Diabetes Epidemiological Study

Yunran Chen

Abstract

We aim to assess the relationship between glucose tolerance (GT) and insulin sensitivity (IS) with adjusting other potential risk factors for diabetes, such as age, blood pressure and obesity measures. We consider logorithm transformation on both GT and IS, and applied a linear regression. Our results suggest a significant negative relation exists between GT and IS. Specifically, with holding other potential factors as constants, we are 95% confident that if (IS + 1) increases by 10%, GT will decrease by 2.17% to 2.85% on average. Model diagnosis and sensitivity analysis are conducted which indicate our model is reasonable and robust.

1. Introduction

Several studies have shown insulin sensitivity (IS) may relate the glucose tolerance (GT)(O'rahilly et al. 1994),(Tripathy et al. 2000). We aim to assess the relationship between GT and IS and quantify such influence adjusting other potential risk factors for diabetes. We obtained data from Insulin Resistance Atherosclerosis Study (IRAS)(Wagenknecht et al. 1995), which contains 868 observations. Each observation includes GT, IS and other potential risk factors for diabetes, such as age, waist-to-hip ratio (WtoHip), body mass index (bmi), diastolic blood pressure (dbp), and systolic blood pressure (sbp).

2. Exploratory Data Analysis

We consider a logarithm transformation for GT since the value is greater than 0 and the logarithm of GT follows normal distribution approximately. Figure 1 presents the scatter plot for log(GT) versus the potential factors and the blue lines present generalized additive model results. It suggests age, waist-to-hip ratio, diastolic blood pressure and systolic blood pressure may have linear relationship to log(GT). Figure 2 suggests linear relationship exists between log(GT) and log(IS), and non-linear relationship exists between log(GT) and bmi. Figure 3 suggests a strong relation between log(IS) and log(GT). Since several predictors are relatively strong related, a check on multicolinearity and variable selection may be needed.

3. Materials and Methods

To address our interests, we consider a linear regression model to quantify the relationship between GT and IS adjusting other potential factors. Based on the EDA result, we consider logarithm transformation for both GT and IS. Since there are 117 observations with IS equaling to 0, we consider log(1+IS) transformation. We first specify a full regression model including all the covariates, then remove covariate `bmi` since the potential effect is not significant as suggested by t-test and does not bring improvement for the model fitting as suggested by BIC. We also apply stepwise regression and lasso, which also suggest removing `bmi` from the model. Our final model is as follows:

$$\log(GT_i) = \beta_0 + \beta_1 \log(1 + IS_i) + \beta_2 Age_i + \beta_3 WtoHip_i + \beta_4 dbp_i + \beta_5 sbp_i + \epsilon_i,$$

where i represent each observation, and we assume $\epsilon_i \sim N(0, \sigma^2)$ (i.i.d).

The estimation of β_1 can quantify the relationship between GT and IS adjusting other potential factors. Model diagnosis and sensitivity analysis are conducted which indicate our model is reasonable and robust.

4. Results

From our coefficient estimation shown in Table 1, IS has significant negative relationship with GT. Specifically, with holding other potential factors as constants, we are 95% confident that if (IS + 1) increases by 10%, GT will decrease by 2.17% to 2.85% on average (calculated by $1.1^{\beta_1} - 1$). Since the high value of GT indicates high risk of diabetes, larger value of IS relate to less risk of diabetes.

Table 1: Estimation for linear regression

	2.5 %	97.5 %	estimate	std.error	statistic	p.value
(Intercept)	4.2373	5.0329	4.6351	0.2027	22.8684	0.0000
Age	0.0022	0.0091	0.0057	0.0017	3.2528	0.0012
dbp	-0.0086	-0.0021	-0.0054	0.0017	-3.2076	0.0014
log(IS + 1)	-0.5411	-0.4414	-0.4912	0.0254	-19.3334	0.0000
sbp	0.0018	0.0054	0.0036	0.0009	3.9325	0.0001
WtoHip	0.2196	0.8717	0.5456	0.1661	3.2848	0.0011

