

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Jason Baumgardner

April 23, 2012

Date

Exploring risk factors for reduced physical activity and functional abilities in adults with hemophilia: a cross-sectional survey from a single large hemophilia treatment center

By

Jason Baumgardner
Master of Science in Public Health

Department of Biostatistics

Lisa Elon, MS, MPH

Committee Chair

Dr. Christine Kempton, MD, MSc

Committee Member

Exploring risk factors for reduced physical activity and functional abilities in adults with hemophilia: a cross-sectional survey from a single large hemophilia treatment center

By

Jason Baumgardner

Bachelor of Science
University of Tennessee
2006

Bachelor of Science
University of Tennessee
2001

Thesis Committee Chair: Lisa Elon, MS, MPH.

Thesis Committee Member: Dr. Christine Kempton, MD, MSc

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in the Department of Biostatistics
2012

Abstract

Exploring risk factors for reduced physical activity and functional abilities in adults with hemophilia: a cross-sectional survey from a single large hemophilia treatment center

By Jason Baumgardner

Physical activity and functional abilities are important to maintaining quality of life in the aging hemophilia population. The risk factors leading to reduced physical activity and functional limitations in people with hemophilia are under-explored. Physical activity and functioning were evaluated in persons with hemophilia at a single large hemophilia treatment center to determine the risk factors for reduced physical activity and functional limitations, and specifically to examine the relationship between joint disease (arthropathy) and reduced physical activity/functional limitations. Data were obtained on 88 patients via the medical record, patient questionnaires and clinical examination. The World Federation of Haemophilia (WFH) orthopedic joint score was used as a measure of arthropathy. The Framingham Physical Activities Index (PAI), the Hemophilia Activities List (HAL), and the Timed Up-and-Go Test (TUG) were used as outcome measures. We used multiple linear regression to assess the relationship between our dependent variables and the covariates. The HAL summary score decreased 5.8 points per 10 point increase in WFH score ($P=.04$), decreased 12.0 points in patients with chronic liver disease ($P=.006$), and decreased 2.5 points for each target joint ($P=.02$). PAI score decreased by 5.5% for every ten years of age ($P<.0001$) and TUG average time increased 5% for every ten years of age ($P=.01$). The present study provides some evidence for the relationship between arthropathy and reduced physical functioning. It appears that improving joint health and preventing chronic liver disease could improve functional independence among adults with hemophilia. However, reducing the risk for falls and increasing overall physical activity will require further understanding of factors that influence these important measures of physical functioning and physical activity.

Exploring risk factors for reduced physical activity and functional abilities in adults with hemophilia: a cross-sectional survey from a single large hemophilia treatment center

By

Jason Baumgardner

Bachelor of Science
University of Tennessee
2006

Bachelor of Science
University of Tennessee
2001

Thesis Committee Chair: Lisa Elon, MS, MPH.

Thesis Committee Member: Dr. Christine Kempton, MD, MSc

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in the Department of Biostatistics
2012

Acknowledgements

Above all, I wish to express my deepest gratitude to Lisa Elon, my indefatigable advisor, for her constant accessibility and patience throughout what must have been an exasperating process in helping me with this report.

I am also greatly indebted to Dr. Christine Kempton for her steady cooperation and expert guidance.

Dr. Patrick Kilgo was kind enough to read a draft of this report and offer criticism, for which I am appreciative.

I also wish to acknowledge Drs. Michael Kutner and Mary Kelley for their advice on statistical problems within their respective areas of expertise.

Finally I am grateful to my classmates for their assistance.

Dedicated to all those living with hemophilia.

Contents

Introduction	1
Methods	2
Subjects.....	2
Independent Variables	2
Dependent Variables.....	3
Data Analysis.....	4
Results	5
Patient Characteristics.....	5
Linear Regression Analysis	5
PAI.....	5
HAL	6
TUG	6
Discussion	6
Summary of Models.....	6
Relationship to Literature	7
Clinical Significance.....	7
Limitations of the Study	8
Strengths of the Study.....	8
Conclusion	9
Summary.....	9
Future Studies	9
References	9
Tables	11
Appendix A – Additional Figures	15
Appendix B – SAS Code	21

Introduction

Joint disease and other clinical complications in patients with hemophilia often lead to reduced physical activity and functional independence, and subsequently to the impairment of the overall quality of life. However, specific research studies investigating the risk factors for reduced physical activity and functioning in adult hemophilia patients are uncommon. Fearn (2010) found moderate impairment of balance, mobility, physical activity and related measures in adult males with hemophilia as compared to healthy controls, with balance and mobility assessed using the Timed Up-and-Go Test (TUG), and physical activity level measured using the Human Activity Profile (HAP) [1]. Using a self-report questionnaire, Sherlock (2010) found lower levels of physical activity among Irish people with hemophilia, as compared to the European Union average, especially in older patients with severe hemophilia [2]. Hassan (2011) demonstrated a highly significant correlation between radiological joint score and the functional independence score in haemophilia (FISH) in adolescent hemophilia patients [3]. Monahan (2011) showed a significant relationship between self-reported physical limitations and multiple characteristics of underlying hemophilia severity including age, frequency of joint bleeds, baseline factor VIII/IX level, and BMI in adolescent boys [4]. Nonwhite race was also associated with each of the poorer functional outcomes in their study. Biere-Rafi (2011) found, in univariate analysis, that obese hemophiliacs had a significantly lower sum score on the Hemophilia Activities Index (HAL), mainly due to impaired lower limb function [5].

Among the many possibilities for measuring physical activity and physical functioning, we examined three separate outcomes – two self-assessment instruments and a performance test. The Framingham Physical Activities Index (PAI) is derived from the physical activities questionnaire developed for the Framingham Study, and provides a measure of the level of daily physical activity [6]. The PAI is considered a reliable measure of metabolic activity (METs), and low scores have been associated with increased risk for cardiovascular disease. The other self-report measure is the Haemophilia Activities List (HAL), which is a self-rated functional ability score for use in adults with

hemophilia. It reflects the subject's ability to complete many activities required for daily living. The Timed Up-and-Go Test is a basic test of functional mobility that has been validated as a general measure of performance in adults, and is predictive of the risk of falling [7].

Once the risk factors for the development of reduced physical activity and functional abilities are understood, guidelines for management of hemophilia can be improved. The objective of the present study was to evaluate demographic and clinical measures as risk factors for reduced physical activity and functional abilities in persons with hemophilia. We anticipated joint disease (as measured by joint score) to be the major predictor of physical activity and physical functioning in patients with hemophilia.

Methods

Subjects: After approval by the Emory University Institutional Review Board, all persons being treated at the Emory Comprehensive Hemophilia Treatment Center who were males with hemophilia A or B (factor VIII or IX activity \leq 40%) and age \geq 25 years were invited to participate via a mailed invitation. Subjects who had not responded to this invitation were approached during their regularly scheduled visit. Informed consent was obtained prior to initiation of any study-related procedure. The first patient was enrolled 4/22/2010 and the last was enrolled 9/1/2011.

Independent Variables: Historical information was obtained from review of the medical record and from the patient using self-reported questionnaires. The following information was included: type and severity of hemophilia, age, race, ethnicity, historical baseline level of factor VIII / IX, number and location of target joints (target joint definition: elbows, shoulders, hips, knees, and ankles having *four bleeds during the previous 6 months or \geq 20 bleeds in lifetime*), history of comorbid conditions such as HIV and hepatitis C diagnosis, chronic liver disease status (definition: current positive hepatitis C viral load or evidence of fibrosis on prior biopsy even if hepatitis C cleared), prior bone fractures, and prior joint surgeries.

Physical examination included measurement of height and weight to calculate BMI and physical assessment of bilateral ankles, knees, and elbows using the orthopedic joint score as recommended by the Orthopedic Advisory Committee of the World Federation of Haemophilia (WFH) [8]. The examination was performed by one of two physical therapists who were familiar with joint measurement and examination. The WFH joint score evaluates joints for evidence of swelling, muscle atrophy, axial contracture, and instability. There are a maximum of 12 points each for the knee and ankle and 10 for the elbow. Hence the maximum score for all six joints is 68 points, with a higher score indicating more severe arthropathy.

Dependent Variables. Physical activity was measured using a self-assessment instrument (PAI). Functional ability was measured using both a self-assessment instrument (HAL) and a performance test (TUG). All three are continuous outcomes. For PAI, the interviewer asks the individual about the average hours of participation in sleep, rest, occupational, and extracurricular activities over a typical 24-hour period, along with the intensity of the activity according to the categories *sleep, sedentary, slight, moderate, and heavy* activity. The PAI composite score is then calculated by summing the number of hours spent in each activity intensity level and multiplying by a respective weight factor derived from the estimated oxygen consumption requirement for each intensity level. Higher scores indicate more activity [6].

For the HAL, an overall score as well as three component scores can be calculated (activities involving the upper extremities, basic activities involving the lower extremities, and complex activities involving the lower extremities). There are seven domains: lying/sitting/kneeling/standing, functions of legs, functions of arms, use of transportation, self-care, household tasks, and leisure activities and sport, with 42 questions overall. Normalized scores for the seven domains, components, and the full questionnaire are obtained, with possible scores ranging from 0 to 100, where 0 represented the worst possible functional status and 100 the best possible functional status [7].

The TUG is performed by having the subject sit in a chair with arms with their back resting on the back of the chair. The subject gets up from the chair, walks three meters away from the chair at a

regular pace, turns around, walks back to the chair and sits down. The time it takes to perform this task is measured. Less than 10 seconds is normal, whereas taking 14 or more seconds has been associated with increased risk for falls [9,10]. The subject was given one untimed practice trial before testing, and the average of the three trials was recorded.

Data Analysis: Analysis proceeded from descriptive statistics of all measures, through bivariate relationships between physical activity/functional measures and each covariate, to the multivariate relationship of physical activity/functional measures with the covariates. The (unadjusted) estimates of the relationship between the continuous measures of physical activity and our covariates were calculated with simple linear regression. To evaluate the independent predictive value of the multiple variables, we then performed a multiple linear regression model with all covariates. A backward elimination strategy was then employed to select a final model. We forced WFH score into each model, even if not significant or a confounder, since we were especially interested in examining the relationship between arthropathy and physical activity/functioning. Variables that were found to be neither confounders nor significant predictors themselves were dropped from the final model. Regression diagnostic techniques to detect issues with residuals, non-linearity, influential observations, and multicollinearity were applied to the full (non-reduced) model and remediation performed as needed. TUG and PAI summary scores were log transformed (using the natural logarithm) in order for the residuals to achieve a Gaussian normal distribution. Diagnostics were again performed on each of the three final models. All analyses were conducted using SAS v9.3 statistical software (SAS Institute Inc., Cary, NC, USA). The criterion set for statistical significance was $\alpha=.05$.

In 20% (n=18) of the 88 patients the total WFH score is not available because of one or more individual joint scores being missing. Most of these patients had undergone a replacement or other major surgical intervention on the joint, and consequently the joint could not be scored. There were other reasons – for example, one patient had had a bleed in the ankle that day, and so the clinical score could not be measured. Using a multiple imputation algorithm, missing values were imputed so that the whole sample could be analyzed. Twenty datasets were imputed and estimates from these datasets were

combined using methods described by McKnight et al. [11]. Values that were missing were assumed to be missing at random.

Results

Patient Characteristics (Table 1). The 88 patients in the study population were all male, ranging from 25 to 85 years of age (median = 41 years). Of the 34% (n=30) who were non-white, 27 were Black, two were Asian and one was American Indian/Alaska Native. One half (n=44) of the patients reported having a prior joint procedure, and 47% (n=41) reported a prior bone fracture. Forty-three percent (n=38) reported having a joint with ≥ 4 bleeds in the last 6 months, and 64% (n=56) reported having a joint with ≥ 20 bleeds in his lifetime – 70% percent (n=62) had at least one such target joint. Thirty percent (n=26) of the patients were HIV positive. Eighty-one percent (n=71) of the patients had hemophilia A and 19% (n=17) had hemophilia B. Fifty-three percent (n=47) had chronic liver disease. WFH score ranged from 1 to 43 (median =12). PAI scores ranged from 25 to 65.9 (median = 31). Fourteen percent (n=12) scored above 38, deemed “active,” and one quarter scored under 28, deemed “sedentary.” HAL scores ranged from 23 to 100 (median = 77). Twenty-four percent (n=21) of the scores were 95 or above, and there were 14 subjects with the maximum score of 100. TUG average time ranged from 5.8 to 22.3 seconds (median = 9.3 seconds). Eight percent (n=7) of patients scored above 14 seconds, deemed at risk for falls.

Linear Regression Analysis.

PAI (Table 2). Only age was significantly associated with PAI ($P<.0001$), with PAI score decreasing by 5.5% for every 10 year increase in age. Although the simple, unadjusted relationship between WFH score and PAI was significant, after controlling for age and confounding variables (target joints, non-white race, and baseline Factor VIII/IX level), WFH score was *not* associated with PAI ($P=.31$). Chronic liver disease and HIV status were also significant according to the simple linear regression, but were not related to PAI after controlling for other variables. Additionally, history of fracture, history of joint procedure, and BMI were neither related to PAI nor were confounding variables.

HAL (Table 3). Together, the covariates WFH score (partial $R^2 = .06$), target joints, and chronic liver disease explained 38% of the observed variation in HAL score. The HAL summary score decreased 5.8 points per 10 point increase in WFH score ($P=.04$), decreased 12.0 points in patients with chronic liver disease ($P=.006$), and decreased 2.5 points for each target joint ($P=.02$). History of joint procedure, baseline Factor VIII/IX level, and HIV status were all significant according to the simple linear regression, but none was significant after controlling for the other covariates. Additionally, age, history of fracture, non-white race, and BMI were neither related to HAL nor were confounding variables.

TUG (Table 4). Only age was significantly associated with TUG ($P=.01$), with TUG average time increasing by 5% for every 10 year increase in age. Although the simple, unadjusted relationship between WFH score and TUG was significant, after controlling for age and confounding variables (history of joint procedure, HIV status, and chronic liver disease), WFH score was *not* associated with TUG ($P=.22$). History of fracture, history of joint procedure, non-white race, baseline Factor VIII/IX level, and BMI were neither related to PAI nor were confounding variables.

Two-way interactions between Age, BMI, and WFH score were considered initially for each outcome, but none were statistically significant.

Discussion

Summary of Models. Higher WFH score was associated with reduced functional ability as measured by the hemophilia-specific questionnaire (HAL). This finding is consistent with clinical observations that arthropathy is associated with reduced physical functioning in hemophilia patients. However, WFH score was not significantly associated with the more general physical activity measurement (PAI), or the objective test of physical functioning (TUG). Similarly, chronic liver disease history and number of target joints were both associated with HAL score, but neither had a significant association with PAI or with TUG. Conversely, the effect of age was significant for PAI and TUG, but not for HAL. While HAL and TUG are assessments of functional ability, PAI is a measure of metabolic activity and has an

element of desire/motivation. Further, TUG is primarily an assessment of lower limb movement, and as such, may be primarily related to lower joints. Non-white race, BMI, prior joint procedure, prior fracture, baseline Factor VIII/IX level, and HIV status were not significantly associated with any of the three outcome measures.

Relationship to Literature. The results of our study confirm some previously identified associations between physical activity and both demographic and clinical risk factors in hemophilia patients but do not corroborate all previously reported associations. We found joint disease, increasing age, and number of target joints (e.g. joint bleeds) to have statistically significant associations with physical activity/functioning by at least one of our outcome measures, which was congruent with previous studies [3,4].

We failed to demonstrate an association between BMI and physical activity, despite the fact that previous studies have shown obesity to be predictive of limitations of joint range of motion and physical functioning in hemophilia patients, that people with hemophilia appear to be more sensitive to overweight-associated arthropathy, and that obese patients had a significantly greater frequency of bleeds [5,12]. We also failed to show a relationship between non-white race and physical activity, which is incongruent with the findings of Monahan et al. [4].

Clinical Significance. As the life expectancy of people with hemophilia approaches that of the general population [13], maintaining physical activity and physical functioning becomes essential to a healthy life. Reduced physical activity threatens physical health by increasing the risk of obesity, heart disease, and reduced bone strength. Furthermore, reduced physical functioning of specific joints threatens one's ability to care for oneself and engage with family and community. It is important to identify those hemophilia patients with impaired functional ability and help them find ways of staying active in order to prevent decline [2].

The reasons for reduced physical activity and physical functioning in patients with hemophilia are multifactorial. Interestingly, the presence of chronic liver disease (partial $R^2 = .09$) and target joints (partial $R^2 = .08$) were each more significantly associated with HAL sum score than was arthropathy

(partial $R^2 = .04$). The TUG was less influenced by the joint score (partial $R^2 = .02$) and more influenced by age (partial $R^2 = .07$). This may reflect the deterioration in muscle strength and balance that occurs with age, irrespective of hemophilia or arthropathy (8% of patients were > 14 seconds and considered at risk for falls). Overall, the cohort is not very active as measured by the PAI, with over one quarter of subjects being sedentary (having a PAI < 28). Few subjects (14%) were considered active (having a PAI > 38). Age is again the main predictor (partial $R^2 = .18$) but the covariates together only explained about 25% of the variation in the PAI, indicating that factors not measured are responsible for a significant amount of the variability – much of the variability in PAI score may have to do with motivation, an expectation of activity, and lifestyle.

