_DUR 1 0 0 0 / ESTIMATE = BOTH;
   * CONTRAST ‘INCRETIN 365-729 DAY DUR VS SULFONYLUREA’ CUMUL\_DUR 0 1 0 0 / ESTIMATE = BOTH;
   * CONTRAST ‘INCRETIN >= 730 DAY DUR VS SULFONYLUREA’ CUMUL\_DUR 0 0 1 0 / ESTIMATE = BOTH;
   * CONTRAST ‘ANY OTHER VS SULFONYLUREA’ CUMUL\_DUR 0 0 0 1 / ESTIMATE = BOTH;

Please provide SAS output in html format, named “Incretins – Pancreatic Cancer – Cumul\_dur Analysis – Site”, replacing “Site” with your site initials.

1. Please report the p-value for the test for heterogeneity for the duration categories of the incretin-based drug variable for the crude and adjusted analyses for in **“CNODES Incretins Phase IIa Results Tables (November 13, 2014).xlsx**”. To obtain this p-value, add the following contrast statement (P-value comes out of the ContrastTest table):
   * CONTRAST ‘DIFFERENCE ACROSS INCRETIN DURATION CATEGORIES’ CUMUL\_DUR 1 0 0 0,

CUMUL\_DUR 0 1 0 0,

CUMUL\_DUR 0 0 1 0;



#### Analysis by cumulative duration of use according to type of incretin-based drug

For sites that have both DPP-4 inhibitors and GLP-1 analogs, the cumulative duration of use analysis will be repeated for each of the types of incretin-based drugs. Thus, this analysis will use the same incretin-based drug categories created in secondary analysis (a) with the same sub-classification of cumulative duration defined in secondary analysis (b). Thus, patients will be classified into one of the 8 mutually exclusive categories defined below:

1. DPP-4 inhibitor cumulative duration of < 365 days (**Cumul\_dur\_type=’1’**), defined by a cumulative duration of DPP-4 inhibitors of less than one year between (and including) study cohort entry and index day.
2. DPP-4 inhibitor cumulative duration of 365-729 days (**Cumul\_dur\_type=’2’**), defined by a cumulative duration of DPP-4 inhibitors of 1 to 2 years between (and including) study cohort entry and index day.
3. DPP-4 inhibitor cumulative duration of ≥730 days (**Cumul\_dur\_type=’3’**), defined by a cumulative duration of DPP-4 inhibitors drugs of two or more years between (and including) study cohort entry and index day.
4. GLP-1 analog cumulative duration of < 365 days (**Cumul\_dur\_type=’4’**), defined by a cumulative duration of GLP-1 analogs of less than one year between (and including) study cohort entry and index day (**Note:** calculated among patients with no ever-use of DPP-4 inhibitors; i.e., among those with **Cumul\_dur\_type≠’1’,** **Cumul\_dur\_type≠’2’,** and **Cumul\_dur\_type≠’3’**).
5. GLP-1 analog cumulative duration of 365-729 days (**Cumul\_dur\_type=’5’**), defined by a cumulative duration of GLP-1 analogs of 1 to 2 years between (and including) study cohort entry and index day (**Note:** calculated among patients with no ever-use of DPP-4 inhibitors; i.e., among those with **Cumul\_dur\_type≠’1’,** **Cumul\_dur\_type≠’2’,** and **Cumul\_dur\_type≠’3’**).
6. GLP-1 analog cumulative duration of ≥730 days (**Cumul\_dur\_type=’6’**), defined by a cumulative duration of GLP-1 analogs of two or more years between (and including) study cohort entry and index day (**Note:** calculated among patients with no ever-use of DPP-4 inhibitors; i.e., among those with **Cumul\_dur\_type≠’1’,** **Cumul\_dur\_type≠’2’,** and **Cumul\_dur\_type≠’3’**).
7. Sulfonylureas (**Cumul\_dur\_type=’7’**) [This will be the reference category], defined as:
   1. No ever-use of DPP-4 inhibitors, or GLP-1 analogs (i.e., **Cumul\_dur\_type≠’1’**, **Cumul\_dur\_type≠’2’**, **Cumul\_dur\_type≠’3’**, **Cumul\_dur\_type≠’4**’, **Cumul\_dur\_type≠’5’,** and **Cumul\_dur\_type≠’6’**); and
8. Ever-use of any of the following drugs between (and including) base cohort entry and (index day – 365 days): ATC A10BB, A10BC;
9. Other anti-diabetic agents (**Cumul\_dur\_type=’8’**):
   1. All patients not meeting one of the other 7 Cumul\_dur\_type categories, i.e., no ever-use of DPP-4 inhibitors, GLP-1 analogs, or sulfonylureas (i.e., those with **Cumul\_dur\_type≠’1’**, **Cumul\_dur\_type≠’2’**, **Cumul\_dur\_type≠’3’**, **Cumul\_dur\_type≠’4**’, **Cumul\_dur\_type≠’5’**, **Cumul\_dur\_type≠’6’**, and **Cumul\_dur\_type≠’7’**).

