

**A PRELIMINARY EVALUATION OF THE ROLE OF
ADIPONECTIN AND LEPTIN IN TYPE 1 DIABETES**

by

Wenxiu Dong

B.S. in Mathematics and Applied Mathematics, China University of Geosciences, China, 2012

Submitted to the Graduate Faculty of

the Department of Biostatistics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH
GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Wenxiu Dong

It was defended on

April 22, 2014

Ingrid Libman, MD, PhD, Assistant Professor, Pediatrics,
Children's Hospital of Pittsburgh of UPMC

Natalie Hecht Baldauff, DO, Fellow, Pediatric Endocrinology,
Children's Hospital of Pittsburgh of UPMC

Francis Pike, PhD, Assistant Professor, Biostatistics, Graduate School of Public Health,
University of Pittsburgh

Evelyn Talbott, DrPH, Professor, Epidemiology, Graduate School of Public Health,
University of Pittsburgh

Thesis Advisor: Vincent C. Arena, PhD, Associate Professor, Biostatistics, Graduate School
of Public Health, University of Pittsburgh

Copyright © by Wenxiu Dong

2014

Vincent C. Arena, PhD

**A PRELIMINARY EVALUATION OF THE ROLE OF
ADIPONECTIN AND LEPTIN IN TYPE 1 DIABETES**

Wenxiu Dong, M.S.

University of Pittsburgh, 2014

ABSTRACT

Type 1 Diabetes (T1D), as a chronic disease, has been an emerging issue in the field of public health. In juvenile T1D, the effect of insulin on adiponectin and leptin has not been well studied nor understood. In this thesis we focused on the question concerning patient's adiponectin and leptin level changes over a 5 month period before and after the start of insulin therapy in children with newly diagnosed T1D. A substantial focus of this project was the assembling of an analytical data file that could be used for the initial analysis and for future in-depth studies once updated data becomes available. Data cleaning and verification were carried out on all data sets. All programming was done using SAS 9.3 (32).

We found no statistically significant difference in adiponectin levels between days 0 and 1, days 0 and 3; however there was a statistically significant increase in adiponectin levels between days 0 and 5 (post insulin therapy). Adiponectin levels remained significantly elevated at the 3 month visit. During the first five days leptin levels did not significantly differ from each other. However, at the 3 month visit the values were significantly higher from baseline.

Adiponectin and leptin levels may serve as a good indicator for insulin sensitivity and provide a better understanding into the relationship with insulin therapy, thereby providing improvements in treatment for juvenile T1D.

TABLE OF CONTENTS

PREFACE.....	VIII
1.0 INTRODUCTION.....	1
2.0 GENERAL DATA MANAGEMENT	3
3.0 SLOW PROGRESSOR STUDY.....	5
3.1.1 Question of interest.....	5
3.1.2 Antibody	5
3.1.3 Assessment of elapsed time since positive antibodies.....	6
3.1.4 Data processing	6
3.1.5 Slow progressor subgroup identified	7
4.0 THE ADIPONECTIN/LEPTIN STUDY	9
4.1.1 Blood draw protocol	9
4.1.2 Windows	10
4.1.3 Target subset	11
4.1.4 Results for adiponectin.....	12
4.1.5 Results for leptin.....	16
5.0 DISCUSSION AND CONCLUSIONS	20
APPENDIX: SAS PROGRAM	21
BIBLIOGRAPHY	31

LIST OF TABLES

Table 3-1 Summary of elapsed years for 354 subset	7
Table 4-1 Summary for adiponectin	13
Table 4-2 Results of paired comparisons for adiponectin	14
Table 4-3 AIC/BIC for adiponectin	15
Table 4-4 Mixed model least squares means	15
Table 4-5 Contrasts to baseline	16
Table 4-6 Summary for leptin.....	17
Table 4-7 Wilcoxon Signed Rank test results for leptin.....	17
Table 4-8 AIC/BIC for leptin.....	18
Table 4-9 Mixed model least square means.....	19
Table 4-10 Contrasts to baseline.....	19

LIST OF FIGURES

Figure 1. Boxplot of adiponectin over time	14
Figure 2. Boxplot of leptin over time.....	18

PREFACE

A special thanks to Dr. Vincent Arena for affording me the opportunity to participate in the Juvenile Diabetes Project (JOD) and gaining practical insight into research issues. All SAS programs were written by myself for this project under the guidance of Dr. Arena.

Interpretation of the statistical analysis was done in collaboration with Drs. Natalie Hecht Baldauff, Arena and Ingrid Libman. Also, a special thanks to Dr. Baldauff who spent valuable time checking the original paper records so we have more accurate data.

Acknowledgement is given to Dr. Dorothy Becker for the use of data from the Juvenile Onset Diabetes Project. This work was supported by National Institutes of Health (NIH) grants R01 DK46864 (D. Becker), Grant Numbers UL1 RR024153 and UL1TR000005 (PCTRC) and the Renziehausen Fund (I. Libman).

1.0 INTRODUCTION

The focus of the thesis was to make a preliminary examination of the relationship of insulin on adiponectin and leptin levels in children newly diagnosed with Type 1 Diabetes (T1D). This project is part of a 38-year prospective study of T1D at Children's Hospital of Pittsburgh Diabetes Center under the direction of Dr. Becker (Juvenile Onset Diabetes Study). A secondary task of this thesis was to identify disease free first degree relatives (FDRs) having positive autoantibodies whose progression to diabetes was greater than ten years or who are still disease-free after ten years. These individuals are referred to as slow progressors. Individuals having two or more positive autoantibodies are at an increased risk of developing T1D and thus, are of particular interest in understanding the etiology and epidemiology of diabetes.