The signs of the relationship (except estimation for effect of dbp) are consistent to other studies(Branchtein et al. 1997),(O'rahilly et al. 1994),(Jaffrain-Rea et al. 2001). Age, dbp, sbp and waist-to-hip ratio also have significant relationship with GT. Holding other potential factors as constant, we are 95% confident that 1 year increase in age relates to GT increases by 0.224% to 0.910%, 1 unit increase of dbp is related to GT decreases by 0.208% to 0.861% on average, 1 unit increase of sbp is related to GT increases by 0.180% to 0.541% on average, 1 unit increase of waist-to-hip ratio is related to GT increases by 24.6% to 139% on average. The model suggests negative relationship between dbp and GT holding other potential factors as constants, which may be contradict to the background knowledge. But this may be reasonable since sbp and dbp have relative strong relation and sbp show relative strong relation to GT as shown in Figure 3. With adjusting the effect of sbp, the relation between dbp and GT may appear to be negative.

Figure 5 present the diagnosis of the model.From the residual plot, we found that the residuals are roughly distributed symmetrically around zero. From the qqplot, the overall residuals satisfy the normal assumption and only several points at the right tail slightly violate the normal assumption. To better capture the potential nonlinear effect, we also consider a generalized additive model (GAM) which is presented in discussion section. Since the GAM does not bring much improvement, and also for simplicity and interpretability, we consider the linear regression as the final model.

5. Sensitivity Analysis

Based on the outlier test, there is no studentized residuals with Bonferroni $p < 0.05$. Therefore, we conducted sensitivity analysis to check the influence of the influential points only [1]. Specifically, we compare the model estimation with or without influential points (45 observations are considered as influential points). Table 2 shows the estimation for model based on data without influential points. Compare Table 1 and Table 2, the confidence intervals are similar, and only the estimation for WtoHip are relatively different. Since we only focus on quantifying the relation between IS and GT, we conclude that our model is robust.

Table 2: Estimation for linear regression based on data without influential points

	2.5 %	97.5 %	estimate	std.error	statistic	p.value
(Intercept)	4.5216	5.2445	4.8830	0.1841	26.5188	0.0000

[1] Defined as the data with cook distance greater than $4/n - p - 1$, where n is the sample size and p is the number of parameters.

	2.5 %	97.5 %	estimate	std.error	statistic	p.value
Age	0.0020	0.0081	0.0050	0.0016	3.2208	0.0013
dbp	-0.0102	-0.0041	-0.0072	0.0016	-4.6006	0.0000
log(IS + 1)	-0.5571	-0.4663	-0.5117	0.0231	-22.1103	0.0000
sbp	0.0026	0.0060	0.0043	0.0009	5.0174	0.0000
WtoHip	0.0543	0.6525	0.3534	0.1524	2.3194	0.0206

6. Discussion

6.1 GAM

We also consider a generalized additive model (GAM) to capture the potential non-linear effect. The final model is as follows:

$$\log(GT_i) = \beta_0 + f_1(log(1 + IS_i)) + f_2(Age_i) + \beta_1 WtoHip_i + \beta_2 dbp_i + \beta_3 sbp_i + \epsilon_i,$$

where f_i is a smooth function provided by R package `mgcv` by default, $\epsilon_i \sim N(0, \sigma^2)$ (i.i.d).

Notice the relationship between `bmi` and GT is not significantly non-zero and thus we exclude the term. And the `WtoHip`, `dbp` and `sbp` have linear relation with `log(GT)` as suggested by GAM. Figure 7 suggests a non-linear relationship between `log(GT)` and `log(1+IS)`. Specifically, as `log(1+IS)` increases, `log(GT)` will decrease linearly at first, but the decreasing rate will gradually slow down especially after $log(1 + IS) > 1.5$. From the diagonosis of GAM as shown in Figure 6, GAM does not bring us much improvement.

6.2 Proportional Odds Model

Since GT can also be used to diagonose type 2 diabetes with $GT < 140$, $140 \leq GT \leq 200$, $GT > 200$ indicating normal, prediabetes and diabetes respectively. Researchers may be interested in quantify the relation between the risk of diabetes and IS. In this case, we consider a proportional odds model as follows:

$$logit(\text{Diabetes}) = \beta_{01} - \beta_1 log(1 + IS) - \beta_2 dbp - \beta_3 Age - \beta_4 sbp - \beta_5 WtoHip$$

$$logit(\text{Diabetes or Prediabetes}) = \beta_{02} - \beta_1 log(1 + IS) - \beta_2 dbp - \beta_3 Age - \beta_4 sbp - \beta_5 WtoHip$$

Notice the relationship between `bmi` and GT is not significantly non-zero and thus we exclude the term. And this model is the best compared to alternative model with other transformation on covariates in terms of the AIC. Table 3 presents the estimation of the potential relationship.