Conditional logistic regression will be used to estimate ORs and corresponding 95% CIs of pancreatic cancer, comparing cumulative duration of use of each type of incretin-based drug (< 365 days, 365-729 days, and ≥730 days) to ever-use of sulfonylureas, which will serve as the reference group. The models will be crude and adjusted for the potential confounders listed in Section IV. Please complete **TABLE 7** with the results of this secondary analysis, using the crude and adjusted models provided for the primary analysis but with:

1. Cumul\_dur\_type as the exposure variable;
2. The reference category set to Cumul\_dur\_type =’7’;
3. Using the following contrast statements:
   1. CONTRAST ‘DPP-4 < 365 DAY DUR VS SULFONYLUREA’ CUMUL\_DUR\_TYPE 1 0 0 0 0 0 0 / ESTIMATE = BOTH;
   2. CONTRAST ‘DPP-4 365-729 DAY DUR VS SULFONYLUREA’ CUMUL\_DUR\_TYPE 0 1 0 0 0 0 0 / ESTIMATE = BOTH;
   3. CONTRAST ‘DPP-5 >= 730 DAY DUR VS SULFONYLUREA’ CUMUL\_DUR\_TYPE 0 0 1 0 0 0 0 / ESTIMATE = BOTH;
   4. CONTRAST ‘GLP-1 < 365 DAY DUR VS SULFONYLUREA’ CUMUL\_DUR\_TYPE 0 0 0 1 0 0 0 / ESTIMATE = BOTH;
   5. CONTRAST ‘GLP-1 365-729 DAY DUR VS SULFONYLUREA’ CUMUL\_DUR\_TYPE 0 0 0 0 1 0 0 / ESTIMATE = BOTH;
   6. CONTRAST ‘GLP-1 >= 365 DAY DUR VS SULFONYLUREA’ CUMUL\_DUR\_TYPE 0 0 0 0 0 1 0 / ESTIMATE = BOTH;
   7. CONTRAST ‘ANY OTHER VS SULFONYLUREA’ CUMUL\_DUR\_TYPE 0 0 0 0 0 0 1 / ESTIMATE = BOTH;

Please provide SAS output in html format, named “Incretins – Pancreatic Cancer – Cumul\_dur\_type Analysis – Site”, replacing “Site” with your site initials.

1. Please report the p-value for the test for heterogeneity for the duration categories of the incretin-based drug variable for the crude and adjusted analyses for in **“CNODES Incretins Phase IIa Results Tables (November 13, 2014).xlsx**”. To obtain this p-value, add the following contrast statements (P-value comes out of the ContrastTest table):
   * CONTRAST ‘DIFFERENCE ACROSS DPP-4 DURATION CATEGORIES’ CUMUL\_DUR\_TYPE 1 0 0 0 0 0 0,

CUMUL\_DUR\_TYPE 0 1 0 0 0 0 0,

CUMUL\_DUR\_TYPE 0 0 1 0 0 0 0;

* + CONTRAST ‘DIFFERENCE ACROSS GLP-1 DURATION CATEGORIES’ CUMUL\_DUR\_TYPE 0 0 0 1 0 0 0,

CUMUL\_DUR\_TYPE 0 0 0 0 1 0 0,

CUMUL\_DUR\_TYPE 0 0 0 0 0 1 0;

#### Analysis by time since initiation of incretin-based drugs

The third secondary analysis will assess whether the risk of pancreatic cancer varies with time since initiation of treatment among ever-users of incretin-based drugs (Figure 4). For this analysis, time since initiation among ever-users will be calculated as follows: (index day – day of follow-up of the first incretin-based drug) + 1

**Note:** The time since initiation analysis reclassifies exposure among patients deemed to be ever-users of incretin-based drugs in the primary analysis.

**Note:** The time since initiation duration will include the one-year lag period. Consequently, please note that the minimum time since initiation is 366 days.

