Stockholms universitet

Using logistic regression to predict the risk of depression among parents who have lost a child through suicide

Ida Hed Myrberg

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2014:5
Matematisk statistik
Juni 2014

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Bachelor Thesis **2014:5**
http://www.math.su.se

Using logistic regression to predict the risk of depression among parents who have lost a child through suicide

Ida Hed Myrberg*

June 2014

Abstract

In this paper we use data from a survey study among 666 parents who have lost a child through suicide between the years 2004-2007. We create a dichotomised indicator variable for the prevalence of moderate to severe depression and, with the purpose of predicting the risk of depression among these parents, we fit two different logistic regression models. The first, and simpler, model is to be used in a direct manner, mainly containing predictors that are easy to measure and known at the time of loss of the child, or at least quite constant over time. The second, and more complex model, contains predictors that may be influenced by mood, and that may be unknown at the time of loss. Hence the second model is harder to interpret, and, rather than to be used in a direct manner like the first model, it has a hypothesis-generating purpose for future research. The two fitted models have good predictive qualities within the dataset, but they remain to be tested on new observations. Some effects may be underestimated due to influential observations.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail:ida.hed.myrberg@gmail.com . Supervisor: Tom Britton.

Sammanfattning

I denna uppsats används data från en enkätstudie gjord bland 666 föräldrar som mist ett barn i självmord mellan åren 2004-2007. Vi skapar en indikatorvariabel för förekomsten av måttlig till svår depression, och i syfte att prediktera depressionsrisken hos dessa föräldrar anpassar vi två olika logistiska regressionsmodeller. Den första och enklare modellen ska kunna användas på ett direkt sätt och innehåller främst prediktorer som är lätta att mäta och kända vid förlusttillfället, eller åtminsone relativt konstanta över tid. Den andra och mer komplexa modellen innehåller prediktorer som lätt påverkas av sinnesstämning och kan vara okända vid förlusten av barnet. Den andra modellen är därmed mer svårtolkad och har, snarare än den första modellens mer praktiska användning, ett hypotesgenererande syfte för framtida forskning. De två anpassade modellerna har god prediktionsförmåga inom det använda datasetet, men det återstår att testa modellerna på nya observationer. Vissa effekter kan vara underskattade på grund av ett antal inflytelserika observationer.

Contents

1 Introduction

In the ages 15 to 49 years, death due to self-harm is the most common cause of death in Sweden. In the year 2012 there were 817 people who committed suicide, of which 257 were between the ages 15 and 34 (Omerov 2014). Each of these deaths is a tragedy for the friends and family of the deceased. Parents are left with a grief characterized by feelings of guilt and shame, and have an increased risk of developing depression (Omerov et al. 2013). It is crucial that the health care increase their knowledge on how to prevent and find the warning signs of possible future psychological morbidity among these parents, and thereby enable them to function in their everyday life.

In this paper we use data from a survey study conducted at Karolinska Institutet among 666 suicide-bereaved (loss due to suicide) parents to find two models for predicting future depression. A number of predictors were measured in the study, for example socioeconomic factors, relationship between parent and child, psychological premorbidity of the parents and earlier suicide attempts and self-injury of the child. We create a dichotomous indicator variable for the presence of depression (moderate to severe), and use purposeful selection to fit two logistic regression models, one for more direct prediction and one for generating questions for future research.

1.1 Aim

The aim of this thesis is to build two different prediction models for the outcome depression (moderate to severe) 2-5 years after loss among parents who have lost a son or daughter through suicide. This is to enable the medical care to obtain knowledge of the risk of depression in an early stage to be able to take preventative measures for high risk parents.

The first model will have predictors that are known at the time of loss, or at least quite constant over time, and easy to measure in the sense that they are not affected by the mood of the parent. The second model will in addition take into account predictors that are not necessarily known at the time of loss, predictors that are more difficult to measure, and affected by the parent's mood.

The purpose of the first model is to be able to estimate the risk of depression for a parent who has recently lost a child through suicide. The purpose of the second model is to provide guidance on additional factors that may be interesting for further investigation in future studies.

1.2 Disposition of the paper

In section 2 we give a brief background to the topic. In section 3 the reader is introduced to the statistical concepts and methods used throughout the paper. The characteristics and manipulations of the data are presented in section 4. Section 5 is a detailed account of the statistical analysis performed. In section 6 we interpret the results, and discuss strengths and weaknesses of results and methods. Conclusions and implications for future research are addressed in section 7.

2 Background

2.1 The study

The data used in this paper is taken from a survey study approved by the Regional Ethics Committee, conducted in 2009 at the Department of Clinical Neuroscience and the Division of Clinical Cancer Epidemiology at Karolinska Institutet. The study had the overall aim of improving the professional care among parents who have lost a child through suicide. More specifically the target population was limited to parents who have lost a 15 to 30 year old son or daughter through suicide between the years 2004 and 2007. The parents had to be Swedish-speaking and born in a Nordic country. The 915 parents who were asked to participate were identified through the Swedish Cause of Death Register and the Multi-generation Register, and questionnaires were sent to parents who gave informed consent for participation after an initial introductory letter and phone call. The data collection took place between August 2009 and December 2010. The response rate was 73 percent, and in total 666 parents returned their questionnaires. (Omerov 2014)

A random sample of 508 non-bereaved parents matched 1:2 for background variables such as age, gender, a child born the same year as the deceased child etc. was drawn through the Swedish Population Register, with a response rate of 74 percent (377 parents). By comparing the bereaved and non-bereaved parents it was shown that bereaved parents had a significantly higher prevalence of depression, although they did not have a higher prevalence of psychological premorbidity (more than ten years before the loss of the child having been diagnosed with and/or received treatment for psychological ill-health, and/or having used medication for anxiety and/or depression). This indicates that the suicide can explain the higher prevalence of depression among the bereaved parents. (Omerov et al. 2013)

2.2 Other studies on the subject

Suicide-bereavement is still a rather unexplored subject, but there are some studies, mostly register-based ones. Several studies have found that suicide-bereaved are more vulnerable to psychological morbidity than non-bereaved and bereaved by natural causes (Bolton et al. 2013, Feigelman et al. 2008, Kessing et al. 2003). Some of the results of the studies are contradictory, but predictors considered to be important for psychological morbidity include age of parent and child, socioeconomic factors such as income and education, relationship of parent and child, marital status, existance of siblings of the deceased child, time since loss, religious beliefs, previous losses, family history of suicide, personal relationships, gender and work status (Dyregrov et al. 2003, Li et al. 2003 & 2005, Stroebe et al. 2007, Feigelman et al. 2008, Bolton et al. 2013).

2.3 Measuring depression

Determining whether or not someone is depressed without personal contact with the person is of course hard, but there are ways to at least provide guidance in whether depression is present. Omerov et al. (2014) use the Patient Health Questionnaire (PHQ-9) which is a widely used screening instrument to detect characteristics associated with depression. For each of the nine questions in the questionnaire the respondents rate themselves on a four point scale assessing how frequently they have experienced symptoms of depression in the last two weeks. The maximum score is 27, and the optimal cutpoint for moderate to severe depression is suggested to be ≥ 10, with a sensitivity ($P(\text{Patient scores } \geq 10 \mid \text{Patient depressed})$) and specificity ($P(\text{Patient scores } < 10 \mid \text{Patient not depressed})$) of 88% (Kroenke et al. 2010).

Respondents who are taking medications for symptoms of depression may not experience these symptoms if the medicine is effective, and hence not meet the requirements for moderate to severe depression in the questionnaire, although they probably are diagnosed with depression. Therefore in this thesis depression is defined as either one (or both) of the following: scoring ≥ 10 in the PHQ-9 or in the last month taking medicine for depression symptoms at least once a week. With this definition about 25% of the suicide-bereaved parents in the study are considered to suffer from depression (moderate to severe).

3 Method

In this section the reader is introduced to the statistical concepts and methods used throughout the paper. The main references for this section are Agresti (2002) and Hosmer et al. (2013). The theory about odds ratios and logistic regression presented in section 3.1 and 3.2 is taken from the second and fifth chapters respectively of Agresti (2002). Readers who want to deepen their understanding of the theory are suggested to read those parts. In section 3.3 the theory of the model selection method *Purposeful selection* is taken from Hosmer et al. (2013). The theory of statistical tests and model diagnostics in section 3.4 is mostly taken from Agresti (2002) and Hosmer et al. (2013).

3.1 Odds ratio

Let $x = (x_1, ..., x_p)$ be a vector of p predictors, and let $\pi(x) = P(Y = 1|(x_1, ..., x_p))$ be the probability of success given x. We now define the odds as

$$\Omega = \frac{P(Y = 1|(x_1, ..., x_p))}{P(Y = 0|(x_1, ..., x_p))} = \frac{\pi(x)}{1 - \pi(x)},$$

the probability of success divided by the probability of failure, given x. Since the success probability takes values between 0 and 1, the odds takes values from zero to plus infinity. The probability of success is greater than the probability of failure if the odds is greater than one (> 1).

We want to investigate if there is a difference in success probability between elements with predictor vector $x^* = (x_1, ..., x_j + 1, ...x_p)$ and elements with predictor vector $x = (x_1, ..., x_j, ...x_p)$. We therefore introduce the odds ratio, which is defined as

$$\theta = \frac{P(Y = 1|(x_1, ..., x_j + 1, ...x_p))/P(Y = 0|(x_1, ..., x_j + 1, ...x_p))}{P(Y = 1|(x_1, ..., x_j, ...x_p))/P(Y = 0|(x_1, ..., x_j, ...x_p))} =$$

$$= \frac{\pi(x^*)/(1 - \pi(x^*))}{\pi(x)/(1 - \pi(x))}. \quad (1)$$

If the odds ratio is larger than one (> 1) elements with predictor vector x^* has a larger odds than elements with predictor vector x, and the interpretation is opposite if the odds ratio is smaller than one (< 1). (Agresti 2002, ch. 2)

An advantage of the odds ratio is that it can be estimated in all types of studies (Agresti 2002, ch. 2, p.45). It also plays an important part in logistic regression, which is introduced in the section below. Another common measure is the *risk ratio* or *relative risk*, which is simply the ratio between the success probability (*risk*) in one group and the corresponding probability in the other. However, this cannot be estimated in all types of studies (Agresti 2002, ch. 2, p.42) and does not have a given connection to logistic regression, and will therefore not be used in this paper.

3.2 Logistic regression

When a response variable is dichotomous a common approach in statistical modeling is to use logistic regression to model the success probability given a number of predictors. A great advantage with logistic regression compared to linear regression is that the predicted probabilities always lie between zero and one, while in linear regression they range from minus infinity to plus infinity.

The link function is the logarithm of the odds, which means that the regression model is linear in the logarithm of the odds. Denote the success probability given the predictor vector $x_i = (x_{i1}, ..., x_{ip})$ by $\pi(x_i) = P(Y = 1|(x_{i1}, ..., x_{ip}))$. Then the odds is given by $\pi(x_i)/(1 - \pi(x_i))$. The logistic regression model, called the logit, is given by

$$logit(\pi(x_i)) = log\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}. \quad (2)$$

If we want to express the model in terms of the success probability $\pi(x_i)$ we rewrite (2) by exponentiating both sides. We get the following expression

$$\frac{\pi(x_i)}{1-\pi(x_i)} = \exp\left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right). \quad (3)$$

We then invert both sides of (3) and solve for $\pi(x_i)$ as follows

$$\frac{1-\pi(x_i)}{\pi(x_i)} = \frac{1}{\exp\left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right)} \iff \frac{1}{\pi(x_i)} = \frac{1 + \exp\left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right)}{\exp\left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right)} \iff$$

$$\iff \pi(x_i) = \frac{\exp(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij})}{1 + \exp\left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right)}.$$

If we let $\exp\left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right)$ approach infinity we see that $\pi(x_i)$ approaches 1. Similarly, if we let $\exp\left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right)$ approach zero, $\pi(x_i)$ approaches zero.

Even though it is easy to express the logit function in terms of the success probability it is hard to interpret how an increase in one of the predictors affects the success probability, given that the other predictors are held constant. It is easier to investigate how it changes the odds. As above, denote the predictor vector by $x_i = (x_{i1}, ..., x_{ip})$ and the predictor vector with a one-unit increment in the predictor x_j by $x_i^* = (x_{i1}, ..., x_{ij} + 1, ..., x_{ip})$. We now use (3) to see how the increment in x_j affects the odds. We get

$$\frac{\pi(x_i^*)}{1-\pi(x_i^*)} = \exp\left(\beta_0 + \beta_1 x_{i1} + ... + \beta_j(x_{ij} + 1) + ... + \beta_p x_{ip}\right) =$$

$$= \exp\left(\beta_0 + \beta_1 x_{i1} + ... + \beta_j x_{ij} + ... + \beta_p x_{ip}\right) \cdot e^{\beta_j} = \frac{\pi(x_i)}{1-\pi(x_i)} \cdot e^{\beta_j}. \tag{4}$$

So the odds is multiplied by a factor e^{β_j} when there is a one-unit increase in the predictor x_j. From (4) we also see that the odds ratio θ (see (1)) can be expressed as e^{β_j}. Specifically, if $\beta_j = 0$, the predictor x_j does not affect the success probability.

3.2.1 Likelihood equations

Let n_i be the number of observations with predictor vector $x_i = (x_{i1}, ..., x_{ip})$, and denote the total number of observations $n = \sum_{i=1}^{I} n_i$ where I is the number of unique predictor vectors. Now we let Y_i be the number of observations with $Y = 1$, i.e. the number of successes. Then Y_i is a binomial random variable with success probability $\pi(x_i)$, so we can write

$$Y_i \sim Bin(n_i, \pi(x_i)),$$

where

$$\pi(x_i) = \frac{\exp\left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right)}{1 + \exp\left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right)}.$$

Note that the random variables $\{Y_1, ..., Y_I\}$ are independent. We denote the parameter vector by $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)^T$. The probability function for a specific Y_i is given by

$$p(y_i; \boldsymbol{\beta}) = \binom{n_i}{y_i} \pi(x_i)^{y_i} (1-\pi(x_i))^{n_i - y_i} \propto \left(\frac{\pi(x_i)}{1-\pi(x_i)}\right)^{y_i} (1-\pi(x_i))^{n_i}.$$

Thus the likelihood function is given by

$$L(\boldsymbol{\beta}; y) = \prod_{i=1}^{I} \left(\frac{\pi(x_i)}{1-\pi(x_i)}\right)^{y_i} (1-\pi(x_i))^{n_i},$$

and we get the log likelihood function

$$l(\boldsymbol{\beta}; y) = log(L(\boldsymbol{\beta}; y)) = \sum_{i=1}^{I} \left(y_i \cdot log\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) + n_i \cdot log(1-\pi(x_i)) \right) =$$

$$= \sum_{i=1}^{I} \left[y_i \left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} \right) - n_i \cdot log\left(1 + \exp\left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right)\right) \right].$$

We get the likelihood equations by setting the partial derivatives $\partial l(\boldsymbol{\beta}; y)/\partial \boldsymbol{\beta}$ equal to zero. These equations are nonlinear in $\boldsymbol{\beta}$ and therefore the Newton-Raphson iterative solution method is used. Maximum likelihood estimates exist since the log-likelihood function is strictly concave. Though estimates may be infinite if there is *complete separation* or *perfect discrimination*, meaning there is no overlap in predictors having $y = 1$ and predictors having $y = 0$ (Agresti 2002, ch. 5). However such situations are rare in practice. If there is *quasi-complete separation*, meaning the overlap described

above is very small, the estimates are not infinite, but they become extremely large (Hosmer et al. 2013, ch. 4.4).

3.3 Purposeful selection

There are many different methods for finding a suitable multivariable logistic regression model. Most statistical software packages have ready-to-use selection algorithms that choose the "best" model for you, according to some chosen criterion. We will not be using any of these. Instead we will be using *purposeful selection*, a method presented thoroughly in chapter 4 of Hosmer et al. (2013). It is a seven-step procedure which is summarized below.

1. Perform univariable logistic regression with one predictor at a time and look at the likelihood ratio statistic (introduced in section 3.4.2). Select the predictors that have a p-value for the LR-statistic less than 25 percent. We will return to the other predictors in step 4.

2. Fit a multivariable logistic regression model including all of the predictors with p-value less than 25 percent. Check the Wald statistic (introduced in section 3.4.3) of the predictors and exclude predictors that are not statistically significant at a standard significance level, such as 5%. Compare the smaller and larger model with a partial likelihood-ratio test.