From our coefficient estimation $\hat{\beta}$ shown in Table 3, IS has significant negative relationship with GT. Specifically, with holding other potential factors as constants, we are 95% confident that if (IS + 1) increases by 10%, the odds of diabetes (versus non-diabetes) and the odds of prediabetes and diabetes versus normal will decrease by 19.7% to 25.0% on average. Notice the underlying assumption for the proportional odds model is that the coefficients describing the relationship between diabetes versus non-diabetes, between diabetes and pre-diabetes versus normal are the same.

Appendix

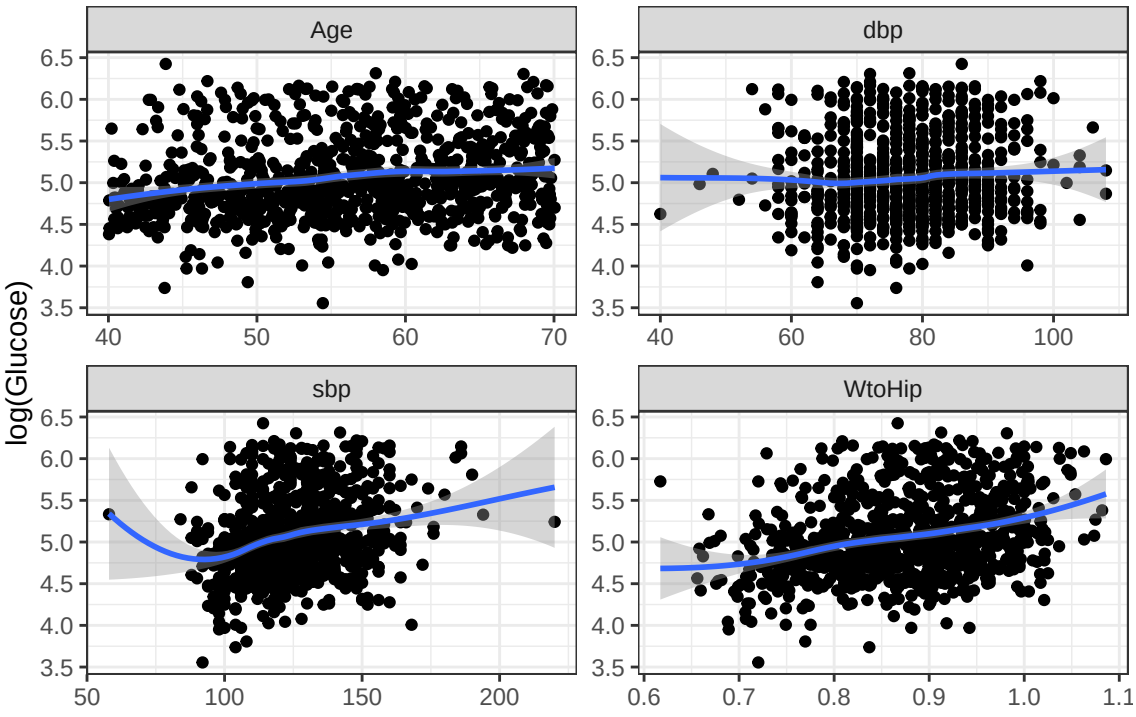

Figure 1: Linear relationship between potential factors to log(GT)

