Dominic D. LaRoche

CPH576D –SAS and Data Management

Final Report

12/06/2014

Patient well-being is an important outcome for any new treatment. I utilized 5 data sets containing information on pancreatectomy patients. Of these cases 92 patients received pancreatectomy surgery and of these 92 surgery patients 63 also received auto-islet transplant. A summary of the patient characteristics for these two treatment groups is provided in tables 1 and 2. I categorized race as either white and non-hispanic or other (table 1). I also collapsed the diagnosis category from 13 categories to 5 since many categories had only a small number of patients (table 2).

Table 1. Patient character

|  | | **Auto-Islet Treatment** | |
| --- | --- | --- | --- |
| **N** | **Y** |
| **Age** | **N** | 27 | 60 |
| **Mean** | 30.81 | 41.88 |
| **Min** | 12.00 | 21.00 |
| **Max** | 46.00 | 66.00 |
| **bmi** | **N** | 27 | 59 |
| **Mean** | 31.85 | 25.63 |
| **Min** | 19.61 | 15.15 |
| **Max** | 56.94 | 50.01 |
| **Duration of Disease (years)** | **N** | 26 | 52 |
| **Mean** | 7.92 | 7.79 |
| **Min** | 0.00 | 1.00 |
| **Max** | 24.00 | 32.00 |
| **Gender** |  |  |  |
| **F** | **N** | 12 | 37 |
| **M** | **N** | 15 | 23 |

Table 2. Prevalence of disease types for each treatment group.

|  | | **Auto-Islet Treatment** | |
| --- | --- | --- | --- |
| **Diagnosis Category** |  | **N** | **Y** |
| **Alcohol** | **N** | 5 | 8 |
| **Familial** | **N** | 1 | 8 |
| **Idiopathic** | **N** | 8 | 31 |
| **Other** | **N** | 11 | 11 |
| **Pancreas Divisum** | **N** | 4 | 5 |

Follow-up assessments of patients’ physical and mental well-being were conducted at baseline, 3 months, and 6 months post-surgery. Physical appeared to increase over these three follow-ups (fig. 1). The same was true for mental composite score (MCS) and physical composite score (PCS) (figs. 2 and 3).

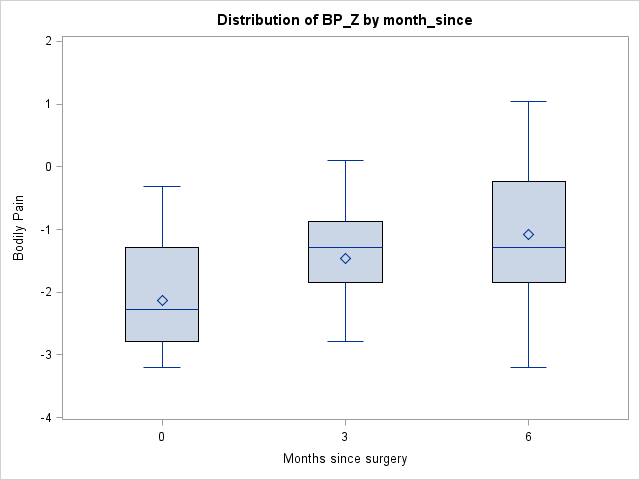


Figure 1. Standardized bodily pain at baseline (0), 3 months, and 6 months post-surgery.

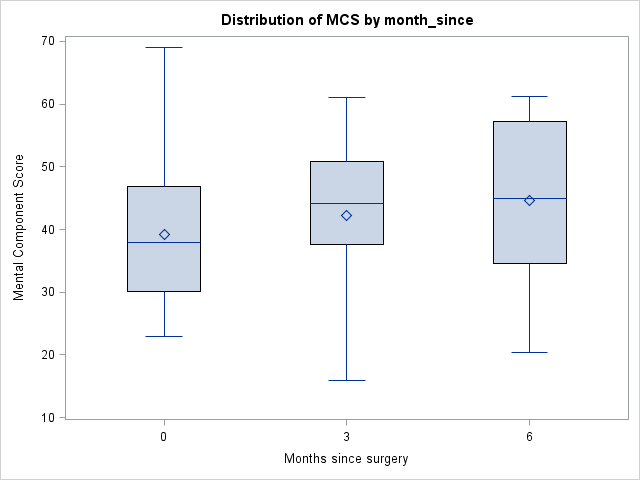


Figure 2. Mental composite score (MCS) at baseline (0), 3 months, and 6 months post-surgery.

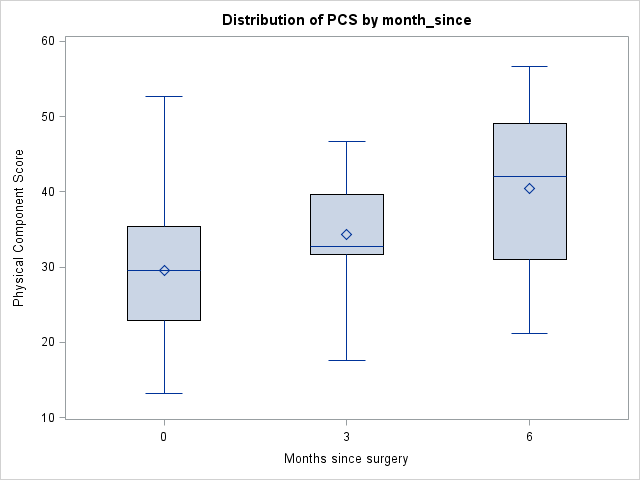


Figure 3. Physical composite score (MCS) at baseline (0), 3 months, and 6 months post-surgery.

I performed a regression analysis to determine if there was a relationship between islet gain (IEQ) and either the body mass index (BMI) or duration of disease of the patient. I modeled IEQ as a function of both BMI and disease duration. The initial model had poor fit and did not show a statistically significant relationship for either of the predictors. However, the removal of two outlying IEQ scores significantly improved model fit and showed that BMI was a significant predictor of IEQ (higher BMI was associated with higher IEQ). The association of IEQ with disease duration was not as strong, negative, and not statistically significant. Parameter estimates are provided in table 3. Figure 4 provides a contour plot of the relationship between the two predictors and IEQ. Figure 5 gives the two bivariate relationships (uncontrolled).