In the past, children presenting with T1D were typically lean or underweight. Examining weight trends over time have shown that the prevalence of being overweight at onset of insulin-treated diabetes in our cohort has tripled from the 1980s to the 1990s (Libman, Pietropaolo et al. 2003, Libman, Pietropaolo et al. 2003). This increased incidence of obesity has also been found in general population. Risk factors such as obesity and insulin-resistance contribute to the development of T1D. In this thesis we use adiponectin as an indicator for insulin resistance and leptin as an indicator for obesity.

The Juvenile Onset Diabetes Study (JOD) at the Pittsburgh Children's Hospital, has been ongoing since 1978. Participants in the study cohort consist of children newly diagnosed with

T1D and their respective family members (mother, father and siblings). The New Onsets are children 18 years and younger who presented to Children's Hospital and were diagnosed with T1D. FDRs are biologically related family members not having indication of diabetes. Serial blood samples were obtained from the New Onsets as well as from the FDRs and frozen for long-term storage. These samples can later be analyzed for specific clinical and research measures depending on the question being posed.

2.0 GENERAL DATA MANAGEMENT

Processing of the data was done prior to any of the analyses. It was necessary to consolidate multiple data sets and abstract the variables of interest to form analytical summary files. In addition the data had to be restructured the data and transform the format from wide (a single record per individual) to long (multiple records per individual). New variables were also created.

Before performing statistical analyses on the data, certain data management was needed. Some of the original data had not been completely verified and edited. In the earlier days of this study, the data was recorded by clinical personnel and research staff and recorded on paper forms. Later it was entered directly into computer data collection forms created in MS ACCESS. There was always a possibility that errors occurred during this process. Thus, data verification and editing was a part of the thesis. Examination of the data was made and tables were created with information about number of observations, mean, standard deviation, median, 25th percentile, 75th percentile, minimum and maximum, etc. for continuous variables. For categorical variables, one-way and two-way tables were created. This was also done for various subgroups of individuals. By summarizing the data we were able to find patients with values that seemed out of range. The next step was to go back to the original paper records and verify that the data was recorded correctly, and that the values were actual outliers. Some data values were identified as errors and updated after checking with the paper records.

Another reason for data verification was to verify that calculations were correctly performed. Some of the recorded information was performed by hand and then entered into the database. Some examples included computation of elapsed times involving date arithmetic. Computed variables were programmed in SAS and verified with the original data. Another issue identified related to not updating the data set when interrelated data was updated. For example, if a blood draw date was changed the elapsed times based on that information also needed to be updated. We identified a number of instances when this occurred. In this project all computations were programmed in SAS.

3.0 SLOW PROGRESSOR STUDY

3.1.1 Question of interest

Researchers have found that FDRs of T1D patients have a higher risk of developing T1D than FDRs who are not related to an individual with T1D. Furthermore, FDRs who are positive to autoantibodies (GAD, IA2, IAA, and ICA human) are at an even greater risk of developing T1D. A recent request has been made by the International Coordinator for the Juvenile Diabetes Research Foundation (JDRF) funded study “What protects islet autoantibody positive individuals who don’t progress” now known as the SNAIL Study. The investigators are interested in finding historical data on people that have had islet autoimmunity for more than 10 years but haven’t (on their last contact) developed any symptoms of insulin dependent diabetes. Once these individuals are identified a comparison can be made to those individuals who are antibody positive but developed disease earlier than ten years. The question is what factors could have protected those individuals from developing T1D.

3.1.2 Antibody

T1D is an autoimmune disease where the body does not recognize islet cell as itself and triggers autoantibodies (antibodies for short) to destroy them. Four types of antibodies are important in

T1D: GAD, IAA, IA2, and ICA human (Libman, Pietropaolo et al. 1998, Pietropaolo, Yu et al. 2005).

3.1.3 Assessment of elapsed time since positive antibodies

Elapsed time is defined as the time interval from the onset of two or more positive autoantibodies to their end date. In the JOD data set all FDR participants were free of insulin dependent T1D when first enrolled. During follow-up a FDR participant may develop insulin dependent diabetes (converters), or may remain disease-free (nonconverters). The end date for converters is defined to be their date when starting insulin treatment. For nonconverters, it is their last contact date.

3.1.4 Data processing

The original source data are maintained in an SQL database and is accessed via MS ACCESS. The necessary tables were first imported into SAS data sets. The data were collected from several time periods for this study corresponding to the different grant funding periods of the JOD Study. Data sets used for the slow progressor study included the participant records for FDRs (containing demographics, medical and follow-up information).

Historic antibody data sets were identified and processed. There was a separate data set for each autoantibody and the relevant data consisted of participant ID, blood draw date, antibody assay value. GAD was considered positive if the value was greater than 0.069; IA2 was positive if the value was greater than 0.032; ICA human was positive if the value was greater than or equal to 5; And, IAA was pre-defined positive or negative in the data set.

Once an individual is positive for a specific antibody, their status does not subsequently change. The definition of having two or more positive antibodies does not have to occur on the same day. For example, a participant testing positive for ICA human on 09 Aug, 2004 and then testing positive for IAA on 01 Aug, 2005 constitutes two positive antibodies. Cumulatively, this participant has tested positive for 2 different antibodies as of 01 Aug 2005.

3.1.5 Slow progressor subgroup identified

Of the 20,706 FDR participants in the JOD Study, 354 have at least 2 antibodies testing positive. Among the 354, 44 have an elapsed time of 10 or more years. The average number of years elapsed is 4.4 with the longest time being 29 years. There were three participants with negative values indicating that they developed positive antibodies after their T1D. All three are converters and their end date is their insulin start date. It is interesting that they tested positive for 2 or more antibodies one year after they started insulin.