3. Compare parameter estimates for predictors included in both the smaller and the larger model. If the parameter estimates are very different it indicates that some of the predictors left out in the previous step should be included again, because they are needed to adjust the effect of other predictors. Hosmer et al. (2013) use $\Delta\hat{\beta}_i = |(\hat{\theta}_i - \hat{\beta}_i)/\hat{\beta}_i| > 20\%$ as an indicator for when estimates are "too different", where $\hat{\beta}_i$ is the estimate for the coefficient of predictor i in the larger model and $\hat{\theta}_i$ is the corresponding estimate in the smaller one.

4. Now it is time to revisit the predictors that were excluded in the first step. Add these one by one to the multivariable model and check the significance with the Wald statistic.

5. Examine the selected predictors from step 4 more closely. The continuous predictors should for example have a linear relationship with the logit, see section 3.4.1 for methods on inverstigating this. The model selected in step 5 is referred to as the *main effects model*.

6. Consider possible interactions between the selected predictors. Hosmer et al. (2013, ch. 4, p. 92) describes interactions as "...an interaction between two variables implies that the effect of each variable is not constant over levels of the other variable." We add interaction terms to the model one by one and check the significance at a standard significance level. All significant interactions are

then added to the model and we investigate if some of them can be excluded, by following step 2 again, but only removing interactions. The model selected in step 6 is referred to as the *preliminary final model*.

7. Check the fit of the final model selected in step 6 (see section 3.4 below).

Hosmer et al. (2013) argue in favor of purposeful selection since it gives the analyst control over every step of the selection process. Knowledge about for example clinically significant predictors, possible confounders and unreasonable interactions is taken into account. With common automatic selection procedures such as backward elimination, stepwise and forward selection, the role of the analyst becomes somewhat redundant. Purposeful selection can under some conditions perform better compared to frequently used automatic selection procedures when it comes to identifying confounders. For large sample sizes the different methods perform roughly equally well, but for smaller sample sizes common in epidemiological and behavioural studies purposeful selection is preferable (Bursac et al. 2008).

3.4 Model fit and diagnostics

A good statistical model describes the data well while still being as simple as possible. There are several different tests and measures to assess the fit and prediction capacity of a logistic regression model, determine whether to keep a predictor, and comparing two models. In this section the ones used in this paper are presented.

3.4.1 Assessing the linearity of continuous predictors

An assumption for the logistic regression model is that the continuous predictors are linear in the log-odds. However, since the response variable can only take two different values, assessing the linearity is not as simple as to plot the response variable against the predictor. Hosmer et al. (2013, ch. 4) introduce a number of methods for dealing with this issue, one of which will be used in this paper. The method will be referred to as the *quantile method*. The data is divided into five different groups based on midpoints of 20%-percentiles for the continuous predictor under investigation. This is a way of transforming the continuous predictor into a categorical predictor. For this new categorical predictor we fit a logistic regression model and find the estimated coefficients of each category. The lowest category is chosen as reference and therefore has a coefficient equal to zero. Now, if the continuous predictor is linear in the log-odds, the estimates of the coefficients should be fairly linear if they are plotted against the midpoints. If the relationship appears not to be linear one can transform the continuous predictor in a suitable way, fit a polynomial or simply use the categorical version of the continuous predictor.

3.4.2 The likelihood-ratio test

The likelihood-ratio test can be used if we want to compare two logistic regression models, the smaller model M_0 to the larger model M_1, to evaluate whether the smaller model holds. We formulate the null hypothesis and alternative hypothesis as

$$H_0 : M_0 \text{ holds}$$

$$H_1 : M_1 \text{ holds but not } M_0.$$

Let L_0 be the maximized likelihood function under the null, and L_1 be the maximized likelihood function under the alternative. Let l_0 and l_1 be the corresponding maximized log-likelihood functions. We now form the test statistic

$$-2 \cdot log\left(\frac{L_0}{L_1}\right) = -2(l_0 - l_1).$$

It can be shown that the likelihood-ratio (LR) statistic asymptotically follows the $\chi^2(df)$-distribution under the null, that is

$$-2 \cdot log\left(\frac{L_0}{L_1}\right) \overset{H_0}{\approx} \chi^2(df),$$

where the degrees of freedom (df) is given by the difference between the number of parameters in M_1 and the number of parameters in M_0. (Agresti 2002, ch. 5)

3.4.3 The Wald test

The Wald test also uses the asymptotic normal distribution of maximum likelihood estimates. Say we have the following hypotheses

$$H_0 : \beta = \beta_0$$

$$H_1 : \beta \neq \beta_0,$$

then the test statistic

$$z = \frac{\hat{\beta} - \beta_0}{SE_{\hat{\beta}}}$$

has an asymptotic standard normal distribution under the null, that is

$$z \overset{H_0}{\approx} N(0, 1).$$

It follows that z^2 has an asymptotic $\chi^2(1)$-distribution under the null. This result can be extended to several dimensions. Let $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)^T$ be the true parameter vector and let $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_p)^T$ be the maximum likelihood estimate of $\boldsymbol{\beta}$. We formulate the hypotheses

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta_0}$$

$$H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta_0}.$$

The test statistic

$$W = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_0})^T [Cov(\hat{\boldsymbol{\beta}})]^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_0})$$

has an asymptotic $\chi^2(df)$-distribution under the null, where the degrees of freedom are given by the number of extra parameters in the alternative hypothesis. (Agresti 2002, ch. 1)

We will often use the Wald χ^2-test to determine whether a predictor should be included in a model. For a categorical predictor with j levels we test if all $j-1$ parameters for the corresponding dummy variables are equal to zero.

3.4.4 The Hosmer-Lemeshow test

When the number of observations with unique predictor vectors are roughly equal to the total number of observations, which is almost always the case when there is one or more continuous predictors, traditional goodness-of-fit tests such as the Deviance or Pearson χ^2-tests (see Hosmer et al., 2013, ch. 5.2.1 for details) do not have asymptotic χ^2-distributions. Hosmer et al. (2013, ch. 5.2.2) present a more suitable goodness-of-fit test for these situations. In short the approach is to group the data into g different groups based on either percentiles of estimated probabilities or fixed values of estimated probabilities. A test statistic $\hat{C}$ is then created based on squared differences between observed and expected frequencies in each group. Asymptotically it can be shown that $\hat{C}$ has a χ^2-distribution with $g-2$ degrees of freedom. In this paper $g = 10$ is used. The interested reader is encouraged to read further details in chapter 5.2.2 in Hosmer et al. (2013).

3.4.5 AIC

The *Akaike information criterion* (AIC) is a measure used to compare different models and is defined as

$$AIC = -2log(L) + 2k,$$

where L is the maximized likelihood function and k is the number of parameters estimates (including the intercept) in the model. There is no statistical test connected to the AIC, instead one compares the value of AIC for different models to find the model with the smallest value. Models with too many parameters are penalized and therefore overfitting can be avoided when the AIC is used (Agresti 2002, ch. 6, p. 216).

3.4.6 ROC, concordance index and R-square measures

A common way of assessing the prediction capacity of a model is to create a *receiver operating characteristic* (ROC) curve. Let

$$\hat{y}_i = \begin{cases} 1 \text{ if } \hat{\pi}(x_i) > \pi_0 \\ 0 \text{ otherwise} \end{cases}$$

be the predicted value of y_i for some cutoff π_0. Now we calculate the *sensitivity* by $P(\hat{y} = 1|y = 1)$ and the *specificity* by $P(\hat{y} = 0|y = 0)$ for all possible values of π_0. Then the ROC-curve is the sensitivity plotted against $1 - specificity$. With the ROC-curve one can examine what sensitivities and specificities are possible to achieve. The larger area under the curve the better the prediction power.

The area under the curve is actually the *concordance index*, denoted by c. If we form all possible pairs of observations (i, j) with $y_i = 1$ and $y_j = 0$, the concordance index is an estimate of the probability that the observation with $y = 1$ have the higher $\hat{\pi}$ as well. We have that $c \in [0.5, 1]$, with $c = 1$ when all observations with $y = 1$ have higher $\hat{\pi}$, and $c = 0.5$ when we may as well have randomly guessed the outcomes (Agresti 2002, ch. 6). Hosmer et al. (2013, ch. 5.2.4, p. 177) provide rough guidelines on how to evaluate the concordance index:

$$\text{If} \begin{cases} 0.5 < c < 0.7 \text{ Poor} \\ 0.7 \leq c < 0.8 \text{ Acceptable} \\ 0.8 \leq c < 0.9 \text{ Excellent} \\ 0.9 \leq c \text{ Outstanding.} \end{cases}$$

A disadvantage with the concordance index is that it does not take into account the number of estimated parameters in the different models. Hence if one only assesses this measure for predictive capacity there is a risk of using a model with too many parameters, overfitted to the data. Therefore we need some kind of adjusted R^2-measure for the amount of explained variation. Mittlböck & Schemper (1996) present the measure

$$R^2_E = 1 - \frac{l_1}{l_0},$$

where l_1 is the maximized log-likelihood for the fitted model and l_0 is the maximized log-likelihood for the model containing only the intercept. Further they introduce the adjusted measure

$$R^2_{E,adj} = 1 - \frac{l_1 - k/2}{l_0 - 1/2},$$

which takes into account the number of estimated parameters (including the intercept) k. See Mittlböck & Schemper (1996) for a theoretical motivation of this adjustment. As in linear regression, $R^2_{E,adj}$ takes values between zero and one.

3.4.7 Outliers and influential observations

To evaluate the fit of a model it is not enough to only look at summary statistics such as the Hosmer-Lemeshow statistic. One must also assess the fit and influence of individual observations. To detect influential observations Hosmer et al. (2013, ch. 5.3) introduce the measurements $\Delta\chi_j^2$ and ΔD_j where the first one is the decrease in the value of the Pearson χ^2 statistic if observation j is deleted and the second one is the similar value for the deviance (for details of Pearson and deviance, see ch. 4 in Agresti 2002). Agresti (2002, ch. 6.2.4) introduce the *confidence interval displacement diagnostic*, here denoted $cidd_j$, which is a measurement of the change in a joint confidence interval for the parameters if observation j is deleted.

Observations can fit poorly without being very influential. To detect such observations one can evaluate the Pearson and deviance residuals (Agresti 2002, ch. 6, p. 220). If we let n_i be the number of elements with predictor vector x_i and y_i be the corresponding number of successes, then the Pearson residual is defined as

$$e_i = \frac{y_i - n_i\hat{\pi}_i}{\sqrt{n_i\hat{\pi}_i(1-\hat{\pi}_i)}}$$

and the deviance residual as

$$d_i = \text{sign}[y_i - n_i\hat{\pi}_i] \times \sqrt{2\left(y_i \cdot log\left(\frac{y_i}{n_i\hat{\pi}_i}\right) + (n_i - y_i) \cdot log\left(\frac{n_i - y_i}{n_i(1-\hat{\pi}_i)}\right)\right)}.$$

Plotting these residuals may give an idea of which observations lack fit, but Agresti stresses that these residuals loose relevance when many of the $n_i = 1$ and that one should not put too much importance into one single residual. There are no definite criteria for when an observation is extreme or too influential, it depends on the specific situation.

4 Data

The data the analysis is based on is survey data collected between August 2009 and December 2010 from 666 suicide-bereaved parents. Characteristics of these parents, such as gender, age, income etc., are presented in Table 1.

The dichotomous response variable is presence of moderate to severe depression (see section 2.3). One of the respondents has missing data for the response variable and is therefore excluded from further analysis. This exclusion should not affect the outcome in any substantial way. Predictors relevant for the models are selected in consultation with specialists on the subject. Ideally the variables selected for the first model should be known at the time of loss and also be easy to measure, and be objective in the sense that the value of a variable is not so much affected by the respondent's troubles remembering and their mood. For the first model 30 predictors are considered, and 16 additional predictors are considered for the second model. For a list of the predictors see Appendix A2.

Since we are dealing with survey data most predictors are nominal or ordinal, except for the continuous predictors age of parent and child, and time since loss of the child.

4.1 Nominal predictors

Several of the predictors are on a nominal scale, which means that there is no logical order of the values. For example a nominal predictor could be the nationality of a person. There is no way to order countries in an ascending or descending manner. Some of the nominal predictors in this paper are dichotomous and some are polychotomous, which means that they have more than two levels.

A way to deal with polychotomous predictors is to introduce a number of dummy (dichotomous) variables. If the predictor has k levels, we select one of the levels as reference and introduce $k-1$ dummy variables for the other levels. Say for example that we set the last level as reference, and call the dummy variables $D_1, ..., D_{k-1}$, then the variable will be coded as in Table 2. This method is called *dummy coding* or *reference cell coding* (Hosmer et al. 2013, ch. 3).

The logistic regression model with the categorical predictor is then given by

$$logit(\pi) = \beta_0 + \sum_{i=1}^{k-1} \beta_i D_i.$$

Though for simplicity, if we denote the categorical predictor x, we write the model as

$$logit(\pi(x)) = \beta_0 + \beta_x x.$$

Table 1: Characteristics of participants

		no. (%)
Total number of parents asked		915
Participants		666 (73)
Gender of parent	Male	283 (42)
	Female	383 (58)
Gender of child	Male	462 (69)
	Female	204 (31)
Marital status	Living with partner	477 (72)
	Living apart	44 (7)
	Single	121 (18)
	Widow/widower	18 (3)
	Data missing	6 (1)
Area of residence	Rural area	162 (24)
	Population less than 10 000	153 (23)
	Population less than 50 000	128 (19)
	Population less than 200 000	117 (18)
	Stockholm/Gothenburg/ Malmö	97 (15)
	Data missing	9 (1)
Education	Elementary school or less	146 (21)
	High school	271 (41)
	University/college ($<$ 3 years)	82 (12)
	University/college ($\geq$ 3 years)	159 (24)
	Data missing	8 (1)
Work status	Employed or self-employed	498 (75)
	Old age pension	59 (9)
	Disability pension	61 (9)
	Unemployment benefits	25 (4)
	Student	4 (0.6)
	Social beneficiary	3 (0.4)
	Other	9 (1)
	Data missing	7 (1)
Income	0-99 000 SEK	34 (5)
	100 000-199 000 SEK	120 (18)
	200 000-399 000 SEK	388 (58)
	$\geq$ 400 000 SEK	109 (16)
	Data missing	15 (2)
	Age of child at loss	**Age of parent**
Mean (std.dev)	23.4 (4.1)	56.3 (6.2)
Median (interquartile range)	23 (7)	56 (8)
Min, max	15, 31	40, 81

4.2 Ordinal predictors

There are a number of ordinal predictors, most of them considered for the second model. With ordinal predictors it is possible to order the levels in an ascending/descending manner, but there is no meaningful way to measure the difference between two levels. An ordinal predictor can for example measure how much the respondent agrees to a statement, with the response alternatives (levels) *1. Not at all, 2. A little, 3. Moderately, 4. A lot.* There is obviously an internal ordering of the levels, but it is not meaningful to numerically measure the difference between two levels.

Table 2: Example of dummy coding of polychotomous variable

	Dummy variables					
Level	D_1	D_2	...	D_j	...	D_{k-1}
1	1	0	...	0	...	0
2	0	1	...	0	...	0
⋮	⋮	⋮	⋱	⋮	⋱	⋮
j	0	0	...	1	...	0
⋮	⋮	⋮	⋱	⋮	⋱	⋮
k-1	0	0	...	0	...	1
k	0	0	...	0	...	0

There are different ways to deal with ordinal predictors in a model. One way is to treat the predictor as if it were numerical, assigning numerical values to the levels. The simplest way of doing this is to assign integer values with the value k for level k. Though if one has reason to believe that there is a greater distance between some of the levels one can assign other values that seem to be more suitable. For example if one believes there is a bigger gap between the levels *Not at all* and *A little* the assigned values could be *1. Not at all, 3. A little, 4. Moderately, 5. A lot.*

An obvious problem with treating an ordinal predictor as numerical is that the assigning of numerical values is arbitrary and would probably be done differently by different persons. Every assignment of numerical values is based on assumptions and there is no neutral or objective way to handle this. To simply use the integer method can seem objective, but this is based on the assumption of equidistance between the levels, an assumption we have no way of controlling the veracity of. Also, treating the predictor as continuous means assuming a linear relationship between the predictor and the log-odds, a relationship that may not be present in reality (Tutz 2012, ch. 4). For this reason we will use another method in dealing with ordinal predictors: to treat them as if they were nominal. This way we do not make any faulty assumptions that cannot be controlled for. A disadvantage of this method is that it does not use the order information, and therefore statistical tests have less power than if the predictors were treated as numerical. The nominal approach is hence more conservative than the numerical.