Limitations of the Study: This study was retrospective in nature. Because of the cross-sectional study design, causality between arthropathy and physical activity cannot be determined. However, it seems unlikely that lack of physical activity or functional ability is a cause of arthropathy rather than arthropathy being a cause of reduced physical activity/functioning. Two of our independent variables are based on questionnaires which ask subjects to recall information from the past – number and location of target joints and prior bone fracture. Also, two of the outcome variables (PAI and HAL) are self-reported measures, and hence are somewhat subjective. Results from a single center may not reflect the situation in the general hemophilia population. The sample size of this study is not large enough to exclude small effects of some variables that in this study were not statistically significant, such as BMI or non-white race.

Strengths of the Study: This study is exploratory in nature and as such is designed to consider multiple risk factors and confounders simultaneously. The use of multiple outcome measures, including two self-reported and one performance measure, provided a more robust assessment of physical activity and functional ability. Moreover, it has been proposed that self-assessment instruments and performance tests measure different constructs and hence both should be considered when clinicians assess a patient's functional abilities [7]. The main predictor variable, the total WFH joint score, is not dependent on

patient recall. Minorities were appropriately represented in the study cohort as the clinic is located in a racially diverse metropolitan area.

Conclusion

Summary. This study provided some evidence for our hypothesis that arthropathy is independently associated with reduced physical functioning in those with hemophilia (as measured by the Hemophilia Activities Index). However, only age was found to be a significant predictor of TUG or of PAI score. From this study it appears that improving joint health and preventing chronic liver disease could impact the ease with which one is able to do activities of daily living, but that reducing the risk for falls and increasing overall physical activity (in order to reduce cardiovascular risk) will require further understanding of factors that influence these important measures of physical functioning and physical activity.

Future Studies. Given that arthropathy was not associated with physical activity (as measured by the Framingham Physical Activities Index), and that physical activity is essential for cardiovascular health, a better understanding of why an estimated 25% of the hemophiliacs were sedentary and only 14% active would be a logical next step.

References

- 1 Fearn M, Hill K, Williams S, Mudge L, Walsh C, McCarthy P *et al.* Balance dysfunction in adults with haemophilia. *Haemophilia* 2010; **16**: 606-14.
- 2 Sherlock E, O'Donnell JS, White B, Blake C. Physical activity levels and participation in sport in Irish people with haemophilia. *Haemophilia* 2010; **16**: e202-9.

- 3 Hassan TH, Badr MA, El-Gerby KM. Correlation between musculoskeletal function and radiological joint scores in haemophilia A adolescents. *Haemophilia* 2011; **17**: 920-925.
- 4 Monahan PE, Baker JR, Riske B, Soucie JM. Physical functioning in boys with hemophilia in the U.S. *Am J Prev Med* 2011; **41**: S360-8.
- 5 Biere-Rafi S, Haak BW, Peters M, Gerdes VEA, Buller HR, Kamphuisen PW. The impairment in daily life of obese haemophiliacs. *Haemophilia* 2011; **17**: 204-208.
- 6 Hannan MT, Felson DT, Anderson JJ, Naimark A. Habitual physical activity is not associated with knee osteoarthritis: the Framingham Study. *J Rheumatology* 1993; **20**: 704-9.
- 7 Van Genderen FR, Westers P, Heijnen L *et al*. Measuring patients' perceptions on their functional abilities: validation of the Haemophilia Activities List. *Haemophilia* 2006; **12**: 36-46.
- 8 Alendort LM, Haschmeyer RH, Pettersson H. A longitudinal study of orthopedic outcomes for severe factor-VIII-deficient haemophiliacs. The Orthopedic Outcome Study Group. *J Intern Med* 1994; **236**: 391-9.
- 9 Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc* 1991; **39**: 142-8
- 10 Shumway-Cook A, Brauer S, Woollacott M. Predicting the probability for falls in community-dwelling older adults using the timed Up & Go test. *Phys Ther* 2000; **80**: 896-903.
- 11 McKnight PE, McKnight KM, Sidani S, Figueredo AJ. *Missing Data*. New York: Guilford Press, 2007.
- 12 Wong TE, Majumdar S, Adams E, Bergman S, Damiano ML, Deutsche J, Recht M. Overweight and obesity in hemophilia: a systematic review of the literature. *Am J Prev Med* 2011; **41**: S369-75.
- 13 Darby SC, Keeling DM, Spooner RJD, Wan Kan S, Giangrande PLF, Collins PW *et al*. The incidence of factor VIII and factor IX inhibitors in the hemophilia population of the UK and their effect on subsequent mortality, 1977-99. *J Thromb Haemost* 2004; **2**: 1047-1054.

Tables

**Table 1. Subject Characteristics; Emory Treatment Center
4/2010-9/2011 (Total N=88)**

Variable	n	(%)	Median	(Centile 25-Centile 75)
Age in years			43.0	(31.9 - 52.4)
Race				
White	58	(66%)		
Black	27	(31%)		
Asian	2	(2%)		
Am. Indian/Alaska Native	1	(1%)		
Ethnicity				
Hispanic	5	(6%)		
Non Hispanic	83	(94%)		
BMI (kg/m ²)			27.0	(23.0 - 29.1)
Prior Joint Procedure	44	(50%)		
Prior Fracture	41	(47%)		
Target Joints				
≥ 4 bleeds in 6 months	38	(43%)		
≥ 20 bleeds in lifetime	56	(64%)		
HIV positive	26	(30%)		
Hemophilia Type				
A	71	(81%)		
B	17	(19%)		
Disease Severity				
Mild	26	(30%)		
Moderate	24	(27%)		
Severe	38	(43%)		
Inhibitor				
Current	5	(6%)		
Ever	5	(6%)		
Chronic Liver Disease	47	(53%)		
Baseline Factor VIII Level (%)			5.5	(.35 - 7.0)
WFH score ¹ (0-68)			14.3	(7 - 22)
PAI score (25-66)			32.5	(27.7 - 35.8)
HAL score (0-100)			72.9	(54 - 94)
TUG average time (seconds)			9.8	(8.0 - 10.7)
¹ n=70 due to missing values				

Table 2. Bivariate and Multivariable Associations of Demographic and Clinical Characteristics with PAI*

Independent Variable	Simple Linear Regression Analysis				Multiple Linear Regression Analysis			
	$\hat{\beta}_0$	$\hat{\beta}_1$	95% Confidence Limits	P Value	R ²	$\hat{\beta}$ Coefficient	95% Confidence Limits	P Value
Intercept	--	--	--	--	--	3.76	(3.64, 3.89)	<.0001
WFH Score ¹	3.57	-.067	(-.105, -.029)	.0007	.13	-.025	(-.074, .024)	.31
Age ²	3.67	-.048	(-.075, -.022)	.0005	.13	-.055	(-.081, -.029)	<.0001
Non-White [⊕]	3.48	-.052	(-.131, .028)	.20	.02	-.033	(-.105, .039)	.36
BMI ³	3.37	.017	(-.015, .050)	.29	.01			
Joint Procedure hx.	3.49	-.054	(-.129, .021)	.15	.02			
Fracture Any Type	3.45	.017	(-.059, .093)	.65	<.01			
Target Joints [⊕]	3.52	-.021	(-.036, -.007)	.005	.09	-.011	(-.030, .007)	.23
Baseline Level ^{4,⊕}	3.44	.021	(-.001, .044)	.064	.04	.018	(-.007, .042)	.15
HIV pos.	3.49	-.088	(-.169, -.007)	.033	.05			
Chronic Liver Dis.	3.51	-.096	(-.169, -.023)	.011	.07			

¹ Per 10 point change, multiple imputation used, R² is avg. across imputations

² Per 10 years change

³ Per 5 point change

⁴ Per 5 percentage points change

*Log scale

Average (of 20 imputed models) adj R²= .25

Multiple Imputation Used for Model

⊕ Confounding variable in multivariable model

Table 3. Bivariate and Multivariable Associations of Demographic and Clinical Characteristics with HAL

Independent Variable	Simple Linear Regression Analysis				Multiple Linear Regression Analysis			
	$\hat{\beta}_0$	$\hat{\beta}_1$	95% Confidence Limits	P Value	R^2	$\hat{\beta}$ Coefficient	95% Confidence Limits	P Value
Intercept	--	--	--	--	--	94.9	(87.3, 102.6)	<.0001
WFH Score ¹	92.3	-12.4	(-16.9, -7.92)	<.0001	.27	-5.76	(-11.4, -.155)	.044
Age ²	81.9	-2.10	(-5.68, 1.48)	.25	.02			
Non-White	73.9	-3.01	(-13.2, 7.16)	.56	<.01			
BMI ³	57.2	2.90	(-1.19, 6.98)	.16	.02			
Joint Procedure hx.	78.4	-11.0	(-20.4, -1.66)	.022	.06			
Fracture Any Type	75.4	-5.39	(-15.0, 4.23)	.27	.01			
Target Joints	85.9	-4.87	(-6.52, -3.22)	<.0001	.29	-2.48	(-4.57, -.395)	.020
Baseline Level ⁴	67.0	5.34	(2.65, 8.04)	.0002	.15			
HIV pos.	76.3	-11.7	(-22.0, -1.41)	.026	.06			
Chronic Liver Dis.	84.2	-21.3	(-29.8, -12.7)	<.0001	.22	-12.0	(-20.6, -3.54)	.006

¹ Per 10 point change, multiple imputation used, R^2 is avg. across imputations
² Per 10 years change
³ Per 5 point change
⁴ Per 5 percentage points change

Average (of 20 imputed models) adj R^2 = .38
Multiple imputation used for model

Table 4. Bivariate and Multivariable Associations of Demographic and Clinical Characteristics with TUG*

Outcome TUG (log scale)

Independent Variable	Simple Linear Regression Analysis				Multiple Linear Regression Analysis			
	β_0	β_1	95% Confidence Limits	P Value	R ²	β Coefficient	95% Confidence Limits	P Value
Intercept	--	--	--	--	--	1.90	(1.72, 2.07)	<.0001
WFH Score ¹	2.12	.085	(.029, .140)	.003	.11	.047	(-.028, .122)	.22
Age ²	2.03	.051	(.013, .089)	.010	.08	.049	(.012, .086)	.011
Non-White	2.24	.037	(-.074, .149)	.51	<.01			
BMI ³	2.33	-.015	(-.060, .031)	.52	<.01			
Joint Procedure hx.	2.21	.080	(-.025, .185)	.13	.03			
Fracture Any Type	2.25	.009	(-.098, .115)	.87	<.01			
Target Joints [⊕]	2.17	.028	(.008, .049)	.008	.08	.018	(-.009, .045)	.20
Baseline Level ⁴	2.27	-.022	(-.054, .010)	.18	.02			
HIV pos. [⊕]	2.24	.019	(-.097, .135)	.75	<.01	-.098	(-.215, .018)	.097
Chronic Liver Dis. [⊕]	2.17	.155	(.054, .256)	.003	.09	.095	(-.017, .206)	.095

¹ Per 10 point change, multiple imputation used, R² is avg. across imputations

² Per 10 years change

³ Per 5 point change

⁴ Per 5 percentage points change

*Log scale

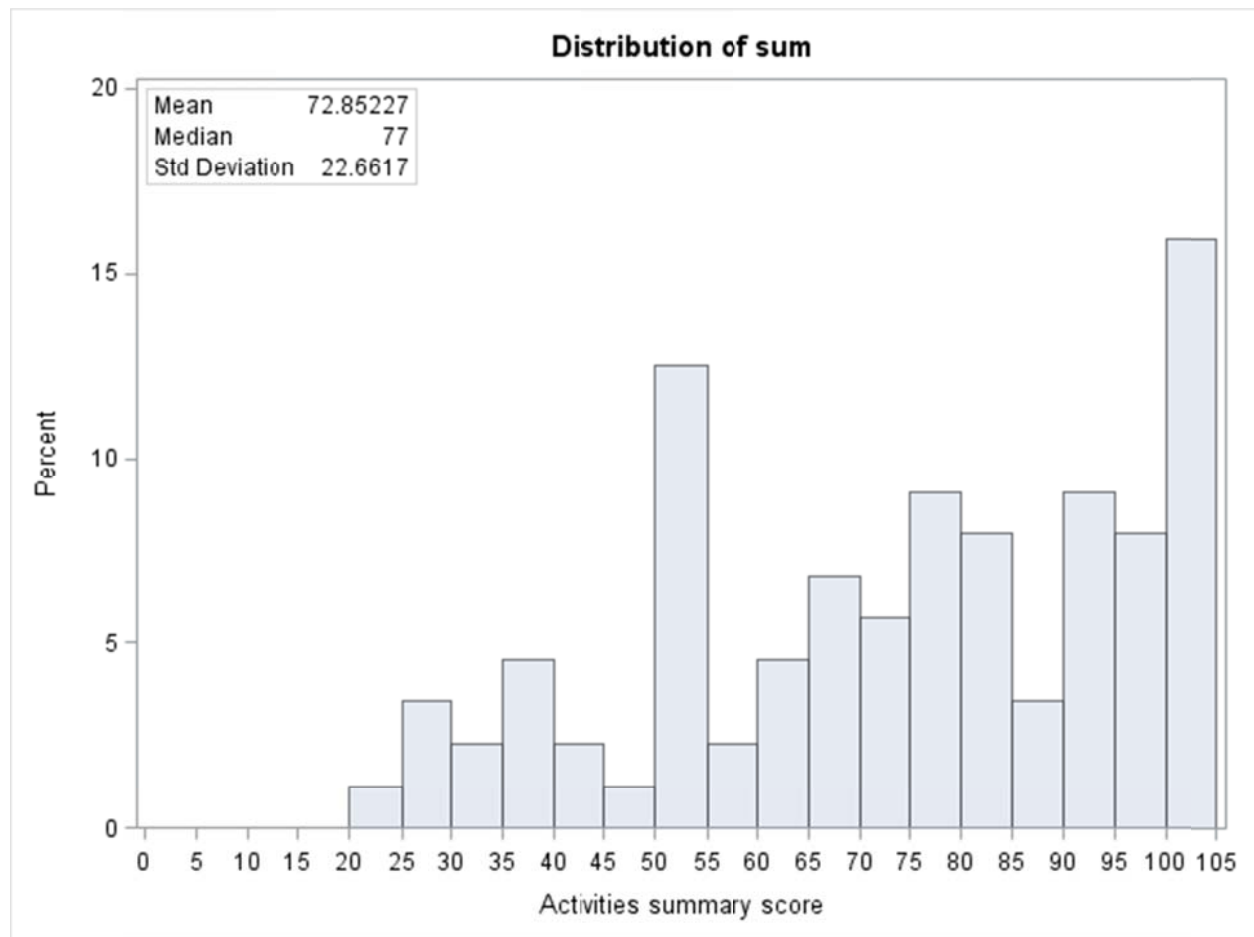
Average (of 20 imputed models) adj R²= .19

Multiple imputation used for model

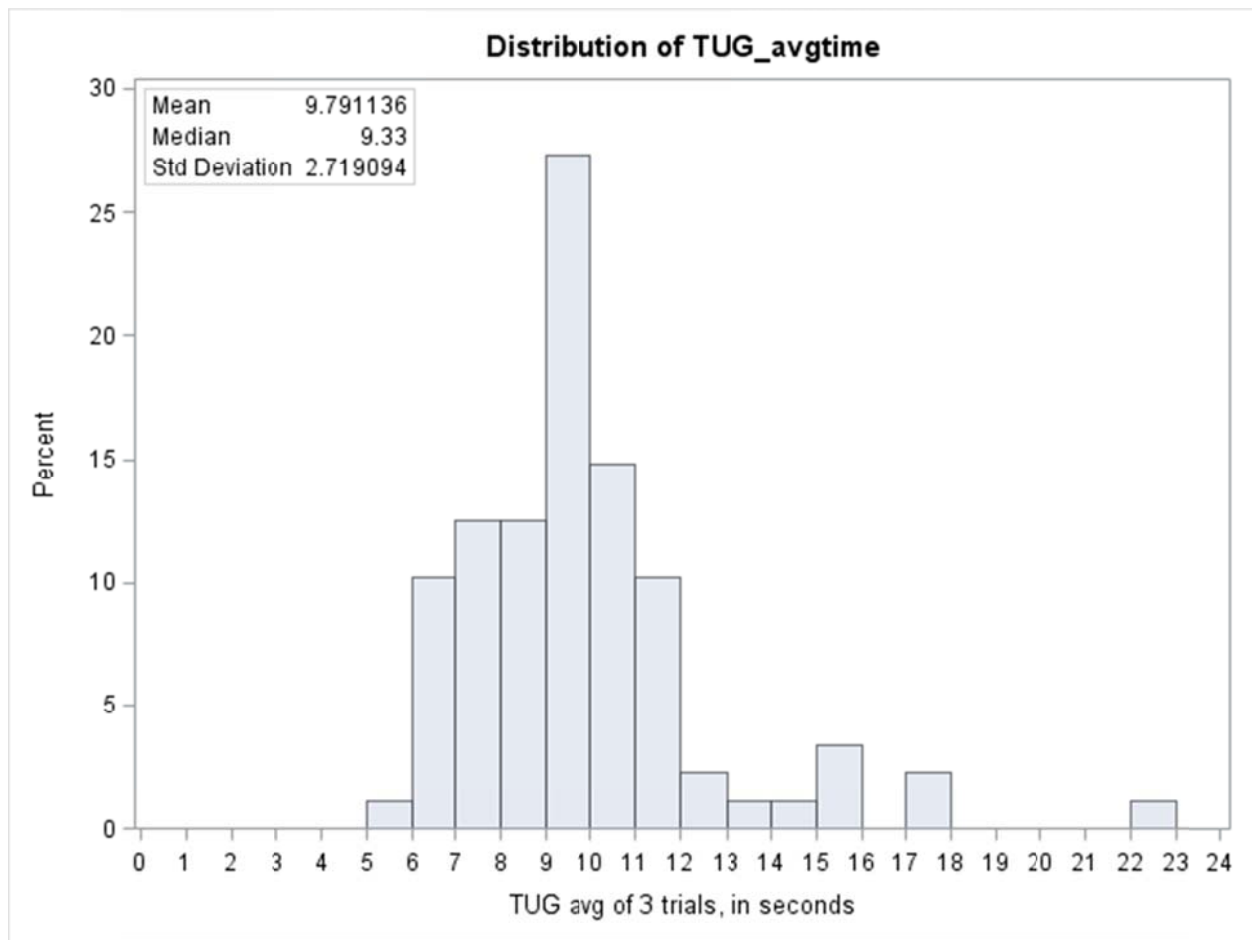
⊕ Confounding variable in multivariable model