Table 3-1 Summary of elapsed years for 354 subset

N	Mean	Std Dev	Minimum	Maximum
354	4.37	4.87	-1.0	29.0

In addition to the 44 confirmed slow progressors, there were 85 individuals who were identified as potentially meeting the 10+ years of elapsed time depending on their disease status once they are contacted. Each of these individuals developed positive antibodies on or before April 1, 2004. However confirmation of their eligibility into the slow progressor could not be made as their last contact information was prior to April 1, 2014. Once these individuals are

contacted, confirmation will be made as to their inclusion or exclusion into the slow progressor study.

4.0 THE ADIPONECTIN/LEPTIN STUDY

4.1.1 Blood draw protocol

The data set used in this study was a subset derived from the JOD cohort and contains 260 new onset T1D subjects. It consists of 130 African American children diagnosed between 1 January 1979 and 31 December 1988 at the Pittsburgh Children's Hospital who were individually matched on sex, age (within 1 year), and diagnosis year (within 1 year) to 130 Caucasians from the Children's Hospital Diabetes Registry (Libman, Pietropaolo et al. 2003). The medical records data set contains patient's information, including age, race, gender and antibodies results, etc. However, blood draw information was not available for all 260 patients. When the blood draw samples were collected, sometimes assays were performed immediately. For other assays, the blood was frozen and stored, and the assays were run at a later date. Since this study began in 1979, and some blood is now unusable and therefore we end up with blood for 208 patients.

The adiponectin/leptin data set contains multiple records per patient. Each record has variables including patient's insulin start data, blood draws date, adiponectin value and leptin value.

The study protocol specified when the blood should be collected. Ideally, there should be 3 blood draws for each patient: one day after the start of insulin, 5 days after the start of insulin and 3 month after the start of insulin. By day 5, most patients went home. This is the reason that

most of the blood draw data occurred during the first 5 days of insulin treatment. The 3 month follow up visit required the patient to come back to the diabetes clinic and have an additional blood draw. The actual date of the 3 month follow up blood draw was quite varied, ranging from two to four months.

4.1.2 Windows

We first merged the medical record data set with the adiponectin/leptin data set containing the multiple blood draws. This enabled the calculation of the elapsed time from insulin start date to blood draw date. The elapsed time was then categorized into one of 11 possible time windows formatted into day 0, day 1, day 2, day3, day4, day 5, month 1, month 2, month 3, month 4 and month 5. Day 0 was the same day as insulin start. Day 1 was one day after insulin start, ... , and day 4 was 4 days after insulin start date. Day 5 was defined as 5 to 14 days after the first insulin injection.

The month measurements were not a specific date but rather an interval of dates. We set month 1 to be the interval [30-15, 30+15) days. Month 2 was the interval [60-15, 60+15) days. Month 3 was the interval [90-15, 90+15) days. Month 4 was the interval [120-15,120+15) days. And, month 5 was [150-15,150+15) days. The maximum elapsed time in our data was 162 days. We then summarized overall adiponectin and leptin levels by elapse time to identify out of range values.

4.1.3 Target subset

The final subset of patients was selected based on having at least one adiponectin or leptin measurement and having complete height, weight, date of birth, race and gender information. There were 171 patients who met these criteria.

After the initial selection, summary statistics were performed on the entire 171 records. We further refined the subset to 156 patients who had at least 3 antibodies tested among GAD, IAA, IA2, ICA human.

In this thesis, we are interested in the patient's adiponectin/leptin changes over time before and after the start of insulin therapy. An issue that arises is that the majority of the patients only have a few measurements. Therefore, we decided to combine consecutive day measurements, and consecutive month measurements, that were not statistically significantly different from each other. We first performed paired t-tests for combinations of any two measurements for both adiponectin and leptin. However, the number of pairs for those t-tests was small. Considering t-test does not perform well with small sample sizes, Q-Q plot was added to further check if the difference of the two measurements follows a normal distribution. If the difference does not violate the normal distribution assumption, the t-test results were used. If the Q-Q plots showed a violation, further analyses were conducted using Wilcoxon Signed Rank test and used.

SAS Macros language was used for the above analysis.

4.1.4 Results for adiponectin

We found an overall positive correlation between adiponectin and leptin values. In addition there was a significant increase in adiponectin as early as 5 days after insulin therapy.

Our aim was to determine if how the level of adiponectin changes after insulin therapy. The paired t-test was performed on the data set containing 171 individual. Table 4-1 shows the basic summary statistics for adiponectin measurement at each time window. We can see that most blood draws were taken on day 0, day 1, day 3, day 5, month 2 and month 3.

As shown in Table 4-2, for the pair-wise comparison, the sample size for each pair is relatively small. And Q-Q plot shows that (day 0, day 1), (day 0, day 3) and (day 1, day 3) are not normally distributed. Paired t-test results were used only for (day 0, day 5) and (day 3, day 5).

Because of small numbers the following comparisons were omitted: day 1 and day 2 (n=1), day 0 and day 4 (n=1), day 1 and day 4 (n=2), day 4 and day 5 (n=1). Month specific comparisons are based on very small sample sizes ($n \leq 3$) and not presented.

By study design patients were to only have a single follow-up visit which was supposed to be at 3 months. However, in practice it ranged from 1 to 5 months. The decision was made to combine the results for adiponectin from months 2 through 4 based on the clinical understanding of the disease pathogenesis. This made the most clinical sense.