Other ways of dealing with ordinal predictors are different kinds of penalized regression, that take into account that the predictor is ordinal and not nominal, but without treating the predictor as continuous. The basic idea is to, instead of maximizing the usual log-likelihood, maximizing a penalized log-likelihood which penalizes large differences between the coefficients of two adjacent response categories. For a more detailed presentation of the theory and comparisons with other methods see Tutz (2012) and Gertheiss & Tutz (2009). Because of time limitations these methods will not be implemented here.

4.3 Categories with few observations

In some of the categorical predictors there are very few observations for one or more of the levels. This can cause problems with too large standard errors. A way to deal with this issue is to merge two or several categories where this is reasonable from a clinical perspective. This is often possible for ordinal predictors, since adjacent categories are related in a natural way, but, based on subject matter knowledge, it can also be possible for nominal predictors. For the predictors *ContactPC* (intensity of contact between parent and child during year before loss), *SuicCYB* (suicide attempts of child during year before loss), *Edu* (parent's level of education), *Work* (parent's work situation) and *Social* (intensity of parent's social life, see Appendix A2 for more detailed descriptions of all predictors), some categories have too few observations (≤ 14 observations) and are therefore merged. In Appendix A1 the old and new categories for those predictors are presented.

4.4 Missing data

A common problem in survey studies is when a respondent skips one or several questions, resulting in missing data. There are different mechanisms associated with missing data. The most stringent assumption about the missingness is *Missing Completely at Random* (MCAR), which indicates that the probability of data being missing for a certain variable is independent of the value of that variable and the values of any other variables. This is the case when a respondent randomly skips a question. A less stringent assumption is *Missing at Random* (MAR), which indicates that the probability of data being missing for a certain variable is independent of the value of that variable, but dependent on the values of one or more of the other variables. For example, this would be the case if skipping a question about income is related to the sex of the respondent, but not to the income itself. When none of the above assumptions are met, data is considered to be *Not Missing at Random* (NMAR). (Allison 2002)

There are several ways of dealing with missing data, many of which assume that data are MCAR or at least MAR. With for example listwise deletion one only analyzes observations with no missing data at all. Assuming data are MCAR this method does not induce bias, since the complete case data can then be considered a random subsample of the full data set. Though with NMAR data this method can induce heavy bias. (Allison 2002)

Parts of the data used in this paper are missing, and because many of the questions in the questionnaire are of a sensitive nature we cannot assume that the missing data are MCAR or even MAR. In total for the 46 predictors considered there is about 1 percent missing data. The predictor with the most missing data have about 7 percent missing, and measure to what degree the parent was prepared that the cause of death could be suicide. Around 15 percent of the observations for the predictors considered for the first model have at least one missing value. The corresponding proportion for the second model is 26 percent. If we compare the respondents with no

missing values for the predictors considered in both the first and second model, with the respondents with at least one missing value, we find that the two groups do not differ in any substantial way when it comes to background variables such as age of parent and child, population size in area of residence, marital status, income, education and work status. The differences for psychological premorbidity and prevalence of depression are also small. In the respondent group with at least one missing value, there is a larger proportion of men, compared to the group with no missing values (48 percent compared to 41 percent), though this difference is not huge either. This simple comparison between the two groups for a number of background variables do not give us any indication that the two groups differ in any substantial way. Though the analysis can of course be deepened, examining for example associations between the missingness of one predictor and levels of other predictors and the response.

In total the amount of missing data is rather small (Bennett 2001). Because of this, and since we found little indication that respondents with missing values and respondents with no missing values differ, we use an imputation method that generally is likely to induce bias. Though in this case the bias is probably not that large. For continuous predictors the median of the existing data is imputed and for ordinal and nominal predictors the mode (most common value) is imputed.

One could argue that the proportion of missing data in itself cannot determine the severeness of the bias, instead one should look at the patterns of the missing data (Dong & Peng 2013). A more reasonable imputation method would then be the nearest neighbor method, which in short imputes a missing data value from an observation that has similar values for the non-missing data (Jönsson & Wohlin 2006). A common way to handle missing data is to perform a sensitivity analysis that compares different imputation methods and assesses the robustness of them (Bennett 2001). A sensitivity analysis is not conducted in this paper because of time constraints.

5 Results

In this section the model selection process is described in detail. First we assess the assumption of linearity for the continuous predictors, then we proceed to the selection and diagnostics of the first model. Lastly we select and evaluate fit and prediction capacity of the second model.

5.1 Assessing the linearity of continuous predictors

To assess the linearity of the continuous predictors *AgeC, AgeP* (age of the child and parent) and *Losstime* (time between loss of child and study participation), we use the quantile method described in section 3.4.1, to create plots for the parameter estimates with 95% Wald confidence intervals, when categorical versions of the continuous predictors are used. The results of this analysis is presented in Figure 1, 2 and 3. Based on these plots we have reason to believe that the continuous predictors are not linear in the log-odds, though for *AgeC* and *AgeP* the confidence intervals are quite wide, and the effect could still be linear. However, we will wait a little with taking measures for these possible violations of the linearity assumption, and evaluate the linearity again in a later stage, to see if the continuous predictors could be linear in a multivariable model.

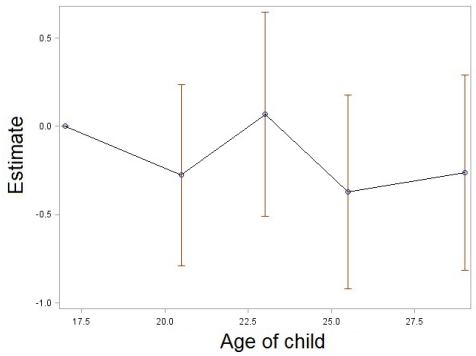

Figure 1: Parameter estimates with 95% Wald CI's plotted against midpoints of quantile groups of AgeC

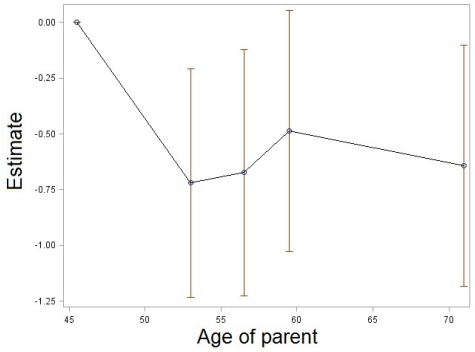

Figure 2: Parameter estimates with 95% Wald CI's plotted against midpoints of quantile groups of AgeP

5.2 Fitting of the first model

It is now time to fit the first model. The aim of this model is described in section 1.1. A list of all predictors considered for this model can be found in Appendix A2.

5.2.1 Purposeful selection of predictors

The model selection strategy will be *purposeful selection*, described in section 3.3. The first step is to fit a univariable logistic regression model to all the predictors one by one, with the response variable *Dep*, which is an indicator for moderate to severe

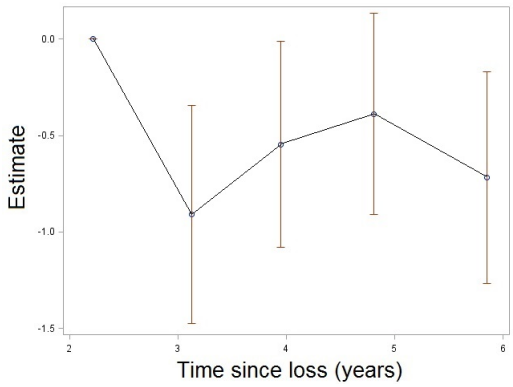

Figure 3: Parameter estimates with 95% Wald CI's plotted against midpoints of quantile groups of Losstime

depression (see section 2.3 for details). If, for example, the predictor is *AgeC* the univariable logistic regression model is given by

$$logit(\pi(AgeC_i)) = \beta_0 + \beta_{AgeC} \cdot AgeC_i.$$

The results of fitting the univariable models can be seen in Appendix A3. We include the predictor *ID* (an arbitrary number given to each respondent) in the analysis because if such predictors become significant it may imply that there are confounding predictors not measured. After fitting the univariable models the predictors that have a p-value for the LR-statistic less than 25 percent are put in a multivariable logistic regression model. Details for the fitting of this model can be found in Appendix A4.

The least significant predictor in this model is *SexC* (sex of the child) with a Wald χ^2 p-value of 0.6221. When *SexC* is deleted from the model an estimate for one of the levels of the predictor *ContactPC* (intensity of contact between parent and child during year before loss) change more than 20 percent (see section 3.3 for a description of the $\Delta\hat{\beta}$-measurement used). However, since *ContactPC* also has a high p-value (=0.5771) we ignore this and proceed with deletion of *ContactPC*. This affects the estimates of *SelfinjCYB* (self-harm of child during year before loss) and *Inc* (parent's income on an ordinal scale). *SelfinjCYB* has a high p-value and will be deleted in a later stage, and for now we ignore the change in the estimate of *Inc*. We now delete the predictor *AgeP* which has the current highest p-value (=0.5091). The estimate of one of the levels of *Inc* now change once again, but we ignore this change and delete *SelfinjCYB* which has a p-value of 0.4408. This changes the estimates of *SuicCE* (child's suicide attempts earlier than year before loss) and *Inc*, but since *SuicCE* has a high p-value (=0.5714) we simply delete it from the model. This does not affect any of the estimates too much. We continue the selection process by deleting *ViolSuic* (if the suicide was violent, p-value of 0.4375) and afterwards delete *Prevloss1* (loss of person close to parent during the ten years before loss, p-value of 0.2747). This does

not affect the other estimates very much. We proceed in deleting the least significant predictor one at a time until all predictors are significant.

In Table 3 the deletion process is summarized with some measures to assess the quality of the models. The first model is the model containing all predictors with univariable p-values less than 25 percent. We see that the AIC decreases up to the deletion of *Biologic* (parent being the biological parent of the child) and that $R^2_{E,adj}$ (see section 3.4.6) increases up to the deletion of *PrevLoss1*. Therefore we continue with the models 1.7 and 1.8 from Table 3 to step 4 of the purposeful selection process, even though not all predictors in these models are statistically significant. We do this because the prediction capacity of the models are considered more important than individual p-values of predictors. Earlier we noted that one parameter estimate of *Inc* changed quite often when other predictors were deleted. If we examine these changes more closely it seems that the estimates of *Inc* only changed along the way of reaching models 1.7 and 1.8, and have not changed very much if we compare models 1.7 and 1.8 with model 1.1. Also the reference category for *Inc* has quite few observations (=34), making the standard errors for the estimates large.

Table 3: Summary of step 2 and 3 in the selection process for the first model

No.	**Predictors deleted from previous model**	**No. of pred-ictors**	**No. of para-meters**	**AIC**	**-2logL**	***c***	$\boldsymbol{R^2_{E,adj}}$	**LR p-value larger vs smaller model**
1.1	None	18	31	680.567	618.567	0.776	0.13454	
1.2	SexC	17	30	678.809	618.809	0.775	0.13555	0.6228
1.3	ContactPC	16	27	674.857	620.857	0.772	0.13682	0.5625
1.4	AgeP	15	26	673.292	621.292	0.772	0.13758	0.5095
1.5	SelfinjCYB	14	23	669.938	623.938	0.768	0.13805	0.4495
1.6	SuicCE	13	21	667.073	625.073	0.767	0.13920	0.5669
1.7	ViolSuic	12	20	665.668	625.668	0.768	**0.13974**	0.4405
1.8	Prevloss1	11	19	**664.734**	626.734	0.766	0.13965	0.3019
1.9	Biologic	10	18	666.174	630.174	0.762	0.13640	0.0636
1.10	EmplTD	9	16	666.664	634.664	0.758	0.13308	0.1059
1.11	Loc	8	12	667.067	643.067	0.746	0.12722	0.0779
1.12	PartnTD	7	11	668.220	646.220	0.739	0.12435	0.0758
1.13	Guardian	6	10	670.220	650.220	0.735	0.12035	0.0455

Before moving on to step 4 of the selection process we reenter the predictors deleted in models 2-6 one by one to see if they may be significant or improve the AIC or $R^2_{E,adj}$ of models 1.7 and 1.8. We also try adding the continuous predictor *AgeP* as a categorical predictor with the method from section 5.1. We find that none of these improve Models 1.7 and 1.8 and so we proceed to the fourth step of the selection process.

In step 4 we evaluate the predictors that had a p-value higher than or equal to 25 percent in the univariable models. We add them one by one to the models selected

in the previous step, and find that none of them are significant or improve the AIC or $R^2_{E,adj}$. We also see if the continuous predictor *AgeC* becomes significant in any of the models if we use it as a categorical predictor, but it does not.

In the fifth step it is time to return to the assumption of linearity for the continuous predictor *Losstime*. Once again we use the quantile method described in section 3.4.1 to create plots for the parameter estimates with 95% Wald confidence intervals, of the categorical version of *Losstime*, call it *LosstimeCat*, in Models 1.7 and 1.8. In Figure 4 we present this plot for Model 1.7 (the plot for Model 1.8 is almost identical and therefore omitted). The predictor *Losstime* does not seem to be linear in either one of the models, even when taking the confidence intervals into account, so we try including *LosstimeCat* in the models, and because of the shape of the plot we also fit a cubic polynomial. The results are summarized in Table 4. Both using the categorical predictors and including quadratic and cubic terms improves the $R^2_{E,adj}$ and AIC of the models a lot. The cubic polynomial method seems to work best, but there is no motivation for this approach from a clinical perspective, and there is a risk that we are overfitting the models to the data. For these reasons we will proceed with Models 1.7a and 1.8a (where *Losstime* is treated as categorical) to step 6 in the selection process.

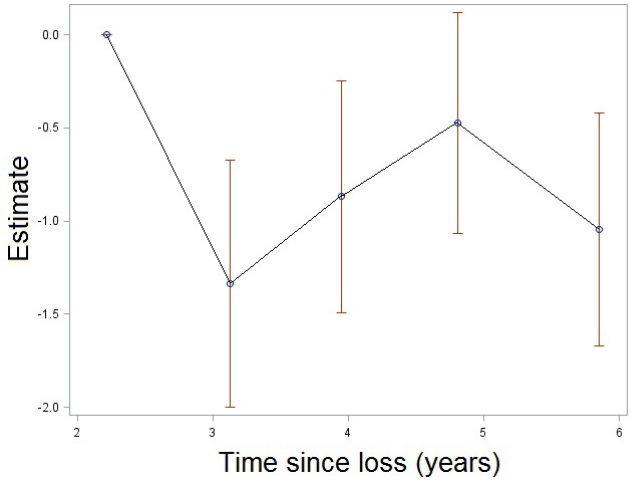

Figure 4: Estimates from Model 1.7 with 95% Wald CI's plotted against midpoints of quantile groups of Losstime

As we reach step 6 in the purposeful selection process we have two models as candidates for the *main effects model*. Model 1.7a contains the predictors *Biologic*, *EmplTD* (parent's employment status at the time of loss), *Guardian* (if the parent was the guardian of the child during most of the upbringing), *Inc*, *Loc* (population in parent's area of residence on an ordinal scale), *LosstimeCat*, *PartnTD* (if the parent was living with a partner at the time of loss), *PremorbP* (psychological premorbidity of the parent), *PrevLoss1*, *PrevLoss3* (parent having suicides in their biological family), *SexP* (sex of parent), and *Work* (parent's work status at time of study). Model 1.8a contains the same predictors except for *PrevLoss1*. Now it is time to investigate

Table 4: Evaluation of transformations of *Losstime* in Models 1.7 and 1.8

No.	**Loss-time**	**No. of parameters**	**AIC**	***c***	$R^2_{E,adj}$	**P-value of Losstime (Cat)**	**P-value of *LossTime*^{2}**	**P-value of *LossTime*^{3}**
1.7	Linear	20	665.668	0.768	0.13974	0.0206		
1.7a	Categ.	23	656.317	0.780	0.15619	0.0005		
1.7b	Cubic	22	652.725	0.783	0.15965	0.0001	0.0002	0.0004
1.8	Linear	19	664.734	0.766	0.13965	0.0187		
1.8a	Categ.	22	655.516	0.777	0.15593	0.0005		
1.8b	Cubic	21	652.123	0.780	0.15912	0.0001	0.0003	0.0005

whether any interactions should be added to the models. For Model 1.7a we have $\binom{12}{2} = 66$ two-factor interactions to consider and $\binom{11}{2} = 55$ of these apply to Model 1.8a. The interactions are added one by one to the different models and a likelihood-ratio test is performed to assess the significance of each interaction. The interactions *SexP*Work* and *LosstimeCat*PartnTD* become significant (LR p-value less than 5%) in both Model 1.7a and 1.8a. They also become significant when included in the same model, for both Model 1.7a and 1.8a. In Table 5 the result of including the interactions is presented. AIC and $R^2_{E,adj}$ for both models improve a lot when the interaction terms are added. Model 1.7c (Model 1.7a with interactions) still has the highest $R^2_{E,adj}$ while Model 1.8c (Model 1.8a with interactions) has the lowest AIC, so we continue with both models to the last step of the purposeful selection. Worth noting is that the p-value for the predictor *PartnTD* increases with around 20 percentage points for both models when the interaction terms are added. This change is not completely unexpected, since the interaction term *LosstimeCat*PartnTD* explains part of the variation previously explained by *PartnTD*. The predictor *PartnTD* will not be deleted since the main effects are considered fixed in this step of the selection process.