Appendix A. Additional Figures

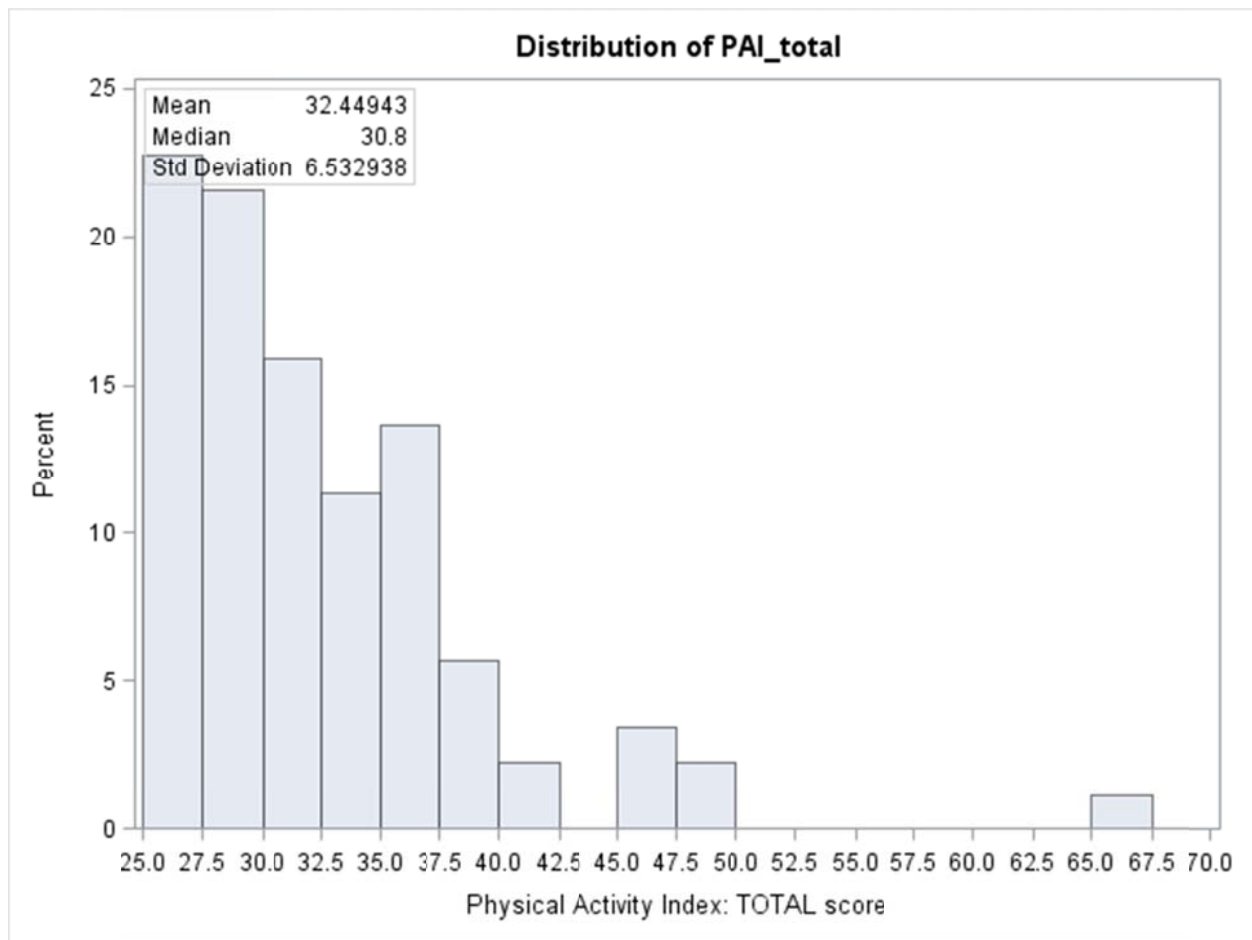
Distribution of HAL



Distribution of TUG

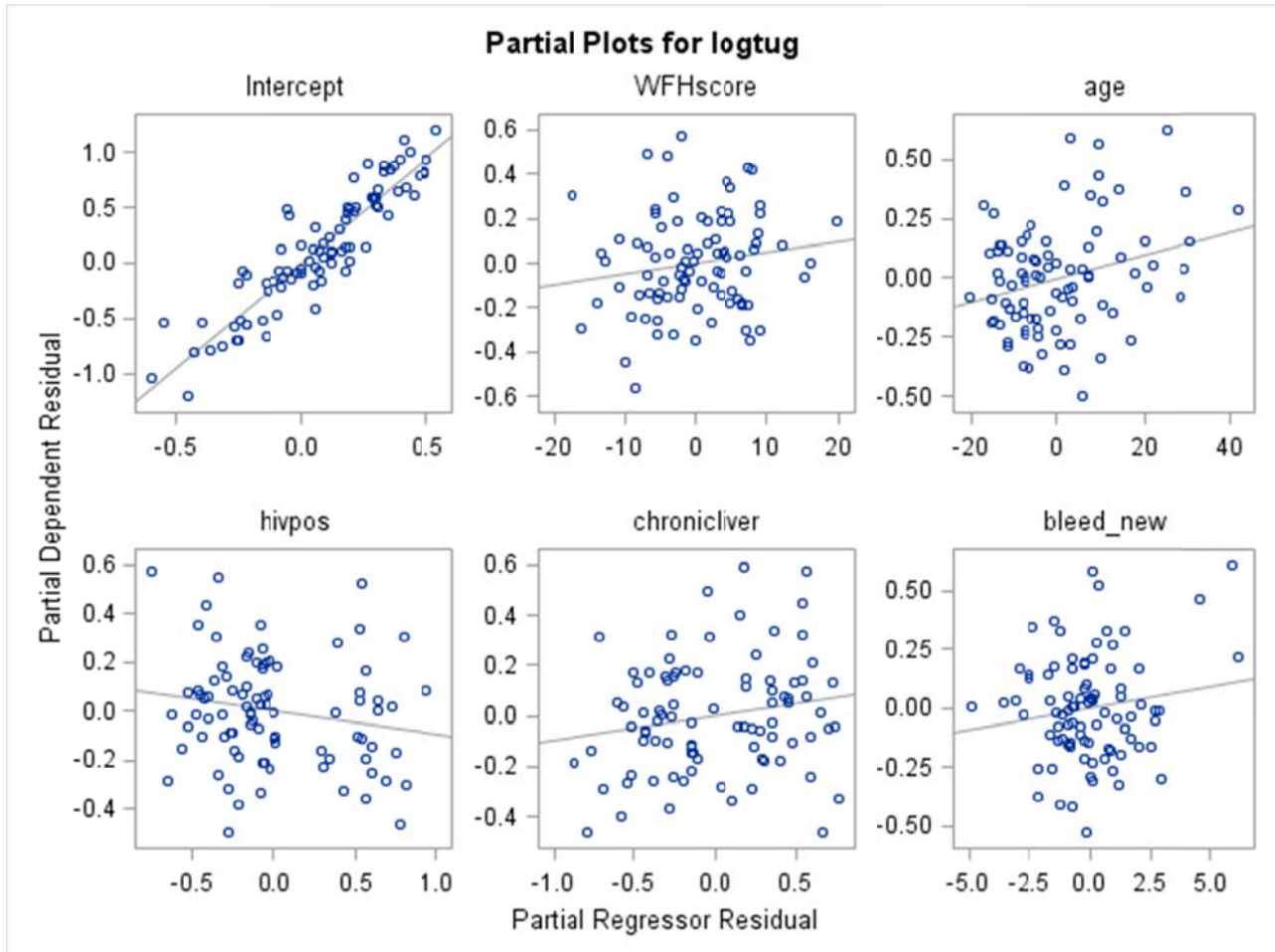


Distribution of PAI



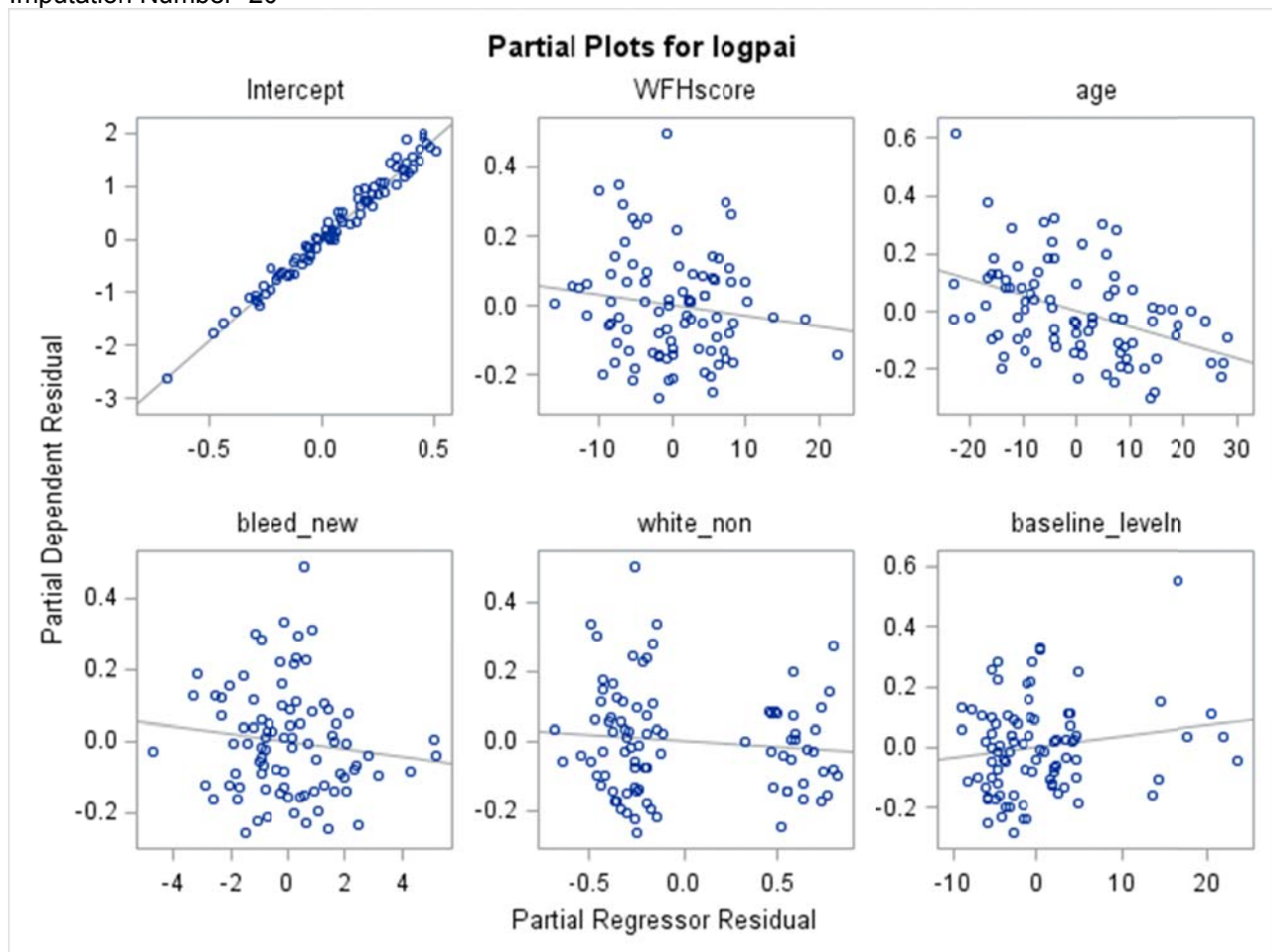
Partial plots – TUG (log scale)

Imputation Number=20



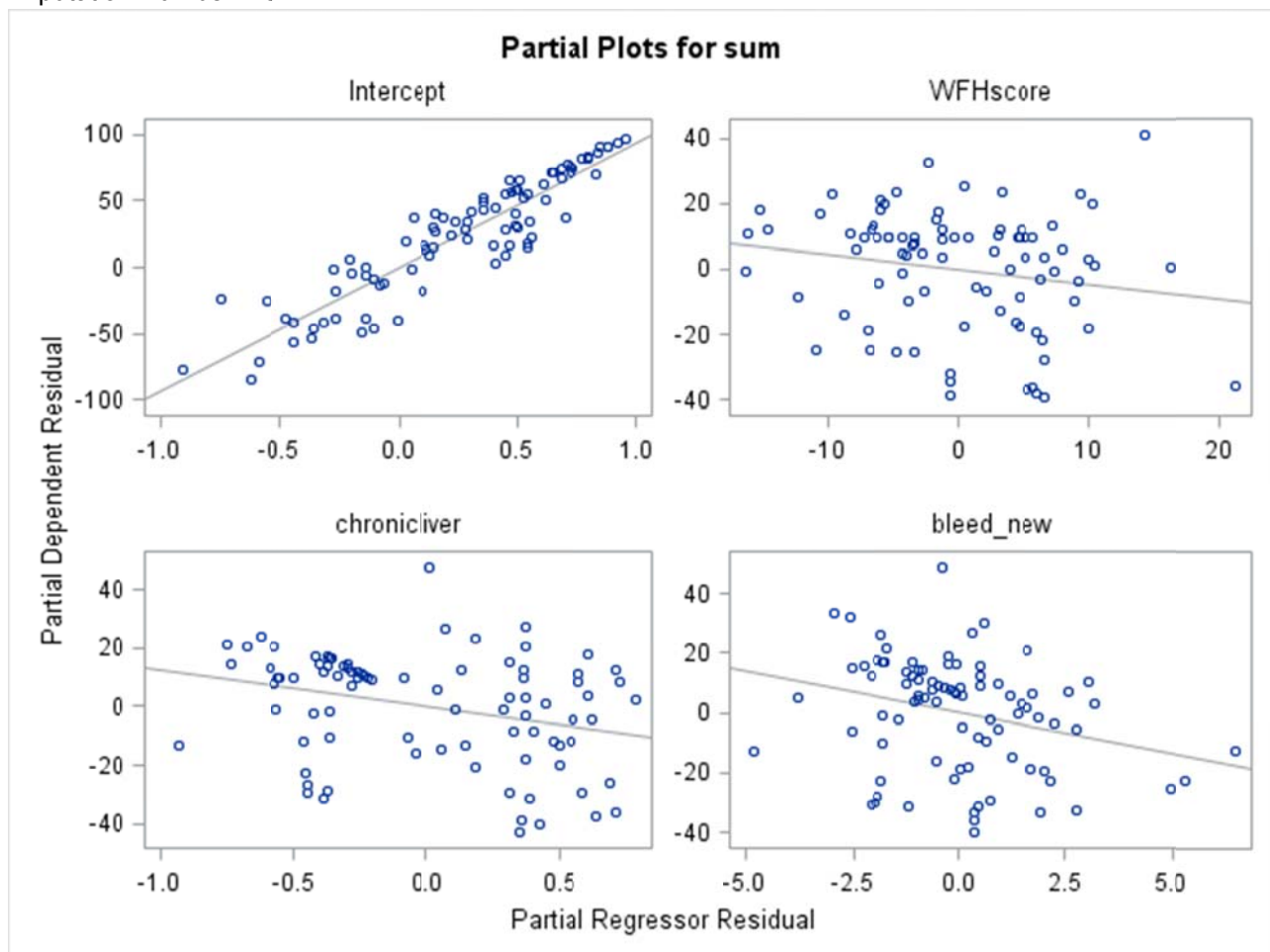
Partial plots – PAI (log scale)

Imputation Number=20



Partial plots – HAL

Imputation Number=20



Appendix B. SAS Code

SAS Program 1

```

libname dec
"\\research.sph.emory.edu\ResearchData\BiosProjs\Hemophilia\Lisa\DATA\Dec
2011 " ;

libname inter
"\\research.sph.emory.edu\ResearchData\BiosProjs\Hemophilia\Lisa\intermediate
datasets\old datasets " ;

libname jason
"\\research.sph.emory.edu\ResearchData\BiosProjs\Hemophilia\Jason" ;

***Use this option to avoid error "format XXXX was not found or could not be
loaded"***;
OPTIONS nofmterr;
***Options for Gplot ***;
SYMBOL INTERPOL=r1 VALUE=dot COLOR=green;
SYMBOL2 INTERPOL=r1 VALUE=star COLOR=red;
SYMBOL3 INTERPOL=r1 VALUE=marker COLOR=blue;

**make format library;
proc format library=LIBRARY;
value steroid 0='LT 3mo'
               1='3-6mo'
               2='>6-12mo'
               3='>12mo';
VALUE TUG      0='NONE'
               1='CRUTCHES'
               2='WALKER'
               3='CANE';

run;

proc contents data=dec.hemo_4;
run;

proc contents data=jason.hemo;
run;

proc freq data=jason.hemo;
table pai_total;
run;