Table 3. Parameter estimates from regression of IEQ and disease duration and BMI.

| **Parameter** | **Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- |
| **Intercept** | 3161.670114 | 81652.50092 | 0.04 | 0.9693 |
| **Disease Duration** | -3131.126704 | 2510.25452 | -1.25 | 0.2183 |
| **BMI** | 7837.833957 | 3032.15972 | 2.58 | 0.0128 |

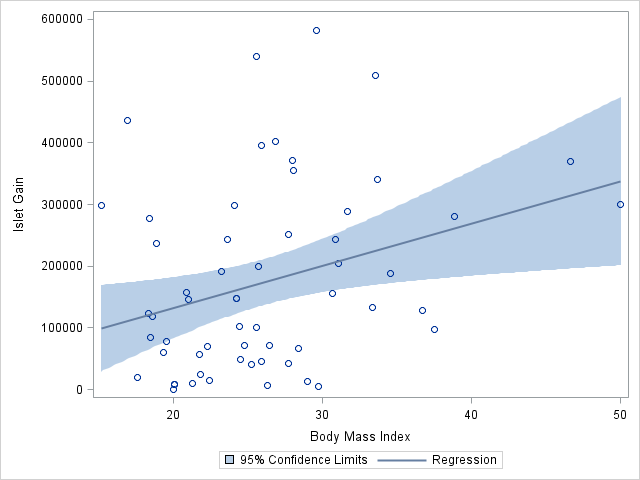
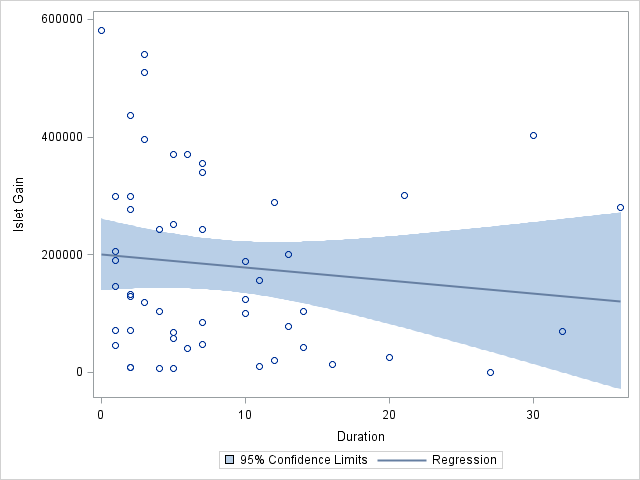
 

Figure 5. Bivariate relationship between BMI and disease duration with IEQ.

I performed a two-sample t-test to determine if there was a significant difference between bodily pain and baseline and 6 months post-surgery. I found an increase of 1.06 standard deviations (p-value=0.02) between baseline and 6 months. However, examination of the distribution suggested the normality assumption may be violated. Therefore, I used a Wilcoxon rank-sum test to verify the results (p-value = 0.0003).

There may have been differential re-admission of the two groups based on auto-islet status. Table 4 gives the overall re-admission status by the two groups. There were several cases where the re-admission date was earlier than the surgery date. These admissions were removed prior to analyses. The re-admission rate appeared to be slightly higher among patients that did not receive auto-islet (79% vs 74%). Among patients who were readmitted there may have been differential admission to the ER for the two auto-islet status groups. Table 5 gives the number of patients readmitted based on auto-islet status by ER status at both 1 month and 6 months post-surgery. At both time points the non-auto-islet patients appeared to have a high ER admission rate than the auto-islet patients.

Table 4.

|  | **Auto-Islet Status** | |
| --- | --- | --- |
| **Readmission Status** | **N** | **Y** |
| **No** | 6 | 16 |
| **Yes** | 23 | 45 |

Table 5. Number of patients readmitted to the hospital by auto-islet status and emergency room status within the first month and within the first 6 months.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Auto-Islet Status** | |
|  | **Emergency Room** | **N** | **Y** |
| **First Month** | **No** | 7 | 11 |
| **Yes** | 15 | 7 |
| **First 6 Months** | **No** | 8 | 19 |
| **Yes** | 15 | 17 |

SAS Code:

\*read in patient data;

**Data** pat;

set final.patient;

run;

**data** diag;

set final.diagnosis;

run;

**data** surg;

set final.surgery;

run;

**data** isl;

set final.islets;

run;

**data** sf;

set final.sf36;

run;

**data** amd;

set final.readmin;

run;

\*data clean-up;

**proc** **freq** data = pat;

table \_character\_;

**run**;

**proc** **means** data= pat nmiss min max mean;

var \_numeric\_;

**run**;

**proc** **freq** data = diag;

table \_character\_;

**run**;

**proc** **means** data= diag nmiss min max mean;

var \_numeric\_;

**run**;

\*change unable to confirm to unknown;

**data** diag;

set diag;

if Pancreatitis1 = 'Unable to confirm' then Pancreatitis1 = 'Unknown';

\*change 9999 to missing;

if Pancreatitis\_yr = **9999** then Pancreatitis\_yr = **.**;

run;

\*remove duplicates;

**proc** **sort** data=diag;

by PatID;

**run**;

**data** diag;

set diag;

by patid;

if first.PatID;

run;

**proc** **means** data= isl nmiss min max mean;

var \_numeric\_;

**run**;

**proc** **univariate** data=isl plot trimmed=(**0.05**);

var ieq;

**run**;

\*negative value and 1 very high outlier;

**data** isl;

set isl;

ieqout = **0**;

if PatID < **1** then delete;

if ieq > **3000000** then ieq = **.**;

if ieq > **750000** then ieqout = **1**;

run;

\*there are still a couple outliers, I labelled them with indicator;

**proc** **sort** data=isl;

by patid;

**run**;

**data** isl;

set isl;

by patid;

if first.PatID;

run;

**proc** **freq** data = surg;

table \_character\_;