Table 5: Evaluation of interactions

No.	**Interactions added**	**No. of parameters**	**AIC**	***c***	$R^2_{E,adj}$
1.7a	None	23	656.317	0.780	0.15619
1.7c	SexP*Work, LosstimeCat*PartnTD	29	646.706	0.805	0.17699
1.8a	None	22	655.516	0.777	0.15593
1.8c	SexP*Work, LosstimeCat*PartnTD	28	646.022	0.803	0.17657

5.2.2 Model diagnostics for the first model

We have two models (Models 1.7c and 1.8c) as candidates for the *preliminary final model*, and it is time to evaluate the fit of these two models. The Hosmer-Lemeshow

χ^2 p-value (see section 3.4.4) is 0.8858 and 0.9281 for Model 1.7c and 1.8c respectively. This indicates good fit for both models, but since the Hosmer-Lemeshow p-value is a summary statistic we must also assess the fit and influence of individual observations. The measurements $\Delta\chi_j^2$, ΔD_j and $cidd_j$ are introduced in section 3.4.7, and we plot these for Model 1.7c in Figures 5 and 6. We see that there are some observations that stand out from the rest, and we take a closer look at observations with $\Delta\chi_j^2 > 15$, $\Delta D_j > 5$ and $cidd_j > 0.4$. There are 11 such observations and if we delete these one at a time 2-9 of the parameter estimates change more than 15 percent. If we delete all of them from Model 1.7c the AIC and $R^2_{E,adj}$ improves a lot, but the results are questionable since quasi-complete separation of data points was detected resulting in unreasonably large estimates (see section 3.2.1). For Model 1.8c the extreme/influential observations are the same as for Model 1.7c with two exceptions, and the results of deleting them are essentially the same as for Model 1.7c. The plots for Model 1.8c are not included since they are almost identical to the plots for Model 1.7c.

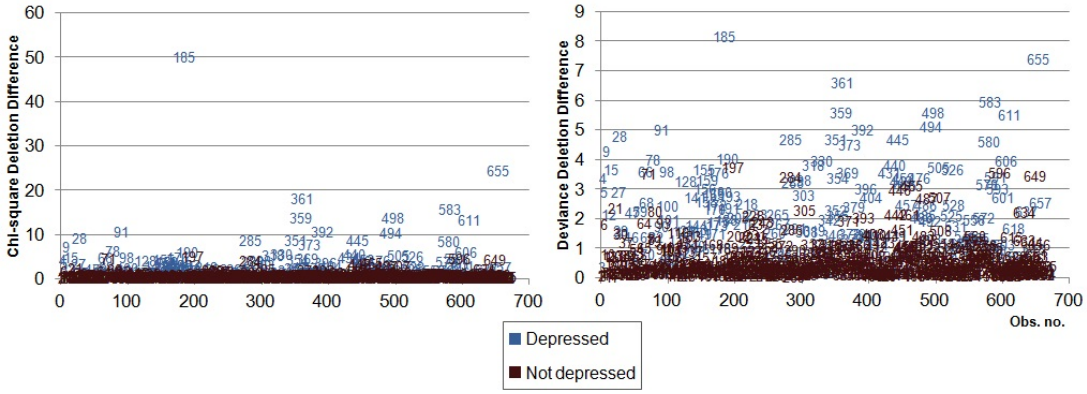

Figure 5: Plot of $\Delta\chi_j^2$ and ΔD_j for Model 1.7c

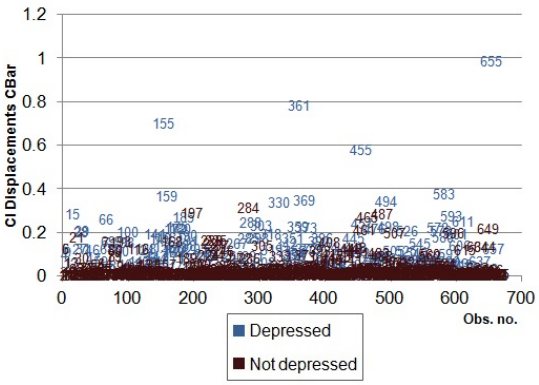

Figure 6: Plot of $cidd_j$ for Model 1.7c

Essentially the same observations that have high $\Delta\chi_j^2$, ΔD_j and $cidd_j$ have large Pearson and deviance residuals in the two models. Plots of these residuals for

Model 1.7c can be found in Appendix A8 (the ones for Model 1.8c are omitted, since they are almost identical to the 1.7c plots). If the extreme/influential observations are erroneous they should be deleted or corrected, however we have no reason to believe that this is the case and hence keep the observations in the models. When the influential observations are included in the models most of the effects become weaker than they would have been if we were to exclude the 11 influential observations, and the effects that are not weakened are only very slightly intensified. Despite the influential observations the fit of the models still looks decent.

We now have two candidates for the final model, Model 1.7c and Model 1.8c. The only difference between them is that Model 1.8c does not contain the predictor *PrevLoss1*. Model 1.7c has the highest $R^2_{E,adj}$ and Model 1.8c has the smallest AIC, although the differences between the models for these two measures are very small (see Table 5). Since the predictor *PrevLoss1* has a high p-value (=0.2527) in Model 1.7c and a simpler model is easier to interpret we will choose Model 8c as our final model.

The selected Model 1.8c contains the predictors *Biologic, EmplTD, Guardian, Inc, Loc, LosstimeCat, PartnTD, PremorbP, PrevLoss3, SexP,* and *Work*, and the interaction terms *SexP*Work* and *LosstimeCat*PartnTD*. It has a concordance index of 0.803, which is considered excellent according to the guidelines provided in section 3.4.6. The corresponding ROC-curve for Model 8c can be found in Appendix A8. In Table 6 the odds ratios significantly different from one for Model 1.8c are presented with 95% Wald confidence intervals. All parameter estimates with p-values, and odds ratios for comparison with reference groups with accompanying confidence intervals for Model 1.8c can be found in Appendix A5. In Figure 7 the predicted probabilities of depression for depressed and non-depressed parents are plotted. We see that the probabilities for the depressed parents in general are higher than for the non-depressed, but there are still depressed parents with low probabilities, as well as non-depressed parents with high probabilities.

Table 6: Significant odds ratios and 95% Wald confidence intervals for Model 1.8c

Label	**OR estimate**	**95% Wald confidence interval**	
Work 2 vs 1 at SexP=2	0.088	0.017	0.464
LossTimeCat 5.85 vs 2.22 at PartnTD=1	0.155	0.042	0.578
LossTimeCat 4.81 vs 2.22 at PartnTD=1	0.313	0.099	0.995
LossTimeCat 5.85 vs 2.22 at PartnTD=2	0.463	0.227	0.945
LossTimeCat 3.95 vs 2.22 at PartnTD=2	0.373	0.181	0.769
LossTimeCat 3.13 vs 2.22 at PartnTD=2	0.145	0.056	0.370
PremorbP 1 vs 2	3.845	2.223	6.652
PrevLoss3 2 vs 1	1.715	1.049	2.802
EmplTD 1 vs 3	2.829	1.333	6.004
Loc 5 vs 1	2.620	1.305	5.258
Loc 2 vs 1	2.680	1.468	4.892

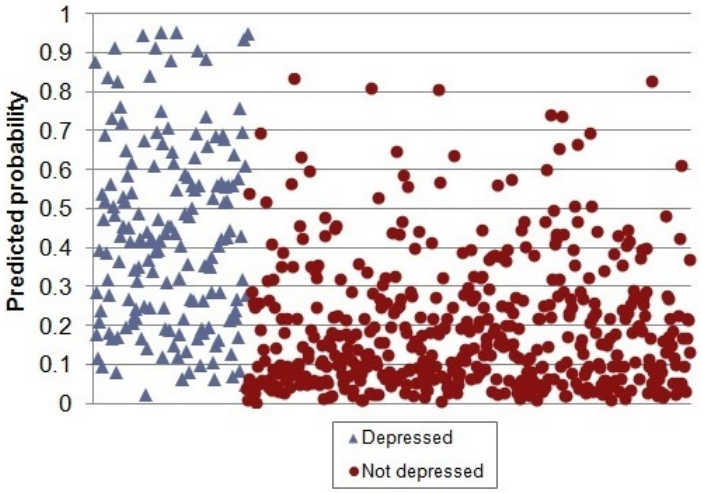

Figure 7: Predicted probabilities for Model 1.8c

5.3 Fitting of the second model

It is now time to fit the second model. The aim of this model is described in section 1.1. For the second model we consider 16 additional predictors to the 30 predictors considered for the first model. A list of all predictors for this model can be found in Appendix A2. Because of the hypothesis-generating and limited practical use of the second model the fitting of it will not be as careful as the fitting of the first model.

5.3.1 Purposeful selection of predictors

The selection strategy for the second model will be the same as for the first model, namely the purposeful selection process described in section 3.3, although we will remove more than one predictor at a time and not examine the $\Delta\hat{\beta}$-changes as closely.

We start by including all predictors with univariable likelihood-ratio p-value less than 25 percent into a multivariable logistic regression model. This model is summarized in Appendix A6. We delete the two predictors with the highest Wald p-values, *SexC* and *EmplUnd* (how understanding the employer of the parent was at the time of loss). We continue deleting two predictors at a time, until the AIC decreases more slowly and $R^2_{E,adj}$ increases more slowly, after which we delete predictors one at a time. The least significant predictor then has a p-value of 0.3216. Along the way there are some quite large changes in the parameter estimates, but many of the predictors affected by this are very insignificant and hence deleted. This deletion process is summarized in Table 7. The first model in the table is the model containing all predictors with univariable p-values less than 25 percent. We see that the AIC decreases up to the deletion of *Work*, and that $R^2_{E,adj}$ increases up to the deletion of *Prep* (if the parent was prepared that the cause of death might be suicide). This makes Model 2.7 and Model 2.10 interesting, but since Model 2.7 has so many insignificant predictors we will continue to the next step with only Model 2.10. Before we proceed

we reenter all the deleted predictors one by one into Model 2.10 to see if they may be significant in this smaller model. None of them are even close to being significant, so we move on to step 4 of the selection process.

Table 7: Summary of step 2 and 3 in the selection process for the second model

No.	**Predictors deleted from previous model**	**No. of predictors**	**No. of parameters**	**AIC**	**-2logL**	***c***	$R^2_{E,adj}$	**LR p-value larger vs smaller model**
2.1	None	30	61	671.864	549.864	0.833	0.18611	
2.2	SexC, EmplUnd	28	56	662.604	550.604	0.831	0.19179	0.9807
2.3	Worry1, Social	26	50	652.602	552.602	0.830	0.19712	0.9199
2.4	SuicCE, Supp4	24	47	647.313	553.313	0.829	0.20017	0.8706
2.5	AgeP, SelfinjCYB	22	43	640.962	554.962	0.828	0.20330	0.8000
2.6	ContactPC, ViolSuic	20	39	636.529	558.529	0.825	0.20388	0.4678
2.7	PrevLoss1	19	38	635.515	559.515	0.825	**0.2039**	0.3207
2.8	Prep	18	35	633.732	563.732	0.823	0.20227	0.239
2.9	PartnTD	17	34	633.174	565.174	0.821	0.20169	0.2298
2.10	Biologic	16	33	**633.127**	567.127	0.819	0.20042	0.1623
2.11	Work	15	31	633.920	571.920	0.814	0.19669	0.0910
2.12	Loc	14	27	634.116	580.116	0.805	0.19110	0.0847
2.13	Feel	13	26	635.292	583.292	0.802	0.18820	0.0747
2.14	Guardian	12	25	636.865	586.865	0.797	0.18478	0.0587
2.15	FysAct	11	21	637.378	595.378	0.791	0.17876	0.0745
2.16	Losstime	10	20	639.027	599.027	0.789	0.17523	0.0561
2.17	PrevLoss3	9	19	640.634	602.634	0.785	0.17176	0.0575
2.18	Inc	8	16	642.503	610.503	0.777	0.16527	0.0488

It is now time to revisit the predictors with univariable p-values greater than 25 percent. We add these one by one to Model 2.10 to see if they can contribute to this multivariable model. We also see if the categorical versions of the predictors *AgeC* and *AgeP* contribute in any way. Since none of these become significant or improve the AIC or $R^2_{E,adj}$ we continue with Model 2.10 to step 5 of the selection process.

In section 5.1 we concluded that the continuous predictor *Losstime* did not seem to be linear in the log-odds. Now we will investigate if there may be a linear relationship in the presence of the other predictors of Model 2.10. We therefore use the quantile method described in section 3.4.1 to create a plot of the parameter estimates with 95% Wald confidence intervals, for the categorical version of *Losstime* (which we call *LosstimeCat*). As seen in Figure 8 the relationship still does not seem to be linear, even when we take the confidence intervals into account. We include the categorical predictor *LosstimeCat* into the model and we also fit a cubic polynomial for the continuous predictor *Losstime*. The results can be seen in Table 8. Even if the cubic polynomial model has the smallest AIC we choose to continue with the model where

Losstime is categorical, since it is easier to interpret and there is not as much risk of overfitting the model to the data. Both the AIC and $R^2_{E,adj}$ improve substantially when we use the categorical version rather than the linear continuous version. Thus Model 2.10g is chosen as our *main effects model*.

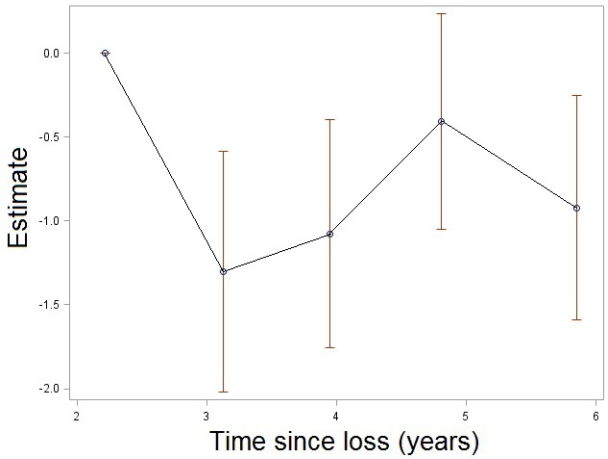

Figure 8: Parameter estimates with 95% Wald CI's from Model 2.10 plotted against midpoints of quantile groups of Losstime

Table 8: Evaluation of transformations of *Losstime* in Models 2.10

No.	**Loss-time**	**No. of parameters**	**AIC**	c	$R^2_{E,adj}$	**P-value of Losstime (Cat)**	**P-value of $Loss\text{-}Time^2$**	**P-value of $Loss\text{-}Time^3$**
2.1	Linear	33	633.127	0.819	0.20042	0.072		
2.10a	Categ.	36	624.113	0.83	0.21642	0.0015		
2.10b	Cubic	35	623.223	0.829	0.21628	0.0006	0.0013	0.0023

We have now reached the sixth step of the purposeful selection process, and it is time to check for interactions. Model 2.10 has 16 predictors and $\binom{16}{2} = 120$ possible two-factor interactions. These are added one by one to Model 2.10. The seven interactions that become significant are *SexP*Work, PrevLoss3*Supp3, PrevLoss3*Feel, PrevLoss3*Worry2, Supp2*Hobby, PrevLoss3*Hobby* and *EmplTD*Hobby*. The interactions *Guardian*Supp3, PremorbP* Supp3, PrevLoss3*FysAct, Supp3*Worry2* and *Guardian*PrevLoss3* are not significant but substantially improve the AIC and/or $R^2_{E,adj}$. All of these 12 interactions are together added to Model 2.10. We then remove insignificant interactions one by one. The result of this is presented in Table 9. We see that Model 2.10d has the smallest AIC, but this model still has several very insignificant interaction terms, so it will not be used. The last model (Model 2.10g) in the table only contains significant interactions, and the AIC and $R^2_{E,adj}$ is only slightly worse than for the other models, so since Model 2.10g is smaller and therefore easier

to interpret, we choose it as our *preliminary final model*. No interactions of a higher order become significant in this model. We note that when the interactions are added the p-value of the predictor *Supp2* (parent participating in talk with professional during first year after loss) increases with almost forty percentage points. It will not be deleted though, since the main effects are considered fixed at this point.