```

```

data jason.hemo;
set jason.hemo;
tothrs=sleep+job_hrs+extra_hrs;
run;

/* Check that PAI total hours is 24 */
ods rtf;
proc freq data=jason.hemo;
table tothrs / nocum nopercent ;
run;
ods rtf close;

proc freq data=dec.hemo_4;
table age /nocum;
run;

proc print data=dec.hemo_4 noobs;
var id age pai_total;
run;

/* PAI by age */

SYMBOL INTERPOL=r1 VALUE=dot COLOR=green;
PROC GPLOT DATA=dec.hemo_4 ;
PLOT pai_total*age ;
RUN;

proc sgscatter data= dec.hemo_4 ;
matrix age pai_total/diagonal=(histogram) ;
run;

proc freq data=jason.hemo;
table race / nocum;
run;

proc freq data=dec.hemo_4;
table hemophiliatype / nocum;
run;

proc freq data=dec.hemo_4;
table hivpos / nocum;
run;

SYMBOL2 /*INTERPOL=r1*/ VALUE=star COLOR=blue;
PROC GPLOT DATA=dec.hemo_4 ;
where race in (1,2);
PLOT pai_total * age =smoke ;
RUN;

PROC SGPLOT DATA= dec.hemo_4 ;
SCATTER X=age Y=pai_total /GROUP=hemophiliatype;

```

```
RUN;
```

```
SYMBOL2 /*INTERPOL=r1*/ VALUE=star COLOR=blue;
PROC GPLOT DATA=dec.hemo_4 ;
PLOT pai_total * age =hivpos ;
RUN;
```

```
proc freq data=dec.hemo_4;
table job / nocum nopercnt;
run;
```

```
/* job by PAI */
proc print data=dec.hemo_4 noobs;
var job pai_total;
run;
```

```
/* TUG */
ods rtf;
proc print data=dec.hemo_4 noobs;
var id trial1_time trial2_time trial3_time avg_time trial1_comm trial2_comm
trial3_comm avg_comm;
run;
ods rtf close;
```

```
/* Print out TUG */
ods rtf;
proc print data=jason.hemo noobs;
var id trial1_time trial2_time trial3_time avg_tug trial1_comm trial2_comm
trial3_comm avg_comm;
run;
ods rtf close;
```

```
/* PAI by TUG */
*SYMBOL INTERPOL=r1 VALUE=dot COLOR=green;
PROC GPLOT DATA=jason.hemo;
PLOT pai_total*avg_tug ;
RUN;

proc sgscatter data= jason.hemo ;
matrix pai_total avg_tug/diagonal=(histogram) ;
run;
```

```
proc contents data=jason.hemo order=varnum ;
run;
```

```
/* PAI by HAL */
```

```

PROC GPLOT DATA=jason.hemo;
PLOT pai_total*sum ;
RUN;

/* TUG by HAL */

PROC GPLOT DATA=jason.hemo;
PLOT avg_tug*sum ;
RUN;

/* HAL and TUG by age */

PROC GPLOT DATA=jason.hemo;
PLOT sum*age;
RUN;

PROC GPLOT DATA=jason.hemo;
PLOT avg_tug*age;
RUN;

/* disease severity */

PROC UNIVARIATE DATA=jason.hemo;
VAR log_baselevel;
*CLASS sex;
HISTOGRAM log_baselevel;
INSET MEAN median ;
RUN;

PROC GPLOT DATA=jason.hemo;
PLOT logbase*sum;
RUN;

PROC GPLOT DATA=jason.hemo;
PLOT avg_tug*logbase;
RUN;

/* Joint Scores */

proc sort data=jason.hemo;
by id;
run;

ods rtf;
proc print data=jason.hemo noobs;
var id wfhscore pettersson wfh_numjoints pettersson_numjoints ;
run;

```



```

ods rtf close;

proc contents data=jason.hemo;
run;

***try to get these two to print right***;
***WFH***;
proc print data=jason.hemo noobs;
var id relbow_totjointscore lelbow_totjointscore rknee_totjointscore
lknee_totjointscore
rankle_totjointscore lankle_totjointscore wfhscore;
run;
***Petterson***;
proc print data=jason.hemo noobs;
var id relbow_totalradioscore lelbow_totalradioscore rknee_totalradioscore
lknee_totalradioscore
rankle_totalradioscore lankle_totalradioscore petterson;
run;
ods rtf close;

***realtioship between wfh (exam) and petterson (radio)***;
ods rtf;
proc sgscatter data= jason.hemo;
matrix wfhscore petterson /diagonal=(histogram) ;
run;
ods rtf close;

data jason.hemo;
set jason.hemo;

array wfh {6} relbow_totjointscore lelbow_totjointscore rknee_totjointscore
lknee_totjointscore
rankle_totjointscore lankle_totjointscore;

wfh_numjoints = 0;
do i=1 to 6;
if wfh {i} ne . then wfh_numjoints+1;
end;

run;

/* Various scatter plots */

proc sgscatter data= jason.hemo;
matrix wfhscore sum /diagonal=(histogram) ;
run;

proc sgscatter data= jason.hemo;
matrix wfhscore tug_avgtime /diagonal=(histogram) ;
run;

```

```
proc sgscatter data= jason.hemo;
matrix petterson wfhscore tug_avgtime sum pai_total/diagonal=(histogram) ;
run;
```

```
proc contents data=jason.hemo ;
run;
```

```
data jason.hemo;
set jason.hemo;
logtug=log(tug_avgtime) ; logsum=log(sum); logwfh=log(wfhscore);
logpett=log(petterson);
run;
```

```
data jason.hemo;
set jason.hemo;
logpai = log(pai_total);
run;
```

```
data jason.hemo;
set jason.hemo;
logage = log(age);
run;
```

```
proc sgscatter data= jason.hemo;
matrix logpett logwfh logtug sum logpai bmi_sas/diagonal=(histogram) ;
run;
```

```
proc sgscatter data= jason.hemo;
matrix logtug sum logpai/diagonal=(histogram) ;
run;
```

```
proc sgscatter data= jason.hemo;
matrix tug_avgtime sum pai_total/diagonal=(histogram) ;
run;
```

```
proc sgscatter data= jason.hemo;
matrix wfhscore logage log_baselevel/diagonal=(histogram) ;
run;
```

```
proc sgscatter data= jason.hemo;
plot logtug*bmi_sas;
run;
```

```
SYMBOL INTERPOL=r1 VALUE=dot COLOR=green;
```

```
PROC GPLOT DATA=jason.hemo ;
PLOT log_baselevel*logwfh;
RUN;
```

```
PROC GGPLOT DATA=jason.hemo ;
PLOT log_baselevel*petterson ;
RUN;
```

```
PROC GGPLOT DATA=jason.hemo ;
PLOT logwfh*log_baselevel;
RUN;
```

```
PROC GGPLOT DATA=jason.hemo ;
PLOT petterson*log_baselevel ;
RUN;
```

```
proc sgscatter data= jason.hemo;
matrix log_baselevel logpai logtug sum /diagonal=(histogram) ;
run;
```

```
SYMBOL INTERPOL=rl VALUE=dot COLOR=green;
PROC GGPLOT DATA=jason.hemo ;
PLOT logpai*log_baselevel log_baselevel*logpai logtug*log_baselevel
log_baselevel*logtug
sum*log_baselevel log_baselevel*sum;
RUN;
```

```
proc freq data=jason.hemo;
table hivpos / nopercnt nocum;
run;
```

```
/* HIV status */
```

```
PROC UNIVARIATE DATA=jason.hemo noprint;
VAR wfhscore logwfh;
CLASS hivpos;
HISTOGRAM wfhscore logwfh;
INSET MEAN median n;
RUN;
```

```
PROC UNIVARIATE DATA=jason.hemo noprint;
VAR petterson;
CLASS hivpos;
HISTOGRAM petterson;
INSET MEAN median n;
RUN;
```

```
PROC UNIVARIATE DATA=jason.hemo noprint;
VAR log_baselevel sum logsum;
CLASS hivpos;
HISTOGRAM log_baselevel sum logsum;
INSET MEAN median n;
```

```

RUN;

SYMBOL INTERPOL=r1 VALUE=dot COLOR=green;
SYMBOL2 INTERPOL=r1 VALUE=dot COLOR=red;
PROC GPLOT DATA=jason.hemo ;
PLOT sum*log_baselevel=hivpos logsum*log_baselevel=hivpos;
RUN;

ods rtf;
proc freq data=jason.hemo;
table hivpos / nopercnt nocum;
run;
proc freq data=jason.hemo;
table hivpos*(race hemophiliatype )/ nocum nopercnt norow;
run;
ods rtf close;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR age logage bmi bmi_sas;
CLASS hivpos;
HISTOGRAM age logage bmi bmi_sas;
INSET MEAN median std n;
RUN;

/* Fracture */

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR sum logsum logage log_baselevel wfhscore logwfh bmi_sas;
CLASS fracture;
HISTOGRAM sum logsum logage log_baselevel wfhscore logwfh bmi_sas;
INSET MEAN median std n;
RUN;

/* bmi */

proc sgscatter data= jason.hemo ;
matrix bmi_sas sum logage log_baselevel wfhscore logwfh /diagonal=(histogram)
;
run;

PROC GPLOT DATA=jason.hemo ;
PLOT (sum logage log_baselevel wfhscore logwfh)*bmi_sas;
RUN;

/* age */

PROC GPLOT DATA=jason.hemo ;
PLOT (sum logsum logpai pai_total logtug tug_avgtime log_baselevel wfhscore
logwfh pettersen)*logage;
RUN;

```

```

/* Joint exams by Physical Activities measures */

PROC Gplot DATA=jason.hemo ;
PLOT (logpai logtug logsum)*logwfh      (logpai logtug logsum)*petterson;
RUN;

/* Prophylaxis */

***current***;
PROC UNIVARIATE DATA=jason.hemo noprint;
VAR logpai logtug logsum logwfh petterson logage log_baselevel;
CLASS current_prophylaxis;
HISTOGRAM logpai logtug logsum logwfh petterson logage log_baselevel;
INSET MEAN median std n;
RUN;

***ever***;
PROC UNIVARIATE DATA=jason.hemo noprint;
VAR logpai logtug logsum logwfh petterson logage log_baselevel;
CLASS ever_prophylaxis;
HISTOGRAM logpai logtug logsum logwfh petterson logage log_baselevel;
INSET MEAN median std n;
RUN;

proc ttest data=jason.hemo ;
var logpai logtug logsum logwfh petterson logage log_baselevel;
CLASS current_prophylaxis;
run;

ods rtf;
proc freq data= jason.hemo;
table hivpos racebw hemophiliatype smoke fracture current_prophylaxis
ever_prophylaxis/ nocum;
run;
ods rtf close;

/* LOOKING FOR INTERACTION WITH HIV STATUS */

***Physical Activity measures by joint scores***;

SYMBOL2 INTERPOL=rl VALUE=star COLOR=blue;
PROC Gplot DATA=jason.hemo;
PLOT sum * logwfh =hivpos      sum      *petterson = hivpos;
RUN; ***No interaction***;

SYMBOL2 INTERPOL=rl VALUE=star COLOR=blue;
PROC Gplot DATA=jason.hemo;
PLOT logpai * logwfh =hivpos      pai_total * logwfh =hivpos      logpai
*petterson = hivpos      pai_total *petterson = hivpos;
RUN; ***follow up with statistical test but doesn't look like interaction
(hiv pos sample small)**;

```

```

SYMBOL2 INTERPOL=rl VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;
PLOT logtug * logwfh =hivpos      tug_avgtime * logwfh =hivpos      logtug
*petterson = hivpos      tug_avgtime *petterson = hivpos;
RUN; ***No interaction***;

***Physical Activity measures by disease severity***;

SYMBOL2 INTERPOL=rl VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;
PLOT sum* log_baselevel =hivpos ;
run; ***YES interaction***;

SYMBOL2 INTERPOL=rl VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;
PLOT pai_total* log_baselevel =hivpos      logpai* log_baselevel =hivpos ;
run; ***YES interaction***;

SYMBOL2 INTERPOL=rl VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;
PLOT logtug* log_baselevel =hivpos      tug_avgtime* log_baselevel
=hivpos ;
run; ***NO interaction***;

***Joint score by disease severity***;

SYMBOL2 INTERPOL=rl VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;
PLOT logwfh* log_baselevel =hivpos      wfhscore* log_baselevel =hivpos
petterson*log_baselevel=hivpos;
run; ***No interaction***;

/*LOOKING FOR INTERACTION WITH RACE*/

data jason.hemo;
set jason.hemo;
if race in (3,5) then racebw = . ;
else racebw=race;
run;

proc freq data=jason.hemo;
table racebw / nocum;
run;

***Physical Activity measures by joint scores***;

SYMBOL2 INTERPOL=rl VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo; where racebw in (1,2);
PLOT sum * logwfh =racebw      sum *petterson = racebw;

```

```

                                ***possible interaction for pettersson by
HAL, follow up with statistical test***;
RUN;

SYMBOL2 INTERPOL=rl VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo; where racebw in (1,2);
PLOT logpai * logwfh =racebw      pai_total * logwfh =racebw      logpai
*petterson = racebw      pai_total *pettersson = racebw;
RUN;

SYMBOL2 INTERPOL=rl VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;  where racebw in (1,2);
PLOT logtug * logwfh =racebw      tug_avgtime * logwfh =racebw      logtug
*petterson = racebw      tug_avgtime *pettersson = racebw;
RUN;

***Physical Activity measures by disease severity***;

SYMBOL2 INTERPOL=rl VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;  where racebw in (1,2);
PLOT sum* log_baselevel =racebw;
run; ***Possible interaction here. Follow up with statistical test***;

SYMBOL2 INTERPOL=rl VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;  where racebw in (1,2);
PLOT pai_total* log_baselevel =racebw      logpai* log_baselevel =racebw ;
run;

SYMBOL2 INTERPOL=rl VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;  where racebw in (1,2);
PLOT logtug* log_baselevel =racebw      tug_avgtime* log_baselevel
=racebw ;
run;

***Joint score by disease severity***;

SYMBOL2 INTERPOL=rl VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;  where racebw in (1,2);
PLOT logwfh* log_baselevel =racebw      wfhscore* log_baselevel =racebw
pettersson*log_baselevel=racebw;
run;

/* LOOKING FOR INTERACTION WITH FRACTURE */

***Physical Activity measures by joint scores***;

SYMBOL2 INTERPOL=rl VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;

```

```

PLOT sum * logwfh = fracture          sum  *petterson = fracture;
*** interaction WFH and HAL***;
RUN;

SYMBOL2 INTERPOL=r1 VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;
PLOT logpai * logwfh =fracture          logpai  *petterson = fracture  ;
***interaction for both ***;
RUN;

SYMBOL2 INTERPOL=r1 VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;
PLOT logtug * logwfh =fracture          logtug  *petterson = fracture  ;
RUN;  ***no interaction***;

***Physical Activity measures by disease severity***;

SYMBOL2 INTERPOL=r1 VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;
PLOT sum* log_baselevel =fracture          logpai* log_baselevel =fracture
logtug* log_baselevel =fracture;
run;  ***Interaction all three***;

***Joint score by disease severity***;

SYMBOL2 INTERPOL=r1 VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;
PLOT logwfh* log_baselevel =fracture      petterson*log_baselevel=fracture;
run;  ***interaction both, weaker on petterson by baselevel***;

/* LOOKING FOR INTERACTION WITH SMOKE */

***Physical Activity measures by joint scores***;

SYMBOL2 INTERPOL=r1 VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;
PLOT sum * logwfh = smoke          sum  *petterson = smoke;

RUN;

SYMBOL2 INTERPOL=r1 VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;
PLOT logpai * logwfh =smoke          logpai  *petterson = smoke  ;

RUN;

SYMBOL2 INTERPOL=r1 VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;
PLOT logtug * logwfh =smoke          logtug  *petterson = smoke  ;
RUN;

```



```

***Physical Activity measures by disease severity***;

SYMBOL2 INTERPOL=r1 VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;
PLOT sum* log_baselevel =smoke      logpai* log_baselevel =smoke
logtug* log_baselevel =smoke;
run;

***Joint score by disease severity***;

SYMBOL2 INTERPOL=r1 VALUE=star COLOR=blue;
PROC GPLOT DATA=jason.hemo;
PLOT logwfh* log_baselevel =smoke    petterson*log_baselevel=smoke;
run;

/* Looking at frequencies of other variables */

proc contents data=jason.hemo ;
run;

proc contents data=jason.hemo order=varnum ;
run;

ods rtf;
***family history**
proc freq data=jason.hemo;
table parent_hipfracture parent_osteo / nocum ;
run;

***bleeds***;
proc freq data=jason.hemo;
table ge4bleeds_any ge20bleeds_any/ nocum;
run;

proc freq data=jason.hemo;
table ge4bleeds_rhip ge4bleeds_lhip ge4bleeds_rknee ge4bleeds_lknee
ge4bleeds_rshoulder ge4bleeds_lshoulder
ge4bleeds_relbow ge4bleeds_lelbow ge4bleeds_rankle ge4bleeds_lankle
ge4bleeds_rother ge4bleeds_lother ge4bleeds_othjointspec/ nocum ;
run;

proc freq data=jason.hemo;
table ge20bleeds_rhip ge20bleeds_lhip ge20bleeds_rknee ge20bleeds_lknee
ge20bleeds_rshoulder ge20bleeds_lshoulder
ge20bleeds_relbow ge20bleeds_lelbow ge20bleeds_rankle ge20bleeds_lankle
ge20bleeds_rother1 ge20bleeds_lother1 ge20bleeds_rother2 ge20bleeds_lother2
ge20bleeds_othjoint1spec ge20bleeds_othjoint2spec/ nocum ;
run;

***conditions***;