Table 4-1 Summary for adiponectin

Variable	N	Mean	Std Dev	Median	Minimum	Maximum	25th Pctl	75th Pctl
adipo day 0	57	14.0	6.99	12.2	3.51	30.1	9.19	18.6
adipo day 1	41	12.5	7.22	10.6	3.37	31.5	6.20	17.0
adipo day 2	2	5.96	0.16	5.96	5.84	6.07	5.84	6.07
adipo day 3	41	18.5	10.5	17.4	2.87	51.4	10.3	22.9
adipo day 4	9	17.1	9.89	19.0	5.96	35.3	6.73	21.4
adipo day 5	38	21.7	9.87	20.6	3.40	40.7	13.8	29.7
adipo month 1	8	27.7	13.5	28.0	10.8	46.6	15.9	38.1
adipo month 2	17	18.8	11.6	16.5	7.68	54.7	11.1	21.8
adipo month 3	20	24.8	12.9	21.0	8.73	55.7	15.6	32.2
adipo month 4	9	21.6	9.60	20.8	8.85	36.9	15.0	30.1
adipo month 5	5	32.3	18.5	25.2	15.3	58.2	18.2	44.9

The initial study was designed to have a baseline measurement of adiponectin and leptin immediately after the first insulin injection and a follow-up measurement which would be taken after about 3 months of insulin therapy. But the data that was collected was sparser than expected. To set up a baseline level, we used the information from the paired t-tests (or Wilcoxon Signed Rank test) to determine which measurements could be collapsed.

After comparing all the paired test (or Wilcoxon Signed Rank test) results, we combined (if multiple values, we averaged) day 0 to day 3 measurements to be the baseline adiponectin level. Day 4 and day 5 were combined (if multiple values, we averaged) to be an immediate

follow-up adiponectin level. The single measure from either month 2, 3 or 4 was used as the 3 month follow-up visit value.

Table 4-2 Results of paired comparisons for adiponectin

adiponectin	Day 0	Day 1	Day 3
Day 1	11(0.07) +	--	--
Day 3	16(1.00) +	1(.)	--
Day 5	22(0.0010) !	4(0.24) !	13(0.0499) !

+ Wilcoxon Signed Rank test

! Paired t-test

Figure 1 show that mean adiponectin levels increase over time. Median value although increased from baseline shows a slight decrease from day 5 to the 3 month visit.

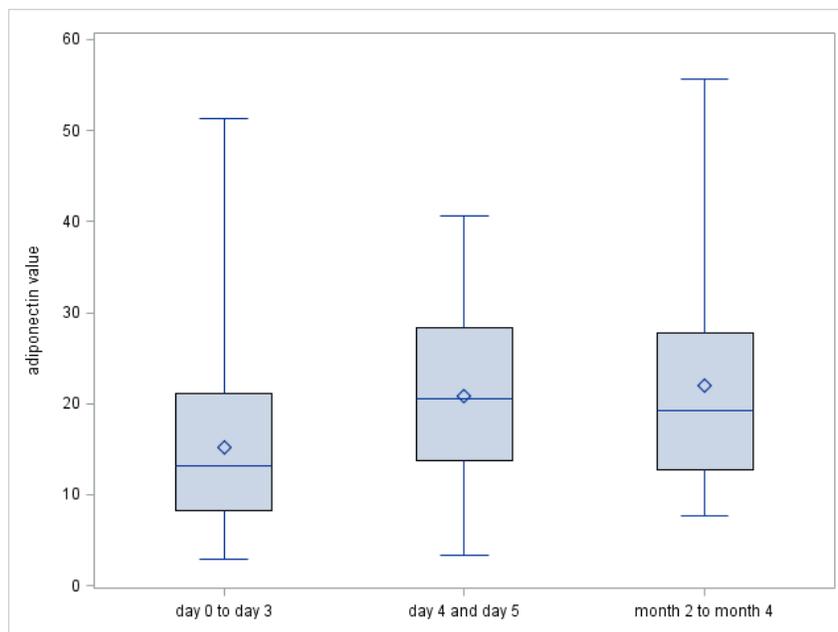


Figure 1. Boxplot of adiponectin over time

A mixed model was fitted using the combined measurements (Wolfinger and Chang 1995). The AIC/BIC criteria was used to determine the form of the covariance matrix (Table 4-3). The AR(1) has the smallest BIC and unstructured has the smallest AIC. AR(1) was used for fitting the mixed model.

The estimated baseline adiponectin level is 15.6, and 20.1 for day 4-5 combined and 21.9 for month 3 based on our model (Table 4-4). After Bonferroni adjustment, day 4-5 is statistically significantly higher than baseline, and month 3 is significantly higher than baseline (Table 4-5).

Table 4-3 AIC/BIC for adiponectin

ADIPO/COV STRUCTURE	AIC	BIC
UNSTRUCTURED	1495.4	1514.2
COMPOUND SYMMETRY	1507.2	1513.5
AR(1)	1500.2	1506.5
Huynh-Feldt	1503.1	1515.7

Table 4-4 Mixed model least squares means

Least Squares Means						
Effect	windows	Estimate	Standard Error	DF	t Value	Pr > t
windows	baseline	15.58	0.89	50	17.59	<.0001
windows	day 4 and 5	20.07	1.19	50	16.85	<.0001
windows	month 3	21.86	1.34	50	16.35	<.0001

Table 4-5 Contrasts to baseline

Differences of Least Squares Means									
Effect	windows	_windows	Estimate	Standard Error	DF	t Value	Pr > t 	Adjustment	Adj P
windows	day 4 5	baseline	4.48	1.13	50	3.98	0.0002	Bonferroni	0.0004
windows	month 3	baseline	6.27	1.44	50	4.35	<.0001	Bonferroni	0.0001

4.1.5 Results for leptin

Table 4-6 shows the summary for leptin. The pattern of available data is similar to that of adiponectin, most observations were recorded on day 0, day 1, day 3, day 5, month 2 and month 3. No significant increase of leptin was found within the first 5 days after insulin therapy.