Table 9: Evaluation of interactions (second model)

No.	**Interactions added**	**No. of parameters**	**AIC**	c	$R^2_{E,adj}$
2.10	None	33	633.127	0.819	0.20042
2.10a	SexP*Work, Guardian*Supp3, PremorbP*Supp3, PrevLoss3*Supp3, PrevLoss3*Feel, PrevLoss3*FysAct, PrevLoss3*Hobby, PrevLoss3*Worry2, EmplTD*Hobby, Supp2*Hobby, Supp3*Worry2, Guardian*PrevLoss3	69	587.559	0.896	0.30909
2.10b	SexP*Work, Guardian*Supp3, PrevLoss3*Supp3, PrevLoss3*Feel, PrevLoss3*FysAct, PrevLoss3*Hobby, PrevLoss3*Worry2, EmplTD*Hobby, Supp2*Hobby, Supp3*Worry2, Guardian*PrevLoss3	68	586.092	0.896	0.30972
2.10c	SexP*Work, Guardian*Supp3, PrevLoss3*Supp3, PrevLoss3*Feel, PrevLoss3*Hobby, PrevLoss3*Worry2, EmplTD*Hobby, Supp2*Hobby, Supp3*Worry2, Guardian*PrevLoss3	64	582.652	0.893	0.30897
2.10d	SexP*Work, Guardian*Supp3, PrevLoss3*Supp3, PrevLoss3*Feel, PrevLoss3*FysAct, PrevLoss3*Hobby, PrevLoss3*Worry2, Supp2*Hobby, Supp3*Worry2, Guardian*PrevLoss3	60	592.381	0.886	0.29068
2.10e	SexP*Work, Guardian*Supp3, PrevLoss3*Supp3, PrevLoss3*Feel, PrevLoss3*Hobby, PrevLoss3*Worry2, Supp2*Hobby, Supp3*Worry2, Guardian*PrevLoss3	56	588.031	0.884	0.29114
2.10f	SexP*Work, PrevLoss3*Supp3, PrevLoss3*Feel, PrevLoss3*Hobby, PrevLoss3*Worry2, Supp2*Hobby, Supp3*Worry2, Guardian*PrevLoss3	55	588.018	0.883	0.28983
2.10g	SexP*Work, PrevLoss3*Feel, PrevLoss3*Hobby, PrevLoss3*Worry2, Supp2*Hobby, Supp3*Worry2, Guardian*PrevLoss3	54	588.868	0.881	0.28736

5.3.2 Model diagnostics for the second model

The Hosmer-Lemeshow χ^2 p-value for Model 2.10g is 0.3751, which is acceptable, but we must also evaluate the fit and influence of individual observations. In Figure 9 and 10 the measurements $\Delta\chi^2_j$, ΔD_j and $cidd_j$ (see section 3.4.7 for descriptions) are plotted. As seen in these plots observation number 665 seems to be extremely influential. If we investigate this observation closer we find that when deleted it changes 20 of the parameter estimates more than 15 percent. We also look closer at observations with $\Delta\chi^2_j > 15$, $\Delta D_j > 5$ and $cidd_j > 0.6$, and find 23 such observations.

Essentially the same observations that have high $\Delta\chi^2_j$, ΔD_j and/or $cidd_j$ have large Pearson and deviance residuals (see Appendix A8 for plots). In Table 10 the

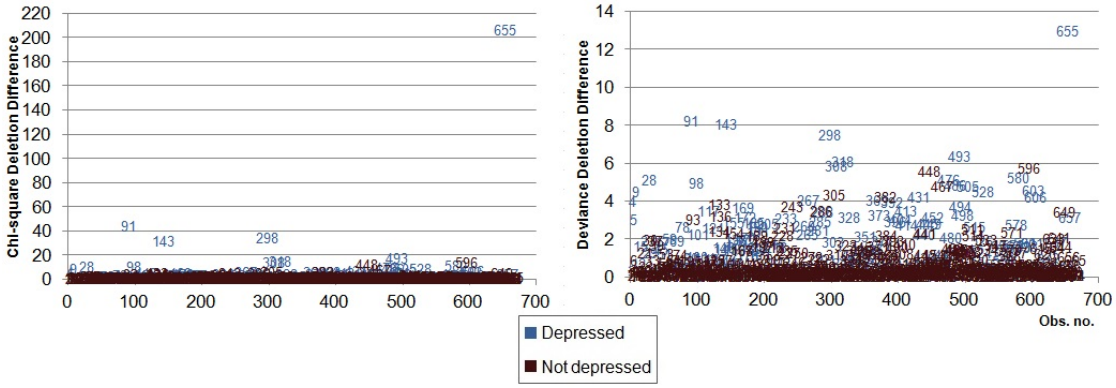

Figure 9: Plot of $\Delta\chi^2_j$ and ΔD_j for Model 2.10g

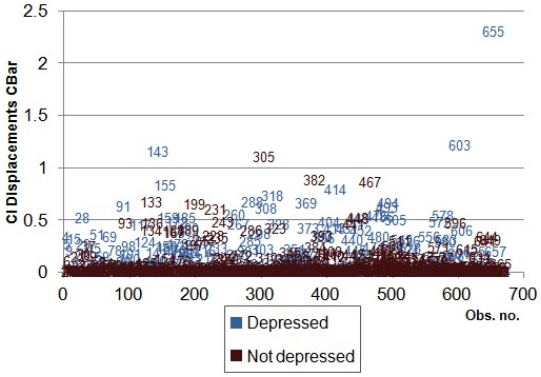

Figure 10: Plot of $cidd_j$ for Model 2.10g

results of fitting a model to the data where observation number 655 is deleted and another model to the data where the 23 influential observations are deleted, are presented. We see that the AIC and $R^2_{E,adj}$ improves drastically when the 23 observations are deleted, but even when only observation number 655 is deleted they improve substantially. However, nothing indicates that there is something wrong with any of the observations, and hence it is not acceptable to omit them from the analysis. As Hosmer et al. writes (2013, ch. 5, p. 199) "One should not simply lift the rug and sweep potentially inconvenient data under it, no matter what affect deletion might have on a fitted model." The consequence of including all the influential observations is for 45 of the 53 parameter estimates (excluding the intercept) a weakening of the effect, and for the other 8 parameters the effect is only very slightly intensified.

We have now chosen Model 2.10g as our second model. Model 2.10g contains the interactions presented in Table 9 and it also includes the 16 main effects *SexP, LosstimeCat, Guardian, PremorbP, PrevLoss3, EmplTD, Loc, Work, Inc*, *Supp2, Supp3* (if parent participated in talk with professional or group talks for bereaved during first year after loss), *Alc* (alcohol consumption of parent on an ordinal scale), *Feel* (if parent has a person to share their inner feelings with), *FysAct* (physical activity of parent on

Table 10: Influential observations refitting of Model 2.10g

Observations deleted	**AIC**	**Hosmer-Lemeshow p-value**	c	$R^2_{E,adj}$
None	588.868	0.3751	0.881	0.28736
Obs. no. 655	575.564	0.8464	0.888	0.30509
23 observations	457.274	0.5822	0.931	0.46269

an ordinal scale), *Hobby* (practicing of a hobby on an ordinal scale) and *Worry2* (the parent worrying during year before loss that the child might commit suicide). The model has a concordance index of 0.881, which is considered excellent according to the guidelines provided in section 3.4.6. The corresponding ROC-curve for Model 2.10g can be found in Appendix A8. In Table 11 the odds ratios significantly different from one (for comparison with reference groups) for Model 2.10g are presented with 95% Wald confidence intervals. All parameter estimates with p-values, and odds ratios for comparison with reference groups with accompanying confidence intervals for Model 2.10g can be found in Appendix A7. In Figure 11 the predicted probabilities for depressed and non-depressed parents are plotted. We see that the probabilities for the depressed parents in general are higher than for the non-depressed, but there are still depressed parents with low probabilities, as well as non-depressed parents with high probabilities. Most of the non-depressed parents have quite low probabilities, while the dispersion is larger for the depressed parents.

Table 11: Significant odds ratios and 95% Wald confidence intervals for Model 2.10g

Label	**OR estimate**	**95% Wald confidence interval**	
Work 2 vs 1 at SexP=2	0.019	0.002	0.166
Feel 1 vs 2 at PrevLoss3=1	2.544	1.297	4.988
Hobby 1 vs 3 at Supp2=1 PrevLoss3=1	6.966	1.733	28.01
Hobby 1 vs 4 at Supp2=1 PrevLoss3=1	3.067	1.181	7.965
Hobby 1 vs 5 at Supp2=2 PrevLoss3=1	13.437	1.961	92.095
Hobby 1 vs 4 at Supp2=1 PrevLoss3=2	15.899	2.83	89.334
Hobby 1 vs 4 at Supp2=2 PrevLoss3=2	8.439	1.575	45.213
Hobby 1 vs 5 at Supp2=2 PrevLoss3=2	12.122	1.04	141.335
Worry2 1 vs 2 at Supp3=1 PrevLoss3=2	0.162	0.036	0.736
Worry2 1 vs 2 at Supp3=1 PrevLoss3=1	7.623	2.121	27.4
Worry2 1 vs 3 at Supp3=1 PrevLoss3=1	17.953	2.731	118.029
Worry2 1 vs 3 at Supp3=2 PrevLoss3=2	0.073	0.005	0.98
Worry2 1 vs 2 at Supp3=2 PrevLoss3=1	7.351	1.318	40.993
Supp2 1 vs 2 at Hobby=3	0.16	0.04	0.644
Supp2 1 vs 2 at Hobby=4	0.34	0.164	0.708
Supp3 1 vs 2 at Worry2=3	0.018	0.002	0.163
Guardian 2 vs 1 at PrevLoss3=1	6.681	1.658	26.917

Continued on next page

Table 11 – continued from previous page

Label	OR estimate	95% Wald confidence interval	
LossTimeCat 5.85 vs 2.22	0.442	0.213	0.916
LossTimeCat 3.95 vs 2.22	0.284	0.134	0.604
LossTimeCat 3.13 vs 2.22	0.213	0.096	0.472
PremorbP 1 vs 2	6.049	3.123	11.717
EmplTD 1 vs 3	3.722	1.546	8.963
Loc 5 vs 1	5.239	2.206	12.447
Loc 4 vs 1	2.674	1.219	5.867
Loc 3 vs 1	2.851	1.335	6.087
Loc 2 vs 1	3.822	1.867	7.82
Alc 1 vs 2	0.384	0.166	0.886
Alc 1 vs 3	0.062	0.015	0.262
FysAct 1 vs 3	7.696	1.966	30.132
FysAct 1 vs 4	4.579	1.392	15.058
FysAct 1 vs 5	4.964	1.442	17.089

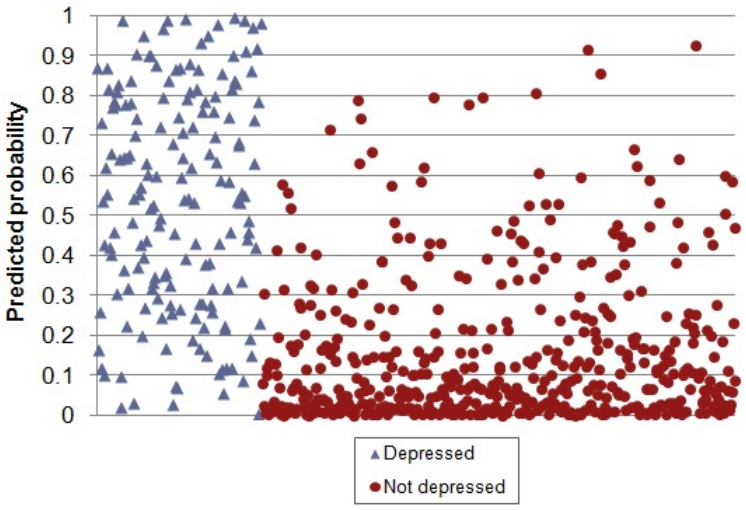

Figure 11: Predicted probabilities for Model 2.10g

6 Discussion

We have now chosen our first and second model. In the sections below the interpretations, strengths and weaknesses of these two models are discussed.

6.1 The first model

The effect of time since loss was not found to be linear as one might expect, and the effect seems to depend on if the parent was living with a partner at the time of loss. In Figure 12 the estimated probabilities for the different loss time categories are plotted dependent on whether the parent was living with a partner (=2) or not (=1), with the other predictors at their reference levels. From the estimates it seems that the first years after loss are harder for those not living with a partner at the time of loss. Though whether the parent was living with a partner at time of loss does not tell us anything about the parent's situation today or even a short time after the loss. This makes it harder to interpret the interaction between *LosstimeCat* and *PartnTD*. Ideally we would have continuous information of the relationship status of the parent, then we would be able to determine the effect of living with a partner and time since loss in a more certain way.

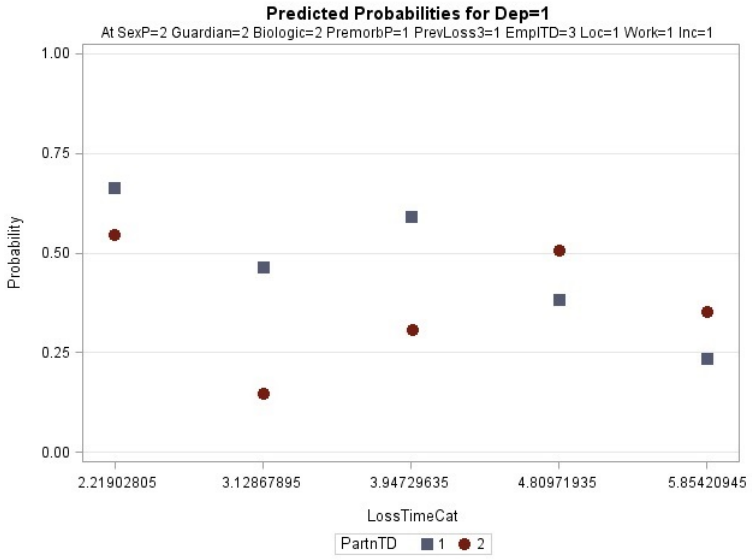

Figure 12: Plot of LosstimeCat by PartnTD

Effects that were somewhat expected are that parents who have been the guardians of their children in most of the upbringing have an increased risk of depression, and the same goes for biological parents. It seems logical that parents having been close to the child feel worse after the loss. Having had psychological morbidity more than ten years before the loss or having biological relatives who have committed suicide also increases the risk of depression, which seems reasonable from a clinical perspective.

An interesting finding in the first model is that for women being retired is associated with a lower risk of depression, compared to women who are employed, self-employed or students. For men we cannot see a similar effect. For both men and women, there is an indication that being unemployed, on sick-leave, or a social beneficiary increases the risk of depression compared to being employed, self-employed or a student, but this effect is not significant. We also find that for parents who are employed, self-employed or students, women have a higher risk of depression than men. For other work categories the effects are not significant and point in different directions. Note that the employment status of the parent is measured at the time of the survey rather than at the time of loss, so it is possible that changes have occurred after the loss, and in the worst case these changes could be related to the depression status of the parent. For example some of the parents may be on sick-leave *because* they are depressed, and then we have reverse causality. For this reason results for employment status should be interpreted cautiously, though we suspect that the work status of parents in general is fairly constant over the years. There is a predictor for employment status at the time of loss called *EmplTD*, but the categories are not exactly as we would want them. Although even for this predictor we find that being unemployed or on full-time sick-leave at the time of loss is associated with an increased risk of depression.

There is an indication that having an income of 200 000 SEK or more is associated with a decreased risk, compared to having an income of less than 100 000 SEK. This effect is not significant, but if we compare the higher income groups to the second highest it does seem that the wealthier parents have a lower risk of depression. Since income is measured at the time of the survey rather than at the time of loss, we should interpret these results cautiously, though we suspect that the income of the parents generally have not changed very much since the time of loss.

Comparing urban and rural areas, it seems that living in a rural area lowers the risk of depression, but not all of the comparisons are significant, and if other pairwise comparisons are made the effect is not always in favour of the less populated area. Also the parent's residence area is not measured at the time of loss, even if we suspect that it is fairly constant over time.

When assessing the fit of the first model we found that there were some observations that seemed to be influential and have poor fit. When these were deleted most of the effects became stronger, and the prediction capacity was improved. Though we have no reason to exclude the observations since we could not find anything erroneous in them. Including them makes the fit questionable, but since effects seem to be underestimated rather than overestimated we will not risk drawing too far-reaching conclusions from the model.