```

```

proc freq data=jason.hemo;
table kidneystones diabetes cushing chronicliver colitis crohns hyperthyroid
hypothyroid wtloss_unintentional wtloss_intentional / nocum;
run;

***other***;
proc freq data=jason.hemo;
table levothyroxin dilantin steroid_gt3mo steroid_totalltime
testosteronereplace jointsurgery / nocum ;
run;

***HIV***;
proc freq data=jason.hemo;
table hivpos_cd4lt200 hivpos_trt / nocum;
run;
ods rtf close;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR lastcd4 ;
HISTOGRAM lastcd4 ;
INSET MEAN median std n;
RUN;

proc freq data=jason.hemo;
table lastviral / nocum;
run; ***see if can change this to numeric variable***;

proc gplot data=jason.hemo;
plot (sum logpai logtug logwfh petterson log_baselevel) *lastcd4 ;
run;

***factor exposure days**;
proc freq data=jason.hemo;
table exact_totfactordays cat_totfactordays/ nocum ;
run;

proc gplot data=jason.hemo;
plot (sum logpai logtug logwfh petterson log_baselevel) * cat_totfactordays;
run;

***inhibitor***;
proc freq data=jason.hemo;
table current_inhibitor date_currrentinhibitor_pos/ nocum;
run;***only five***;

***hepatitis C***;
proc freq data=jason.hemo;
table hepc_antibody hepc_viralload antihcvtrt antihcvtrt_type hepc_genotype
liverbiopsy liverbiopsy_finding /nocum;
run;

***looking at age distributions by disease***;

```

```
PROC UNIVARIATE DATA=jason.hemo noprint;
VAR age;
CLASS hivpos ;
HISTOGRAM age;
INSET min median max n;
RUN;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR age;
CLASS hepc_antibody ;
HISTOGRAM age;
INSET min median max n;
RUN;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR age;
CLASS hepc_viralload ;
HISTOGRAM age;
INSET min median max n;
RUN;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR age;
CLASS chronicliver ;
HISTOGRAM age;
INSET min median max n;
RUN;

***some interactions ***;

proc gplot data=jason.hemo;
plot logtug*age=hivpos;
run;

proc gplot data=jason.hemo;
plot logwfh*age=hivpos;
run;

proc gplot data=jason.hemo;
plot wfhscore*age=hivpos;
run;

proc gplot data=jason.hemo;
plot petterson*age=hivpos;
run;

proc gplot data=jason.hemo;
plot log_baselevel*age=hivpos;
run;

proc gplot data=jason.hemo;
plot (wfhscore petterson)*age=chronicliver;
run;
```

```

***Joint procedures***;

proc freq data=jason.hemo;
table joint_procedure_any / nocum;
run;

Proc freq data=jason.hemo;
table arthro_synovectomy open_synovectomy radio_synovectomy replacement
othinvasive / nocum;
run;

Proc freq data=jason.hemo;
table arthro_synovectomy_any open_synovectomy_any radio_any replacement_any
othinvasive_any / nocum;
run;

proc print data=jason.hemo noobs;
var id jointsurgery joint_procedure_any arthro_synovectomy open_synovectomy
radio_synovectomy replacement othinvasive;
run;

proc print data=jason.hemo noobs;
var id jointsurgery joint_procedure_any arthro_synovectomy_any
open_synovectomy_any radio_any replacement_any othinvasive_any;
run; ***use this one which corrects for the "any" button not being checked
on the entry form***;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR logwfh wfhscore petterson;
CLASS joint_procedure_any;
HISTOGRAM logwfh wfhscore petterson;
INSET MEAN median n ;
RUN; ***big differences especially WFH***;
***might want to look at replacement separately on these***;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR age;
CLASS joint_procedure_any;
HISTOGRAM age;
INSET MEAN median n ;
RUN; ***big difference***;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR log_baselevel;
CLASS joint_procedure_any;
HISTOGRAM log_baselevel;
INSET MEAN median n ;
RUN; ***big difference***;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR sum logpai logtug;
CLASS joint_procedure_any;
HISTOGRAM sum logpai logtug;
INSET MEAN median n ;
RUN; ***slight differences ***;

```

```

***Fractures (from fracture form)***;

proc freq data=jason.hemo;
table fracture_spontan_any fracture_lowcat_any fracture_highcat_any
fracture_anytype fracture/ nocum ;
run;

proc print data=jason.hemo noobs;
var id fracture fracture_anytype fracture_spontan_any fracture_lowcat_any
fracture_highcat_any ;
run;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR logwfh petterson;
CLASS fracture_anytype;
HISTOGRAM logwfh petterson;
INSET MEAN median n ;
RUN; ***WFH is the only one that looks significant***;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR age logage;
CLASS fracture_anytype;
HISTOGRAM age logage;
INSET MEAN median n ;
RUN;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR log_baselevel;
CLASS fracture_anytype;
HISTOGRAM log_baselevel;
INSET MEAN median n ;
RUN;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR sum logpai logtug;
CLASS fracture_anytype;
HISTOGRAM sum logpai logtug;
INSET MEAN median n ;
RUN;

proc gplot data=jason.hemo;
plot (logpai logtug sum)*log_baselevel=fracture_anytype ;
run;

proc gplot data=jason.hemo;
plot (logwfh petterson)*log_baselevel=fracture_anytype ;
run;

proc gplot data=jason.hemo;
plot (logpai logtug sum)*logwfh=fracture_anytype (logpai logtug
sum)*petterson=fracture_anytype ;
run;

proc gplot data=jason.hemo;

```

```

plot (logwfh petterson)*logage = fracture_anytype ;
run;

proc gplot data=jason.hemo;
plot log_baselevel*logage = fracture_anytype ;
run;

proc gplot data=jason.hemo;
plot (logpai logtug sum)*logage = fracture_anytype ;
run;

***Joint Bleeds ***;

***ge4 bleeds in 6 months***;
PROC UNIVARIATE DATA=jason.hemo noprint;
VAR logwfh petterson;
CLASS ge4bleeds_any;
HISTOGRAM logwfh petterson;
INSET MEAN median n ;
RUN;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR age logage;
CLASS ge4bleeds_any;
HISTOGRAM age logage;
INSET MEAN median n ;
RUN;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR log_baselevel;
CLASS ge4bleeds_any;
HISTOGRAM log_baselevel;
INSET MEAN median n ;
RUN;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR sum logpai logtug;
CLASS ge4bleeds_any;
HISTOGRAM sum logpai logtug;
INSET MEAN median n ;
RUN;

***ge 20 bleeds in lifetime***;
PROC UNIVARIATE DATA=jason.hemo noprint;
VAR logwfh petterson;
CLASS ge20bleeds_any;
HISTOGRAM logwfh petterson;
INSET MEAN median n ;
RUN;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR age logage;
CLASS ge20bleeds_any;

```

```

HISTOGRAM age logage;
INSET MEAN median n ;
RUN;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR log_baselevel;
CLASS ge20bleeds_any;
HISTOGRAM log_baselevel;
INSET MEAN median n ;
RUN;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR sum logpai logtug;
CLASS ge20bleeds_any;
HISTOGRAM sum logpai logtug;
INSET MEAN median n ;
RUN;

/* Other categorical comparisons */

***Disease Factor *** ;

***hep c and chronic liver disease***;
ods rtf;
proc freq data=jason.hemo;
table (hepc_antibody hepc_viralload)* chronicliver / nopercnt norow nocol
chisq expected;
run;
***HIV and hepatitis C***;
proc freq data=jason.hemo;
table (hepc_antibody hepc_viralload)* hivpos / nopercnt norow nocol chisq
expected;
run;
***HIV and chronic liver***;
proc freq data=jason.hemo;
table chronicliver* hivpos / nopercnt norow nocol chisq expected ;
run;

*** Joints Factor ***;

***joint bleeds and joint procedures***;
proc freq data=jason.hemo;
table (ge4bleeds_any ge20bleeds_any)* joint_procedure_any / relrisk nopercnt
nocol ;
run;
***joint bleeds and fracture***;
proc freq data=jason.hemo;
table (ge4bleeds_any ge20bleeds_any)* fracture_anytype / relrisk nopercnt
nocol ;
run;
***joint procedures and fracture***;
proc freq data=jason.hemo;
table joint_procedure_any* fracture_anytype / nopercnt norow nocol chisq
expected ;

```

```

run;
ods rtf close;

proc freq data=jason.hemo;
table fracture*fracture_anytype / nopercnt norow nocol;
run;

***Check across the two factors - HIV by fracture***;
proc freq data=jason.hemo ;
table hivpos* fracture_anytype / relrisk nopercnt nocol ;
run;

***compare some of the other categorical variables***;

/* Descriptive statistics for Table 1 */

proc means data=jason.hemo  N NMISS MEAN STD MIN Q1 MEDIAN Q3 MAX MAXDEC=2;
var /*pai_total logpai tug_avgtime logtug*/ sum;
run;

proc means data=jason.hemo  N NMISS MEAN STD MIN Q1 MEDIAN Q3 MAX MAXDEC=2;
var wfhscore logwfh pettersen;
run;

proc means data=jason.hemo  N NMISS MEAN STD MIN Q1 MEDIAN Q3 MAX MAXDEC=2;
var age logage bmi_sas baseline_leveln log_baselevel;
run;

proc means data=jason.hemo  N NMISS MEAN STD MIN Q1 MEDIAN Q3 MAX MAXDEC=2;
var lastcd4;
run;

proc freq data=jason.hemo;
table cat_totfactordays / nocum ;
run;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR pai_total sum tug_avgtime logtug;
HISTOGRAM pai_total sum tug_avgtime logtug;
RUN;

PROC UNIVARIATE DATA=jason.hemo noprint;
VAR baseline_leveln log_baselevel;
HISTOGRAM baseline_leveln log_baselevel;
RUN;

```



```
PROC UNIVARIATE DATA=jason.hemo noprint;
VAR lastcd4 ;
HISTOGRAM lastcd4 ;
RUN;
```

```
PROC UNIVARIATE DATA=jason.hemo noprint;
VAR age logage bmi_sas ;
HISTOGRAM age logage bmi_sas ;
RUN;
```

```
/* Some Regressions */
```

```
proc reg data=jason.hemo;
model petterson=log_baselevel ;
run;
```

```
proc reg data=jason.hemo;
model logwfh=log_baselevel ;
run;
```

```
proc reg data=jason.hemo;
model petterson=logwfh;
run;
```

```
proc reg data=jason.hemo;
model logtug=logage;
run;
```

```
proc reg data=jason.hemo;
model sum=logtug;
run;
```

```
proc reg data=jason.hemo;
model log_baselevel=cat_totfactordays;
run;
```

```
proc reg data=jason.hemo;
model petterson=age;
*where chronicliver = 1;
run;
```

```
proc reg data=jason.hemo;
model petterson=age;
where hivpos = 1;
run;
```

```
proc reg data=jason.hemo;
model wfhscore=age;
where hivpos = 1;
run;
```

```

proc reg data=jason.hemo;
model sum=log_baselevel;
where hivpos = 1;
run;
proc gplot data=jason.hemo;
plot sum*log_baselevel = hivpos ;
run;

/* joint score and HAL interaction by BMI category (obesity) */
;
proc gplot data=jason.hemo;
plot sum*(wfhscore pettersen);
where bmicat=' <18.5';
run;
proc gplot data=jason.hemo;
plot sum*(wfhscore pettersen);
where bmicat='18.5 to LT25';
run;
proc gplot data=jason.hemo;
plot sum*(wfhscore pettersen);
where bmicat='25 to <30';
run;
proc gplot data=jason.hemo;
plot sum*(wfhscore pettersen);
where bmicat='>=30';
run;

proc freq data=jason.hemo ;
table bmicat/ nocum ;
run;

data jason.hemo;
set jason.hemo;
if bmi_sas=. then bmicat2='0';
else if bmi_sas<25 then bmicat2='1';
else if 25<=bmi_sas<30 then bmicat2='2';
else if bmi_sas>=30 then bmicat2='3';
run;

proc freq data=jason.hemo;
table bmicat2 / nocum;
run;

proc gplot data=jason.hemo;
plot sum*wfhscore ;
where bmicat2='1';
run;
proc gplot data=jason.hemo;
plot sum*wfhscore;
where bmicat2='2';
run;
proc gplot data=jason.hemo;
plot sum*wfhscore;
where bmicat2='3';

```

```
run;

proc freq data=jason.hemo;
table bmicat2 / nocum;
run;

proc gplot data=jason.hemo;
plot sum*petterson ;
where bmicat2='1';
run;
proc gplot data=jason.hemo;
plot sum*petterson;
where bmicat2='2';
run;
proc gplot data=jason.hemo;
plot sum*petterson;
where bmicat2='3';
run;

proc freq data=jason.hemo;
table bmicat2 / nocum;
run;

***use these ***;

***HAL***;
proc gplot data=jason.hemo;
plot sum*wfhscore=bmicat2;
run;
proc gplot data=jason.hemo;
plot sum*petterson=bmicat2;
run;

***TUG***;
proc gplot data=jason.hemo;
plot tug_avvertime*wfhscore=bmicat2;
run;
proc gplot data=jason.hemo;
plot tug_avvertime*petterson=bmicat2;
run;

***PAI***;
proc gplot data=jason.hemo;
plot pai_total*wfhscore=bmicat2;
run;
proc gplot data=jason.hemo;
plot pai_total*petterson=bmicat2;
run;
```

SAS Program 2

```

libname jason "H:\Thesis" ;

***Use this option to avoid error "format XXXX was not found or could not be
loaded"***;
OPTIONS nofmterr;

proc freq data=jason.hemo;
table /*ge20bleeds_any joint_procedure_any hivpos hepc_antibody
hepc_viralload chronicliver fracture_anytype racebw*/ bmicat2/ nocum;
run;

proc print data=jason.hemo noobs;
var id sum;
run;

***fixing mis-coded values of HAL (sum) ***;
data jason.hemo;
set jason.hemo;
if id="077" then sum=65;
if id="080" then sum=90;
if id="082" then sum=76;
run;

data jason.outmi;
set jason.outmi;
if id="077" then sum=65;
if id="080" then sum=90;
if id="082" then sum=76;
run;

proc freq data=jason.outmi;
by _imputation_;
table sumorder / nocum;
run;

/* Continuous Covariates */

proc reg data=jason.hemo;
model sum=baseline_leveln / clb;
run;
proc reg data=jason.hemo;
model tug_avgtime=baseline_leveln / clb;
run;
proc reg data=jason.hemo;
model pai_total=baseline_leveln / clb;

```

```

run;

proc reg data=jason.hemo;
model sum=wfhscore / clb;
run;
proc reg data=jason.hemo;
model tug_avgtime=wfhscore / clb;
run;
proc reg data=jason.hemo;
model pai_total=wfhscore / clb;
run;

proc reg data=jason.hemo;
model sum=petterson / clb;
run;
proc reg data=jason.hemo;
model tug_avgtime=petterson / clb;
run;
proc reg data=jason.hemo;
model pai_total=petterson / clb;
run;

proc reg data=jason.hemo;
model sum=age / clb;
run;
proc reg data=jason.hemo;
model tug_avgtime=age / clb;
run;
proc reg data=jason.hemo;
model pai_total=age/ clb;
run;

proc reg data=jason.hemo;
model sum=bmi_sas/ clb;
run;
proc reg data=jason.hemo;
model tug_avgtime=bmi_sas / clb;
run;
proc reg data=jason.hemo;
model pai_total=bmi_sas / clb;
run;

/* Binary covariates */

proc reg data=jason.hemo;
model sum=joint_procedure_any / clb;
run;
proc reg data=jason.hemo;
model tug_avgtime=joint_procedure_any/ clb;
run;
proc reg data=jason.hemo;
model pai_total=joint_procedure_any/ clb;
run;

proc reg data=jason.hemo;

```

```

model sum=hivpos/ clb;
run;
proc reg data=jason.hemo;
model tug_avgtime=hivpos/ clb;
run;
proc reg data=jason.hemo;
model pai_total=hivpos/ clb;
run;

proc reg data=jason.hemo;
model sum=hepc_antibody/ clb;
run;
proc reg data=jason.hemo;
model tug_avgtime=hepc_antibody/ clb;
run;
proc reg data=jason.hemo;
model pai_total=hepc_antibody/ clb;
run;

proc reg data=jason.hemo;
model sum=hepc_viralload/ clb;
run;
proc reg data=jason.hemo;
model tug_avgtime=hepc_viralload/ clb;
run;
proc reg data=jason.hemo;
model pai_total=hepc_viralload/ clb;
run;

proc reg data=jason.hemo;
model sum=chronicliver / clb;
run;
proc reg data=jason.hemo;
model tug_avgtime=chronicliver / clb;
run;
proc reg data=jason.hemo;
model pai_total=chronicliver / clb;
run;

proc reg data=jason.hemo;
model sum=fracture_anytype / clb;
run;
proc reg data=jason.hemo;
model tug_avgtime=fracture_anytype / clb;
run;
proc reg data=jason.hemo;
model pai_total=fracture_anytype / clb;
run;

/* Create dummy variables for bmicat2 variable (3 categories bmi) Note this
changes the copy of the dataset
that is on my H drive */

proc means data=jason.hemo;