For leptin, none of the paired difference followed a normal distribution. T-test analyses were not appropriate. Instead Wilcoxon Signed Rank test was used (Table 4-7). Leptin levels increase during the first 3 days then decrease from day 3 to day 5. But none of the paired results reached statistical significance. Thus, results from days 0 to 5 were collapsed together (if multiple measures, we averaged) to derive the baseline measurement for leptin. A significant increase in leptin ($p=0.0016$; $n=23$ Wilcoxon Signed Rank test) was noted from baseline to 3 month visit (Figure 2).

Table 4-6 Summary for leptin

Variable	N	Mean	Std Dev	Median	Minimum	Maximum	25th Pctl	75th Pctl
leptin day 0	54	5.78	12.0	1.20	0.29	66.5	0.72	2.45
leptin day 1	40	4.17	7.94	1.21	0.50	30.2	0.50	2.43
leptin day 2	3	0.47	0.05	0.50	0.42	0.50	0.42	0.50
leptin day3	40	3.67	6.15	1.23	0.50	29.2	0.55	3.27
leptin day 4	9	4.63	9.56	1.44	0.50	30.0	0.73	2.43
leptin day 5	39	4.25	7.10	1.71	0.44	39.5	0.79	5.10
leptin month 1	17	10.9	16.2	3.45	0.50	65.6	2.02	13.3
leptin month 2	19	14.7	13.5	12.2	0.54	50.4	2.11	24.9
leptin month 3	20	11.7	15.2	3.87	1.01	61.7	1.53	16.8
leptin month 4	9	7.62	9.77	3.50	0.50	30.0	2.03	7.92
leptin month 5	7	7.82	7.81	4.39	0.33	22.3	1.82	12.2

Table 4-7 Wilcoxon Signed Rank test results for leptin

Leptin	Day 0	Day 1	Day 3
Day 1	10(0.10)	--	--
Day 3	14(0.80)	1(.)	--
Day 5	20(0.9)	4(0.5)	13(1.00)

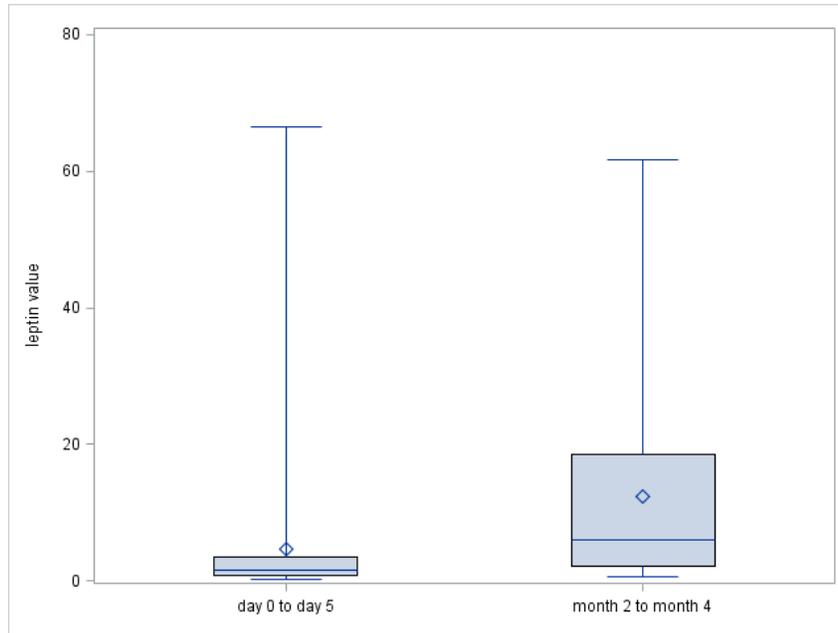


Figure 2. Boxplot of leptin over time

The AIC/BIC are the same for unstructured and Huynh-Feldt (Table 4-8). Mixed model was fitted using unstructured covariance matrix. The estimated baseline leptin level is 4.6 and 12.7 for 3 month visit (Table 4-9). And month 3 leptin level is statistically significantly higher than baseline (Table 4-10).

Table 4-8 AIC/BIC for leptin

LEPTIN/COV STRUCTURE	AIC	BIC
UNSTRUCTURED	1316.5	1325.9
COMPOUND SYMMETRY	1329.1	1335.4
AR(1)	1329.1	1335.4
Huynh-Feldt	1316.5	1325.9

Table 4-9 Mixed model least square means

Least Squares Means						
Effect	windows	Estimate	Standard Error	DF	t Value	Pr > t
windows	baseline	4.62	0.78	153	5.93	<.0001
windows	month 3	12.68	1.99	153	6.37	<.0001

Table 4-10 Contrasts to baseline

Differences of Least Squares Means							
Effect	windows	_windows	Estimate	Standard Error	DF	t Value	Pr > t
windows	month 3	baseline	8.06	2.06	153	3.92	0.0001

5.0 DISCUSSION AND CONCLUSIONS

In the JOD cohort we were able to identify 44 slow progressors who had more than ten year elapse from the time they developed two or more positivity antibodies. An additional 85 potential slow progresses were identified and will be contacted in hopes that they have not developed insulin dependent diabetes. Data from these individuals will be contrasted to those who developed disease sooner than ten years. This may provide information as to what protects some individuals with positive antibodies from rapidly developing insulin depend diabetes.