6.2 The second model

In the second model there are four interactions between having a family history of suicide and the predictors *Feel, Hobby, Worry2* and *Guardian*. This implies that the effect of these four predictors depend on the parent having a family history of suicide.

For parents having no such family history it seems that not having a person to share their inner feelings with increases the risk of depression. For parents having had suicide in the family the effect seems to be reverse, although this is not statistically significant. It is important to note that the parent having or not having anyone to share their inner feelings with reflects the parent's situation at the time of the survey rather than at the time of loss, which means that the possible prevalence of depression can have influenced the parent's perception of having people to share their feelings with. Therefore the association may be overestimated.

Having a hobby is associated with a decreased risk of depression for parents without a family history of suicide. For parents with a family history of suicide the effect of having a hobby is incoherent and the confidence intervals are wide, though the significant effects go in the same direction as for parents without a family history of suicide. For these effects we once again have the problem of the predictor reflecting the situation at the time of the survey rather than at the time of loss, and it is very likely that the practicing of a hobby is influenced by the possible prevalence of depression. Hence we cannot for sure say that practicing a hobby has a protective effect based on these data. We have a similar problem for physical activity − it is associated with a lower risk of depression, but there may be reverse causality. We still have reason to believe that both practicing a hobby and engaging in a physical activity has a protective effect on the psychological health, although our data may have overestimated this effect.

Excessive drinking is associated with the prevalence of depression, but here too there may be a case of reverse causality − alcohol consumption may increase because of depression.

There are some predictors for which the effects point in different directions and are hard to interpret. The parent worrying during the year before loss that the child might commit suicide can be both good and bad for the risk of depression. For parents having a family history of suicide it seems that worrying increases the risk of depression, while for parents not having a family history of suicide it seems to decrease the risk, though the confidence intervals for the odds ratios are wide and the effects are somewhat incoherent. Perhaps there are other factors connected to the worrying that are not measured. A parent having worried may be more prepared for the loss of the child and therefore handling the grief better, while at the same time long-term worrying may have weakened the psychological health of the parent, making them more susceptible to psychological morbidity.

Having been the guardian of the child during most of the upbringing is associated with an increased risk of depression for parents not having a family history of suicide,

while the effect seems to be reverse for parents having a family history of suicide, though the second effect is not significant. We will not speculate in the reason for this, only note that it may be interesting to investigate further in future studies.

The effect of having talked to a professional during the first year after loss or participated in group talks for bereaved is counterintuitive since it seems to increase the risk of depression. Though we have strong reason to believe that the parents seeking help and support after the loss are the ones that are the most psychologically affected by it, and therefore it is not unlikely that they are depressed some years later.

In the discussion above about the first model we concluded that including extreme and/or influential observations result in underestimating rather than overestimating the effects, and the same goes for the second model. Hence even though the fit is not the best, we do not risk drawing too far-reaching conclusions.

6.3 Comparison between the first and second model

The first model is smaller, with only two interaction terms and 11 main effects, while the second model is larger with 7 interactions and 16 main effects. It is therefore harder to interpret the effects in the second model, and some effects are counterintuitive. Both models produce a wide range of predicted probabilities, see Figure 7 and 11, though the second model seems better at discriminating between depressed and non-depressed parents, which is confirmed by the concordance index. In Table 12 the AIC, concordance index and $R^2_{E,adj}$ of the two models are compared. The second model performs better than the first for all of these measures, although the first model still has a good concordance index. If we assess the ROC-curves in Appendix A8 it seems that we can achieve a higher sensitivity and specificity with the second model compared to the first. We must keep in mind that for the second model there may be reverse causality for some predictors, and the purpose of the second model is not to predict the risk of depression in a direct manner. Though if associations with possible reverse causality are examined in future research, a prediction model containing some predictors of the second model, with a better prediction capacity than the first model, may be achievable.

Table 12: AIC, concordance index and $R^2_{E,adj}$ of the first and second model

Model	**AIC**	$\boldsymbol{c}$	$\boldsymbol{R^2_{E,adj}}$
First model (1.8c)	646.022	0.803	0.17657
Second model (2.10g)	588.868	0.881	0.28736

It would be interesting to see if some predictors have different effects in the first model compared to the second model. Could it be that some predictors reduce the risk of depression in the first model and increase the risk in the second? Since different predictors are included in different interaction terms in the first compared to the second model, we cannot compare all the estimates directly. For example the

predictor *PrevLoss3* is included in four interactions in the second model and therefore the effect of it cannot be expressed in a single parameter estimate for comparison with the first model, even though it is a dichotomous predictor. The parameter estimate for *PrevLoss3* in the second model is only valid on its own when the interaction predictors are at their reference levels. Therefore is it hard to compare the effect of *PrevLoss3* in the first compared to the second model. Though the predictors that are included in the same interactions for both models, and the predictors that are not included in interactions in any of the models, can be compared pretty straight forward. One of the levels of two of these predictors have a positive effect in one model and a negative effect in the other, but both effects are close to zero and insignificant. For the other comparable effects the first model has the smaller effect in 12 of the 13 possible comparisons. Thus the effects in the first model discussed above are enhanced in the presence of the additional predictors in the second model.

The predictor *PartnTD*, that measures whether the parent was living with a partner at the time of loss, is included in the first model, but not in the the second. However, in the second model, the predictor *Feel*, that measures whether the parent has someone to share their inner feelings with is included. It could be that *Feel* explains most of the variation previously explained by *PartnTD*. The predictor *Biologic*, measuring whether the parent is the biological parent of the child, is included in the first but not the second model. We cannot find another predictor in the second model that seems to "replace" *Biologic* in the same way as for *PartnTD* and *Feel*. There are only a few parents that are not biological parents of their children, so in the presence of the variables in the second model, perhaps *Biologic* simply did not explain enough of the variation.

There were some predictors that we expected would have an effect, but did not become significant in any of the models. These were age of the parent and child, religious beliefs, and the presence of siblings to the deceased child. It is possible that the effect of these predictors is better explained by other predictors. For example the parent's age became significant in the univariable analysis, but not in the multivariable. It could be that the effect of age is better explained by the parent's work status. Though the presence of siblings did not become significant in either the univariable or the multivariable analysis, and the same goes for the predictor for believing in God. Only a few of the deceased children did not have siblings, so perhaps we would have found an effect if there were more participants in the study. Regarding the parents' religious beliefs the parents were simply asked if they believed in God, which could be perceived as exclusionary of other beliefs than Christianity, hence there is a risk that parents with other beliefs were not detected.

For both models we based the prediction capacity on the concordance index and $R^2_{E,adj}$. There are more methods of evaluating the prediction capacity, for example to randomly divide the data set into two parts, fit a model for one part and check how well it fits the other part. This method was not implemented here since the number of predictors is rather high and the number of observations is not huge, hence the power

would be reduced too much. When refraining from implementing this method we must be aware that there is a risk of overfitting the models to the data, implying that predictions for new observations may not be the best. Though, as we noted when discussing the two models in the sections above, the effects may be somewhat underestimated because of the influential observations, a fact that speaks against overfitting.

Both models use the logit link, but there are several other possible choices that still keep the fitted probabilities within the interval [0,1], such as the probit link which use the standard normal cumulative distribution function $\Phi(\cdot)$ according to $\pi(x) = \Phi(\alpha+\beta x)$. We have chosen the logit link because it is customary in the epidemiological and behavioural research fields and it conveniently translates into odds-ratios that are easy to interpret.

6.4 Parents of the same child

The data for this paper is collected from all parents who have lost a child through suicide between the years 2004 and 2007 (except for the nonresponse), so there will inevitably be some parents connected to the same child. Since the data is completely anonymized there is no way of distinguishing these pairwise relationships, or even knowing how many there are. Hypothetically this could result in a number of pairwise observations being very similar and influencing estimates too much. However, this does not have to be the case. There are predictors that will be identical for parents of the same child, such as time since loss and age of the child, but most of the predictors are not child-specific, and the relationship between a parent and a child is individual for each parent. Also it is not at all sure that being parents of the same child has the same effect on the development of depression for each parent. Of course it would be desirable to be able to map these pairwise relationships, but even though we cannot, we have little reason to believe that the effect of them is severe for the outcome of this analysis.

6.5 Unit and item nonresponse

The data in this paper is based on a survey study where the nonresponse rate was 27 percent. Nonresponse does not have to impair the quality of a study. If the nonrespondents do not differ in any significant way from the respondents the sample is still representative for the population and inferences are possible to make. Though if the nonrespondents do differ significantly from the respondents, the effect on the quality of the study can be severe. In this study we lack information about the nonrespondents regarding background variables such as age, sex, socioeconomic factors etc. It would be desirable to have such information to enable a simple comparison between respondents and nonrespondents. We have reason to believe that the prevalence of depression is underestimated, since many of the parents who declined to participate stated psychological distress or ill-health as reason (Omerov 2014). This could indicate that our estimates are quite cautious and conservative, but it does not have to since it

all depends on how the predictors are distributed among the nonrespondents compared to the respondents.

Biemer & Lyberg (2003, ch. 3) propose a number of methods for reducing nonresponse, including gifts or economic compensation, lessen the response burden for nonrespondents and recontacting nonrespondents. However, for ethical reasons, none of these methods can be applied in this study. If a parent declines to participate, their reasons cannot be questioned, and it would be disrespectful and insensitive to try to pursue the parent into participating.

Apart from unit nonresponse (when a respondent declines participating) there were some item nonresponse (when a respondent skips certain questions) which was discussed in detail in section 4.4. The imputation method used was mode and median imputation, a method that only take into account the data structure on a general level. Other methods were discussed but not implemented, partly because of time constraints. It would have been desirable to perform a sensitivity analysis based on a few other imputation methods.

A way to completely avoid nonresponse is to conduct a register based study, which also has the advantage of providing more objectively measured data. For example child-specific predictors measuring the child's previous contact with the health care would be reported by the health care, rather than reflecting the parent's perception which may not be completely accurate. Though a register based study has the disadvantage of not being able to measure predictors of a more personal nature, such as the relationship between parent and child, feelings, religious beliefs etc. Also the indicator for depression would be based on medical diagnoses and use of medication, thus lacking information on parents that have not been diagnosed but are depressed according to the patient health questionnaire.

7 Conclusion and future perspectives

With the purpose of predicting the risk of depression among parents who have lost a child through suicide, we have used purposeful selection to fit two different prediction models. The first and simpler model contains predictors that are easy to measure, and are for the most part known at the time of loss. This model is designed to enable the health care to estimate the risk of depression for parents who have recently lost a child. The second and more complex model contains predictors that are influenced by mood, and may be unknown at the time of loss. This model has a hypothesis-generating purpose, and has less practical use than the first one. Both models have good predictive qualities within the dataset, though they remain to be tested on new observations. Some effects may be underestimated due to influential observations.

We found interesting effects in both models, and some of these effects raise further questions. There are some implications for future research and some for further analysis of these specific data. We have previously discussed the need for a sensitivity analysis of the imputation method, and for a way to control for parents of the same child. It would also be interesting to more closely compare the group of bereaved parents to the non-bereaved control group of 377 parents, by fitting a logistic regression model to each group. The predictors used would be the ones measured for both groups, hence predictors specific to the loss would not be possible to include. It would then be investigated if the same predictors become significant in both groups and if the effects point in the same direction.

In the second model we found some associations that may express reverse causality, such as alcohol consumption, physical activity and practicing a hobby. It is not surprising that these predictors are associated with the prevalence of depression, but there is a need to look into the directions of the associations. Does for example practicing a hobby decrease the risk of depression for bereaved parents, or is the only effect that depression takes away the energy and will to practice a hobby? We also found that talking to a professional or participating in group talks for bereaved was associated with an increased risk of depression, which at a first look is very counterintuitive, but is probably due to the fact that the parents who feel the worst seek support. It would be interesting for the health care to look into the true effect of group talks for bereaved, talking to professionals and other means of support, perhaps by studying the same parents over several years. Also, with a longitudinal study, interactions with time since loss and other predictors may be discovered and more closely examined. However, a longitudinal study investigating effects of means of support would require an early contact with bereaved parents and perhaps even a randomized trial, which may not be possible due to ethical reasons.

Suicide-bereaved parents are more susceptible to psychological morbidity than both non-beraved and bereaved by natural causes, and if the health care is able to identify high risk parents, preventative measures can be taken. Essentially this boils down to the parents' possibilities of leading a functioning everyday life. It is therefore crucial that research in this field is continued and deepened.

Acknowledgements

First of all I would like to thank Tommy Nyberg, my supervisor at Karolinska Institutet, for being so generous with your time, commitment and valuable guidance, and for introducing me to this very important and interesting subject. Thank you Pernilla Omerov for letting me use your data, and for your time and great knowledge. A big thank you to the Division of Clinical Cancer Epidemiology at Karolinska Institutet for being so welcoming. Thank you Tove Bylund Grenklo − without you this paper would not have been written. I also wish to thank my supervisor Tom Britton for advice and support. Last but not least, thank you Gustav Sandberg and Mika Romanus for taking time to read and give critique.

References

Agresti, A. (2002): *Categorical Data Analysis.* Wiley Series in Probability and Statistics

Allison, P.D. (2002): *Missing data.* Thousand Oaks: SAGE Publications, Inc. doi: http://dx.doi.org/10.4135/9781412985079

Bennett, D.A. (2001): How can I deal with missing data in my study?. *Australian and New Zealand Journal of Public Health*, 25: 464–469. doi: 10.1111/j.1467-842X.2001.tb 00294.x

Biemer, P.P., Lyberg, L.E. (2003): *Introduction to survey quality.* Hoboken, NJ: John Wiley & Sons, Inc.

Bolton, J.M., Au, W., Leslie, W.D., et al. (2013): Parents Bereaved by Offspring Suicide: A Population-Based Longitudinal Case-Control Study. *JAMA Psychiatry.* 2013;70(2):158-167. doi:10.1001/jamapsychiatry.2013.275

Bursac, Z., Gauss, C.H., Williams, D.K. (2008): Purposeful selection of variables in logistic regression. *Source Code for Biology and Medicine.* doi:10.1186/1751-0473-3-17

Dong, Y., Peng, C.Y.J. (2013): Principled missing data methods for researchers. *SpringerPlus* 2013 2:222. doi:10.1186/2193-1801-2-222

Dyregrov, K., Nordanger, D., Dyregrov, A. (2003): Predictors od psychosocial distress after suicide, SIDS and accidents. *Death Studies*, 27: 1437165, doi: 10.1080/074811803 90136973

Feigelman, W., Jordan, J.R., Gorman, B.S. (2008): How they died, time since loss, and bereavement outcomes. *OMEGA*, Vol. 58(4) 251-273, Inc. doi: 10.2190/OM.58.4.a

Gertheiss, J., Tutz, G. (2009): Penalized Regression with Ordinal Predictors. *International Statistical Review*, 77, 3, 345–365 doi:10.1111/j.1751-5823.2009.00088.x

Hosmer, D.W., Lemeshow, S., Sturdivant, R.X. (2013): *Applied Logistic Regression.* Wiley Series in Probability and Statistics

Jönsson, P., Wohlin, C. (2006): Benchmarking k-Nearest Neighbour Imputation with Homogeneous Likert Data. *Empirical Software Engineering: An International Journal*, Vol. 11, No. 3, pp. 463-489. doi:10.1007/s10664-006-9001-9

Kessing, L.V., Agerbo, E., Mortensen, P.B. (2003): Does the impact of major stressful life events on the risk of developing depression change throughout life?. *Cambridge University Press, Psychological Medicine*, 2003, 33, 1177–1184. doi: 10.1017/S0033291703007852

Kroenke, K., Spitzer, R.L., Williams, J.B.W., Löwe, B. (2010): The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. *General Hospital Psychiatry*, 32(4), 345-359. doi:10.1016/j.genhosppsych.2010.03.006

Li, J., Laursen, T.M., Precht, D.H, Olsen, J., Mortensen, P.B. (2005): Hospitalization for Mental Illness among Parents after the Death of a Child. *New England Journal of Medicine*; 352:1190-1196, March 24, 2005, doi:10.1056/NEJMoa033160

Li, J., Precht, D.H., Olsen, J. (2003): Mortality in parents after death of a child in Denmark: a nationwide follow-up study. *The Lancet*, Volume 361, Issue 9355, Pages 363 - 367, 1 February 2003, doi:10.1016/S0140-6736(03)12387-2

Mittlböck, M., Schemper, M. (1996): Explained variation for logistic regression. *Statistics in Medicine*, Volume 15, Issue 19, pages 1987–1997, 15 October 1996, doi: 10.1002/(SICI)1097-0258(19961015)15:19<1987::AID-SIM318>3.0.CO;2-9

Omerov, P., Steineck, G., Nyberg, T., et al. (2013): Psychological morbidity among suicide-bereaved and non-bereaved parents: a nationwide population survey. *BMJ Open* 2013;3;e003108. doi:10.1136/bmjopen-2013-003108

Omerov, P. (2014): *Parents who have lost a son or daughter through suicide – towards improved care and restored psychological health.* Karolinska Institutet, uri: http://hdl.handle.net/10616/41852

Stroebe, M., Schut, H., Stroebe, W. (2007): Health outcomes of bereavement. *The Lancet*, Volume 370, Issue 9603, Pages 1960 - 1973, 8 December 2007, doi:10.1016/S0140-6736(07)61816-9

Tutz, G. (2012): *Regression for Categorical Data.* Cambridge University Press

Appendix

A1 Merging of categories

In Table 13 the merging of some categories is presented, see section 4.3 for more details.