```

```

var bmi_sas ;
run;

data jason.hemo;
set jason.hemo;
x1=0; x2=0;
if bmicat2=2 then x1=1;
else if bmicat2=3 then x2=1;
run;

proc means data=jason.hemo mean;
var pai_total sum tug_avgtime;
class bmicat2;
run;

proc sort data=jason.hemo;
by bmi_sas;
run;

proc print data=jason.hemo noobs;
var bmi_sas bmicat2 x1 x2;
run;

***regression for bmicat2***;
proc reg data=jason.hemo;
model sum=x1 x2 / clb;
run;
proc reg data=jason.hemo;
model tug_avgtime=x1 x2 / clb;
run;
proc reg data=jason.hemo;
model pai_total=x1 x2 / clb;
run;

proc gplot data=jason.hemo;
plot (pai_total sum tug_avgtime)*bmi_sas=bmicat2;
run;

**joint bleeds***;

proc sort data=jason.hemo;
by id;
run;

proc print data=jason.hemo noobs;
var id b4 ge4bleeds_rshoulder ge4bleeds_lshoulder ge4bleeds_relbow
ge4bleeds_lelbow /* upper extremity */
ge4bleeds_rhip ge4bleeds_lhip ge4bleeds_rknee ge4bleeds_lknee
ge4bleeds_rankle ge4bleeds_lankle /* lower extremity */;
run;

proc print data=jason.hemo noobs;
var id b20 ge20bleeds_rshoulder ge20bleeds_lshoulder ge20bleeds_relbow
ge20bleeds_lelbow /* upper extremity */;

```

```

ge20bleeds_rhip ge20bleeds_lhip ge20bleeds_rknee ge20bleeds_lknee
ge20bleeds_rankle ge20bleeds_lankle /* lower extremity */;
run;

proc freq data=jason.hemo;
table ge4bleeds_any ge20bleeds_any;
run;

data jason.hemo;
set jason.hemo;
b4=0; b20=0;
if ge4bleeds_any="YES" then b4=1;
if ge20bleeds_any="YES" then b20=1;
run;

proc reg data=jason.hemo;
model sum=b4 / clb;
run;
proc reg data=jason.hemo;
model tug_avgtime=b4 / clb;
run;
proc reg data=jason.hemo;
model pai_total=b4 / clb;
run;

proc reg data=jason.hemo;
model sum=b20 / clb;
run;
proc reg data=jason.hemo;
model tug_avgtime=b20 / clb;
run;
proc reg data=jason.hemo;
model pai_total=b20 / clb;
run;

***Creat bleed_new variable ***;
data jason.hemo;
set jason.hemo;
if sum(ge4bleeds_rshoulder, ge20bleeds_rshoulder) >0 then bleed_rshoulder =
1; else bleed_rshoulder = 0;
if sum(ge4bleeds_lshoulder, ge20bleeds_lshoulder) >0 then bleed_lshoulder =
1; else bleed_lshoulder = 0;
if sum(ge4bleeds_relbow, ge20bleeds_relbow) >0 then bleed_relbow = 1;
else bleed_relbow = 0;
if sum(ge4bleeds_lelbow, ge20bleeds_lelbow) >0 then bleed_lelbow = 1;
else bleed_lelbow = 0;
if sum(ge4bleeds_rhip, ge20bleeds_rhip) >0 then bleed_rhip = 1;
else bleed_rhip = 0;
if sum(ge4bleeds_lhip, ge20bleeds_lhip) >0 then bleed_lhip = 1;
else bleed_lhip = 0;
if sum(ge4bleeds_rknee, ge20bleeds_rknee) >0 then bleed_rknee = 1;
else bleed_rknee= 0;
if sum(ge4bleeds_lknee, ge20bleeds_lknee) >0 then bleed_lknee = 1;
else bleed_lknee = 0;
if sum(ge4bleeds_rankle, ge20bleeds_rankle) >0 then bleed_rankle = 1;
else bleed_rankle= 0;

```



```

if sum(ge4bleeds_lankle,    ge20bleeds_lankle) >0    then bleed_lankle = 1;
else bleed_lankle = 0;

bleed_new = sum( bleed_rshoulder, bleed_lshoulder, bleed_relbow,
bleed_lelbow, bleed_rhip, bleed_lhip, bleed_rknee, bleed_lknee, bleed_rankle,
bleed_lankle) ;

run;

proc freq data=jason.hemo;
table bleed_new / nocum;
run;

proc print data=jason.hemo noobs;
var id bleed_rshoulder bleed_lshoulder bleed_relbow bleed_lelbow bleed_rhip
bleed_lhip bleed_rknee bleed_lknee bleed_rankle bleed_lankle bleed_new;
run;

***disease severity***;

data jason.hemo;
set jason.hemo;
if baseline_leveln<1    then baseline_cat='1';
else if 1<=baseline_leveln<5 then baseline_cat='2';
else if baseline_leveln>=5    then baseline_cat='3';
run;

proc freq data=jason.hemo;
table baseline_cat / nocum;
run;

proc freq data=jason.hemo;
table dz_severity / nocum; ***this is the original variable in the dataset
but 5
is included in the middle category. ***;
run;

data jason.hemo;
set jason.hemo;
z1=0; z2=0;
if baseline_cat=1 then z2=1;
else if baseline_cat=2 then z1=1;
run;

proc means data=jason.hemo mean;
var pai_total sum tug_avgtime;
class baseline_cat;
run;

proc sort data=jason.hemo;
by baseline_leveln;
run;

proc print data=jason.hemo noobs;
var baseline_leveln baseline_cat z1 z2;

```

```

run;

***regression for baseline_cat***;
proc reg data=jason.hemo;
model tug_avgtime=z1 z2 / clb;
run;
proc reg data=jason.hemo;
model sum=z1 z2 / clb;
run;
proc reg data=jason.hemo;
model pai_total=z1 z2 / clb;
run;

***race white vs nonwhite***;
proc freq data=jason.hemo;
table race / nocum;
run;

data jason.hemo;
set jason.hemo;
if race=2 then white_non=0;
else white_non=1;
run;

proc freq data=jason.hemo;
table white_non / nocum;
run;

proc reg data=jason.hemo;
model sum=white_non/ clb;
run;
proc reg data=jason.hemo;
model tug_avgtime=white_non/ clb;
run;
proc reg data=jason.hemo;
model pai_total=white_non/ clb;
run;

/* Get the means */

proc means data=jason.hemo mean;
var pai_total sum tug_avgtime;
class joint_procedure_any ;
run;

proc means data=jason.hemo mean;
class joint_procedure_any ;
var pai_total sum tug_avgtime;
run;

```

```

proc means data=jason.hemo mean;
class hivpos ;
var pai_total sum tug_avgtime;
run;

/* Look at histograms */

PROC UNIVARIATE DAta=jason.hemo noprint;
VAR wfhscore;
*CLASS sex;
HISTOGRAM wfhscore;
INSET MEAN median n std;
RUN;

PROC UNIVARIATE DAta=jason.hemo noprint;
VAR petterson;
*CLASS sex;
HISTOGRAM petterson;
INSET MEAN median n std;
RUN;

/* Another look at Tug */

proc freq data=jason.hemo;
table tug_avgtime / nopercnt;
run;

proc means data=jason.hemo  N MEAN STD MIN Q1 MEDIAN Q3 MAX MAXDEC=2;
var tug_avgtime;
run;

PROC UNIVARIATE DAta=jason.hemo noprint;
VAR tug_avgtime;
HISTOGRAM tug_avgtime;
INSET q1 median q3 ;
RUN;

*****
IMPUTATION
*****;

/* Missing Data Stuff */

proc sort data=jason.hemo;
by id;
run;

proc print data=jason.hemo noobs;
var id wfhscore /*petterson hepc_viralload*/;
run;

```

```

***we observe joint procedure highly correlated to whether joint score is
missing***;
proc print data=jason.hemo;
var joint_procedure_any wfhscore petterson;
run;

proc print data=jason.outmi (obs=100);
var _imputation_ petterson petterson_any;
run;

proc freq data=jason.hemo;
tables wfhscore*petterson*hepc_viralload /list missing nocum nopercnt;
run;
proc freq data=jason.outmi;
tables petterson*petterson_any/list missing nocum nopercnt;
run;

proc freq data=jason.hemo;
table petterson / nocum;
run;

proc print data=jason.hemo;
var id petterson petterson_any;
run;

/*Create training dataset */

data jason.hemo;
set jason.hemo;
select=ranuni(0);
run;

proc sort data=jason.hemo;
by select;
run;

proc print data =jason.hemo ;
var id select ;
run;

data jason.training ;
set jason.hemo ;
if _n_ < 61;
run;

proc sort data=jason.training;
by id;
run;

proc print data=jason.training;
var id ;
run;
***We decided not to use the training dataset (obs=60) after all***;

```

```

/* Create the imputed datasets */

proc mi data=jason.hemo seed=0 out=outmi nimpute=20; ***make 20 imputed
datasets from original dataset***
***throw in everything for imputation purposes;
var wfhscore petterson age joint_procedure_any hivpos hepc_antibody
hepc_viralload chronicliver
fracture_anytype bleed_new white_non baseline_leveln bmi_sas
/*wfh omit left elbow*/
relbow_totjointscore rknee_totjointscore lknee_totjointscore
rankle_totjointscore lankle_totjointscore
/*petterson omit left elbow*/
relbow_totalradioscore rknee_totalradioscore lknee_totalradioscore
rankle_totalradioscore lankle_totalradioscore
;
run;

***make dataset permanent***;
data jason.outmi;
set outmi;
run;

proc print data=jason.outmi(obs=100);
var _imputation_ id wfhscore petterson /*hepc_viralload*/ ;
*where id='047';
run;

proc means data=jason.outmi mean stderr ;
class _imputation_ ;
var wfhscore petterson /*hepc_viralload*/;
run;

*proc contents data=jason.outmi;
*run;

proc gplot data=jason.outmi;
plot petterson*petterson_any;
*plot wfhscore*petterson;
run;

*****
SLR Analysis - Outcome HAL (SUM)
*****;

/* need to do multiple imputation for wfhscore due to missing values */

ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;

```

```

by _imputation_;
model sum= /* outcome HAL (SUM) */ wfhscore ;
run;
quit;

ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=86;
modeleffects intercept wfhscore ;
run;
quit;
ods select all;

proc print data=pet;
var parm estimate LCLMean UCLMean probt;
run;

proc means data=fitstat mean;
where label2= "R-Square";
var nvalue2;
run;

/* Other variables with proc reg (change variable each time) */

ods output FitStatistics = fitstat;
proc reg data=jason.hemo ;
model sum= /* outcome HAL (SUM) */ bmi_sas / clb ;
run;
quit;

*****
Model Selection - Outcome SUM (HAL)
*****;

PROC UNIVARIATE DATA=jason.hemo /*noprint*/ plots ;
VAR sum;
*CLASS sex;
HISTOGRAM sum;
INSET MEAN q1 mean median q3;
RUN;

proc freq data=jason.hemo;
table sum / nocum;
run;

proc sort data=jason.outmi;
by _imputation_ id;
run;

*data jason.outmi;
*set jason.outmi;
*sumsq=sum*sum;

```

```

*run;

/* Example of how to do weighted least squares */

proc reg data=a1;
model diast=age;
output out=a2 r=resid;

data a2; set a2;
absr=abs(resid); sqrr=resid*resid;

proc gplot data=a2;
plot (resid absr sqrr)*age;

proc reg data=a2;
model absr=age;
output out=a3 p=shat;

data a3; set a3;
wt=1/(shat*shat);

proc reg data=a3;
model diast=age / clb;
weight wt;
run;

*****;

***Doing WEIGHTED LEAST SQUARES for heteroskedasticity ***;

/* Do mlr for FULL MODEL (maximum model) on each imputed dataset
   Also doing regression diagnostics on full model */

*****;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi covout;
by _imputation_;
model sum= /* outcome SUM */
          wfhscore /*pettersen*/ age joint_procedure_any hivpos
chronicliver fracture_anytype bleed_new white_non baseline_leveln bmi_sas
          /*agewfh agebmi wfhbmi /*interaction terms */
/partial vif collinoint ;
*interaction: test agewfh=agebmi=wfhbmi=0; ***chunk test for interaction
terms***;
output out=resid rstudent=rstudents ;
run;
quit;
ods select all;
***chunk test for interaction terms not significant, drop interaction terms
and proceed. ***;

data resid;

```

```

set resid;
absr=abs(rstudents);
sqrr=rstudents*rstudents;

ods select none;
ods output FitStatistics = fitstat;
proc reg data=resid covout;
by _imputation_;
model absr= /* outcome SUM */
          wfhscore /*petterson*/ age joint_procedure_any hivpos
chronicliver fracture_anytype bleed_new white_non baseline_leveln bmi_sas
/partial vif collinoint ;
output out=resid2 p=shat ;
run;
quit;
ods select all;

data resid2;
set resid2;
wt=1/(shat*shat);
*****;

*ods select none;
ods output FitStatistics = fitstat;
proc reg data=resid2 outest=outreg covout;
by _imputation_;
model sum= /* outcome SUM */
          wfhscore /*petterson*/ age joint_procedure_any hivpos
chronicliver fracture_anytype bleed_new white_non baseline_leveln bmi_sas
/partial vif collinoint /*clb*/;
output out=resid3 rstudent=rstudents2 cookd=cooks h=leverage Dffits=Dffits;
weight wt;
run;
quit;
*ods select all;

***look at normality of studentized residuals***;
proc univariate plot data=resid3 normal ;
by _imputation_ ;
where _imputation_ in (20);
var rstudents2;
run;

***look at studentized residuals and outlier criteria ***;
proc print data=resid3 noobs;
by _imputation_ ;
where _imputation_ in (20);
var id sum rstudents2 cooks leverage dffits;
run;

*** getting average R-square ***;
proc means data=fitstat min max mean;
where label2= "R-Square";
var nvalue2;

```



```

run;

*** mi analyze to do parameter estimation from aggregated results***;
proc mianalyze data=outreg edf=77; ***edf = complete data ddf = 88- n
parameters estimated ***;
modeleffects intercept wfhscore /*petterson*/ age joint_procedure_any hivpos
chronicliver fracture_anytype bleed_new white_non baseline_leveln bmi_sas
/*agewfh agebmi wfhbmi /*interaction terms */;
run;

/* NEW PLAN - Doing ordinal LOGISTIC regression based on (rough) quartiles
of HAL outcome */

***checking if scores above 85 make a difference ***;
data jason.hemo;
set jason.hemo;
if sum = 100 then sumtest=1;
else if sum > 84 then sumtest =0;
else sumtest = . ;
run;

proc freq data=jason.hemo;
table sumtest / nocum;
run;

proc ttest data=jason.hemo;
where sumtest ne . ;
var wfhscore ;
class sumtest;
run;

SYMBOL INTERPOL=r1 VALUE=dot COLOR=green;
SYMBOL2 VALUE=plus COLOR=blue;
SYMBOL3 VALUE=plus COLOR=red;
PROC GPLOT DATA=jason.hemo ;
PLOT sum*wfhscore = sumtest;
RUN;

SYMBOL INTERPOL=r1 VALUE=dot COLOR=green;
SYMBOL2 VALUE=plus COLOR=blue;
SYMBOL3 VALUE=plus COLOR=red;
PROC GPLOT DATA=jason.hemo ;
PLOT logsum*wfhscore = sumtest; /*try log transformation */
RUN;

***Create ordinal variable for sum (HAL) based on quartiles***;
data jason.hemo;
set jason.hemo;
if sum < 54 then sumorder=1;

```

```

else if sum < 77 then sumorder=2;
else if sum < 95 then sumorder= 3;
else sumorder= 4;
run;

proc freq data=jason.hemo;
table sumorder / nocum;
run;

data jason.outmi;
set jason.outmi;
if sum < 54 then sumorder=1;
else if sum < 77 then sumorder=2;
else if sum < 95 then sumorder= 3;
else sumorder= 4;
run;

proc freq data=jason.outmi;
table sumorder* _imputation_ / nocum norow nopercnt nocol;
run;

/* Do mlr for FULL MODEL (maximum model) on each imputed dataset
   Also doing regression diagnostics on full model */

ods select none;
ods output FitStatistics = fitstat;
proc logistic data=jason.outmi outest=outreg covout plots=ALL ;
by _imputation_;
model sumorder = /* outcome HAL - Note using ordinal version based on
quartiles*/
    wfhscore /*petterson*/ age joint_procedure_any hivpos
chronicliver fracture_anytype bleed_new white_non baseline_leveln bmi_sas
    /*agewfh agebmi wfhbmi /*interaction terms */ / rsq ;
/*interaction: test agewfh=agebmi=wfhbmi=0*/; ***chunk test for
interaction***;
output out=resid    pred=p ;
run;
ods select all;
***chunk test for interaction not significant, no interaction ***;

*proc print data=fitstat noobs;
*where _imputation_ in (20);
*run;

***look at diagnostics for cumulative probability model, per Paul Allison
page 142 ***;
proc print data=resid noobs;
by _imputation_ ;
where _imputation_ in (20);
var id sumorder /*note using ordinal version of sum*/ _level_ p ;
run;