Time since initiation will then be classified into one of the four mutually exclusive categories defined below:

1. Incretin-based drug time since initiation of 366-729 days (**Time\_init=’1’**), defined by a time since initiation of incretin-based drugs of one year to two years in the period between (and including) study cohort entry and index day.
2. Incretin-based drug time since initiation of ≥730 days (**Time\_init=’2’**), defined by a time since initiation of incretin-based drugs of two or more years in the period between (and including) study cohort entry and index day.
3. Sulfonylureas (**Time\_init=’3’**) [This will be the reference category], defined as:
4. Never use of incretin-based drugs (i.e., **Time\_init≠’1’** and **Time\_init≠’2’**); and
5. Ever-use of any of the following drugs between (and including) base cohort entry and (index day – 365 days): ATC A10BB, A10BC;
6. Other anti-diabetic agents (**Time\_init=’4’**):
7. All patients not meeting one of the other 3 Time\_init categories, i.e., no use ever of incretin-based drugs or sulfonylureas (those with **Time\_init≠’1’**, **Time\_init≠’2’**, and **Time\_init≠’3’)**

Conditional logistic regression will be used to estimate ORs and corresponding 95% CIs of pancreatic cancer, comparing time since initiation of incretin-based drugs (366-729 days, and ≥730 days) to ever-use of sulfonylureas, which will serve as the reference group. The models will be crude and adjusted for the potential confounders listed in Section IV. Please complete **TABLE 8** with the results of this secondary analysis, using the crude and adjusted models provided for the primary analysis but with:

1. Time\_init as the exposure variable;
2. The reference category set to Time\_init =’3’;
3. Using the following contrast statements:
   * CONTRAST ‘INCRETIN INIT TIME 366-729 DAYS VS SULFONYLUREA’ TIME\_INIT 1 0 0 / ESTIMATE = BOTH;
   * CONTRAST ‘INCRETIN INIT TIME >= 730 DAYS VS SULFONYLUREA’ TIME\_INIT 0 1 0 / ESTIMATE = BOTH;
   * CONTRAST ‘ANY OTHER VS SULFONYLUREA’ TIME\_INIT 0 0 1 / ESTIMATE = BOTH;

Please provide SAS output in html format, named “Incretins – Pancreatic Cancer – Time Init Analysis – Site”, replacing “Site” with your site initials.

1. Please report the p-value for the test for heterogeneity for the time-since-initiation categories of the incretin-based drug variable for the crude and adjusted analyses for in **“CNODES Incretins Phase IIa Results Tables (November 13, 2014).xlsx**”. To obtain this p-value, add the following contrast statement (P-value comes out of the ContrastTest table):
   * CONTRAST ‘DIFFERENCE ACROSS INCRETIN INITIATION TIME CATEGORIES’ TIME\_INIT 1 0 0,

TIME\_INIT 0 1 0;

#### Analysis by time since initiation according to type of incretin-based drug

For sites that have both DPP-4 inhibitors and GLP-1 analogs, the time since initiation analysis will be repeated for each of the types of incretin-based drugs. Thus, this analysis will use the same incretin-based drug categories created in secondary analysis (a) with the same definition of time since initiation defined in secondary analysis (d). Thus, patients will be classified into one of the 6 mutually exclusive categories defined below:

1. DPP-4 inhibitor time since initiation of 366-729 days (**Time\_init\_type=’1’**), defined by a time since initiation of DPP-4 inhibitors of one year to two years in the period between (and including) study cohort entry and index day.
2. DPP-4 inhibitor time since initiation of ≥730 days (**Time\_init\_type=’2’**), defined by a time since initiation of DPP-4 inhibitors of two or more years in the period between (and including) study cohort entry and index day.
3. GLP-1 analog time since initiation of 366-729 days (**Time\_init\_type=’3’**), defined by a time since initiation of GLP-1 analogs of one year to two years in the period between (and including) study cohort entry and index day (**Note:** calculated among patients with no ever-use of DPP-4 inhibitors; i.e., among those with **Time\_init\_type≠’1’**, and **Time\_init\_type≠’2’)**.
4. GLP-1 analog time since initiation of ≥730 days (**Time\_init\_type=’4’**), defined by a time since initiation of GLP-1 analogs of two or more years in the period between (and including) study cohort entry and index day (**Note:** calculated among patients with no ever-use of DPP-4 inhibitors; i.e., among those with **Time\_init\_type≠’1’**, and **Time\_init\_type≠’2’).**
5. Sulfonylureas (**Time\_init\_type=’5’**) [This will be the reference category], defined as:
6. No ever-use of incretin-based drugs (i.e., **Time\_init\_type≠’1’**, **Time\_init\_type≠’2’**, **Time\_init\_type≠’3’**, and **Time\_init\_type≠’4’**); and
7. Ever-use of any of the following drugs between (and including) base cohort entry and (index day – 365 days): ATC A10BB, A10BC.
8. Other anti-diabetic agents (**Time\_init\_type=’6’**):
   1. All patients not meeting one of the other 5 Time\_init\_type categories, i.e., no ever-use of incretin-based drugs and sulfonylureas (those with **Time\_init\_type≠’1’**, **Time\_init\_type≠’2’**, **Time\_init\_type≠’3’**, **Time\_init\_type≠’4’**, and **Time\_init\_type≠’5’**).