**run**;

**proc** **means** data= surg nmiss min max mean;

var \_numeric\_;

**run**;

**proc** **univariate** data=surg plot trimmed=**0.05**;

var ht\_pre wt\_pre;

histogram;

**run**;

\*a couple high wieght outliers and one low hieght outlier;

\*check bivariate outliers;

**proc** **sgplot** data=surg;

scatter x=ht\_pre y=wt\_pre;

**run**;

\*thought I would see a tighter relationship there!;

\* I don't think there are any observations extreme enough to merit removal;

**proc** **freq** data = amd;

table \_character\_;

**run**;

**proc** **means** data= amd nmiss min max mean;

var \_numeric\_;

**run**;

\*consolodate admission reasons;

**data** amd;

set amd;

if Admission\_reason1 = 'Gastrointestinal bleeding' then Admission\_reason1 = 'GI Bleeding';

if Admission\_reason1 = 'Hypoglyxemia' then Admission\_reason1 = 'Hypoglycemia';

if Admission\_reason1 = 'Incisonal hernia' then Admission\_reason1 = 'Incisional hernia';

if Admission\_reason1 = 'Nausea&Vomiting' or Admission\_reason1 = 'Vomiting' then Admission\_reason1 = 'Nausea/Vomiting';

if Admission\_reason1 = 'Pain' then Admission\_reason1 = 'Abdominal Pain';\*not sure if this is right - I would check with PI;

run;

**proc** **means** data= sf nmiss min max mean;

var \_numeric\_;

**run**;

\*merge data;

**proc** **sort** data = pat;

by patid;

**run**;

**proc** **sort** data = diag;

by patid;

**run**;

**data** mdat;

merge pat diag;

by patid;

run;

**proc** **sort** data=mdat;

by patid;

**run**;

**proc** **sort** data=surg;

by patid;

**run**;

**data** mdat;

merge mdat surg;

by patid;

run;

**proc** **sort** data=mdat;

by patid;

**run**;

**proc** **sort** data=isl;

by patid;

**run**;

**data** mdat;

merge mdat isl;

by patid;

run;

\* amd and sf have multiple obs per pat;

\*number of pts with panx and also with isl?;

**proc** **freq** data=mdat;

table autoislettx;

**run**;

\*92 pts received surgery and 63 also rcvd isl;

\*calculate duration of the disease from diag date and surgery date;

**data** mdat;

set mdat;

dur = year(surgery\_dt)-pancreatitis\_yr;

\*classify as white or other;

race = '';

if (white = **1**) and (hispanic ne **1**) then race = 'White,Non-hispanic';

if (white = **0**) then race = 'Other';

\*calculate BMI;

bmi = wt\_pre/(ht\_pre\*\***2**);

\*simplify diagnosis;

diagsimp= '';

if pancreatitis1 ne '' then diagsimp = 'Other';

if pancreatitis1 = 'Alcohol' then diagsimp = 'Alcohol';

if pancreatitis1 = 'Idiopathic' then diagsimp = 'Idiopathic';

if pancreatitis1 = 'Pancreas Divisum' then diagsimp = 'Pancreas Divisum';

if pancreatitis1 = 'Familial' then diagsimp = 'Familial';

run;

\*Summarize two groups;

**proc** **tabulate** data = mdat;

label autoislettx = 'Auto-Islet Treatment';

label dur = 'Duration of Disease (years)';

var age bmi dur;

class autoislettx gender diagsimp race;

table age\*(N mean min max) BMI\*(N mean min max) dur\*(N mean min max) Gender\*N, autoislettx;

**run**;

**proc** **tabulate** data = mdat;

label autoislettx = 'Auto-Islet Treatment';

label diagsimp = 'Diagnosis Category';

class autoislettx diagsimp;

table diagsimp\*N, autoislettx;

**run**;

\*Thanks for this code!;

**data** sfz;

set sf;

/\*Standardize \*/

PF\_Z = (PF - **84.52404**) / **22.89490** ;

RP\_Z = (RP - **81.19907**) / **33.79729** ;

BP\_Z = (BP - **75.49196**) / **23.55879** ;

GH\_Z = (GH - **72.21316**) / **20.16964** ;

MH\_Z = (MH - **74.84212**) / **18.01189** ;

RE\_Z = (RE - **81.29467**) / **33.02717** ;

SF\_Z = (SF - **83.59753**) / **22.37642** ;

VT\_Z = (VT - **61.05453**) / **20.86942** ;

/\*create physical and mental health component score \*/

PCS = (PF\_Z \* **0.42402**) + (RP\_Z \* **0.35119**) + (BP\_Z \* **0.31754**) +

(GH\_Z \* **0.24954**) + (MH\_Z \* -**.22069**) + (RE\_Z \* -**.19206**) +

(SF\_Z \* -**.00753**) + (VT\_Z \* **0.02877**);

MCS = (PF\_Z \* -**.22999**) + (RP\_Z \* -**.12329**) + (BP\_Z \* -**.09731**) +

(GH\_Z \* -**.01571**) + (MH\_Z \* **0.48581**) + (RE\_Z \* **0.43407**) +

(SF\_Z \* **0.26876**) + (VT\_Z \* **0.23534**);

/\* create the score \*/

PCS = **50** + (PCS \* **10**);

MCS = **50** + (MCS \* **10**);

drop pf--rp\_z;

drop gh\_z--vt\_z;

run;

\*create separate data for each visit to retain ptid date, BP\_Z, MCS and PCS;

**proc** **sort** data=sfz;

by patid;

**run**;

**proc** **sort** data = mdat;

by patid;

**run**;

\*merge into long form;

**data** longm;

merge mdat sfz;

by patid;

run;

\*calculate months since surgery;

**data** longm;

set longm;

month\_since = round( (fu\_dt-surgery\_dt)/**30.5**,**1**);

run;

**proc** **sort** data =longm;

by month\_since;

**run**;

**proc** **boxplot** data = longm;

label month\_since = 'Months since surgery';

label BP\_Z = 'Bodily Pain';

where month\_since in(**0**,**3**,**6**);

plot BP\_Z\*month\_since;