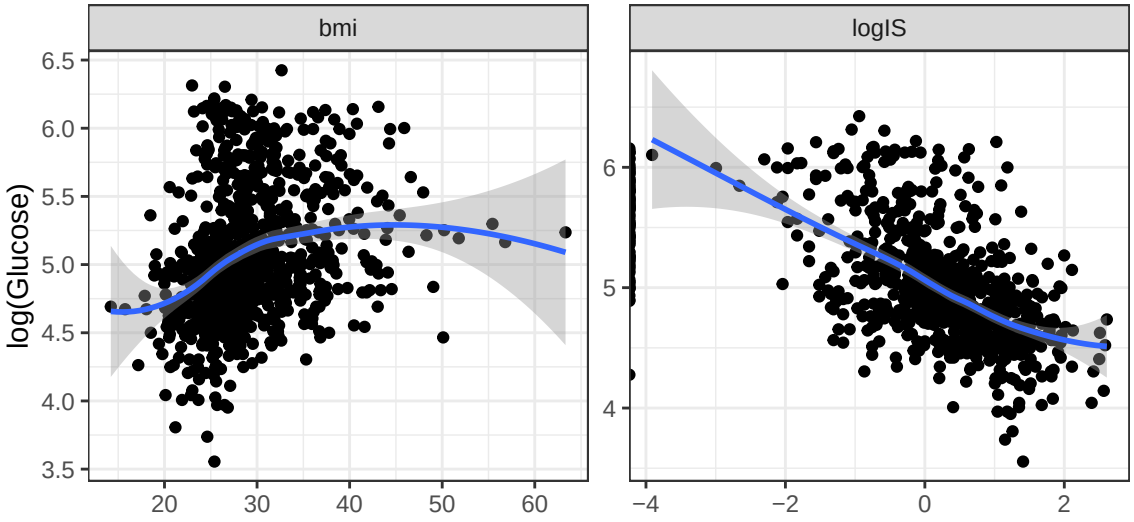

Figure 2: Non-linear relationship between potential factors to log(GT)