Conditional logistic regression will be used to estimate ORs and corresponding 95% CIs of pancreatic cancer, comparing time since initiation of each type of incretin-based drug (366-729 days and ≥730 days) to ever-use of sulfonylureas, which will serve as the reference group. The models will be crude and adjusted for the potential confounders listed in Section IV. Please complete **TABLE 9** with the results of this secondary analysis, using the crude and adjusted models provided for the primary analysis but with:

1. Time\_init\_type as the exposure variable;
2. The reference category set to Time\_init\_type =’5’;
3. Using the following contrast statements:
   * CONTRAST ‘DPP-4 INIT TIME 366-729 DAYS VS SULFONYLUREA’ TIME\_INIT\_TYPE 1 0 0 0 0 / ESTIMATE = BOTH;
   * CONTRAST ‘DPP-4 INIT TIME >= 730 DAYS VS SULFONYLUREA’ TIME\_INIT\_TYPE 0 1 0 0 0 / ESTIMATE = BOTH;
   * CONTRAST ‘GLP-1 INIT TIME 366-729 DAYS VS SULFONYLUREA’ TIME\_INIT\_TYPE 0 0 1 0 0 / ESTIMATE = BOTH;
   * CONTRAST ‘GLP-1 INIT TIME >= 730 DAYS VS SULFONYLUREA’ TIME\_INIT\_TYPE 0 0 0 1 0 / ESTIMATE = BOTH;
   * CONTRAST ‘ANY OTHER VS SULFONYLUREA’ TIME\_INIT\_TYPE 0 0 0 0 1 / ESTIMATE = BOTH;

Please provide SAS output in html format, named “Incretins – Pancreatic Cancer – Time\_Init\_Type Analysis – Site”, replacing “Site” with your site initials.

1. Please report the p-value for the test for heterogeneity for the time-since-initiation categories of the incretin-based drug variable for the crude and adjusted analyses for in **“CNODES Incretins Phase IIa Results Tables (November 13, 2014).xlsx**”. To obtain this p-value, add the following contrast statements (P-value comes out of the ContrastTest table):
   * CONTRAST ‘DIFFERENCE ACROSS DPP-4 INITIATION TIME CATEGORIES’ TIME\_INIT\_TYPE 1 0 0 0 0,

TIME\_INIT\_TYPE 0 1 0 0 0;

* + CONTRAST ‘DIFFERENCE ACROSS GLP-1 INITIATION TIME CATEGORIES’ TIME\_INIT\_TYPE 0 0 1 0 0,

TIME\_INIT\_TYPE 0 0 0 1 0;

### 3. Sensitivity Analyses

We will conduct four sensitivity analyses to assess the robustness of the findings. In the first two, the lag periods will be varied to account for uncertainty in the length of the possible latency period, as well as to assess the impact of protopathic bias. In the third, we will use a stricter exposure definition. In the fourth, we will examine the impact of sulfonylurea use between base cohort entry and study cohort entry.

1. First, we will vary the lag period prior to index day using a lag of 6 months rather than one year. Consequently, in this analysis, repeat the primary analyses described in section V.1 but only considering the anti-diabetic prescriptions between, and including, base cohort entry date and (INDEX DAY – 180 days). Please report your results in **TABLE 10**. Please provide SAS output in html format, named “Incretins – Pancreatic Cancer – Lag\_6\_Mo Analysis – Site”, replacing “Site” with your site initials.

**Note:** The start of follow-up remains study cohort entry + 365 days to minimize protopathic bias.

1. Second, we will vary the lag period prior to index day using a lag of 2 years. Consequently, in this analysis, repeat the primary analyses described in section V.1 but only considering the anti-diabetic prescriptions between, and including, base cohort entry date and (INDEX DAY – 730 days). For this analysis, the start of follow-up (T0) will be set to study cohort entry + 730 days, and all patients who exit the cohort prior to T0 will be excluded (**Note**: this analysis will be restricted to cases and matched controls with at least 730 days of follow-up; there is no need to resample controls for this analysis).

Please report your results in **TABLE 11**. Please provide SAS output in html format, named “Incretins – Pancreatic Cancer – Lag\_2Yr Analysis – Site”, replacing “Site” with your site initials.

**Note**: Not all sites will have sufficient follow-up time to conduct this analysis. This analysis should only be conducted by sites with at least 730 days of follow-up. Those who are unable to participate in this analysis should report this in the site-specific protocol deviations form.

1. Third, the primary analysis will be repeated using a stricter exposure definition consisting of receiving *at least 4 prescriptions within any 12-month period* (i.e., using a moving, time-dependent 12-month window). This criterion will minimize the inclusion of non-persistent users (and those who may have used these drugs sporadically), whose infrequent exposure to these drugs is unlikely to have any biological effect. Please report your results in **TABLE 12**. Please provide SAS output in html format, named “Incretins – Pancreatic Cancer – Intensity Analysis – Site”, replacing “Site” with your site initials.