**run**;

**proc** **boxplot** data = longm;

label month\_since = 'Months since surgery';

label mcs = 'Mental Component Score';

where month\_since in(**0**,**3**,**6**);

plot mcs\*month\_since;

**run**;

**proc** **boxplot** data = longm;

label month\_since = 'Months since surgery';

label pcs = 'Physical Component Score';

where month\_since in(**0**,**3**,**6**);

plot pcs\*month\_since;

**run**;

\*this trend doesn't look so convincing when you look at all time points

\*relationship between ieq and dur or bmi;

**proc** **glm** data= mdat;

model ieq = dur bmi ;

**run**;

\*no significant relationship try without outliers;

**proc** **glm** data= mdat;

where ieqout = **0**;

model ieq = dur bmi ;

**run**;

\*after removing outlying ieq values there is a relationship with BMI;

**proc** **sgplot** data = mdat;

label ieq = 'Islet Gain';

label bmi = 'Body Mass Index';

where ieqout = **0**;

reg x=bmi y=ieq/ CLM

CLMATTRS=(CLMLINEATTRS=

(COLOR=Green PATTERN= ShortDash));

**run**;

**proc** **sgplot** data = mdat;

label ieq = 'Islet Gain';

label dur = 'Duration';

where ieqout = **0**;

reg x=dur y=ieq / CLM

CLMATTRS=(CLMLINEATTRS=

(COLOR=Green PATTERN= ShortDash));

**run**;

\*Test for difference in bodily pain between baseline and 6 months;

**data** longm;

set longm;

if month\_since = **0** then tmp = 'Baseline';

if month\_since = **6** then tmp = '6 Months';

run;

**proc** **ttest** data= longm;

class tmp;

var BP\_Z;

**run**;

\*normality is slightly suspect so check with rank-sum test;

**proc** **npar1way** wilcoxon correct=no data=longm;

class tmp;

var BP\_Z;

**run**;

\*even more significant;

\*readmission rates for the two groups overall, at 1 month and 6months;

\*merge mdat and readmit data;

**proc** **sort** data=mdat;

by patid;

**run**;

**proc** **sort** data=amd;

by patid;

**run**;

**data** ldat2;

merge mdat amd;

by patid;

run;

**data** ldat2;

set ldat2;

day\_since = readm\_dt-surgery\_dt;

month\_since = (day\_since)/**30.5**;

run;

\*limit ti just cases with a surgery;

**data** ldat2;

set ldat2;

if surgery\_dt = **.** then delete;

\*remove admission dates before surgery;

if (read\_dt ne **.**) and (readm\_dt < surgery\_dt) then delete;

run;

\*check for any recurrenc by limiting to first recurence;

**proc** **sort** data =ldat2;

by patid readm\_dt;

**run**;

**data** ldat2\_first;

set ldat2;

if first.patid;

by patid;

run;

**data** ldat2\_first;

set ldat2\_first;

if readm\_dt ne **.** then read = 'Yes';

if readm\_dt = **.** then read = 'No';

run;

**proc** **tabulate** data = ldat2\_first;

label autoislettx = 'Auto-Islet Status';

label read = 'Readmission Status';

class read autoislettx;

table read, autoislettx\*N;

**run**;

\*limit to only readmited patients to compare ER and overall beteen the two groups;

**data** ldat2\_ronly;

set ldat2;

if readm\_dt = **.** then delete;

run;

\*limit to only cases where readmits happen in first month;

**proc** **sort** data= ldat2\_ronly;

by patid;

**run**;

**data** ldat2\_1month;

set ldat2\_ronly;

if month\_since > **1** then delete;

\*eliminate repeated visits in the first month;

if first.patid;

by patid;

run;

**proc** **format**;

value er\_fmt

**0** = 'No'

**1** = 'Yes';

**run**;

**proc** **tabulate** data = ldat2\_1month;

label autoislettx = 'Auto-Islet Status';

format er er\_fmt.;

class er autoislettx;

table er, autoislettx;

**run**;

\*limit to readmits to before 6 months;

**data** ldat2\_6month;

set ldat2\_ronly;

if month\_since > **6** then delete;

\*eliminate repeated visits in the first month;

if first.patid;

by patid;

run;

**proc** **tabulate** data = ldat2\_6month;

label autoislettx = 'Auto-Islet Status';

format er er\_fmt.;

class er autoislettx;

table er, autoislettx;

**run**;

SAS Log:

467 \*read in patient data;

468 Data pat;

469 set final.patient;

470 run;

NOTE: There were 94 observations read from the data set FINAL.PATIENT.

NOTE: The data set WORK.PAT has 94 observations and 9 variables.

NOTE: DATA statement used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

471

472 data diag;

473 set final.diagnosis;

474 run;

NOTE: There were 106 observations read from the data set FINAL.DIAGNOSIS.

NOTE: The data set WORK.DIAG has 106 observations and 3 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

475

476 data surg;

477 set final.surgery;

478 run;

NOTE: There were 90 observations read from the data set FINAL.SURGERY.

NOTE: The data set WORK.SURG has 90 observations and 7 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

479

480 data isl;

481 set final.islets;

482 run;

NOTE: There were 67 observations read from the data set FINAL.ISLETS.

NOTE: The data set WORK.ISL has 67 observations and 2 variables.

NOTE: DATA statement used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

483

484 data sf;

485 set final.sf36;

486 run;

NOTE: There were 189 observations read from the data set FINAL.SF36.

NOTE: The data set WORK.SF has 189 observations and 10 variables.

NOTE: DATA statement used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

487

488 data amd;

489 set final.readmin;

490 run;

NOTE: There were 350 observations read from the data set FINAL.READMIN.

NOTE: The data set WORK.AMD has 350 observations and 5 variables.

NOTE: DATA statement used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

491

492 \*data clean-up;

493 proc freq data = pat;

494 table \_character\_;

495 run;

NOTE: There were 94 observations read from the data set WORK.PAT.

NOTE: PROCEDURE FREQ used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

496

497 proc means data= pat nmiss min max mean;

498 var \_numeric\_;

499 run;

NOTE: There were 94 observations read from the data set WORK.PAT.