Table 13: Merging of categories

Name	**Description**	**Old coding**	**New coding**
ContactPC	Did parent and child have contact during year before loss?	1=No 2=Occasionally 3=Less than monthly 4=Monthly 5=Weekly 6=Daily	1=No/Occationally/ Less than monthly 2=Monthly 3=Weekly 4=Daily
SuicCYB	Did child attempt suicide during year before loss?	1=Do not know 2=No 3=Yes, but no contact with medical care 4=Yes, and contact with medical care one or several times	1=Do not know 2=No 3=Yes, with or without contact with medical care
Edu	Education of parent (today)	1=Less than elementary school 2=Elementary school 3=High school 4=University/college ($<$ 3 years) 4=University/college ($\geq$ 3 years)	1=Elementary school or less 2=High school 3=University/college ($<$ 3 years) 4=University/college ($\geq$ 3 years)
Work	Parent's work situation today	1=Employed 2=Old age pension 3=Disability pension 4=Unemployment benefits 5=Student 6=Social beneficiary 7=Other	1=Employed/Self-employed/Student 2=Old age pension 3=Disability pension/Unemployment benefits/Social beneficiary/Other
Social	Has parent met friends or aqcuaintances and/or practiced activity with others during last year?	1=No 2=Less than monthly 3=Monthly 4=Weekly 5=Daily	1=No/Less than monthly 2=Monthly 3=Weekly 4=Daily

A2 List of variables

Table 14 lists all variables considered when building the first model, along with type of variable and possible values. It also includes the response variable *Dep*. Table 15 lists the additional predictors considered when building the second model, along with type of variable and possible values.

Table 14: List of variables, first model (including response)

Name	**Description**	**Type**	**Coding**
AgeC	Age of child at time of death, takes integer values from 15 to 31	Cont.	
AgeP	Age of parent at time of questionnaire, takes integer values from 40 to 81	Cont.	
Biologic	Is parent the biologic parent of the child?	Dichot.	1=No 2=Yes
ContactPC	Did parent and child have contact during year before loss?	Ord.	1=No/Occationally/ Less than monthly 2=Monthly 3=Weekly 4=Daily
Dep	Indicator for depression (response variable)	Dichot.	1=Depressed 2=Not
Edu	Education of parent (today)	Ord.	1=Elementary school or less 2=High school 3=University/college ($<$ 3 years) 4=University/college ($\geq$ 3 years)
EmplTD	Was parent employed at time of death?	Polychot.	1=Parent was on full time sick leave at time of loss 2=Parent was unemployed at time of loss 3=Other
God	Does parent believe in God?	Dichot.	1=No 2=Yes
Guardian	Was parent the guardian of the child during most of the upbringing?	Dichot.	1=No 2=Yes
ID	Individual survey ID-number of parent, Takes integer values from 13 to 1936	Cont.	
Inc	Parent's income today	Ord.	1=0-99 000SEK 2=100 000-199 000SEK 3=200 000-399 000SEK 4=400 000+SEK

Continued on next page

Table 14 – continued from previous page

Name	Description	Type	Coding
LivePC	Did parent and child live together during year before loss?	Ord.	1=No 2=Part-time 3=Full time
Loc	Parent's area of residence today	Ord.	1=Rural area 2=Population less than 10000 3=Pop. less than 50000 4=Pop. less than 200000 5=Stockholm/Gothenburg/ Malmö
Losstime	Time (in years) since loss of child at time of questionnaire, takes values from 1.7 to 6.5	Cont.	
PartnTD	Did the parent live with a partner at time of loss?	Dichot.	1=No 2=Yes
PremorbP	Did parent suffer from psychological morbidity more than ten years before loss?	Dichot.	1=Yes 2=No
PrevLoss1	Has parent lost an important person during the 10 years before loss?	Dichot.	1=No 2=Yes
PrevLoss3	Have any of the parent's biological relatives commited suicide?	Dichot.	1=No 2=Yes
PsychCare CEYB	Did the child have contact with psychiatric care?	Polychot.	1=Do not know 2=No 3=Contact, but not admitted to clinic 4=Contact and admitted to clinic year before suicide and/or earlier
SelfInjCYB	Did the child self-injure him-/herself during year before death?	Polychot.	1=Do not know 2=No 3=Yes, but no contact with medical care 4=Yes, and contact with medical care one or several times
SexC	Gender of child	Dichot.	1=Male 2=Female
SexP	Gender of parent	Dichot.	1=Male 2=Female
Siblings	Did the deceased child have siblings?	Dichot.	1=No 2=Yes, one or more
SuicCE	Did child attempt suicide before year before loss?	Polychot.	1=Do not know 2=No 3=Yes, one or several
SuicCYB	Did child attempt suicide during year before loss?	Polychot.	1=Do not know 2=No 3=Yes, with or without contact with medical care

Continued on next page

Table 14 – continued from previous page

Name	Description	Type	Coding
SwedP	Is parent born in Sweden?	Dichot.	1=No 2=Yes
SwedPP	Are parent's parents born in Sweden?	Dichot.	1=No 2=Yes
View	Did the parent view the body in a formal setting?	Dichot.	1=No 2=Yes
ViolSuic	Was the suicide violent?	Dichot.	1=Yes 2=No
Wittn	Did parent wittness suicide/find the body/see child at scene of suicide?	Dichot.	1=Yes 2=No
Work	Parent's work situation today	Polychot.	1=Employed/Self-employed/ Student 2=Old age pension 3=Disability pension/ Unemployment benefits/ Social beneficiary/Other

Table 15: List of variables, second model

Name	Description	Type	Coding
Alc	Drinking habits of parent, evaluated with AUDIT questionnaire, maximum score is 40	Ord.	1=AUDIT score 0-7 2=AUDIT score 8-15 3=AUDIT score 16+
EmplUnd	Did the employer of the parent show understanding during first year after loss?	Ord.	1=No 2=A little 3=Moderately 4=A lot
Feel	Has parent got anyone to share his/her inner feelings with?	Dichot.	1=No 2=Yes
FysAct	Has parent performed in any fysical activity more than 30 minutes during last year?	Ord.	1=No 2=Less than monthly 3=Monthly 4=Weekly 5=Daily
Hobby	Has parent practiced a hobby during last year?	Ord.	1=No 2=Less than monthly 3=Monthly 4=Weekly 5=Daily
Prep	When parent was informed of child's death, was parent prepared that it could be suicide?	Ord.	1=No 2=A little 3=Moderately 4=A lot
PrevLoss2	Has parent lost an important person after loss of child?	Dichot.	1=No 2=Yes

Continued on next page

Table 15 – continued from previous page

Name	Description	Type	Coding
RelatPCYB	Did parent and child have a good relationship the year before loss?	Ord.	1=No 2=A little 3=Moderately 4=A lot
Social	Has parent met friends or aqcuaintances and/or practiced activity with others during last year?	Ord.	1=No/Less than monthly 2=Monthly 3=Weekly 4=Daily
Supp1	Did the parent meet a professional to discuss possible reasons for the suicide?	Dichot.	1=No 2=Yes
Supp2	Did parent participate in talk with professional during year after loss?	Dichot.	1=No 2=Yes
Supp3	Did parent participate in group talks for bereaved?	Dichot.	1=No 2=Yes
Supp4	Has parent been in contact with association for bereaved?	Dichot.	1=No 2=Yes
TalkTD	Did the parent talk about the suicide to a person close to him/her at time of death?	Ord.	1=No 2=A little 3=Moderately 4=A lot
Worry1	Did the parent worry about psychic health of child during year before loss?	Ord.	1=No 2=A little 3=Moderately 4=A lot
Worry2	Was parent worried that the child might attempt suicide during year before loss?	Ord.	1=No 2=A little 3=Moderately 4=A lot

A3 Univariable logistic regression

The results from the fitting of the univariable logistic regression models are presented in Table 16 with LR statistics, and corresponding p-values. Every row thus represents the result of fitting a simple logistic regression model to that predictor and the response variable *Dep* as exemplified below with the predictor $AgeC$:

$$logit(\pi(AgeC_i)) = \beta_0 + \beta_{AgeC} \cdot AgeC_i.$$

Table 16: Univariable logistic regression

Name	**Df**	**LR**	**P-value**
Predictors considered for the first model:			
AgeC	1	0.6115	0.4342
AgeP	1	4.7228	0.0298
Biologic	1	7.7470	0.0054
ContactPC	3	6.8282	0.0776
Edu	3	3.6557	0.3011
EmplTD	2	26.4661	< 0.0001
God	1	0.7885	0.3745
Guardian	1	3.6767	0.0552
ID	1	1.0858	0.2974
Inc	3	32.4421	< 0.0001
LivePC	2	2.7690	0.2504
Loc	4	5.6436	0.2274
Losstime	1	4.3222	0.0376
PartnTD	1	9.0435	0.0026
PremorbP	1	43.5043	< 0.0001
PrevLoss1	1	1.5181	0.2179
PrevLoss3	1	5.9424	0.0148
PsychCareCEYB	3	0.2490	0.9693
SelfInjCYB	3	6.8218	0.0778
SexC	1	1.6979	0.1926
SexP	1	21.2813	< 0.0001
Siblings	1	0.0047	0.9453
SuicCYB	2	0.2276	0.8925
SuicCE	2	4.3812	0.1118
SwedP	1	0.0003	0.9872
SwepPP	1	1.0137	0.3140
View	1	0.1134	0.7363
ViolSuic	1	1.9017	0.1679
Wittn	1	0.1719	0.6784
Work	2	30.1394	< 0.0001
Additional predictors considered for the second model:			
Alc	2	9.3791	0.0092

Continued on next page

Table 16 – continued from previous page

Name	Df	LR	P-value
EmplUnd	4	9.7599	0.0447
Feel	1	8.7116	0.0032
FysAct	4	19.8676	0.0005
Hobby	4	26.7207	< 0.0001
Prep	3	8.6536	0.0343
PrevLoss2	1	0.6190	0.4314
RelatPCYB	3	0.6119	0.8937
Social	3	12.4072	0.0061
Supp1	1	0.1771	0.6739
Supp2	1	14.8107	0.0001
Supp3	1	9.0435	0.0026
Supp4	1	12.4522	0.0004
TalkTD	3	4.0625	0.2548
Worry1	3	5.7735	0.1232
Worry2	3	8.4627	0.0374

A4 Step 2 multivariable model (first model)

In Table 17 the results from fitting the multivariable model containing all predictors with univariable LR p-value less than 25 percent are presented. This is part of step 2 in the purposeful selection process for fitting the first model, see section 5.2.1.

Table 17: Summary of model containing predictors with univariable LR p-value less than 25%

Predictor	**Df**	**Wald p-value**
AgeP	1	0.596
Biologic	1	0.111
ContactPC	3	0.5827
EmplTD	2	0.1167
Guardian	1	0.1387
Inc	3	0.013
Loc	4	0.0904
Losstime	1	0.0266
PartnTD	1	0.093
PremorbP	1	< 0.0001
PrevLoss1	1	0.3687
PrevLoss3	1	0.0516
SelfinjCYB	3	0.5884
SexC	1	0.6221
SexP	1	0.1893
SuicCE	2	0.477
ViolSuic	1	0.3432
Work	2	0.1104

Fit and prediction measures

AIC	c	$R^2_{E,adj}$
680.567	0.776	0.13454

A5 Parameter estimates and odds ratios for Model 1.8c

In Table 18 the parameter estimates, standard errors and Wald p-values for Model 1.8c are presented. In Table 19 odds ratios with 95% Wald confidence intervals are presented.

Table 18: Estimates and p-values for Model 1.8c

Parameter	**Level**	**Level**	**Estimate**	**Standard Error**	**Wald p-value for level estimate**	**DF for predictor**	**Wald p-value for predictor**
Intercept			0.1706	0.6175	0.7823		
SexP*Work	1	3	0.4915	0.5867	0.4022	2	0.0384
SexP*Work	1	2	2.405	0.9612	0.0123		
LosstimeCat* PartnTD	5.85	1	-1.0926	0.7607	0.1509	4	0.0114
LosstimeCat* PartnTD	4.81	1	-1.0088	0.6876	0.1424		
LosstimeCat* PartnTD	3.95	1	0.669	0.7595	0.3784		
LosstimeCat* PartnTD	3.13	1	1.1053	0.734	0.1321		
SexP	1		-0.6839	0.2749	0.0128	1	0.0128
LosstimeCat	5.85		-0.7691	0.3635	0.0344	4	0.0003
LosstimeCat	4.81		-0.1513	0.3571	0.6718		
LosstimeCat	3.95		-0.9868	0.3696	0.0076		
LosstimeCat	3.13		-1.9339	0.4794	< 0.0001		
Guardian	1		-0.7333	0.4773	0.1245	1	0.1245
Biologic	1		-1.3113	0.779	0.0923	1	0.0923
PremorbP	2		-1.3468	0.2796	< 0.0001	1	< 0.0001
PartnTD	1		0.5176	0.4874	0.2883	1	0.2883
PrevLoss3	2		0.5393	0.2506	0.0314	1	0.0314
EmplTD	1		1.0398	0.384	0.0068	2	0.0247
EmplTD	2		0.4147	0.5547	0.4547		
Loc	5		0.963	0.3555	0.0067	4	0.0133
Loc	4		0.4231	0.3453	0.2205		
Loc	3		0.5126	0.3247	0.1145		
Loc	2		0.9858	0.307	0.0013		
Work	3		0.1755	0.3828	0.6467	2	0.0106
Work	2		-2.4307	0.8484	0.0042		
Inc	4		-0.9051	0.594	0.1276	3	0.0038
Inc	3		-0.09	0.4899	0.8542		
Inc	2		0.6683	0.5028	0.1838		

Table 19: Odds ratio estimates and confidence intervals for Model 1.8c

Label	**OR estimate**	**95% Wald confidence interval**	
Work 3 vs 1 at SexP=1	1.948	0.699	5.43
Work 2 vs 1 at SexP=1	0.975	0.327	2.906
Work 3 vs 1 at SexP=2	1.192	0.563	2.524
Work 2 vs 1 at SexP=2	0.088	0.017	0.464
LosstimeCat 5.85 vs 2.22 at PartnTD=1	0.155	0.042	0.578
LosstimeCat 4.81 vs 2.22 at PartnTD=1	0.313	0.099	0.995
LosstimeCat 3.95 vs 2.22 at PartnTD=1	0.728	0.197	2.687
LosstimeCat 3.13 vs 2.22 at PartnTD=1	0.437	0.143	1.337
LosstimeCat 5.85 vs 2.22 at PartnTD=2	0.463	0.227	0.945
LosstimeCat 4.81 vs 2.22 at PartnTD=2	0.86	0.427	1.731
LosstimeCat 3.95 vs 2.22 at PartnTD=2	0.373	0.181	0.769
LosstimeCat 3.13 vs 2.22 at PartnTD=2	0.145	0.056	0.37
Guardian 1 vs 2	0.48	0.188	1.224
Biologic 1 vs 2	0.269	0.059	1.241
PremorbP 1 vs 2	3.845	2.223	6.652
PrevLoss3 2 vs 1	1.715	1.049	2.802
EmplTD 1 vs 3	2.829	1.333	6.004
EmplTD 2 vs 3	1.514	0.51	4.49
Loc 5 vs 1	2.62	1.305	5.258
Loc 4 vs 1	1.527	0.776	3.004
Loc 3 vs 1	1.67	0.883	3.155
Loc 2 vs 1	2.68	1.468	4.892
Inc 4 vs 1	0.405	0.126	1.296
Inc 3 vs 1	0.914	0.35	2.388
Inc 2 vs 1	1.951	0.728	5.226

A6 Step 2 multivariable model (second model)

In Table 20 the results from fitting the multivariable model containing all predictors with univariable LR p-value less than 25 percent are presented. This is part of step 2 in the purposeful selection process for fitting the second model, see section 5.3.1.