**Note:** The date of ever-use of incretin-based drugs is defined by the date of prescription/dispensation of the fourth prescription within the 12-month window.

**Note:** To be considered exposed to incretin-based drugs in this analysis, none of the qualifying four prescriptions (in the 12-month window) can occur in the 365-day lag period.

**Note:** Patients with <4 prescriptions of incretin-based drugs during a 12-month window (and not meeting the stricter exposure criteria in another 12-month window) will be classified as either ever-users of sulfonylureas or ever-users of other anti-diabetic drugs. As with the primary analysis, these exposure categories are defined hierarchically.

**Note:** The minimum number of prescriptions required for the ever-exposed to sulfonylureas and ever-exposed to other-anti-diabetic drug categories remains 1.

**Note:** Patients who only receive incretin-based drugs (i.e., do not receive sulfonylureas or any other type of anti-diabetic drug) but do not meet the stricter exposure category should have their exposure classified as “other anti-diabetic drugs”. While it would be optimal to classify these patients as “non-users”, there will likely be few patients meeting this criterion, and the inclusion of such a category will thus result in convergence issues.

1. Fourth, we will sub-classify ever-users of sulfonylureas to examine the impact of exposure between base cohort entry and study cohort entry. To do so:
   1. Reclassify exposure using the following hierarchical exposure classification:
   2. Incretin-based drugs (**Sulf\_Exposure=’1’**), defined by ever-use of any of the following drugs between (and including) study cohort entry and (index day – 365 days):
   * DPP-4 inhibitors [sitagliptin, vildagliptin, saxagliptin][ATC A10BH, A10BD07-A10BD13]; or
   * GLP-1 analogs [exenatide, liraglutide][ATC A10BX04, A10BX07];
   1. Sulfonylureas initiated prior to study cohort entry (**Sulf\_Exposure=’2’**), defined as:
   * No ever-use of incretin-based drugs (i.e., those with **Sulf\_Exposure≠’1’**); and
   * Ever-use of any of the following drugs between (and including) base cohort entry and (but excluding) study cohort entry: ATC A10BB, A10BC
   1. Sulfonylureas initiated at or after study cohort entry (**Sulf\_Exposure=’3’**) [This will be the reference category], defined as:
   * No ever-use of incretin-based drugs and no use of sulfonylureas prior to study cohort entry (i.e., those with **Sulf\_Exposure≠’1’** and **Sulf\_Exposure≠’2’**); and
   * Ever-use of any of the following drugs between (and including) study cohort entry and (index day – 365 days): ATC A10BB, A10BC
   1. Other anti-diabetic agents (**Sulf\_Exposure=’4’**):
   * All patients with never-use of incretin-based drugs and never-use of sulfonylureas (i.e., those classified as **Sulf\_Exposure≠’1’**, **Sulf\_Exposure≠’2’**, and **Sulf\_Exposure≠’3’**)
   1. Please complete **TABLE 13** with the results of this sensitivity analysis, using the crude and adjusted models provided for the primary analysis but with:
2. Sulf\_Exposure as the exposure variable;
3. The reference category set to Sulf\_Exposure =’3’;
4. Using the following contrast statements:
   * CONTRAST ‘INCRETINS VS POST-STUDY COHORT ENTRY SULF’ SULF\_EXPOSURE 1 0 0 / ESTIMATE = BOTH;
   * CONTRAST ‘PRE-STUDY COHORT ENTRY SULF VS POST-STUDY COHORT ENTRY SULF’ SULF\_EXPOSURE 0 1 0 / ESTIMATE = BOTH;
   * CONTRAST ‘ANY OTHER VS POST-STUDY COHORT ENTRY SULF’ SULF\_EXPOSURE 0 0 1 / ESTIMATE = BOTH;

Please provide SAS output in html format, named “Incretins – Pancreatic Cancer – Sulfo\_Hx Analysis – Site”, replacing “Site” with your site initials.

### 4. Reduced Model Analyses

Some sites will have a small number of cases in their case-control analyses. For this reason, we will repeat our primary analysis using two reduced models. These analyses are to be completed by all sites, regardless of their number of cases.

**Note:** Sites for which models do not converge (either full or reduced models) must note this information in the site-specific protocol deviations (Appendix I).