NOTE: PROCEDURE MEANS used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

500

501 proc freq data = diag;

502 table \_character\_;

503 run;

NOTE: There were 106 observations read from the data set WORK.DIAG.

NOTE: PROCEDURE FREQ used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

504 proc means data= diag nmiss min max mean;

505 var \_numeric\_;

506 run;

NOTE: There were 106 observations read from the data set WORK.DIAG.

NOTE: PROCEDURE MEANS used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

507 \*change unable to confirm to unknown;

508 data diag;

509 set diag;

510 if Pancreatitis1 = 'Unable to confirm' then Pancreatitis1 = 'Unknown';

511 \*change 9999 to missing;

512 if Pancreatitis\_yr = 9999 then Pancreatitis\_yr = .;

513 run;

NOTE: There were 106 observations read from the data set WORK.DIAG.

NOTE: The data set WORK.DIAG has 106 observations and 3 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

514 \*remove duplicates;

515 proc sort data=diag;

516 by PatID;

517 run;

NOTE: There were 106 observations read from the data set WORK.DIAG.

NOTE: The data set WORK.DIAG has 106 observations and 3 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

518 data diag;

519 set diag;

520 by patid;

521 if first.PatID;

522 run;

NOTE: There were 106 observations read from the data set WORK.DIAG.

NOTE: The data set WORK.DIAG has 104 observations and 3 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

523

524 proc means data= isl nmiss min max mean;

525 var \_numeric\_;

526 run;

NOTE: There were 67 observations read from the data set WORK.ISL.

NOTE: PROCEDURE MEANS used (Total process time):

real time 0.03 seconds

cpu time 0.01 seconds

527 proc univariate data=isl plot trimmed=(0.05);

528 var ieq;

529 run;

NOTE: PROCEDURE UNIVARIATE used (Total process time):

real time 0.17 seconds

cpu time 0.04 seconds

530 \*negative value and 1 very high outlier;

531 data isl;

532 set isl;

533 ieqout = 0;

534 if PatID < 1 then delete;

535 if ieq > 3000000 then ieq = .;

536 if ieq > 750000 then ieqout = 1;

537 run;

NOTE: There were 67 observations read from the data set WORK.ISL.

NOTE: The data set WORK.ISL has 66 observations and 3 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

538 \*there are still a couple outliers, I labelled them with indicator;

539 proc sort data=isl;

540 by patid;

541 run;

NOTE: There were 66 observations read from the data set WORK.ISL.

NOTE: The data set WORK.ISL has 66 observations and 3 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

542 data isl;

543 set isl;

544 by patid;

545 if first.PatID;

546 run;

NOTE: There were 66 observations read from the data set WORK.ISL.

NOTE: The data set WORK.ISL has 63 observations and 3 variables.

NOTE: DATA statement used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

547

548 proc freq data = surg;

549 table \_character\_;

550 run;

NOTE: There were 90 observations read from the data set WORK.SURG.

NOTE: PROCEDURE FREQ used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

551 proc means data= surg nmiss min max mean;

552 var \_numeric\_;

553 run;

NOTE: There were 90 observations read from the data set WORK.SURG.

NOTE: PROCEDURE MEANS used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

554 proc univariate data=surg plot trimmed=0.05;

555 var ht\_pre wt\_pre;

556 histogram;

557 run;

NOTE: PROCEDURE UNIVARIATE used (Total process time):

real time 0.49 seconds

cpu time 0.18 seconds

558 \*a couple high wieght outliers and one low hieght outlier;

559 \*check bivariate outliers;

560 proc sgplot data=surg;

561 scatter x=ht\_pre y=wt\_pre;

562 run;

WARNING: SASUSER.TEMPLAT is not a template store! It will be ignored.

NOTE: PROCEDURE SGPLOT used (Total process time):

real time 0.12 seconds

cpu time 0.01 seconds

NOTE: There were 90 observations read from the data set WORK.SURG.

563 \*thought I would see a tighter relationship there!;

564 \* I don't think there are any observations extreme enough to merit removal;

565

566

567 proc freq data = amd;

568 table \_character\_;

569 run;

NOTE: There were 350 observations read from the data set WORK.AMD.

NOTE: PROCEDURE FREQ used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

570 proc means data= amd nmiss min max mean;

571 var \_numeric\_;

572 run;

NOTE: There were 350 observations read from the data set WORK.AMD.

NOTE: PROCEDURE MEANS used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

573 \*consolodate admission reasons;

574 data amd;

575 set amd;

576 if Admission\_reason1 = 'Gastrointestinal bleeding' then Admission\_reason1 = 'GI Bleeding';

577 if Admission\_reason1 = 'Hypoglyxemia' then Admission\_reason1 = 'Hypoglycemia';

578 if Admission\_reason1 = 'Incisonal hernia' then Admission\_reason1 = 'Incisional hernia';

579 if Admission\_reason1 = 'Nausea&Vomiting' or Admission\_reason1 = 'Vomiting' then

579! Admission\_reason1 = 'Nausea/Vomiting';

580 if Admission\_reason1 = 'Pain' then Admission\_reason1 = 'Abdominal Pain';\*not sure if this is

580! right - I would check with PI;

581 run;

NOTE: There were 350 observations read from the data set WORK.AMD.

NOTE: The data set WORK.AMD has 350 observations and 5 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

582

583 proc means data= sf nmiss min max mean;

584 var \_numeric\_;

585 run;

NOTE: There were 189 observations read from the data set WORK.SF.

NOTE: PROCEDURE MEANS used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

586

587 \*merge data;

588 proc sort data = pat;

589 by patid;

590 run;

NOTE: There were 94 observations read from the data set WORK.PAT.

NOTE: The data set WORK.PAT has 94 observations and 9 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time 0.01 seconds

cpu time 0.00 seconds

591 proc sort data = diag;

592 by patid;

593 run;

NOTE: There were 104 observations read from the data set WORK.DIAG.