```

```

proc sort data=resid;
by sumorder;
run;

SYMBOL /*INTERPOL=r1*/ VALUE=dot COLOR=green;
PROC GPLOT DATA=resid ;
where _imputation_ in (20);
by _level_ ;
PLOT p*sumorder ;
RUN;

PROC UNIVARIATE DATA=resid noprint;
*where _imputation_ in (20);
VAR p;
CLASS _level_;
by sumorder;
HISTOGRAM p;
INSET MEAN median n ;
RUN;

proc freq data=resid;
table sumorder*_level_ / norow nocol nopercnt nocum;
run;

*** mi analyze to do parameter estimation from aggregated results***;
ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=77 ;
modeleffects /*intercept*/ wfhscore /*petterson*/ age joint_procedure_any
hivpos chronicliver fracture_anytype bleed_new white_non baseline_leveln
bmi_sas
/*agewfh agebmi wfhbmi /* interaction terms */;
run;
quit;

data pet;
set pet;
odds_ratio=exp(estimate);
run;

proc print data=pet;
var parm estimate odds_ratio StdErr probt;
run;

/* Now proceed with model selection in a backward fashion */

*** 1. removing bmi_sas (highest p-value) ***;

ods select none;
ods output FitStatistics = fitstat;
proc logistic data=jason.outmi outest=outreg covout plots=ALL ;
by _imputation_;

```

```

model sumorder = /* outcome HAL - Note using ordinal version based on
quartiles*/
                    wfhscore /*petterson*/ age joint_procedure_any
hivpos chronicliver fracture_anytype bleed_new white_non baseline_leveln
/*bmi_sas*/
/ rsq ;
run;
ods select all;

*** mi analyze to do parameter estimation from aggregated results***;
ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=78 ;
modeleffects /*intercept*/ wfhscore /*petterson*/ age joint_procedure_any
hivpos chronicliver fracture_anytype bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
quit;

data pet;
set pet;
odds_ratio=exp(estimate);
run;

proc print data=pet;
var parm estimate odds_ratio StdErr probt;
run;
***no confounding, drop bmi_sas***;

***2. removing age (highest p-value) ***;

ods select none;
ods output FitStatistics = fitstat;
proc logistic data=jason.outmi outest=outreg covout plots=ALL ;
by _imputation_;
model sumorder = /* outcome HAL - Note using ordinal version based on
quartiles*/
                    wfhscore /*petterson age */ joint_procedure_any
hivpos chronicliver fracture_anytype bleed_new white_non baseline_leveln
/*bmi_sas*/
/ rsq ;
run;
ods select all;

*** mi analyze to do parameter estimation from aggregated results***;
ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=79 ;
modeleffects /*intercept*/ wfhscore /*petterson age */ joint_procedure_any
hivpos chronicliver fracture_anytype bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
quit;

data pet;

```

```

set pet;
odds_ratio=exp(estimate);
run;

proc print data=pet;
var parm estimate odds_ratio StdErr probt;
run;
***no confounding, drop age***;

***3. removing hivpos (highest p-value) ***;

ods select none;
ods output FitStatistics = fitstat;
proc logistic data=jason.outmi outest=outreg covout plots=ALL ;
by _imputation_;
model sumorder = /* outcome HAL - Note using ordinal version based on
quartiles*/
                wfhscore /*petterson age */ joint_procedure_any
/*hivpos*/ chronicliver fracture_anytype bleed_new white_non baseline_leveln
/*bmi_sas*/
/ rsq ;
run;
ods select all;

*** mi analyze to do parameter estimation from aggregated results***;
ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=80 ;
modeleffects /*intercept*/ wfhscore /*petterson age */ joint_procedure_any
/*hivpos*/ chronicliver fracture_anytype bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
quit;

data pet;
set pet;
odds_ratio=exp(estimate);
run;

proc print data=pet;
var parm estimate odds_ratio StdErr probt;
run;
***no confounding, drop hivpos***;

***4. removing baseline_leveln (highest p-value) ***;

ods select none;
ods output FitStatistics = fitstat;
proc logistic data=jason.outmi outest=outreg covout plots=ALL ;
by _imputation_;

```

```

model sumorder = /* outcome HAL - Note using ordinal version based on
quartiles*/
                    wfhscore /*petterson age */ joint_procedure_any
/*hivpos*/ chronicliver fracture_anytype bleed_new white_non /*
baseline_leveln bmi_sas*/
/ rsq ;
run;
ods select all;

*** mi analyze to do parameter estimation from aggregated results***;
ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=81 ;
modeleffects /*intercept*/ wfhscore /*petterson age */ joint_procedure_any
/*hivpos*/ chronicliver fracture_anytype bleed_new white_non /*
baseline_leveln bmi_sas*/;
run;
quit;

data pet;
set pet;
odds_ratio=exp(estimate);
run;

proc print data=pet;
var parm estimate odds_ratio StdErr probt;
run;
***no confounding, drop baseline_leveln *** ;

***5. removing fracture_anytype (highest p-value) ***;

ods select none;
ods output FitStatistics = fitstat;
proc logistic data=jason.outmi outest=outreg covout plots=ALL ;
by _imputation_;
model sumorder = /* outcome HAL - Note using ordinal version based on
quartiles*/
                    wfhscore /*petterson age */ joint_procedure_any
/*hivpos*/ chronicliver /*fracture_anytype*/ bleed_new white_non /*
baseline_leveln bmi_sas*/
/ rsq ;
run;
ods select all;

*** mi analyze to do parameter estimation from aggregated results***;
ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=82 ;
modeleffects /*intercept*/ wfhscore /*petterson age */ joint_procedure_any
/*hivpos*/ chronicliver /*fracture_anytype*/ bleed_new white_non /*
baseline_leveln bmi_sas*/;
run;
quit;

```

```

data pet;
set pet;
odds_ratio=exp(estimate);
run;

proc print data=pet;
var parm estimate odds_ratio StdErr probt;
run;
***no confounding , drop fracture_anytype ***;

***6. removing joint_procedure_any (highest p-value) ***;

ods select none;
ods output FitStatistics = fitstat;
proc logistic data=jason.outmi outest=outreg covout plots=ALL ;
by _imputation_;
model sumorder = /* outcome HAL - Note using ordinal version based on
quartiles*/
                    wfhscore /*petterson age joint_procedure_any
hivpos*/ chronicliver /*fracture_anytype*/ bleed_new white_non /*
baseline_leveln bmi_sas*/
/ rsq ;
output out=resid    pred=p ;
run;
ods select all;

***look at diagnostics for cumulative probability model, per Paul Allison
page 142 ***;
proc print data=resid noobs;
by _imputation_ ;
where _imputation_ in (20);
var id sumorder /*note using ordinal version of sum*/ _level_ p ;
run;

proc sort data=resid;
by _level_;
run;

PROC UNIVARIATE DATA=resid noprint;
*where _imputation_ in (20);
VAR p;
CLASS sumorder;
by _level_;
HISTOGRAM p;
INSET MEAN median n ;
RUN;

*** mi analyze to do parameter estimation from aggregated results***;
ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=83 ;

```

```

modeleffects /*intercept*/ wfhscore /*petterson age joint_procedure_any
hivpos*/ chronicliver /*fracture_anytype*/ bleed_new white_non /*
baseline_leveln bmi_sas*/;
run;
quit;

data pet;
set pet;
odds_ratio=exp(estimate);
L_odds=exp(LCLMean); U_odds=exp( UCLMean) ;
run;

proc print data=pet;
var parm estimate odds_ratio L_odds U_odds StdErr probt;
run;
*** no confounding, drop joint_procedure_any***;

***7. removing white_non (highest p-value) ***;

ods select none;
ods output FitStatistics = fitstat;
proc logistic data=jason.outmi outest=outreg covout plots=ALL ;
by _imputation_;
model sumorder = /* outcome HAL - Note using ordinal version based on
quartiles*/
wfhscore /*petterson age joint_procedure_any
hivpos*/ chronicliver /*fracture_anytype*/ bleed_new /* white_non
baseline_leveln bmi_sas*/
/ rsq ;
run;
ods select all;

*** mi analyze to do parameter estimation from aggregated results***;
ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=84 ;
modeleffects /*intercept*/ wfhscore /*petterson age joint_procedure_any
hivpos*/ chronicliver /*fracture_anytype*/ bleed_new /* white_non
baseline_leveln bmi_sas*/;
run;
quit;

data pet;
set pet;
odds_ratio=exp(estimate);
run;

proc print data=pet;
var parm estimate odds_ratio StdErr probt;
run;
***confounds chronicliver and causes wfhscore to be non-significant.
put back white_non. ***;

```



```

/* STOP
   Final model is 6. */

/* NEW PLAN AGAIN
   Going back to treating HAL as a continuous outcome, same as PAI and TUG

Do mlr for FULL MODEL (maximum model) on each imputed dataset
   Also doing regression diagnostics on full model */

*ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model sum= /* outcome HAL (SUM) */
          wfhscore /*petterson*/ age joint_procedure_any hivpos
chronicliver fracture_anytype bleed_new white_non baseline_leveln bmi_sas
          /*agewfh agebmi wfhbmi /*interaction terms */
/ partial vif collinoint;
/*interaction: test agewfh=agebmi=wfhbmi=0; ***chunk test for
interaction***;*/
output out=resid rstudent=rstudents cookd=cooks h=leverage Dffits=Dffits;
run;
quit;
*ods select all;
***chunk test for interaction terms not significant, drop interaction terms
and proceed***;

***look at normality of studentized residuals***;
proc univariate plot data=resid normal ;
by _imputation_ ;
where _imputation_ in (20);
var rstudents;
run;

***look at studentized residuals and outlier criteria ***;
proc print data=resid noobs;
by _imputation_ ;
where _imputation_ in (20);
var id logpai /*note using LOG PAI*/ rstudents cooks leverage dffits;
run;

*** getting average R-square ***;
proc means data=fitstat min max mean;
where label2= "R-Square";
var nvalue2;
run;

*** mi analyze to do parameter estimation from aggregated results***;
ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=77 ;

```

```

modeleffects intercept wfhscore /*petterson*/ age joint_procedure_any hivpos
chronicliver fracture_anytype bleed_new white_non baseline_leveln bmi_sas
/*agewfh agebmi wfhbmi /* interaction terms */;
run;
quit;

proc print data=pet;
var parm estimate LCLMean UCLMean probt;
run;

/* Now proceed with model selection doing backward elimination */

*** 1. removing joint_procedure_any (highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model sum= /* outcome HAL (SUM) */
           wfhscore /*petterson*/ age /*joint_procedure_any*/
hivpos chronicliver fracture_anytype bleed_new white_non baseline_leveln
bmi_sas;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

*** mi analyze to do parameter estimation from aggregated results***;
ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=78 ;
modeleffects intercept wfhscore /*petterson*/ age /*joint_procedure_any*/
hivpos chronicliver fracture_anytype bleed_new white_non baseline_leveln
bmi_sas;
run;
quit;

proc print data=pet;
var parm estimate LCLMean UCLMean probt;
run;
***leave out joint_procedure_any***;

*** 2. removing bmi_sas (highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model sum= /* outcome HAL (SUM) */
           wfhscore /*petterson*/ age /*joint_procedure_any*/
hivpos chronicliver fracture_anytype bleed_new white_non baseline_leveln
/*bmi_sas*/;

```

```

run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

*** mi analyze to do parameter estimation from aggregated results***;
ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=79 ;
modeleffects intercept wfhscore /*petterson*/ age /*joint_procedure_any*/
hivpos chronicliver fracture_anytype bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
quit;

proc print data=pet;
var parm estimate LCLMean UCLMean probt;
run;
***leave out bmi_sas***;

*** 3. removing fracture_anytype (highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model sum= /* outcome HAL (SUM) */
wfhscore /*petterson*/ age /*joint_procedure_any*/
hivpos chronicliver /*fracture_anytype*/ bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

*** mi analyze to do parameter estimation from aggregated results***;
ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=80 ;
modeleffects intercept wfhscore /*petterson*/ age /*joint_procedure_any*/
hivpos chronicliver /*fracture_anytype*/ bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
quit;

proc print data=pet;
var parm estimate LCLMean UCLMean probt;
run;
***leave out fracture_anytype**;
```

```

*** 4. removing hivpos (highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model sum= /* outcome HAL (SUM) */
           wfhscore /*petterson*/ age /*joint_procedure_any
hivpos*/ chronicliver /*fracture_anytype*/ bleed_new white_non
baseline_leveln /*bmi_sas*/;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

*** mi analyze to do parameter estimation from aggregated results***;
ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=81 ;
modeleffects intercept wfhscore /*petterson*/ age /*joint_procedure_any
hivpos*/ chronicliver /*fracture_anytype*/ bleed_new white_non
baseline_leveln /*bmi_sas*/;
run;
quit;

proc print data=pet;
var parm estimate LCLMean UCLMean probt;
run;
***leave out hivpos***;

*** 5. removing baseline_leveln (highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model sum= /* outcome HAL (SUM) */
           wfhscore /*petterson*/ age /*joint_procedure_any
hivpos*/ chronicliver /*fracture_anytype*/ bleed_new white_non
/*baseline_leveln bmi_sas*/;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

*** mi analyze to do parameter estimation from aggregated results***;
ods output ParameterEstimates= pet;

```

```

proc mianalyze data=outreg edf=82 ;
modeleffects intercept wfhscore /*petterson*/ age /*joint_procedure_any
hivpos*/ chronicliver /*fracture_anytype*/ bleed_new white_non
/*baseline_leveln bmi_sas*/;
run;
quit;

proc print data=pet;
var parm estimate LCLMean UCLMean probt;
run;
***confounding, especially age (.35), leave in baseline_leveln***;

*** 6. removing white_non (next highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model sum= /* outcome HAL (SUM) */
wfhscore /*petterson*/ age /*joint_procedure_any
hivpos*/ chronicliver /*fracture_anytype*/ bleed_new /*white_non*/
baseline_leveln /*bmi_sas*/;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

*** mi analyze to do parameter estimation from aggregated results***;
ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=82;
modeleffects intercept wfhscore /*petterson*/ age /*joint_procedure_any
hivpos*/ chronicliver /*fracture_anytype*/ bleed_new /*white_non*/
baseline_leveln /*bmi_sas*/;
run;
quit;

proc print data=pet;
var parm estimate LCLMean UCLMean probt;
run;
***confounding, leave in white_non***;

*** 7. removing age (next highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model sum= /* outcome HAL (SUM) */

```

```

                                wfhscore /*petterson age joint_procedure_any hivpos*/
chronicliver /*fracture_anytype*/ bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

*** mi analyze to do parameter estimation from aggregated results***;
ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=82;
modeleffects intercept wfhscore /*petterson age joint_procedure_any hivpos*/
chronicliver /*fracture_anytype*/ bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
quit;

proc print data=pet;
var parm estimate LCLMean UCLMean probt;
run;

*** 8. try dropping baseline_leveln, white_non, and age all at once ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model sum= /* outcome HAL (SUM) */
                                wfhscore /*petterson age joint_procedure_any hivpos*/
chronicliver /*fracture_anytype*/ bleed_new /*white_non baseline_leveln
bmi_sas*/
/ partial vif collinoint;
output out=resid rstudent=rstudents cookd=cooks h=leverage Dffits=Dffits;
run;
quit;
ods select all;

*** getting average R-square ***;
proc means data=fitstat min max mean;
where label2= "Adj R-Sq";
var nvalue2;
run;

proc print data=fitstat;
run;

***look at normality of studentized residuals***;
proc univariate plot data=resid normal ;
by _imputation_ ;
where _imputation_ in (20);

```

```

var rstudents;
run;

***look at studentized residuals and outlier criteria ***;
proc print data=resid noobs;
by _imputation_ ;
where _imputation_ in (20);
var id logpai /*note using LOG PAI*/  rstudents cooks leverage dffits;
run;

*** mi analyze to do parameter estimation from aggregated results***;
ods output ParameterEstimates= pet;
proc mianalyze data=outreg  edf=84;
modeleffects intercept wfhscore /*petterson age joint_procedure_any hivpos*/
chronicliver /*fracture_anytype*/ bleed_new /*white_non baseline_leveln
bmi_sas*/;
run;
quit;

proc print data=pet;
var parm estimate LCLMean UCLMean probt;
run;

*** 9. try adding back white_non ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi  outest=outreg covout;
by _imputation_;
model sum= /* outcome HAL (SUM) */
          wfhscore /*petterson age joint_procedure_any hivpos*/
chronicliver /*fracture_anytype*/ bleed_new white_non /* baseline_leveln
bmi_sas*/;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

*** mi analyze to do parameter estimation from aggregated results***;
ods output ParameterEstimates= pet;
proc mianalyze data=outreg  edf=83;
modeleffects intercept wfhscore /*petterson age joint_procedure_any hivpos*/
chronicliver /*fracture_anytype*/ bleed_new white_non /*baseline_leveln
bmi_sas*/;
run;
quit;

proc print data=pet;
var parm estimate LCLMean UCLMean probt;