Table 20: Summary of model containing predictors with univariable LR p-value less than 25%

Predictor	**Df**	**Wald p-value**
SexP	1	0.3103
AgeP	1	0.7134
SexC	1	0.8658
Losstime	1	0.0709
Guardian	1	0.103
Biologic	1	0.1883
PremorbP	1	< 0.0001
ContactPC	3	0.4169
SelfinjCYB	3	0.6268
SuicCE	2	0.6866
ViolSuic	1	0.3964
PartnTD	1	0.1631
PrevLoss1	1	0.3803
PrevLoss3	1	0.054
EmplTD	2	0.1351
Loc	4	0.0839
Work	2	0.2066
Inc	3	0.0217
Supp2	1	0.0067
Supp3	1	0.1958
Supp4	1	0.7853
Alc	2	0.0022
EmplUnd	4	0.9504
Feel	1	0.1203
FysAct	4	0.094
Hobby	4	0.063
Prep	3	0.3599
Social	3	0.798
Worry1	3	0.8318
Worry2	3	0.0545

Fit and prediction measures

AIC	c	$R^2_{E,adj}$
671.864	0.833	0.18611

A7 Parameter estimates and odds ratios for Model 2.10g

In Table 21 the parameter estimates, standard errors and Wald p-values for Model 2.10g are presented. In Table 22 odds ratios with 95% Wald confidence intervals are presented.

Table 21: Estimates and p-values for Model 2.10g

Predictor	**Level**	**Level**	**Estimate**	**Standard Error**	**Wald p-value for level estimate**	**Df for predictor**	**Wald p-value for predictor**
Intercept			-1.1732	1.0286	0.254		
SexP*Work	1	3	1.0104	0.6758	0.1349	2	0.0032
SexP*Work	1	2	3.8781	1.2088	0.0013		
PrevLoss3* Feel	2	1	-1.8643	0.843	0.027	1	0.027
PrevLoss3* Hobby	2	2	0.8329	1.0076	0.4085	4	0.0057
PrevLoss3* Hobby	2	3	1.5927	1.0503	0.1294		
PrevLoss3* Hobby	2	4	-1.6454	0.8656	0.0573		
PrevLoss3* Hobby	2	5	0.103	1.1659	0.9296		
PrevLoss3* Worry2	2	2	3.8495	0.9439	<0.0001	3	0.0005
PrevLoss3* Worry2	2	3	2.1892	1.3051	0.0935		
PrevLoss3* Worry2	2	4	0.556	0.7342	0.4489		
Supp2*Hobby	2	2	-0.2921	0.858	0.7335	4	0.0295
Supp2*Hobby	2	3	1.3898	0.919	0.1304		
Supp2*Hobby	2	4	0.6334	0.7	0.3656		
Supp2*Hobby	2	5	-2.1086	1.1091	0.0573		
Supp3*Worry2	2	2	0.0363	0.998	0.9709	3	0.022
Supp3*Worry2	2	3	3.3216	1.1822	0.005		
Supp3*Worry2	2	4	-0.5095	0.6253	0.4152		
Guardian* PrevLoss3	2	2	-2.7041	1.323	0.041	1	0.041
SexP	1		-0.9713	0.3214	0.0025	1	0.0025
LosstimeCat	5.85		-0.8164	0.3716	0.028	4	0.0013
LosstimeCat	4.81		-0.5967	0.3654	0.1024		
LosstimeCat	3.95		-1.2583	0.3845	0.0011		
LosstimeCat	3.13		-1.5451	0.4054	0.0001		

Continued on next page

Table 21 – continued from previous page

Predictor	Level	Level	Estimate	Standard Error	Wald p-value for level estimate	Df for predictor	Wald p-value for predictor
Guardian	2		1.8993	0.711	0.0076	1	0.0076
PremorbP	1		1.7998	0.3373	<0.0001	1	<0.0001
PrevLoss3	2		2.888	1.5032	0.0547	1	0.0547
EmplTD	1		1.3143	0.4484	0.0034	2	0.0069
EmplTD	2		-0.4497	0.6735	0.5044		
Loc	5		1.6562	0.4415	0.0002	4	0.0011
Loc	4		0.9836	0.4009	0.0142		
Loc	3		1.0476	0.387	0.0068		
Loc	2		1.3407	0.3654	0.0002		
Work	3		-0.0902	0.4471	0.8402	2	0.0015
Work	2		-3.9524	1.101	0.0003		
Inc	4		-1.2823	0.6715	0.0562	3	0.0072
Inc	3		-0.5293	0.5663	0.35		
Inc	2		0.3539	0.5704	0.535		
Supp2	2		0.4441	0.5898	0.4515	1	0.4515
Supp3	2		0.6918	0.3543	0.0509	1	0.0509
Alc	2		0.9573	0.4267	0.0249	2	0.0002
Alc	3		2.7762	0.7327	0.0002		
Feel	1		0.9337	0.3436	0.0066	1	0.0066
FysAct	2		-1.3448	0.7162	0.0604	4	0.0615
FysAct	3		-2.0407	0.6964	0.0034		
FysAct	4		-1.5214	0.6074	0.0123		
FysAct	5		-1.6021	0.6308	0.0111		
Hobby	2		-0.0632	0.5918	0.915	4	0.0285
Hobby	3		-1.9411	0.71	0.0063		
Hobby	4		-1.1208	0.4868	0.0213		
Hobby	5		-0.4894	0.6926	0.4798		
Worry2	2		-2.0312	0.6527	0.0019	3	0.0003
Worry2	3		-2.8878	0.9608	0.0027		
Worry2	4		0.2005	0.3556	0.5729		

Table 22: Odds ratio estimates and confidence intervals for Model 2.10g

Label	**OR estimate**	**95% Wald confidence interval**	
Work 3 vs 1 at SexP=1	2.51	0.778	8.097
Work 2 vs 1 at SexP=1	0.928	0.261	3.299
Work 3 vs 1 at SexP=2	0.914	0.38	2.195
Work 2 vs 1 at SexP=2	0.019	0.002	0.166
Feel 1 vs 2 at PrevLoss3=1	2.544	1.297	4.988
Feel 1 vs 2 at PrevLoss3=2	0.394	0.087	1.796
Hobby 1 vs 2 at Supp2=1 PrevLoss3=1	1.065	0.334	3.397
Hobby 1 vs 3 at Supp2=1 PrevLoss3=1	6.966	1.733	28.01
Hobby 1 vs 4 at Supp2=1 PrevLoss3=1	3.067	1.181	7.965
Hobby 1 vs 5 at Supp2=1 PrevLoss3=1	1.631	0.42	6.34
Hobby 1 vs 2 at Supp2=2 PrevLoss3=1	1.427	0.367	5.551
Hobby 1 vs 3 at Supp2=2 PrevLoss3=1	1.736	0.466	6.462
Hobby 1 vs 4 at Supp2=2 PrevLoss3=1	1.628	0.512	5.182
Hobby 1 vs 5 at Supp2=2 PrevLoss3=1	13.437	1.961	92.095
Hobby 1 vs 2 at Supp2=1 PrevLoss3=2	0.463	0.066	3.271
Hobby 1 vs 3 at Supp2=1 PrevLoss3=2	1.417	0.155	12.971
Hobby 1 vs 4 at Supp2=1 PrevLoss3=2	15.899	2.83	89.334
Hobby 1 vs 5 at Supp2=1 PrevLoss3=2	1.472	0.175	12.387
Hobby 1 vs 2 at Supp2=2 PrevLoss3=2	0.62	0.093	4.139
Hobby 1 vs 3 at Supp2=2 PrevLoss3=2	0.353	0.051	2.466
Hobby 1 vs 4 at Supp2=2 PrevLoss3=2	8.439	1.575	45.213
Hobby 1 vs 5 at Supp2=2 PrevLoss3=2	12.122	1.04	141.335
Worry2 1 vs 2 at Supp3=1 PrevLoss3=2	0.162	0.036	0.736
Worry2 1 vs 3 at Supp3=1 PrevLoss3=2	2.011	0.2	20.209
Worry2 1 vs 4 at Supp3=1 PrevLoss3=2	0.469	0.118	1.862
Worry2 1 vs 2 at Supp3=1 PrevLoss3=1	7.623	2.121	27.4
Worry2 1 vs 3 at Supp3=1 PrevLoss3=1	17.953	2.731	118.029
Worry2 1 vs 4 at Supp3=1 PrevLoss3=1	0.818	0.408	1.643
Worry2 1 vs 2 at Supp3=2 PrevLoss3=2	0.157	0.019	1.293
Worry2 1 vs 3 at Supp3=2 PrevLoss3=2	0.073	0.005	0.98
Worry2 1 vs 4 at Supp3=2 PrevLoss3=2	0.781	0.17	3.588
Worry2 1 vs 2 at Supp3=2 PrevLoss3=1	7.351	1.318	40.993
Worry2 1 vs 3 at Supp3=2 PrevLoss3=1	0.648	0.132	3.19
Worry2 1 vs 4 at Supp3=2 PrevLoss3=1	1.362	0.462	4.015
Supp2 1 vs 2 at Hobby=1	0.641	0.202	2.038
Supp2 1 vs 2 at Hobby=2	0.859	0.254	2.902
Supp2 1 vs 2 at Hobby=3	0.16	0.04	0.644
Supp2 1 vs 2 at Hobby=4	0.34	0.164	0.708
Supp2 1 vs 2 at Hobby=5	5.283	0.831	33.576
Supp3 1 vs 2 at Worry2=1	0.501	0.25	1.003
Supp3 1 vs 2 at Worry2=2	0.483	0.076	3.048

Continued on next page

Table 22 – continued from previous page

Label	OR estimate	95% Wald confidence interval	
Supp3 1 vs 2 at Worry2=3	0.018	0.002	0.163
Supp3 1 vs 2 at Worry2=4	0.833	0.304	2.287
Guardian 2 vs 1 at PrevLoss3=1	6.681	1.658	26.917
Guardian 2 vs 1 at PrevLoss3=2	0.447	0.05	3.962
LosstimeCat 5.85 vs 2.22	0.442	0.213	0.916
LosstimeCat 4.81 vs 2.22	0.551	0.269	1.127
LosstimeCat 3.95 vs 2.22	0.284	0.134	0.604
LosstimeCat 3.13 vs 2.22	0.213	0.096	0.472
PremorbP 1 vs 2	6.049	3.123	11.717
EmplTD 1 vs 3	3.722	1.546	8.963
EmplTD 2 vs 3	0.638	0.17	2.388
Loc 5 vs 1	5.239	2.206	12.447
Loc 4 vs 1	2.674	1.219	5.867
Loc 3 vs 1	2.851	1.335	6.087
Loc 2 vs 1	3.822	1.867	7.82
Inc 4 vs 1	0.277	0.074	1.034
Inc 3 vs 1	0.589	0.194	1.787
Inc 2 vs 1	1.425	0.466	4.357
Alc 1 vs 2	0.384	0.166	0.886
Alc 1 vs 3	0.062	0.015	0.262
FysAct 1 vs 2	3.837	0.943	15.62
FysAct 1 vs 3	7.696	1.966	30.132
FysAct 1 vs 4	4.579	1.392	15.058
FysAct 1 vs 5	4.964	1.442	17.089

A8 Residual plots and ROC-curves

In Figure 13 and 14 Pearson and deviance residuals are plotted for Model 1.7c and 2.10g respectively. Figure 15 and 16 are ROC-curves for Model 1.8c and 2.10g respectively.

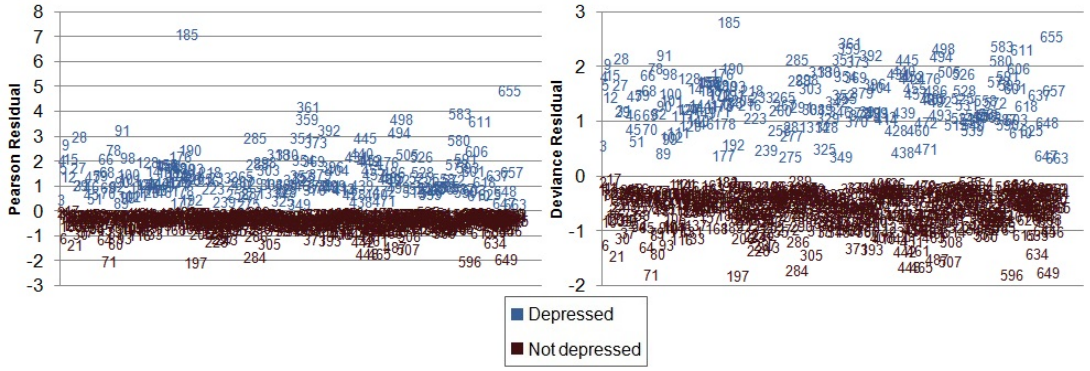

Figure 13: Plot of Pearson and deviance residuals for Model 1.7c

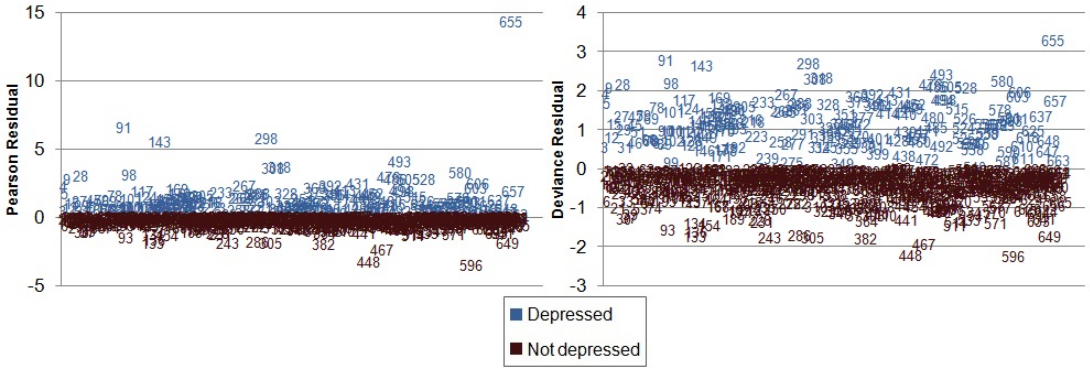

Figure 14: Plot of Pearson and deviance residuals for Model 2.10g

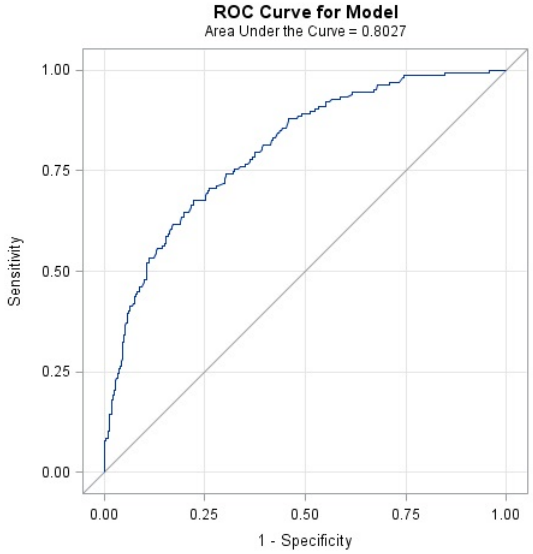

Figure 15: ROC-curve for Model 1.8c

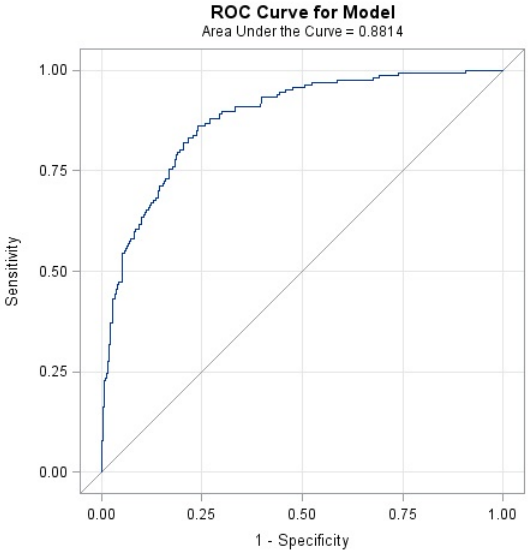

Figure 16: ROC-curve for Model 2.10g