NOTE: The data set WORK.DIAG has 104 observations and 3 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

594 data mdat;

595 merge pat diag;

596 by patid;

597 run;

NOTE: There were 94 observations read from the data set WORK.PAT.

NOTE: There were 104 observations read from the data set WORK.DIAG.

NOTE: The data set WORK.MDAT has 108 observations and 11 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

598

599 proc sort data=mdat;

600 by patid;

601 run;

NOTE: There were 108 observations read from the data set WORK.MDAT.

NOTE: The data set WORK.MDAT has 108 observations and 11 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

602 proc sort data=surg;

603 by patid;

604 run;

NOTE: There were 90 observations read from the data set WORK.SURG.

NOTE: The data set WORK.SURG has 90 observations and 7 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

605 data mdat;

606 merge mdat surg;

607 by patid;

608 run;

NOTE: There were 108 observations read from the data set WORK.MDAT.

NOTE: There were 90 observations read from the data set WORK.SURG.

NOTE: The data set WORK.MDAT has 108 observations and 17 variables.

NOTE: DATA statement used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

609

610 proc sort data=mdat;

611 by patid;

612 run;

NOTE: There were 108 observations read from the data set WORK.MDAT.

NOTE: The data set WORK.MDAT has 108 observations and 17 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

613 proc sort data=isl;

614 by patid;

615 run;

NOTE: There were 63 observations read from the data set WORK.ISL.

NOTE: The data set WORK.ISL has 63 observations and 3 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

616 data mdat;

617 merge mdat isl;

618 by patid;

619 run;

NOTE: There were 108 observations read from the data set WORK.MDAT.

NOTE: There were 63 observations read from the data set WORK.ISL.

NOTE: The data set WORK.MDAT has 110 observations and 19 variables.

NOTE: DATA statement used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

620

621 \* amd and sf have multiple obs per pat;

622

623 \*number of pts with panx and also with isl?;

624 proc freq data=mdat;

625 table autoislettx;

626 run;

NOTE: There were 110 observations read from the data set WORK.MDAT.

NOTE: PROCEDURE FREQ used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

627 \*92 pts received surgery and 63 also rcvd isl;

628

629 \*calculate duration of the disease from diag date and surgery date;

630 data mdat;

631 set mdat;

632 dur = year(surgery\_dt)-pancreatitis\_yr;

633 \*classify as white or other;

634 race = '';

635 if (white = 1) and (hispanic ne 1) then race = 'White,Non-hispanic';

636 if (white = 0) then race = 'Other';

637 \*calculate BMI;

638 bmi = wt\_pre/(ht\_pre\*\*2);

639 \*simplify diagnosis;

640 diagsimp= '';

641 if pancreatitis1 ne '' then diagsimp = 'Other';

642 if pancreatitis1 = 'Alcohol' then diagsimp = 'Alcohol';

643 if pancreatitis1 = 'Idiopathic' then diagsimp = 'Idiopathic';

644 if pancreatitis1 = 'Pancreas Divisum' then diagsimp = 'Pancreas Divisum';

645 if pancreatitis1 = 'Familial' then diagsimp = 'Familial';

646 run;

NOTE: Missing values were generated as a result of performing an operation on missing values.

Each place is given by: (Number of times) at (Line):(Column).

18 at 632:7 27 at 632:23 19 at 638:13 19 at 638:21

NOTE: There were 110 observations read from the data set WORK.MDAT.

NOTE: The data set WORK.MDAT has 110 observations and 23 variables.

NOTE: DATA statement used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

647 \*Summarize two groups;

648 proc tabulate data = mdat;

649 label autoislettx = 'Auto-Islet Treatment';

650 label dur = 'Duration of Disease (years)';

651 var age bmi dur;

652 class autoislettx gender diagsimp race;

653 table age\*(N mean min max) BMI\*(N mean min max) dur\*(N mean min max) Gender\*N, autoislettx;

654 run;

NOTE: There were 110 observations read from the data set WORK.MDAT.

NOTE: PROCEDURE TABULATE used (Total process time):

real time 0.03 seconds

cpu time 0.01 seconds

655

656 proc tabulate data = mdat;

657 label autoislettx = 'Auto-Islet Treatment';

658 label diagsimp = 'Diagnosis Category';

659 class autoislettx diagsimp;

660 table diagsimp\*N, autoislettx;

661 run;

NOTE: There were 110 observations read from the data set WORK.MDAT.

NOTE: PROCEDURE TABULATE used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

662

663 \*Thanks for this code!;

664 data sfz;

665 set sf;

666 /\*Standardize \*/

667 PF\_Z = (PF - 84.52404) / 22.89490 ;

668 RP\_Z = (RP - 81.19907) / 33.79729 ;

669 BP\_Z = (BP - 75.49196) / 23.55879 ;

670 GH\_Z = (GH - 72.21316) / 20.16964 ;

671 MH\_Z = (MH - 74.84212) / 18.01189 ;

672 RE\_Z = (RE - 81.29467) / 33.02717 ;

673 SF\_Z = (SF - 83.59753) / 22.37642 ;

674 VT\_Z = (VT - 61.05453) / 20.86942 ;

675 /\*create physical and mental health component score \*/

676 PCS = (PF\_Z \* 0.42402) + (RP\_Z \* 0.35119) + (BP\_Z \* 0.31754) +

677 (GH\_Z \* 0.24954) + (MH\_Z \* -.22069) + (RE\_Z \* -.19206) +

678 (SF\_Z \* -.00753) + (VT\_Z \* 0.02877);

679 MCS = (PF\_Z \* -.22999) + (RP\_Z \* -.12329) + (BP\_Z \* -.09731) +

680 (GH\_Z \* -.01571) + (MH\_Z \* 0.48581) + (RE\_Z \* 0.43407) +

681 (SF\_Z \* 0.26876) + (VT\_Z \* 0.23534);

682

683 /\* create the score \*/

684 PCS = 50 + (PCS \* 10);

685 MCS = 50 + (MCS \* 10);

686 drop pf--rp\_z;

687 drop gh\_z--vt\_z;

688 run;

NOTE: Missing values were generated as a result of performing an operation on missing values.