We examined adiponectin and leptin levels both pre and post insulin therapy and found that by the month 3 visit levels have increased from baseline. Adiponectin significantly increased from baseline as soon as 5 day after the start of insulin therapy and continued to increase through the 3-month follow-up. In constrast, leptin did not increase significantly for the first 5 days of post insulin therapy. However, by the 3-month follow-up leptin increased significantly from baseline.

APPENDIX

SAS PROGRAM

```

/*****
/*
/*
/*      adiponectin_leptin_2013_11_16_vca.sas      */
/*      Wenxiu_Dong SAS 9.3(32)                    */
/*
/*
/*****

libname diab "f:\diabetes project";
filename jodf "f:\diabetes project\jod_formats_2014_01_02.sas";

filename alsum "f:\diabetes project\thesis table.doc";

%include jodf;** use external .sas file to import formats;

/*****
/*      DATA CLEANING      */
/*****

data md;
  set diab.hala_data_2013_10_25;

  * change the format of dob, dtinsstart, ddx from datetime19. to date9.;
  dob = datepart(dob) ;
  attrib dob format = date9.;

  dtinsstart = datepart(dtinsstart) ;
  attrib dtinsstart format = date9.;

  ddx = datepart(ddx) ;
  attrib ddx format = date9.;

  /* calculate bmi */
  bmi = wtdx / (htdx/100)**2;

  /* calculating number of tested antibodies among gad,ia2,ica_human,iaa
```

```

        and number of positive among those tested */
tested4 = n(gad_positive, ia2_positive, ica_human_positive, iaa_positive);
pos4 = sum(gad_positive, ia2_positive, ica_human_positive, iaa_positive);
if pos4 gt 2 then pos4 = 3;

/* calculate variable only_icar_positive, for people who have >=3
antibodies tested*/
/* if all were negative, then only_icar_positive = 1 if ica_rat_positive
=1*/
/* if at least one was positive, then only_icar_positive =0*/
if tested4 ge 3 & pos4 = 0 & ica_rat_positive = 1 then only_icar_positive
= 1;
if tested4 ge 3 & pos4 > 0 then only_icar_positive = 0;
if tested4 ge 3 & pos4 =0 & ica_rat_positive=0 then only_icar_positive =
0;

if calc_dq2 = 0 and calc_dq8 = 0 then dq2_dq8 = 0 ;
if calc_dq2 = 1 and calc_dq8 = 0 then dq2_dq8 = 1 ;
if calc_dq2 = 2 and calc_dq8 = 0 then dq2_dq8 = 2 ;
if calc_dq2 = 0 and calc_dq8 = 1 then dq2_dq8 = 3 ;
if calc_dq2 = 0 and calc_dq8 = 2 then dq2_dq8 = 4 ;
if calc_dq2 = 1 and calc_dq8 = 1 then dq2_dq8 = 5 ;

label tmp_id = " ID in Excel file"
pid = " Participant ID"
gender = "Gender"
race = "Race"
medrecno = "Medical record number"
an = "Acanthosis Nigricans"
ketonuria = "Ketonuria"
tanner_group = "Tanner Group"
dob = "Birth date"
agedx = "Age at dx"
bmi = "calculated BMI"
AlCDX = "Alc Dx"
AlCM3 = "Alc month 3"
calc_dq2 = "Number of DQ2 alleles"
calc_dq8 = "Number of DQ8 alleles"
calc_nonasp = "ASP haplotype"
gad_positive = "positive GAD"
ia2_positive = "positive IA2"
iaa_positive = "positive IAA"
ica_human_positive = "positive ICA human"
ica_rat_positive = "Positive ICA rat"
only_icar_positive = "Only ICA rat positive"
nonASP_status = "ASP status"
dq2_dq8 = "DQ2/DQ8 haplotype"
dq2_dq8_b = "DQ2/DQ8 consolidated"
tested4 = "number of antibodies tested among 4"
pos4 = "number of positive antibodies among the 4 tested"
;
format race lrace.
gender lgender.

calc_nonasp lcalc_nonasp.
nonasp_status lasp.

```

```

        an lyesno.
        dka lyesno.
        fam lyesno.
        ketonuria lyesno.

        dq2_dq8 ldq2_dq8_a.
        dq2_dq8 ldq2_dq8_b.

        gad_positive l_neg_pos.
        ia2_positive l_neg_pos.
        iaa_positive l_neg_pos.
        ica_human_positive l_neg_pos.
        ica_rat_positive l_neg_pos.
        only_icar_positive l_neg_pos.

;

run;

* read in dataset containing adiponectin and leptin levels, could have
multiple records per person;
data adle;
    set diab.adipo_leptin_2014_01_21;

    /* create adipo and leptin ratio */
    al_ratio = adiponectin / leptin;
    label al_ratio = "ratio of adiponectin to leptin";
run;

*sorting data by patient id for merging purpose;
proc sort data = md;by pid;run;
proc sort data = adle;by pid;run;

/*createting format name intl for variable windows*/
proc format;
    value intl
        0 = "day 0"
        1 = "day 1"
        2 = "day 2"
        3 = "day 3"
        4 = "day 4"
        5 = "day 5"
        6 = "1 month"
        7 = "2 month"
        8 = "3 month"
        9 = "4 month"
        10 = "5 month"
;
run;

* merging the two datasets by patient id,md_ap is long form data, one record
for each blooddraw, multiple records for one person;
data md_ap;
    merge md adle;