1. Reduced Model #1:
   * Combine the ‘neuropathy’, ‘renal disease’, ‘retinal disorders’, and ‘peripheral arteriopathy’ variables into a single ‘microvascular complications’ variable.
   * Check distributions of ‘number of hospitalization episodes of care’, ‘number of unique non-anti-diabetic medications’, and ‘pre-study cohort entry anti-diabetic medications’ variables. If any have no obvious outliers, include these variables as continuous variables.
   * Repeat the primary analysis using this reduced list of covariates
   * Complete **TABLE 14**.
   * Please provide SAS output in html format, named “Incretins – Pancreatic Cancer – Reduced Analysis 1 – Site”, replacing “Site” with your site initials.
2. Reduced Model #2:
   * Repeat the primary analysis only including exposure and the following covariates:
     1. Deyo version of the Charlson comorbidity index
     2. History of acute or chronic pancreatitis
   * Complete **TABLE 15**.
   * Please provide SAS output in html format, named “Incretins – Pancreatic Cancer – Reduced Analysis 2 – Site”, replacing “Site” with your site initials.

**Note:** Please see Dropbox folder ‘Incretins Project Team\Analytic Protocols\’ in SAS macros.zip (See: Deyo Charlson Comorbidity Score calculation.sas, Deyo Charlson Comorbidity Score ICD9.sas, and Deyo Charlson Comorbidity Score ICD10.sas) for a SAS Macro that implements the Deyo version of the Charlson comorbidity index. Also, please see the corresponding paper in Dropbox folder ‘Incretins Project Team\Analytic Protocols\’ entitled ‘Quan 2005 Medical Care.pdf ’.

**Note:** Because type 2 diabetes is included in the Deyo score, and all patients in this cohort necessarily have the disease, please remove MILDDIAB when calculating the score but keep SEVDIAB.

**Note:** If your site’s ICD-9 classification is limited to 3 digits, you will not be able to differentiate between MILDDIAB and SEVDIAB. In this case, you should assume that ICD 9 code 250 is MILDDIAB. However, to reduce the potential residual confounding due to this lack of granularity, please do a site-specific protocol deviation in which you use the ‘microvascular complications’ variable (created for Reduced Model #1) to identify those with SEVDIAB. Please record this deviation in the appropriate appendix.

The macro that defines the comorbidities based on ICD-9 codes (‘charlson\_score\_ICD9’) has been modified to allow the user to input the granularity of the diagnosis codes used in their hospital data and in their medical claims data.  Note the additional parameters in the macro ‘hosp\_granularity’ and MSP\_granularity’ in the macro call:

%macro charlson\_score\_ICD9(

var\_ID=,           /\*name of patient ID variable\*/

hosp\_dat=,         /\*name of the hospital diagnostic dataset\*/

hospdx\_var=,       /\*name of the diagnostic variables recording ICD9 codes in hospital diagnostic data\*/

n\_hospdx\_var=,     /\*number of the diagnostic variables recording ICD9 codes in hospital diagnostic data\*/

hosp\_granularity=, /\*ICD9 diagnosis codes granularity on hospital diagnostic data\*/

MSP\_dat=,          /\*name of the medical service plan dataset\*/

MSPdx\_var=,        /\*name of the diagnostic variables recording ICD9 codes in medical service plan data\*/

n\_MSPdx\_var=,      /\*number of the diagnostic variables recording ICD9 codes in medical service plan data\*/

MSP\_granularity=   /\*ICD9 diagnosis codes granularity on medical service plan data\*/

);

**Note:**  The macro that defines the comorbidities based on ICD-10 codes (‘charlson\_score\_ICD10’) does not contain those additional parameters.

The outputted datasets from these macros will now contain a variable indicating the source of the data (from\_file=’hsp’ or ‘med’). If your medical claims diagnosis codes have a granularity of 3 then you will need to check those with mild diabetes who have microvascular complications to confirm that the patient should actually be defined as having severe diabetes (if from\_file=’med’ and disease=’MILDDIAB’ and microvasc\_comp=’1’ then disease=’SEVDIAB’;).  Exclude all mild diabetes (if disease=’MILDDIAB’ then delete;)

**Note:** Please do not truncate codes for which 4 and 5 digits are available.

**Note:** To be consistent with covariate definitions elsewhere in the protocol, the Deyo score should consider codes recorded ever before (and including) the date of study cohort entry rather than restricting to codes in the year before (and including) study cohort entry.

### Quality Assurance Analyses

Using the simulated dataset provided in Dropbox (‘Incretins Project Team\Incretins – Simulated Data – 2014-10-16.csv’), please apply the exposure definitions listed in Sections III and V and complete the Excel Workbook entitled **“CNODES Incretins Simulation Tables (October 20, 2014).xlsx”** These analyses should be completed and uploaded to your site-specific Dropbox folder (with your site initials added to the file name) prior to the completion of all case-control analyses. To simplify this exercise somewhat, we have limited drugs (variable rx) to other anti-diabetics (rx=1), sulfonylureas (rx=2), DPP-4 inhibitors (rx = 3), and GLP-1 analogs (rx=4).