```

```

run;

*** 10. try adding back baseline_leveln ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model sum= /* outcome HAL (SUM) */
           wfhscore /*petterson age joint_procedure_any hivpos*/
chronicliver /*fracture_anytype*/ bleed_new /*white_non */ baseline_leveln
/*bmi_sas*/;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

*** mi analyze to do parameter estimation from aggregated results***;
ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=83;
modeleffects intercept wfhscore /*petterson age joint_procedure_any hivpos*/
chronicliver /*fracture_anytype*/ bleed_new /*white_non *//baseline_leveln /*
bmi_sas*/;
run;
quit;

proc print data=pet;
var parm estimate LCLMean UCLMean probt;
run;

*** 11. try adding back age ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model sum= /* outcome HAL (SUM) */
           wfhscore /*petterson *//age /*joint_procedure_any
hivpos*/ chronicliver /*fracture_anytype*/ bleed_new /*white_non
baseline_leveln bmi_sas*/;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

*** mi analyze to do parameter estimation from aggregated results***;

```



```

ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=83;
modeleffects intercept wfhscore /*petterson */age /*joint_procedure_any
hivpos*/ chronicliver /*fracture_anytype*/ bleed_new /*white_non
baseline_leveln bmi_sas*/;
run;
quit;

proc print data=pet;
var parm estimate LCLMean UCLMean probt;
run;

/* Final Model is 8. */

*****
SLR Analysis - Outcome PAI (Framingham)
*****;

/* need to do multiple imputation for wfhscore due to missing values */

ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model logpai= /* outcome LOG PAI */ wfhscore ;
run;
quit;

ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=86;
modeleffects intercept wfhscore ;
run;
quit;
ods select all;

proc print data=pet;
var parm estimate LCLMean UCLMean probt;
run;

proc means data=fitstat mean;
where label2= "R-Square";
var nvalue2;
run;

```

```

/* Other variables with proc reg (change variable each time) */

ods output FitStatistics = fitstat;
proc reg data=jason.hemo ;
model logpai= /* outcome LOG PAI */ chronicliver / clb ;
run;
quit;

*****
Model Selection - Outcome PAI (Framingham)
*****;

/* Do mlr for FULL MODEL (maximum model) on each imputed dataset
   Also doing regression diagnostics on full model */

ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model logpai= /* outcome PAI - note changed to LOG PAI for normality*/
              wfhscore /*pettersen*/ age joint_procedure_any hivpos
              chronicliver fracture_anytype bleed_new white_non baseline_leveln bmi_sas
              /*agewfh agebmi wfhbmi /*interaction terms */
/ partial vif collinoint;
*where id ne "049"; ***exclude outlier id = 49 ***;
*interaction: test agewfh=agebmi=wfhbmi=0; ***chunk test for interaction***;
output out=resid rstudent=rstudents cookd=cooks h=leverage Dffits=Dffits;
run;
quit;
ods select all;
***chunk test for interaction terms not significant, drop interaction terms
and proceed***;

***look at normality of studentized residuals***;
proc univariate plot data=resid normal ;
by _imputation_ ;
where _imputation_ in (20);
var rstudents;
run;

***look at studentized residuals and outlier criteria ***;
proc print data=resid noobs;
by _imputation_ ;
where _imputation_ in (20);
var id logpai /*note using LOG PAI*/ rstudents cooks leverage dffits;
run;

*** getting average R-square ***;
proc means data=fitstat min max mean;
where label2= "R-Square";
var nvalue2;

```

```

run;

*** mi analyze to do parameter estimation from aggregated results***;
proc mianalyze data=outreg edf=77 ;
modeleffects intercept wfhscore /*petterson*/ age joint_procedure_any hivpos
chronicliver fracture_anytype bleed_new white_non baseline_leveln bmi_sas
/*agewfh agebmi wfhbmi /* interaction terms */;
run;

/* Now proceed with model selection in a backward fashion */

*** 1. removing chronic_liver (highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model logpai= /* outcome PAI */
                wfhscore /*petterson*/ age joint_procedure_any hivpos
/*chronicliver*/ fracture_anytype bleed_new white_non baseline_leveln
bmi_sas;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

proc mianalyze data=outreg edf=78;
modeleffects intercept wfhscore /*petterson*/ age joint_procedure_any hivpos
/*chronicliver*/ fracture_anytype bleed_new white_non baseline_leveln
bmi_sas;
run;
***no confounding, leave out chronic_liver***;

*** 2. removing bmi_sas (highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model logpai= /* outcome PAI */
                wfhscore /*petterson*/ age joint_procedure_any hivpos
/*chronicliver*/ fracture_anytype bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";

```

```

*var nvalue2;
*run;

proc mianalyze data=outreg edf=79;
modeleffects intercept wfhscore /*petterson*/ age joint_procedure_any hivpos
/*chronicliver*/ fracture_anytype bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
***no confounding, leave out bmi_sas***;

*** 3. removing hivpos (highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model logpai= /* outcome PAI */
wfhscore /*petterson*/ age joint_procedure_any /*
hivpos chronicliver*/ fracture_anytype bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

proc mianalyze data=outreg edf=80;
modeleffects intercept wfhscore /*petterson*/ age joint_procedure_any /*
hivpos chronicliver*/ fracture_anytype bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
***no confounding, leave out hivpos***;

*** 4. removing fracture_anytype (highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model logpai= /* outcome PAI */
wfhscore /*petterson*/ age joint_procedure_any /*
hivpos chronicliver fracture_anytype*/ bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

```

```

proc mianalyze data=outreg edf=81;
modeleffects intercept wfhscore /*petterson*/ age joint_procedure_any /*
hivpos chronicliver fracture_anytype*/ bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
***no confounding, leave out fracture_anytype***;

*** 5. removing joint_procedure_any (highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model logpai= /* outcome PAI */
                wfhscore /*petterson*/ age /*joint_procedure_any
hivpos chronicliver fracture_anytype*/ bleed_new white_non baseline_leveln
/*bmi_sas*/
/ partial vif collinoint;
output out=resid rstudent=rstudents cookd=cooks h=leverage Dffits=Dffits;
run;
quit;
ods select all;

***look at normality of studentized residuals***;
proc univariate plot data=resid normal ;
by _imputation_ ;
where _imputation_ in (20);
var rstudents;
run;

***look at studentized residuals and outlier criteria ***;
proc print data=resid noobs;
by _imputation_ ;
where _imputation_ in (20);
var id logpai /*note using LOG PAI*/ rstudents cooks leverage dffits;
run;

proc means data=fitstat min max mean;
where label2= "Adj R-Sq";
var nvalue2;
run;

proc mianalyze data=outreg edf=82;
modeleffects intercept wfhscore /*petterson*/ age /*joint_procedure_any
hivpos chronicliver fracture_anytype*/ bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
***no confounding, leave out joint_procedure_any***;

*** 6. removing white_non (highest p-value) ***;
ods select none;

```

```

ods output FitStatistics = fitstat;
proc reg data=jason.outmi  outest=outreg covout;
by _imputation_;
model logpai= /* outcome PAI */
                wfhscore /*petterson*/ age /*joint_procedure_any
hivpos chronicliver fracture_anytype*/ bleed_new /*white_non*/
baseline_leveln /*bmi_sas*/;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

proc mianalyze data=outreg edf=83;
modeleffects intercept wfhscore /*petterson*/ age /*joint_procedure_any
hivpos chronicliver fracture_anytype*/ bleed_new /*white_non*/
baseline_leveln /*bmi_sas*/;
run;
***confounds wfhscore (.21) , leave in white_non for now***;

*** 7. removing bleed_new (next highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi  outest=outreg covout;
by _imputation_;
model logpai= /* outcome PAI */
                wfhscore /*petterson*/ age /*joint_procedure_any
hivpos chronicliver fracture_anytype bleed_new*/ white_non baseline_leveln
/*bmi_sas*/;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

proc mianalyze data=outreg edf=83;
modeleffects intercept wfhscore /*petterson*/ age /*joint_procedure_any
hivpos chronicliver fracture_anytype bleed_new*/ white_non baseline_leveln
/*bmi_sas*/;
run;
***confounds wfhscore (.64) AND causes wfhscore an dbaseline_leveln to go
from not significant to significant (or almost)***
    leave in bleed_new ***;

*** 8. removing baseline_leveln (next highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;

```

```

proc reg data=jason.outmi  outest=outreg covout;
by _imputation_;
model logpai= /* outcome PAI */
                wfhscore /*petterson*/ age /*joint_procedure_any
hivpos chronicliver fracture_anytype*/ bleed_new white_non /*baseline_leveln
bmi_sas*/;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

proc mianalyze data=outreg edf=83;
modeleffects intercept wfhscore /*petterson*/ age /*joint_procedure_any
hivpos chronicliver fracture_anytype*/ bleed_new white_non /*baseline_leveln
bmi_sas*/;
run;
***confounds wfhscore and bleed_new, which also becomes almost significant.
    leave in baseline_leveln (besides baseline_leveln is almost significant in
7.***;

/* STOP
    Final model is 5. */

***Extra step -
    check what happens if remove bleed_new, white_non, and baselin_leveln all
together (just leave age and wfhscore) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi  outest=outreg covout;
by _imputation_;
model logpai= /* outcome PAI */
                wfhscore /*petterson*/ age /*joint_procedure_any
hivpos chronicliver fracture_anytype bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
quit;
ods select all;

proc means data=fitstat min max mean;
where label2= "R-Square";
var nvalue2;
run;

proc mianalyze data=outreg edf=85;
modeleffects intercept wfhscore /*petterson*/ age /*joint_procedure_any
hivpos chronicliver fracture_anytype bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
***these three as a group were confounders
    put back in, final model still #5 ***;

```

```

*****
SLR Analysis - Outcome TUG
*****;

/* need to do multiple imputation for wfhscore due to missing values */

ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model logtug= /* outcome LOG TUG */ wfhscore ;
run;
quit;

ods output ParameterEstimates= pet;
proc mianalyze data=outreg edf=86;
modeleffects intercept wfhscore ;
run;
quit;
ods select all;

proc print data=pet;
var parm estimate LCLMean UCLMean probt;
run;

proc means data=fitstat mean;
where label2= "R-Square";
var nvalue2;
run;

/* Other variables with proc reg (change variable each time) */

ods output FitStatistics = fitstat;
proc reg data=jason.hemo ;
model logtug= /* outcome LOG TUG */ bmi_sas / clb ;
run;
quit;

*****
Model Selection - Outcome TUG
*****;

/* Do mlr for FULL MODEL (maximum model) on each imputed dataset
Also doing regression diagnostics on full model */

/* Create interaction terms in outmi dataset */

```



```

data jason.outmi;
set jason.outmi;
agewfh=age*wfhscore ;
agebmi=age*bmi_sas;
wfhbmi=wfhscore*bmi_sas;
run;

ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model logtug= /* outcome TUG - changed to LOG TUG to get normality*/
             wfhscore /*petterson*/ age joint_procedure_any hivpos
chronicliver fracture_anytype bleed_new white_non baseline_leveln bmi_sas
             /*agewfh agebmi wfhbmi /*interaction terms*/
/partial vif collinoit;
output out=resid rstudent=rstudents cookd=cooks h=leverage Dffits=Dffits;
/*Interaction: test agewfh=agebmi=wfhbmi=0*/;
run;
quit;
ods select all;
***chunk test for interaction is not significant, drop interaction terms and
proceed***;

***look at normality of studentized residuals***;
proc univariate plot data=resid normal ;
by _imputation_ ;
where _imputation_ in (20);
var rstudents;
run;

***look at studentized residuals and outlier criteria ***;
proc print data=resid noobs;
by _imputation_ ;
where _imputation_ in (20);
var id logtug /*note changed to logtug*/ rstudents cooks leverage dffits;
run;

*** getting average R-square ***;
proc means data=fitstat min max mean;
where label2= "R-Square";
var nvalue2;
run;

*** mi analyze to do parameter estimation from aggregated results***;
proc mianalyze data=outreg edf=77;
modeleffects intercept wfhscore /*petterson*/ age joint_procedure_any hivpos
chronicliver fracture_anytype bleed_new white_non baseline_leveln bmi_sas
/* agewfh agebmi wfhbmi /*interaction terms */ ;
run;

```

```

/* Now proceed with model selection by backward elimination */

*** 1. removing bmi_sas (highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model logtug= /* outcome TUG */
              wfhscore /*petterson*/ age joint_procedure_any hivpos
chronicliver fracture_anytype bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

proc mianalyze data=outreg edf=78;
modeleffects intercept wfhscore /*petterson*/ age joint_procedure_any hivpos
chronicliver fracture_anytype bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
***no confounding, leave out bmi_sas***;

*** 2. removing fracture_anytype (highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model logtug= /* outcome TUG */
              wfhscore /*petterson*/ age joint_procedure_any hivpos
chronicliver /*fracture_anytype*/ bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

proc mianalyze data=outreg edf=79;
modeleffects intercept wfhscore /*petterson*/ age joint_procedure_any hivpos
chronicliver /*fracture_anytype*/ bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
***no confounding, leave out fracture_anytype***;

```

```

*** 3. removing joint_procedure_any (highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model logtug= /* outcome TUG */
              wfhscore /*petterson*/ age /*joint_procedure_any*/
hivpos chronicliver /*fracture_anytype*/ bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

proc mianalyze data=outreg edf=80;
modeleffects intercept wfhscore /*petterson*/ age /*joint_procedure_any*/
hivpos chronicliver /*fracture_anytype*/ bleed_new white_non baseline_leveln
/*bmi_sas*/;
run;
***no confounding, leave out joint_procedure_any***;

*** 4. removing baseline_leveln (highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model logtug= /* outcome TUG */
              wfhscore /*petterson*/ age /*joint_procedure_any*/
hivpos chronicliver /*fracture_anytype*/ bleed_new white_non
/*baseline_leveln bmi_sas*/;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

proc mianalyze data=outreg edf=81;
modeleffects intercept wfhscore /*petterson*/ age /*joint_procedure_any*/
hivpos chronicliver /*fracture_anytype*/ bleed_new white_non
/*baseline_leveln bmi_sas*/;
run;
***no confounding, leave out baseline_leveln***;

```

```

*** 5. removing white_non (highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model logtug= /* outcome TUG */
              wfhscore /*petterson*/ age /*joint_procedure_any*/
hivpos chronicliver /*fracture_anytype*/ bleed_new /*white_non
baseline_leveln bmi_sas*/
/partial vif collinoint;
output out=resid rstudent=rstudents cookd=cooks h=leverage Dffits=Dffits;
run;
quit;
ods select all;

***look at normality of studentized residuals***;
proc univariate plot data=resid normal ;
by _imputation_ ;
where _imputation_ in (20);
var rstudents;
run;

***look at studentized residuals and outlier criteria ***;
proc print data=resid noobs;
by _imputation_ ;
where _imputation_ in (20);
var id logtug /*note changed to logtug*/ rstudents cooks leverage dffits;
run;

proc means data=fitstat min max mean;
where label2= "Adj R-Sq";
var nvalue2;
run;

proc mianalyze data=outreg edf=82;
modeleffects intercept wfhscore /*petterson*/ age /*joint_procedure_any*/
hivpos chronicliver /*fracture_anytype*/ bleed_new /*white_non
baseline_leveln bmi_sas*/;
run;
***no confounding, leave out white_non***;

*** 6. removing bleed_new (highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model logtug= /* outcome TUG */
              wfhscore /*petterson*/ age /*joint_procedure_any*/
hivpos chronicliver /*fracture_anytype*/ bleed_new white_non baseline_leveln
bmi_sas*/;
run;
quit;
ods select all;

```

```

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

proc mianalyze data=outreg edf=83;
modeleffects intercept wfhscore /*petterson*/ age /*joint_procedure_any*/
hivpos chronicliver /*fracture_anytype bleed_new white_non baseline_leveln
bmi_sas*/;
run;
***confounds wfh & chronic liver, and makes both go from not significant to
significant, especially wfh.
    put back in bleed_new. ***;

*** 7. removing hivpos (next highest p-value) ***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;
model logtug= /* outcome TUG */
                wfhscore /*petterson*/ age /*joint_procedure_any
hivpos*/ chronicliver /*fracture_anytype*/ bleed_new /*white_non
baseline_leveln bmi_sas*/;
run;
quit;
ods select all;

*proc means data=fitstat min max mean;
*where label2= "R-Square";
*var nvalue2;
*run;

proc mianalyze data=outreg edf=83;
modeleffects intercept wfhscore /*petterson*/ age /*joint_procedure_any
hivpos*/ chronicliver /*fracture_anytype*/ bleed_new /*white_non
baseline_leveln bmi_sas*/;
run;
***confounding with wfhscore and chronicliver, and causes chronic liver to be
further from significant,
    put back hivpos. ***;

/* STOP
    Final model is 5. */

*** Extra step
    Check what happens if drop hivfpos, chronicliver, and bleed_new all
together (just leave wfhscore and age)***;
ods select none;
ods output FitStatistics = fitstat;
proc reg data=jason.outmi outest=outreg covout;
by _imputation_;

```

```
model logtug= /* outcome TUG */
              wfhscore /*pettersen*/ age /*joint_procedure_any
hivpos chronicliver /*fracture_anytype bleed_new /*white_non baseline_leveln
bmi_sas*/;
run;
quit;
ods select all;

proc means data=fitstat min max mean;
where label2= "R-Square";
var nvalue2;
run;

proc mianalyze data=outreg edf=85;
modeleffects intercept wfhscore /*pettersen*/ age /*joint_procedure_any
hivpos chronicliver /*fracture_anytype bleed_new /*white_non baseline_leveln
bmi_sas*/;
run;
***these three as a group were confounders, put back in
   final model still #5. ***
```