```

```

by pid;

* calculating elapsed time since diagnosis
  elapsed time = blooddrawdate - date of diagnosis;
etime = blooddrawdate - ddx;
label etime = "elapsed time since diagnosis";

* create new variable windows to have the following values;

  if -3 lt etime le 0 then windows = 0; * windows is 0 when blood draw
happened same time or before diagnosis;
  if etime = 1 then windows = 1; * windows is 1 when blood draw
happened one day after diagnosis;
  if etime = 2 then windows = 2;
  if etime = 3 then windows = 3;
  if etime = 4 then windows = 4;
  if 5 le etime lt 15 then windows = 5; *Oct,3,2013. include blood draw at
day6 - day14 into day5 category;

  if 15 le etime lt 45 then windows = 6; * windows is 6 when blood draw
happened during the first month;
  if 45 le etime lt 75 then windows = 7; * blood draw during month 2;
  if 75 le etime lt 105 then windows = 8; * blood draw during month 3;
  if 105 le etime lt 135 then windows = 9; * blood draw during month 4;
  if etime ge 135 then windows = 10; * blood draw during month 5;
attrib windows format = int1.;
run;

/*****
      reshape data from long form to wide form
*****/

* 1-step sort data by pid;
proc sort data = md_ap out = md_ap_sort;by pid;run;

* 2-step;
data adle_wide;
  set md_ap_sort;
  if windows = . then delete;
  if adiponectin < 0 then adiponectin = .;
  if leptin < 0 then leptin = .;

  by pid;

  keep pid adiponectin0 - adiponectin10 leptin0 - leptin10 al_ratio0 -
al_ratio10;
  retain adiponectin0 - adiponectin10 leptin0 - leptin10 al_ratio0 -
al_ratio10;

  array adipo_a(0:10) adiponectin0 - adiponectin10;
  array leptin_a(0:10) leptin0 - leptin10;
  array ratio_a(0:10) al_ratio0 - al_ratio10;

  if first.pid then
do;
  do i = 0 to 10;

```

```

        adipo_a (i) =. ;
        leptin_a (i) =. ;
        ratio_a (i) = .;
    end;
end;

adipo_a(windows) = adiponectin;
leptin_a(windows) = leptin;
ratio_a(windows) = al_ratio;

if last.pid then output;
run;

proc sort data = adle_wide;by pid;run;

/*****
reshape data from long form to wide form_end
*****/

data wide;
    set adle_wide;
    by pid;

    /* create adiponectin0123 to be the mean of day0-day3
       create adiponectin45 to be the mean level for day4 and day5
       create adiponectin_m to be the mean level for month1 - month5 */
    adiponectin_0123 =
mean(adiponectin0,adiponectin1,adiponectin2,adiponectin3);
    adiponectin_45 = mean(adiponectin4, adiponectin5);
    adiponectin_m =
mean(adiponectin6,adiponectin7,adiponectin8,adiponectin9,adiponectin10);
    adiponectin_m234 = mean(adiponectin7,adiponectin8,adiponectin9);

    n_adiponectin =
n(adiponectin0,adiponectin1,adiponectin2,adiponectin3,adiponectin4,adiponecti
n5,
adiponectin6,adiponectin7,adiponectin8,adiponectin9,adiponectin10);

    /* create day0123 to be the mean of day0-day3
       create day45 to be the mean level for day4 and day5 */
    leptin0123 = mean(leptin0,leptin1,leptin2,leptin3);
    leptin45 = mean(leptin4, leptin5);
    leptin_d = mean(leptin0,leptin1,leptin2,leptin3,leptin4,leptin5);

    leptin_m = mean( leptin6,leptin7,leptin8,leptin9,leptin10);
    leptin_m234 = mean(leptin7,leptin8,leptin9);

    n_leptin = n(leptin0,leptin1,leptin2,leptin3,leptin4,leptin5,
leptin6,leptin7,leptin8,leptin9,leptin10);

```

```

ratio_d =
mean(al_ratio0,al_ratio1,al_ratio2,al_ratio3,al_ratio4,al_ratio5);
ratio_m = mean(al_ratio6,al_ratio7,al_ratio8,al_ratio9,al_ratio10);
ratio_m234 = mean(al_ratio7,al_ratio8,al_ratio9);

label adiponectin0123 = "avg adipo from day0 to day3"
adiponectin45 = "avg adipo from day4 to day5"
adiponectin_m = "avg adipo for all 5 months"
adiponectin_m234 = "avg adipo for month 234"
n_adiponectin = "number of adiponectin measures"

leptin0123 = "avg leptin from day0 to day3"
leptin45 = "avg leptin from day4 to day5"
leptin_d = "avg leptin for all the first five days"
leptin_m = "avg leptin for all 5 months"
leptin_m234 = "avg leptin for month 234"
n_leptin = "number of leptin measures"
;
run;

proc sort data = wide;by pid;run;

data md_wd;
merge md wide;
by pid;
run;

data wd_sub;
set md_wd;
if (n_adiponectin + n_leptin)=. or n(htdx,wtdx,dob,race,gender)<5 then
delete;
run;

data group;
set md_ap;
keep pid gender race ovrwtdx tanner_groups nonasp_status tested4 pos4
windows;
run;

proc sort data = group;by pid;run;

data sp_plot;
merge adle group;
by pid;
run;

*Generate wide format dataset wide_sub;
data diab.wide_sub;
set wd_sub;
run;

ods graphics on;