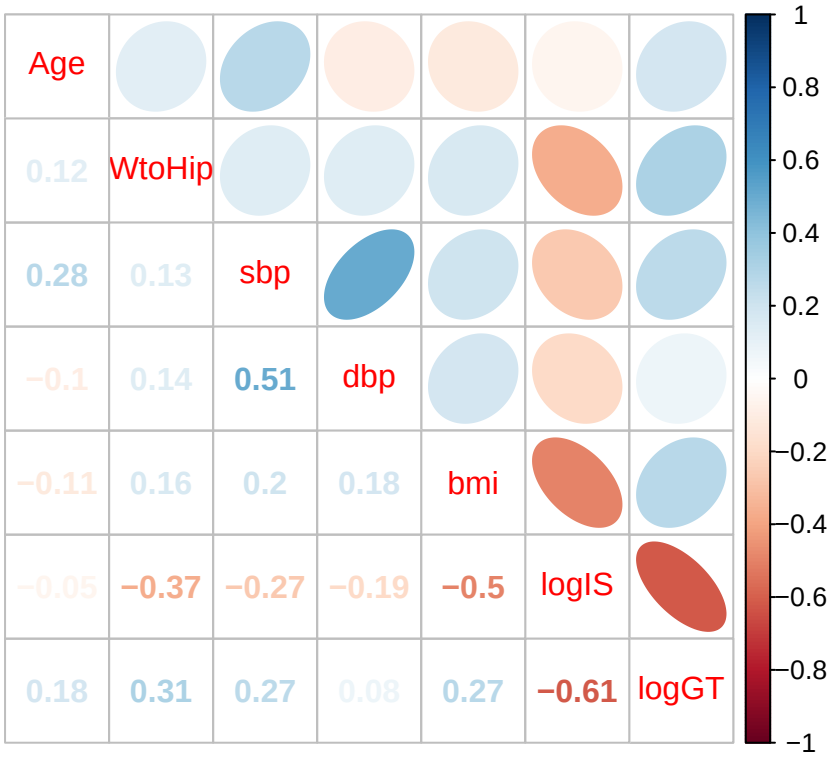

Figure 3: Linear correlation matrix

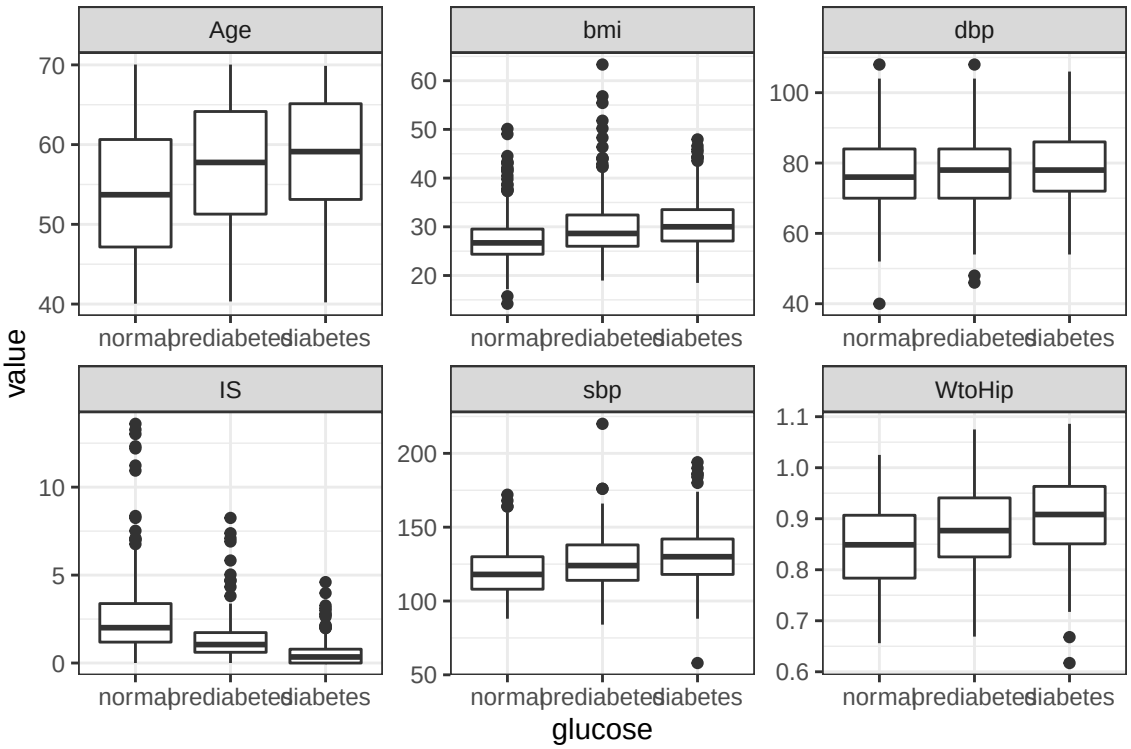

Figure 4: Relationship between potential factors to GT

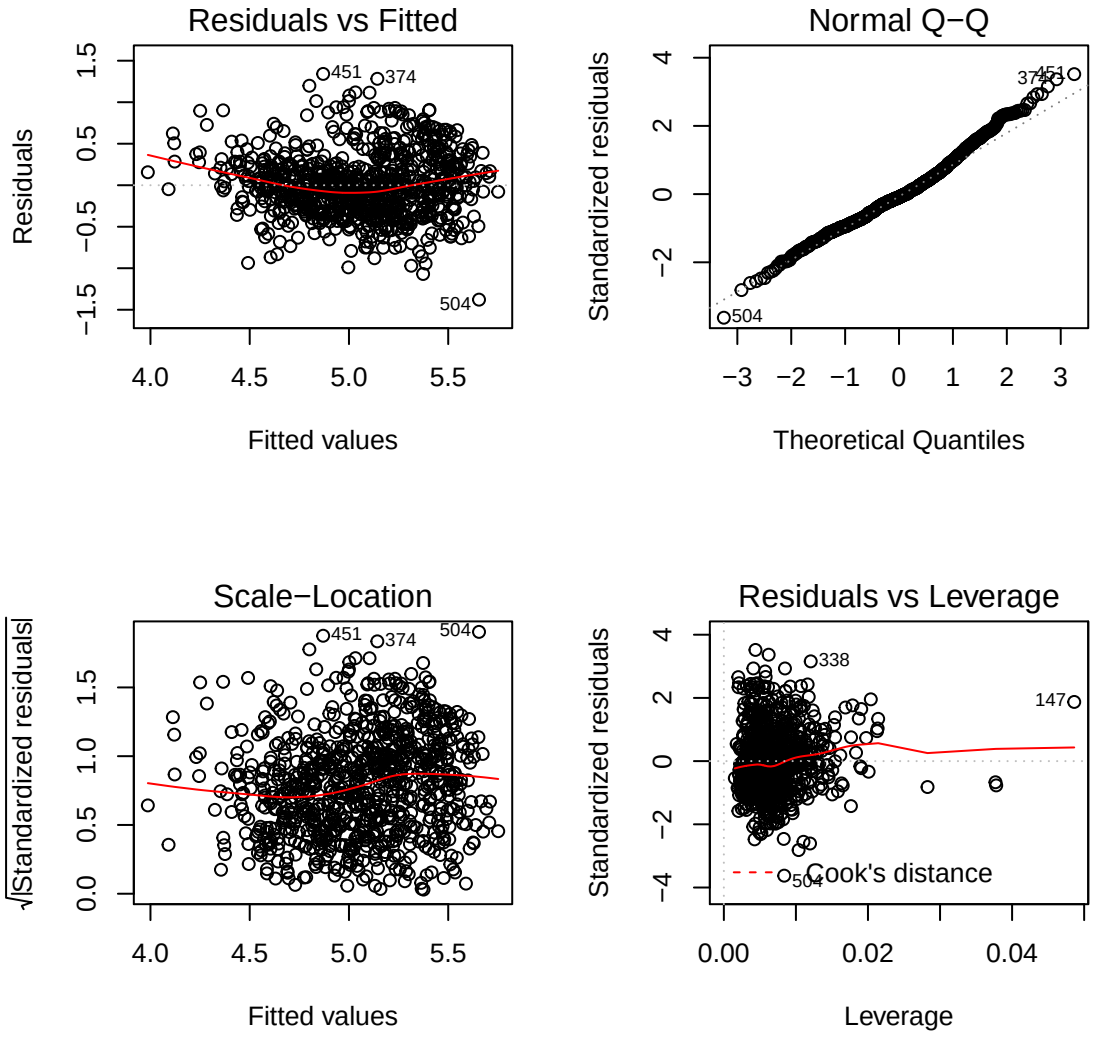

Figure 5: Model Diagnosis

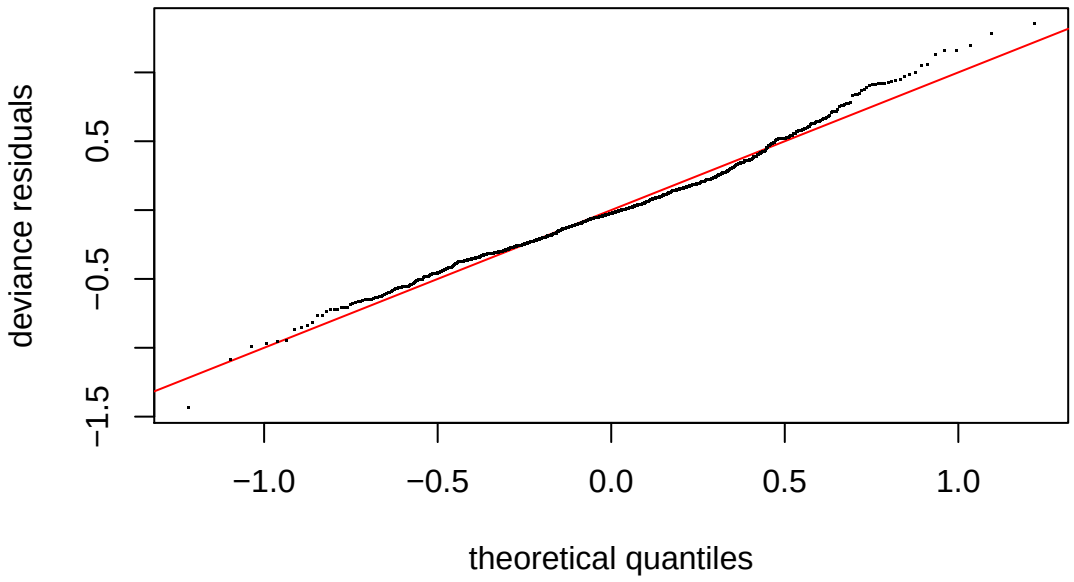

Figure 6: QQ-plot for GAM

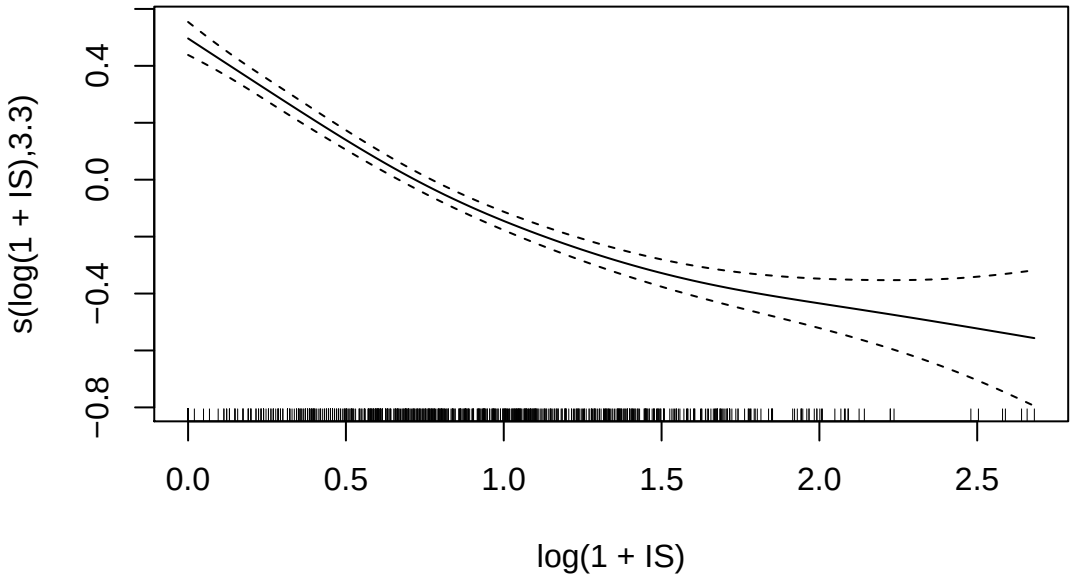

Figure 7: Smoothing function for IS

Table 3: Estimation for proportional odds regression

	2.5 %	97.5 %	estimate	std.error	statistic
Age	-0.0679	-0.0290	-0.0483	0.0099	-4.8714
dbp	0.0077	0.0443	0.0259	0.0093	2.7781
log(1 + IS)	2.3048	3.0152	2.6525	0.1811	14.6464
sbp	-0.0272	-0.0073	-0.0172	0.0051	-3.3719