Each place is given by: (Number of times) at (Line):(Column).

1 at 668:16 1 at 669:16 1 at 672:16 1 at 676:35 1 at 679:35 1 at 684:20

1 at 685:20

NOTE: There were 189 observations read from the data set WORK.SF.

NOTE: The data set WORK.SFZ has 189 observations and 5 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

689

690 \*create separate data for each visit to retain ptid date, BP\_Z, MCS and PCS;

691 proc sort data=sfz;

692 by patid;

693 run;

NOTE: There were 189 observations read from the data set WORK.SFZ.

NOTE: The data set WORK.SFZ has 189 observations and 5 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

694 proc sort data = mdat;

695 by patid;

696 run;

NOTE: There were 110 observations read from the data set WORK.MDAT.

NOTE: The data set WORK.MDAT has 110 observations and 23 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

697

698 \*merge into long form;

699 data longm;

700 merge mdat sfz;

701 by patid;

702 run;

NOTE: MERGE statement has more than one data set with repeats of BY values.

NOTE: There were 110 observations read from the data set WORK.MDAT.

NOTE: There were 189 observations read from the data set WORK.SFZ.

NOTE: The data set WORK.LONGM has 237 observations and 27 variables.

NOTE: DATA statement used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

703

704 \*calculate months since surgery;

705 data longm;

706 set longm;

707 month\_since = round( (fu\_dt-surgery\_dt)/30.5,1);

708 run;

NOTE: Missing values were generated as a result of performing an operation on missing values.

Each place is given by: (Number of times) at (Line):(Column).

48 at 707:15 48 at 707:28 48 at 707:40

NOTE: There were 237 observations read from the data set WORK.LONGM.

NOTE: The data set WORK.LONGM has 237 observations and 28 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

709

710 proc sort data =longm;

711 by month\_since;

712 run;

NOTE: There were 237 observations read from the data set WORK.LONGM.

NOTE: The data set WORK.LONGM has 237 observations and 28 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

713 proc boxplot data = longm;

714 label month\_since = 'Months since surgery';

715 label BP\_Z = 'Bodily Pain';

716 where month\_since in(0,3,6);

717 plot BP\_Z\*month\_since;

718 run;

NOTE: Processing beginning for PLOT statement number 1.

NOTE: There were 91 observations read from the data set WORK.LONGM.

WHERE month\_since in (0, 3, 6);

NOTE: PROCEDURE BOXPLOT used (Total process time):

real time 0.18 seconds

cpu time 0.09 seconds

719 proc boxplot data = longm;

720 label month\_since = 'Months since surgery';

721 label mcs = 'Mental Component Score';

722 where month\_since in(0,3,6);

723 plot mcs\*month\_since;

724 run;

NOTE: Processing beginning for PLOT statement number 1.

NOTE: There were 91 observations read from the data set WORK.LONGM.

WHERE month\_since in (0, 3, 6);

NOTE: PROCEDURE BOXPLOT used (Total process time):

real time 0.17 seconds

cpu time 0.07 seconds

725 proc boxplot data = longm;

726 label month\_since = 'Months since surgery';

727 label pcs = 'Physical Component Score';

728 where month\_since in(0,3,6);

729 plot pcs\*month\_since;

730 run;

NOTE: Processing beginning for PLOT statement number 1.

NOTE: There were 91 observations read from the data set WORK.LONGM.

WHERE month\_since in (0, 3, 6);

NOTE: PROCEDURE BOXPLOT used (Total process time):

real time 0.17 seconds

cpu time 0.07 seconds

731 \*this trend doesn't look so convincing when you look at all time points

732

733 \*relationship between ieq and dur or bmi;

734 proc glm data= mdat;

735 model ieq = dur bmi ;

736 run;

737 \*no significant relationship try without outliers;

NOTE: PROCEDURE GLM used (Total process time):

real time 0.63 seconds

cpu time 0.09 seconds

738 proc glm data= mdat;

739 where ieqout = 0;

740 model ieq = dur bmi ;

741 run;

742 \*after removing outlying ieq values there is a relationship with BMI;

NOTE: PROCEDURE GLM used (Total process time):

real time 0.65 seconds

cpu time 0.09 seconds

743 proc sgplot data = mdat;

744 label ieq = 'Islet Gain';

745 label bmi = 'Body Mass Index';

746 where ieqout = 0;

747 reg x=bmi y=ieq/ CLM

748 CLMATTRS=(CLMLINEATTRS=

749 (COLOR=Green PATTERN= ShortDash));

750 run;

WARNING: SASUSER.TEMPLAT is not a template store! It will be ignored.

NOTE: PROCEDURE SGPLOT used (Total process time):

real time 0.14 seconds

cpu time 0.04 seconds

NOTE: There were 63 observations read from the data set WORK.MDAT.

WHERE ieqout=0;

751 proc sgplot data = mdat;

752 label ieq = 'Islet Gain';

753 label dur = 'Duration';

754 where ieqout = 0;

755 reg x=dur y=ieq / CLM

756 CLMATTRS=(CLMLINEATTRS=

757 (COLOR=Green PATTERN= ShortDash));

758 run;

WARNING: SASUSER.TEMPLAT is not a template store! It will be ignored.

NOTE: PROCEDURE SGPLOT used (Total process time):

real time 0.14 seconds

cpu time 0.04 seconds

NOTE: There were 63 observations read from the data set WORK.MDAT.

WHERE ieqout=0;

759

760 \*Test for difference in bodily pain between baseline and 6 months;

761 data longm;

762 set longm;

763 if month\_since = 0 then tmp = 'Baseline';

764 if month\_since = 6 then tmp = '6 Months';

765 run;

NOTE: There were 237 observations read from the data set WORK.LONGM.

NOTE: The data set WORK.LONGM has 237 observations and 29 variables.

NOTE: DATA statement used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

766

767 proc ttest data= longm;

768 class tmp;

769 var BP\_Z;

770 run;

NOTE: PROCEDURE TTEST used (Total process time):

real time 0.42 seconds

cpu time 0.18 seconds

771 \*normality is slightly suspect so check with rank-sum test;

772 proc npar1way wilcoxon correct=no data=longm;

773 class tmp;

774 var BP\_Z;

775 run;

NOTE: There were 237 observations read from the data set WORK.LONGM.