```

```

/*title "paired ttest result for every month and day combination "*/
proc ttest data = adle_wide;
    paired (adiponectin0-adiponectin9)*(adiponectin1-adiponectin10);
run;

/*title "paired ttest result for every combination of leptin"*/
proc ttest data = adle_wide;
    paired (leptin0-leptin9)*(leptin1-leptin10);
run;

/*title "checking normality of adiponectin"*/
proc univariate data = md_wd normal;
    class windows;
    var adiponectin;
    where adiponectin >0;
run;
/** windows = 0, 1, 3, 5, 6, 7, 8, 10 violates normality assumption;*/
/** generate qq-plot of the difference. graphic for 10 and above pairs;*/

ods html close;
options nodate nonumber orientation = portrait ;
ods noproctitle;
ods rtf file = "d:\academic-related\diabetes project\summary statistics of
medical record.doc" bodytitle;
ods trace on;

/*title "positive disconcordance"*/
proc print data = md_wd;
    where icaronly = 1 or only_icar_positive = 1;
    var pid icaronly only_icar_positive tested4 pos4 ica_rat_positive;
run;

/*title "negative disconcordance"*/
proc print data = md_wd;
    where icaronly = 0 & only_icar_positive = .;
    var pid icaronly only_icar_positive tested4 pos4 ica_rat_positive;
run;

/*title "two patients ASP status are non-ASP"*/
/*title2 "ASP haplotype are neither non-ASP/non-ASP nor non-ASP/ASP"*/
proc print data = md_wd;
    where nonasp_status = 1 & calc_nonasp <0;
    var pid calc_nonasp nonasp_status;
run;

/*title "leptin outliers"*/
proc print data = md_ap;
    var pid gender race leptin tested4 pos4;
    where leptin >40 & tested4 >2;
run;

proc sort data = sp_plot out = sp_gender;by windows;run;

/*title "patients response"*/
proc sgpanel data = sp_gender;
    where adiponectin >0 & tested4>2;

```

```
panelby pos4/rows=1;
series y= adiponectin x=windows/group=pid;
loess y = adiponectin x=windows/nomarkers lineattrs=(color=black thickness
= 2);
run;
proc sgpanel data = sp_gender;
where leptin >0 & tested4>2;
panelby pos4/rows=1;
series y= leptin x=windows/group=pid;
loess y = leptin x=windows/nomarkers lineattrs=(color=black thickness =
2);
run;
```

```
/*SET UP DATA FOR MIXED MODEL FOR ADIPONECTIN AND LEPTIN IN ONE DATA STEP*/
```

```
data amixed(keep = pid windows adipo)
  lmixed(keep = pid windows leptin);

set tmp;

windows = 1; adipo = adiponectin_0123; output amixed;
windows = 2; adipo = adiponectin_45; output amixed;
windows = 3; adipo = adiponectin_m234; output amixed;

windows = 1; leptin = leptin_d; output lmixed;
windows = 3; leptin = leptin_m234; output lmixed;

format windows rm.;

run;
```

```
proc format;
  value rm
    1 = 'baseline'
    2 = 'day 4 and 5'
    3 = 'month 3'
  ;
run;
```

```
ods trace on;
```

```
%macro mixed(var);
```

```
ods select
  fitstatistics
  ;
proc mixed data = amixed;
  title "AIC/BIC for adipo using &var";
  class pid windows;
  model adipo = windows;
  repeated windows/ type = &var sub = pid;
```

```
run;
```

```
ods select
  fitstatistics
  ;
proc mixed data = lmixed;
  title "AIC/BIC for leptin using &var";
  class pid windows;
  model leptin = windows;
  repeated windows/type = &var sub = pid;
```

```
run;
```

```
%mend;
```

```
/*DETERMINE COVARIANCE STRUCTURE*/
```

```

%mixed(un);
%mixed(ar(1));
%mixed(cs);
%mixed(hf);

ods trace off;

ods rtf file = alsum;

/*FIT MIXED MODEL FOR ADIPONECTIN USING AR(1)*/
proc mixed data = amixed;
  title "mixed model for adipo using ar(1)";
  class pid windows;
  model adipo = windows;
  repeated windows/ type = ar(1) sub = pid;

  lsmeans windows/ diff = control('baseline') adjust = bon;

run;

proc mixed data = lmixed;
  title "mixed model for leptin using unstructured";
  class pid windows;
  model leptin = windows;
  repeated windows/type = un sub = pid;

  lsmeans windows/ diff = control('baseline');

run;

ods rtf close;

```

BIBLIOGRAPHY

Libman, I. M., M. Pietropaolo, S. A. Arslanian, R. E. LaPorte and D. J. Becker (2003). "Changing prevalence of overweight children and adolescents at onset of insulin-treated diabetes." Diabetes care **26**(10): 2871-2875.

Libman, I. M., M. Pietropaolo, S. A. Arslanian, R. E. LaPorte and D. J. Becker (2003). "Evidence for heterogeneous pathogenesis of insulin-treated diabetes in black and white children." Diabetes care **26**(10): 2876-2882.

Libman, I. M., M. Pietropaolo, M. Trucco, J. S. Dorman, R. E. LaPorte and D. Becker (1998). "Islet cell autoimmunity in white and black children and adolescents with IDDM." Diabetes Care **21**(11): 1824-1827.

Pietropaolo, M., S. Yu, I. M. Libman, S. L. Pietropaolo, K. Riley, R. E. LaPorte, A. L. Drash, S. Mazumdar, M. Trucco and D. J. Becker (2005). "Cytoplasmic islet cell antibodies remain valuable in defining risk of progression to type 1 diabetes in subjects with other islet autoantibodies." Pediatric diabetes **6**(4): 184-192.

Wolfinger, R. and M. Chang (1995). Comparing the SAS GLM and MIXED procedures for repeated measurements analysis. SAS User's Group International - SUGI SAS Institute Inc.