WtoHip	-3.9180	-0.2629	-2.0853	0.9316	-2.2385

References

Branchtein, Leandro, Maria Inês Schmidt, Sotero S Mengue, Ângela J Reichelt, Maria Cristina G Matos, and Bruce B Duncan. 1997. "Waist Circumference and Waist-to-Hip Ratio Are Related to Gestational Glucose Tolerance." *Diabetes Care* 20 (4). Am Diabetes Assoc: 509–11.

Jaffrain-Rea, M-L, C Moroni, R Baldelli, C Battista, P Maffei, Massimo Terzolo, M Correra, et al. 2001. "Relationship Between Blood Pressure and Glucose Tolerance in Acromegaly." *Clinical Endocrinology* 54 (2). Wiley Online Library: 189–95.

O'rahilly, S, H Gray, A Hattersley, and A Vaag. 1994. "Insulin Resistance as the Major Cause of Impaired Glucose Tolerance: A Self-Fulfilling Prophesy?" *The Lancet* 344 (8922). Elsevier: 585–89.

Tripathy, Devjit, Martin Carlsson, Peter Almgren, Bo Isomaa, Marja-Riitta Taskinen, Tiinamaija Tuomi, and Leif C Groop. 2000. "Insulin Secretion and Insulin Sensitivity in Relation to Glucose Tolerance: Lessons from the Botnia Study." *Diabetes* 49 (6). Am Diabetes Assoc: 975–80.

Wagenknecht, Lynne E, Elizabeth J Mayer, Marian Rewers, Steven Haffner, Joseph Selby, Gerald M Borok, Leora Henkin, et al. 1995. "The Insulin Resistance Atherosclerosis Study (Iras): Objectives, Design, and Recruitment Results." *Annals of Epidemiology* 5 (6). Elsevier: 464–72.