**Note:** Please assume that all prescriptions are have a duration of 90 days.

**Note:** For this analysis, assume that incretin-based drugs became available January 1st, 2002.

**Note:** In the simulation study, there are no cases or controls; exposure prevalence is to be estimated overall.

**Note:** The table numbers in this Excel Workbook are numbered identically to the analyses described in this protocol, with “a” representing the table for summary statistics and “b” representing the table for the first 50 observations.

**Note:** Some patients will enter the base cohort but will not enter the study cohort. These patients should be listed as “Excluded” in the patient-level “b” tables.

**Note:** The reduced model analyses and strict exposure definition analyses do not need to be completed in the quality assurance analyses.

**Note:** In the “b” tables, please report the exposure categories using the text descriptions listed in the corresponding “a” tables. For the duration analyses, please list as “incretins\_dur\_cat”, with incretins replaced by DPP-4 or GLP-1 in the type specific analyses and “cat” replaced by the duration category in the table (e.g., incretins\_dur\_365, incretins\_dur\_366\_730, etc.). For the time-since-initiation analyses, please replace the “dur” with “time”.

# APPENDIX I: Site-specific protocol deviations

**Along with your Flow chart and Tables, please upload your site-specific protocol deviations to Dropbox using the provided protocol deviations entry form. The completed form must be verified and approved by the site liaison prior to submission.**

# APPENDIX II: Amendments

**Note:** Amendments in the analytical protocol are indicated in **blue** font.

**Version 1.2:**

* + - 1. Sex was added as a matching variable to the analytical protocol.
      2. A footnote was added to Table 1 of Section II (Pancreatic Cancer Case-Control Selection) to clarify the matching on follow-up duration.
      3. Sites with less than 730 days of follow-up should not participate in the second sensitivity analysis. This has been clarified in Section V.3.2. (Sensitivity Analyses).
      4. Quality assurance analyses to assess the primary and secondary exposure definitions were added as Section V.5.
      5. Minor modifications were made to the Excel Workbook, which is now named **“CNODES Incretins Phase IIa Results Tables (October 16, 2014).xlsx”.**

**Version 1.3:**

* + - 1. For the quality assurance analyses (Section V.5), please assume that all prescriptions were for 90 days.
      2. For the quality assurance analyses (Section V.5), some patients will enter the base cohort but will not enter the study cohort. These patients should be listed as “Excluded” in the patient-level “b” tables.
      3. The reduced model analyses and strict exposure definition analyses have been removed from the quality assurance analyses (Section V.5) and **“CNODES Incretins Simulation Tables (October 20, 2014).xlsx”** workbook.
      4. Some cells were incorrectly locked in the Excel workbooks. This issue has been fixed in both the **“CNODES Incretins Phase IIa Results Tables (October 20, 2014).xlsx”** and **“CNODES Incretins Simulation Tables (October 20, 2014).xlsx”** workbooks**.**
      5. For all duration and time-since-initiation analyses (Tables 6-9), two columns (1 for the crude model and 1 for the adjusted model) have been added to the **“CNODES Incretins Phase IIa Results Tables (October 20, 2014).xlsx**” to allow for the reporting of the type III analysis p-value for the incretin-based drug variable. Corresponding notes have been added to the secondary analyses section (V.2) requesting this information.
      6. The definition of cumulative duration provided in the Glossary of Terms was corrected.
      7. Question and Answer #1 added to the protocol.
      8. Naming procedures for exposure classification in the “b” tables of the quality assurance analyses were added to the protocol (Section V.5).

**Version 1.4:**

* + - 1. Contrast statements have been added to all duration and time-since-initiation analyses (sections V.2.b-V.2.e) to estimate p-value for the test of heterogeneity. The descriptions of p-values in Tables 6-9 have been clarified in **“CNODES Incretins Phase IIa Results Tables (October 24, 2014).xlsx**”, and previously included references to “type III analysis” have been removed from the protocol and Excel workbook.
      2. A note was added to section V.2.b to clarify how to handle prescriptions that extend beyond the index day.

**Version 1.5:**

* + - 1. Table 1 in the Excel Workbook (See file **“CNODES Incretins Phase IIa Results Tables (November 13, 2014).xlsx”** located in Dropbox) was updated to show that cases with no available controls are to be excluded from the analyses. A corresponding note has been added to Section II.
      2. The maximum caliper for matching on duration of diabetes has been increased to 365 days in Sections II.

**Version 1.6:**

* + - 1. Notes were added to section IV clarifying the weighting of controls in tables 2 and 3.
      2. The Deyo version of the Charlson Comorbidities Index has replaced the Romano version in Reduced Model #2 in section V.4. A new SAS macro as well as a copy of the Quan 2005 Medical Care paper that converts the score from ICD-9 to ICD-10 codes have been provided in Dropbox.
      3. Notes were added to clarify how the Deyo score should be implemented into our diabetic cohorts and how sites with only 3 digits should adapt the score in section V.4
      4. Site-specific protocol deviations section was revised and a protocol deviations form was added to Dropbox.

**Version 1.7**

* + - 1. The note in section IV was changed because the weighting applies to table 2 only and not table 3.
      2. Additional information regarding the Deyo macro was added to section V.4. Revised macros were placed in the SAS macros.zip file in Dropbox (programs are called Deyo Charlson Comorbidity Score ICD9.sas, Deyo Charlson Comorbidity Score ICD10.sas, and Deyo Charlson Comorbidity Score calculation.sas). The macros allow for the specification of the granularity of codes for each data source. Please do not truncate codes for which 4 and 5 digits are available.
      3. Question and Answer #2-3 were added to the protocol.
      4. Since pre-cohort entry anti-diabetic medications are highly correlated with anti-diabetic medications used during follow-up (i.e., exposure), pre-cohort entry anti-diabetic medications should not be included as covariates in the primary analysis (they will be included in a sensitivity analyses). A note clarifying this issue has been added to section V.1.
      5. A note was added to specify that the categorical version of continuous variables should be included as covariates in the primary analysis (section V.1).
      6. A note was added to section V.1 to specify that matching variables need not be included in the models.

**Version 1.8:**

* + - 1. Please disregard amendment #25. Since the anti-diabetic medication variable examines number of drugs rather than the specific drugs, the correlation is not a problem. The previously inserted clarifying note has been removed.
      2. When calculating the Deyo score, comorbidities should be measured using diagnoses ever before (and including) the date of cohort entry rather than restricting to the year before (and including) study cohort entry. A clarifying note was added to section V.4, and revised macros have been posted in the SAS macros.zip file in Dropbox (programs are called Deyo Charlson Comorbidity Score ICD9.sas, Deyo Charlson Comorbidity Score ICD10.sas, and Deyo Charlson Comorbidity Score calculation.sas).
      3. A note in section V.2.b was expanded.
      4. The definition of ever use was clarified throughout the protocol.

# APPENDIX III: Questions & Answers

**Q1. In the simulated data analysis, there is no base cohort entry date or study cohort entry date either. We can determine the base cohort entry date to be the first prescription date for each person, and also determine the study cohort entry date as the first different type of prescription that occurs after January 1, 2002 (simulated data incretins start date). Is this what you’d like done?**

A1. Yes, that is correct.

**Q2. If a site does not have any use of GLP-1, how should the tables that have GLP1 (table 5, 7, 9 in pancreatic cancer) be filled in?**

A2. If a site does not have any GLP-1 use, please leave Tables 5, 7, and 9 blank (otherwise they will simply be duplicates of Tables 4, 6, and 8).

**Q3. The protocol indicated that we should put all confounders in the logistic regression model,  but for those variables that have both continuous and categorical, including: *age, number of hospitalization episodes of care , number of unique non-anti-diabetic medications, and pre-study cohort entry anti-diabetic medications,*  shall I include them in the adjust logistic model in both continuous and categorical? If not, which type shall I use?**

A.3 Since age is a matching variable, it does not need to be included in the model.  For the other variables listed below, please use the categorical versions.  The continuous versions will be used as part of Reduced Model #1, which also collapses individual complications as a composite microvascular complications variable.

# APPENDIX IV: Glossary of Terms

1. **Base Cohort Entry:** all patients newly treated with non-insulin anti-diabetic agents (first ever prescription of non-insulin anti-diabetic drug)
2. **Cumulative Duration:** cumulative duration of use among ever-users will be calculated by summing the days’ supply from all incretin-based drug prescriptions from, and including study cohort entry date until, and including, index day.
3. **Duration of follow-up:** 
   * Calculation: (Index day) – (Study Cohort Entry Date + 365) +1
4. **Duration of treated diabetes for each patient** :
   * Calculation: (Study Cohort Entry) –(Base Cohort Entry) +1
5. **Index Day:** 
   * Cases: The day of follow-up on which the hospital admission or medical service diagnosis of pancreatic cancer occurs.   
     **Note:** This event is defined by days of follow-up, not calendar time.
   * Controls: will be assigned the index day of their respective cases
6. **Initiation of follow up time (to):** 
   * Calculation: Study Cohort Entry + 365
   * Person-time at risk starts here
7. **Study Cohort Entry:** 
   * 1) All patients newly treated with non-insulin anti-diabetic agents (first ever prescription of non-insulin anti-diabetic drug) during the year or any time after incretin-based drugs entered the market up until June 20, 2014; OR
   * 2) Those who added on or switched to an anti-diabetic agent not previously used in their treatment history after incretin-based drugs entered the market up until June 20, 2014
   * **Note:** Base cohort entry and study cohort entry can occur simultaneously and thus be the same day