NOTE: PROCEDURE NPAR1WAY used (Total process time):

real time 0.17 seconds

cpu time 0.06 seconds

776 \*even more significant;

777

778 \*readmission rates for the two groups overall, at 1 month and 6months;

779 \*merge mdat and readmit data;

780 proc sort data=mdat;

781 by patid;

782 run;

NOTE: Input data set is already sorted, no sorting done.

NOTE: PROCEDURE SORT used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

783 proc sort data=amd;

784 by patid;

785 run;

NOTE: There were 350 observations read from the data set WORK.AMD.

NOTE: The data set WORK.AMD has 350 observations and 5 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

786 data ldat2;

787 merge mdat amd;

788 by patid;

789 run;

NOTE: MERGE statement has more than one data set with repeats of BY values.

NOTE: There were 110 observations read from the data set WORK.MDAT.

NOTE: There were 350 observations read from the data set WORK.AMD.

NOTE: The data set WORK.LDAT2 has 381 observations and 27 variables.

NOTE: DATA statement used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

790

791 data ldat2;

792 set ldat2;

793 day\_since = readm\_dt-surgery\_dt;

794 month\_since = (day\_since)/30.5;

795 run;

NOTE: Missing values were generated as a result of performing an operation on missing values.

Each place is given by: (Number of times) at (Line):(Column).

63 at 793:21 63 at 794:26

NOTE: There were 381 observations read from the data set WORK.LDAT2.

NOTE: The data set WORK.LDAT2 has 381 observations and 29 variables.

NOTE: DATA statement used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

796

797 \*limit ti just cases with a surgery;

798 data ldat2;

799 set ldat2;

800 if surgery\_dt = . then delete;

801 \*remove admission dates before surgery;

802 if (read\_dt ne .) and (readm\_dt < surgery\_dt) then delete;

803 run;

NOTE: Variable read\_dt is uninitialized.

NOTE: There were 381 observations read from the data set WORK.LDAT2.

NOTE: The data set WORK.LDAT2 has 340 observations and 30 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

804

805 \*check for any recurrenc by limiting to first recurence;

806 proc sort data =ldat2;

807 by patid readm\_dt;

808 run;

NOTE: There were 340 observations read from the data set WORK.LDAT2.

NOTE: The data set WORK.LDAT2 has 340 observations and 30 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

809 data ldat2\_first;

810 set ldat2;

811 if first.patid;

812 by patid;

813 run;

NOTE: There were 340 observations read from the data set WORK.LDAT2.

NOTE: The data set WORK.LDAT2\_FIRST has 90 observations and 30 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

814

815 data ldat2\_first;

816 set ldat2\_first;

817 if readm\_dt ne . then read = 'Yes';

818 if readm\_dt = . then read = 'No';

819 run;

NOTE: There were 90 observations read from the data set WORK.LDAT2\_FIRST.

NOTE: The data set WORK.LDAT2\_FIRST has 90 observations and 31 variables.

NOTE: DATA statement used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

820

821 proc tabulate data = ldat2\_first;

822 label autoislettx = 'Auto-Islet Status';

823 label read = 'Readmission Status';

824 class read autoislettx;

825 table read, autoislettx\*N;

826 run;

NOTE: There were 90 observations read from the data set WORK.LDAT2\_FIRST.

NOTE: PROCEDURE TABULATE used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

827

828 \*limit to only readmited patients to compare ER and overall beteen the two groups;

829 data ldat2\_ronly;

830 set ldat2;

831 if readm\_dt = . then delete;

832 run;

NOTE: There were 340 observations read from the data set WORK.LDAT2.

NOTE: The data set WORK.LDAT2\_RONLY has 318 observations and 30 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

833 \*limit to only cases where readmits happen in first month;

834 proc sort data= ldat2\_ronly;

835 by patid;

836 run;

NOTE: There were 318 observations read from the data set WORK.LDAT2\_RONLY.

NOTE: The data set WORK.LDAT2\_RONLY has 318 observations and 30 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

837 data ldat2\_1month;

838 set ldat2\_ronly;

839 if month\_since > 1 then delete;

840 \*eliminate repeated visits in the first month;

841 if first.patid;

842 by patid;

843 run;

NOTE: There were 318 observations read from the data set WORK.LDAT2\_RONLY.

NOTE: The data set WORK.LDAT2\_1MONTH has 40 observations and 30 variables.

NOTE: DATA statement used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

844

845 proc format;

846 value er\_fmt

847 0 = 'No'

848 1 = 'Yes';

NOTE: Format ER\_FMT is already on the library.

NOTE: Format ER\_FMT has been output.

849 run;

NOTE: PROCEDURE FORMAT used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

850

851

852 proc tabulate data = ldat2\_1month;

853 label autoislettx = 'Auto-Islet Status';

854 format er er\_fmt.;

855 class er autoislettx;

856 table er, autoislettx;

857 run;

NOTE: There were 40 observations read from the data set WORK.LDAT2\_1MONTH.

NOTE: PROCEDURE TABULATE used (Total process time):

real time 0.03 seconds

cpu time 0.03 seconds

858

859 \*limit to readmits to before 6 months;

860 data ldat2\_6month;

861 set ldat2\_ronly;

862 if month\_since > 6 then delete;

863 \*eliminate repeated visits in the first month;

864 if first.patid;

865 by patid;

866 run;

NOTE: There were 318 observations read from the data set WORK.LDAT2\_RONLY.

NOTE: The data set WORK.LDAT2\_6MONTH has 59 observations and 30 variables.

NOTE: DATA statement used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

867

868 proc tabulate data = ldat2\_6month;

869 label autoislettx = 'Auto-Islet Status';

870 format er er\_fmt.;

871 class er autoislettx;

872 table er, autoislettx;

873 run;

NOTE: There were 59 observations read from the data set WORK.LDAT2\_6MONTH.

NOTE: PROCEDURE TABULATE used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds