# Sensitivity analysis for incomplete data: region of uncertainty

# Sensitivity analysis for incomplete data: region of uncertainty

**Article** · October 2004

CITATIONS
0

READS
19

**3 authors**, including:

Geert Molenberghs
Universiteit Hasselt and University of Leuven
**923** PUBLICATIONS **21,045** CITATIONS

SEE PROFILE

Els Goetghebeur
Ghent University
**130** PUBLICATIONS **2,498** CITATIONS

SEE PROFILE

# Sensitivity Analysis for Incomplete Data: Region of Uncertainty

Geert Molenberghs Michael G. Kenward Els Goetghebeur

September 21, 2012

**Abstract**

Classical inferential procedures induce conclusions from a set of data to a population of interest, accounting for the imprecision resulting from the stochastic component of the model. Less attention is devoted to the uncertainty arising from (unplanned) incompleteness in the data. Through the choice of an identifiable model for non-ignorable non-response, one narrows the possible data generating mechanisms to the point where inference only suffers from imprecision. Some proposals have been made for assessment of sensitivity to these modelling assumptions; many are based on fitting several plausible but competing models. We propose an approach which identifies and incorporates both sources of *uncertainty* in inference: *imprecision* due to finite sampling and *ignorance* due to incompleteness. The developments focus on contingency tables, and are illustrated using data from a psychiatric study and data from a Slovenian plebiscite.

*Keywords:* Contingency Table; Missing At Random; Overspecified Model; Saturated Model

## 1 Introduction

The problem of analysing data sets from which observations are missing is a common one, and the reasons for data being missing are many and varied. In this setting two main problems need addressing. The first, accommodating the lack of balance induced, has to a great extent been answered. Modern statistical tools are not as dependent on simple data structures as was the case before the computer became ubiquitous. The second problem is more fundamental in nature and is far from admitting a straightforward solution. How should one approach statistical inference that accommodates the possible, but unknown, behaviour of the unobserved data ?

Rubin (1976) provided one of the first systematic studies of this issue, and we use his terminology for classifying different classes of processes that give rise to missing values. A process is said to be *missing completely at random* (MCAR) if the probability of an observation being missing is independent of both unobserved and observed data and *missing at random* (MAR) if, conditional on the observed data, the probability is independent of the unobserved data. A

process that is neither MCAR nor MAR is termed *non-random* (MNAR). For likelihood inference MCAR and MAR missing value processes are said to be *ignorable* when the parameters governing the measurement and missing value processes are functionally independent, while a non-random process is non-ignorable. The importance of the MCAR and MAR processes is that, given all the observed data, there remains no dependence of the likelihood on unobserved data and, broadly, inferences can be made that *ignore* the missing value process. This begs the question of whether one can reasonably make assumptions of MCAR and MAR. Sometimes a study design provides the justification (Murray and Findlay 1988), but typically this is not the case and the incomplete data under analysis can never alone answer the question of whether or not a missing value process is non-random. This paper is concerned with how one might approach inference when the possibility of a non-random missingness process cannot be ruled out on *a priori* grounds.

It is not difficult to formulate models for the MNAR setting, i.e., that allow explicit or implicit dependence of the probability of an outcome being missing on the unobserved value of that outcome. For any given model, the likelihood for the observed data is obtained by integrating over the distribution of the missing data and conventional inference proceeds accordingly. Diggle and Kenward (1994) and Molenberghs, Kenward, and Lesaffre (1997) use such models for the analysis respectively of continuous and ordinal longitudinal data with dropout. Little (1995) provides a review of such approaches. Because such modelling assumptions cannot be examined using the data alone, a characteristic form of ambiguity resides in the resulting inferences. This ambiguity manifests itself in a somewhat different fashion in the continuous and discrete data settings. In the former, conclusions that might be drawn about the missing value process may be critically dependent on assumptions such as the shape of the distribution of the unobserved data and the model chosen for the missing value process, see for example Little and Rubin (1987, section 11.4) and Kenward (1998). Likelihoods may be very flat, or multimodal, indicating little information on the missing value process as represented in the chosen model. In the categorical case it is easy to show that models with an entirely different interpretation at the complete data level can show the same deviance (or even saturate the same data); models that saturate the degrees of freedom can yield a boundary solution and strictly positive deviance in both the complete data counts and some of the parameter estimates, while unconstrained solutions may be invalid; non-unique solutions can

be obtained (Molenberghs, Goetghebeur, and Lipsitz 1997).

Such problems with MNAR models do not imply however that they are of no value. In the first place, many of these issues apply equally well to MAR models which have no *a priori* justification: an MAR model can usually be formulated as a special, often implausible, member of a general family of MNAR models. It might be argued then, that one role of MNAR models is to convey information on the inadequacy of simpler and more convenient assumptions rather than telling us about the plausibility of particular MNAR assumptions, which in a strict sense they cannot do. For example, if the addition of a single parameter to an MAR model produces a very large change in the likelihood, one that is inconsistent with random variation, then structure has been identified in the observed data that the MAR model does not accommodate. The interpretation of this structure still presents problems however. Just because the extended model happens to be an instance of an MNAR model, this does not imply necessarily that the structure lacking is associated with the missing value process. This is an example of the ambiguity generated by the missing data setting: the structure identified may be real but there will be a range of models formulated in terms of the complete (but partially unobserved) data that can be used to explain it at the observed data level. In addition, extended models may have exactly the same or very similar likelihoods than the more parsimonious MAR model, even though the effect on key parameters (such as the treatment effect among the complete cases), can be dramatic.

A natural way to proceed is to acknowledge the inherent ambiguity and explore the range of inferences that are consistent with the gap in our knowledge. Essentially this is a form of sensitivity analysis. While there is a formal mathematical statistical framework for imprecision (variance, standard errors, sampling distributions, confidence intervals, hypothesis tests and so on) most implementations of sensitivity analysis have remained *ad hoc*. The Bayesian framework provides one possible route (Draper 1995) and this is being explored elsewhere for the present context. The goal of this paper is to develop a frequentist based framework in which general sensitivity concepts can be formalized and further developed. To this end, a language will be proposed to describe *ignorance* (due to incompleteness of the data) in addition to the familiar *imprecision* (due to finite sampling) and to comine both into *uncertainty*. We focus here on the multinomial contingency table setting, for which interesting and practically useful results can be

obtained in a reasonably direct fashion. We indicate how extensions to more complicated data settings might be developed. In the next section two sets of data are introduced. In Section 3 the concepts of *imprecision*, *ignorance* and *uncertainty* are developed and in Section 4 formulated as a general principle. Application of the principle in the contingency table setting leads to the need to enumerate data sets that are consistent with an extended class of models and with the observed data. This is dealt with in Section 5, where the extended model and its overspecified likelihood emerge as fundamental concept. Three approaches to the overspecified likelihood are explored in Section 6 and a brief comparison of the approaches is given in Section 7. The methodology is applied in Sections 8 and 9. Formal tools to assess the quality of the estimators are discussed in Section 10.

# 2 Examples

## 2.1 The Psychiatric Study

**Example 1** *The data come from a multicentre study involving 315 patients that were treated by fluvoxamine for psychiatric symptoms described as possibly resulting from a dysregulation of serotonine in the brain. The data are discussed in Molenberghs and Lesaffre (1994), Kenward, Lesaffre, and Molenberghs (1994), Molenberghs, Kenward, and Lesaffre (1997), and Michiels and Molenberghs (1997).*

After recruitment to the study, the patient was assessed at four visits. The therapeutic effect and the extent of side effects were scored at each visit on an ordinal scale. Side effects are coded as (1) none; (2) not interfering with functionality; (3) interfering significantly with functionality; (4) side effects surpasses the therapeutic effect. Similarly, the effect of therapy is: (1) no improvement or worsening; (2) minimal improvement; (3) moderate improvement and (4) important improvement. Thus, a side effect occurs if new symptoms occur while there is therapeutic effect if old symptoms disappear. 299 patients have at least one measurement, including 242 completers.

Here, we analyze a dichotomized version of one outcome of the data, where category 1 is contrasted with the others. We will only look at the results on the first and on the last measurement occasion. All four response patterns are seen, although the majority are completers or with people seen at the first but not at the last occasion. The observed data are given in Table 1.

Table 1: Data From Psychiatric Study.

Side Effects.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 89 | 13 | | 26 | | 2 | 0 | 14 |
| 57 | 65 | | 49 | | | | |

Therapeutic Effect.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 11 | 1 | | 7 | | 0 | 2 | 14 |
| 124 | 88 | | 68 | | | | |

### 2.2 The Slovenian Plebiscite

**Example 2** *Rubin, Stern, and Vehovar (1995) studied data from a plebiscite organized in Slovenia on the secession from the former Yugoslavia, in which the Slovenians overwhelmingly voted for independence.*

It was deemed useful to anticipate results of the plebiscite from additional questions in the Slovenian Public Opinion (SPO) Survey, carried out four weeks prior. The main questions were: (1) Are you in favour of Slovenian independence ? (2) Are you in favour of Slovenia's secession from Yugoslavia ? (3) Will you attend the plebiscite ? Questions (1) and (2) are different since independence would have been possible in confederal form as well. Question (3) is highly relevant since not attending was treated as an effective NO to question (1). The data are presented in Table 2. Full details on the study are provided in Rubin, Stern, and Vehovar (1995), who considered both ignorable and non-ignorable models. The ignorable models outperformed the non-ignorable one in the sense that they were much closer to the results of the (true) plebiscite. This clearly calls for a sensitivity analysis (Kenward 1998).

## 3 Imprecision, Ignorance and Uncertainty

To introduce a formal distinction between *statistical imprecision* which is due to the stochastic component in the experimental data, and *statistical ignorance*, which is due to the incompleteness of the observations, we consider a simple example first.

Table 2: Data form the Slovenian Public Opinion Survey.

| | | Independence | | |
|---|---|---|---|---|
| Secession | Attendance | Yes | No | $*$ |
| Yes | Yes | 1191 | 8 | 21 |
| | No | 8 | 0 | 4 |
| | $*$ | 107 | 3 | 9 |
| No | Yes | 158 | 68 | 29 |
| | No | 7 | 14 | 3 |
| | $*$ | 18 | 43 | 31 |
| $*$ | Yes | 90 | 2 | 109 |
| | No | 1 | 2 | 25 |
| | $*$ | 19 | 8 | 96 |

Table 3: Theoretical Distribution Over Completed Cells.

| | |
|---|---|
| $\pi_{111}$ | $\pi_{112}$ |
| $\pi_{121}$ | $\pi_{122}$ |

| | |
|---|---|
| $\pi_{011}$ | $\pi_{012}$ |
| $\pi_{021}$ | $\pi_{022}$ |

Table 4: Theoretical Distribution Over Observed Cells.

| | |
|---|---|
| $\pi_{111}$ | $\pi_{112}$ |
| $\pi_{121}$ | $\pi_{122}$ |

| |
|---|
| $\pi_{01+}$ |
| $\pi_{02+}$ |

**Example 3** *Consider a two-way contingency table in which binary outcomes* $(Y_{i1}, Y_{i2})$ *for subject* $i = 1, \ldots, N$ *are classified. Suppose further that some contribute only to the first measurement* $(R_i = 0)$ *while the remaining ones are observed on both variables* $(R_i = 1)$*. In a selection model, the cell probabilities* $p_{ij}$ $(i, j = 1, 2)$ *are modeled, together with the dropout probabilities* $q_{r|ij}$ $(r = 0, 1)$*. They are grouped into vectors* $\boldsymbol{\theta} = (\boldsymbol{p}^T, \boldsymbol{q}^T)^T$*. The full data probabilities are given by* $\pi_{rij} = p_{ij} q_{r|ij}$*. They are displayed in Table 3. The distribution of the observable data is displayed in Table 4.*

Assume that $(Y_{i1}, Y_{i2}, R_i)$ follows a multinomial sampling scheme. The methods described here extend to general patterns of missingness (e.g., all four patterns in the psychiatric study),

Table 5: Dropout Models Corresponding to The Setting of Table 3. All Measurement Models Are Saturated.

| Dropout Models | | | | |
|---|---|---|---|---|
| Model | $q_{r\|ij}$ | # Par. | Observed d.f. | Complete d.f. |
| 1. MCAR | $q_r$ | 4 | Non-saturated | Non-saturated |
| 2. MAR | $q_{r\|i}$ | 5 | Saturated | Non-saturated |
| 3. Protective | $q_{r\|j}$ | 5 | Saturated | Non-saturated |
| 4. MNAR I | $\text{logit}(q_{r\|ij}) = \alpha + \beta_i + \gamma_j$ | 6 | Overspecified | Non-saturated |
| 5. $M_{\text{sat}}$ | $q_{r\|ij}$ | 7 | Overspecified | Saturated |

multicategorical and higher-way cross-classifications (e.g., the Slovenian plebiscite in Table 2).

A sample from Table 4 produces empirical proportions representing the $\pi'$s with error. This imprecision disappears as the sample size tends to infinity. What remains is ignorance regarding the redistribution of $\pi_{01+}$ and $\pi_{02+}$ over the second outcome value. Indeed, Table 4 leaves constrained ignorance regarding the values of $\pi_{0ij}$ $(i, j = 1, 2)$ and hence regarding any derived parameter of scientific interest. For such parameter, $\psi$ say, a region of possible values which is consistent with Table 4 is called the region of ignorance. Analogously an observed incomplete table leaves ignorance regarding the would-be observed complete table, which in turn leaves imprecision regarding the true complete probabilities. The region of estimators for $\widehat{\psi}$ consistent with the observed data provides an estimated region of ignorance. The $(1 - \alpha)100\%$ *region of uncertainty* is a larger region in the spirit of a the confidence region, designed to capture the combined effects of imprecision and ignorance. We will propose various methods to estimate ignorance and uncertainty. In Section 10, we define properties which such a region is desired to have.

In standard statistical practice, ignorance is hidden in the consideration of a single identified model. Table 5 shows several models for data sampled from Table 4. Not all of them are identified. All measurement models $p_{ij}$ are left unconstrained. Model $M_{\text{sat}}$ (Model 5), saturating the full data degrees of freedom, consists of 3 measurement parameters and 4 dropout parameters. There are only 5 observed degrees of freedom, rendering this model overspecified when fitted to the observed data. Typically, identifiability restrictions are imposed in an MAR model (Model 2), a *protective* model (Model 3) where dropout depends on the potentially unobserved second measurement, but not on

Table 6: Identifiable Models, Fitted To Monotone Patterns From Psychiatric Study.

| | (1,1) | | (1,0) | | deviance |
|---|---|---|---|---|---|
| Side Effects | 83.7 | 12.2 | 28.0 | 4.1 | 495.8 |
| Model 1 (MCAR) | 59.9 | 68.3 | 20.0 | 22.9 | |
| Side Effects | 89.0 | 13.0 | 22.7 | 3.3 | 494.4 |
| Model 2 (MAR) | 57.0 | 65.0 | 22.9 | 26.1 | |
| Side Effects | 89.0 | 13.0 | 18.6 | 7.4 | 494.4 |
| Model 3 (Protective) | 57.0 | 65.0 | 11.9 | 37.1 | |
| Therapeutic Effect | 13.0 | 1.2 | 4.4 | 0.4 | 386.5 |
| Model 1 (MCAR) | 122.7 | 87.1 | 41.1 | 29.2 | |
| Therapeutic Effect | 11.0 | 1.0 | 6.4 | 0.6 | 385.8 |
| Model 2 (MAR) | 124.0 | 88.0 | 39.8 | 28.2 | |
| Therapeutic Effect | 11.0 | 1.0 | 7.1 | -0.1 | 385.8 |
| Model 3 (Protective, Unconstr.) | 124.0 | 88.0 | 80.5 | -12.5 | |
| Therapeutic Effect | 11.6 | 1.0 | 6.4 | 0.0 | 385.8 |
| Model 3 (Protective, Constr.) | 123.4 | 88.0 | 68.5 | 0.0 | |

the first one, or an MCAR model (Model 1). The first two of these saturate the observed degrees of freedom, in contrast with $M_{\text{sat}}$ which is overspecified for the observed degrees of freedom and saturates the complete data degrees of freedom. Model 4 is neither identifiable nor fully saturated.

Table 6 shows the predicted complete tables for Models 1, 2, and 3, fitted to the monotone patterns of the psychiatric study (Table 1). The effect of ignorance is clearly seen by comparing the MAR and protective models: they provide a substantially different prediction for the partially observed table. In addition, the protective model produces a boundary solution, or an invalid solution if the parameters are left unconstrained, for therapeutic effect. We will analyze these data further in Section 8.

## 4 General Principle

We start from the classical approach of fitting a single identifiable model $M_0$ to incomplete data (e.g., Models 1–3 in Table 5). Maximum likelihood estimation produces a parameter estimate $\hat{\boldsymbol{\pi}}$

along with measures of imprecision (estimated standard errors). From $\hat{\boldsymbol{\pi}}$ two predicted contingency tables can be derived as in Table 3.

The fitted complete tables collapse back to fitted values for the incomplete Table 4. Contrasting the latter with the observed data shows the goodness-of-fit of model $M_0$. If there is substantial lack of fit, the original model $M_0$ needs to be reconsidered. Lack of fit has strong bearings on imprecision and, since we want to focus on ignorance, we will assume the fit is acceptable. Assume then that model $M_0$ has adequately captured the structure of the incomplete tables.

One can range through all possible complete tables, compatible with the $M_0$ predicted incomplete table. We call the tables "$M_0-$compatible" and we denote the set by $\mathcal{S}(M_0)$. They collapse back to the same fitted incomplete table which we denote by $\tau_0$. To each table, an extended model $M^*$ will be fitted. For a saturated model, tables and models are in one-one relation.

An insightful way of finding $\mathcal{S}(M_0)$, sketched in Section 5, is by direct enumeration. Equivalently, the modelling process can be expressed in terms of the complete data likelihood. Alternatively, the model $M^*$ can be fitted directly to the observed data. When it is overspecified on the observed data, it leads to a *range* of parameters maximizing the observed data likelihood. We will show that this range coincides with the collection $M^*(\mathcal{S}(M_0))$ when $M^*$ is saturated, but that the equivalence no longer holds in case $M^*$ is not saturated. We conclude that the possibly overspecified model $M^*$ and its associated likelihood are very natural concepts. Section 6 proposes three methods to determine the set of maxima from the overspecified likelihood. The first starts by applying an identifiable transformation, and then determines the set by means of a one-to-many mapping (Section 6.1). It is related to method of moment ideas. Directly solving the score equations is discussed in Section 6.2. Finally, the overspecification can be removed by considering a minimal set of parameters $\boldsymbol{\lambda}$, conditional upon which the others, $\boldsymbol{\mu}$, are identified. We term $\boldsymbol{\lambda}$ the sensitivity parameter and $\boldsymbol{\mu}$ the estimable parameter. This method is described in Section 6.3.

## 5 Enumeration Method

A range of models which extend $M_0$ and are plausible or cannot be excluded on practical grounds are assembled into a class $M^*$. Our goal is to examine their fits to $M_0$-compatible tables and assess their influence on the estimated main parameter(s). This results in the set of complete tables

$M^*(\mathcal{S}(M_0))$. To ensure that $M^*$ adequately reproduces $\tau_0$ when fitted to each $M_0$-compatible table its elements have to be sufficiently elaborate. When $M^*$ is saturated, the model reproduces the (completed) data exactly and hence $M^*(\mathcal{S}(M_0)) = \mathcal{S}(M_0)$.

It is possible to consider the predicted incomplete table as a set of linear constraints for the predicted complete tables, together with the requirement that each predicted complete table must correspond to a valid probability model. One way to express these ideas more formally is to use upper and lower probabilities, related to belief functions (Shafer 1976). Let $\Omega = \{\omega_1, \omega_2, \ldots\}$ be the *frame*, playing the role of the outcome space in probability theory. Informally, LOW($A$) and UP($A$) are defined respectively as lower and upper bounds on the probability that the outcome will be an element of $A$. In analogy with the probability mass function (p.m.f.), the concept of *mass function* is introduced. Whereas a p.m.f. is defined over $\Omega$: $p : \Omega \longrightarrow [0, 1]$, the mass function is defined over power set $2^\Omega$:

$$m : 2^\Omega \longrightarrow [0, 1] \tag{5.1}$$

and has to satisfy $m(\phi) = 0$ and $\sum_{A\subset\Omega} m(A) = 1$.

All $A$ such that $A \subset B$ *provide evidence for* $B$ and hence support the lower probability in $B$. All $A$ such that $A \cap B \neq \phi$ *are compatible with* $B$ and construct the upper probability of $B$. These statements motivate the following formal definitions. The lower and upper functions are given by

$$\text{LOW}(B) = \sum_{A\subset B} m(A)$$

and the plausibility function by

$$\text{UP}(B) = \sum_{A\cap B\neq\phi} m(A).$$

**Example 4 ($2 \times 1$ contingency table with supplemental count)** *Consider a single binary measurement*

*for which $r$ successes (coded as 1) and $n - r$ failures (coded as 2) are observed, with an additional $N - n$ missing observations. The mass, lower, and upper functions are given by:*

$$\begin{array}{rclcrclcrcl} m(\{1\}) & = & r/N & & LOW(\{1\}) & = & r/N & & UP(\{1\}) & = & (r+N-n)/N \\ m(\{2\}) & = & (n-r)/N & & LOW(\{2\}) & = & (n-r)/N & & UP(\{2\}) & = & (N-r)/N \\ m(\{1,2\}) & = & (N-n)/N & & & & & & & & \end{array}$$

Next, consider the setting of the psychiatric study (Table 1):

| | |
|---|---|
| $a_{11}$ | $a_{12}$ |
| $a_{21}$ | $a_{22}$ |

| |
|---|
| $b_1$ |
| $b_2$ |

| | |
|---|---|
| $c_1$ | $c_2$ |

| |
|---|
| $d$ |

It follows that the mass function is given by

$$\begin{aligned} m(11) &= a_{11} & m(\{11,12\}) &= b_1 \\ m(12) &= a_{12} & m(\{21,22\}) &= b_2 \\ m(21) &= a_{21} & m(\{11,21\}) &= c_1 \\ m(22) &= a_{22} & m(\{12,22\}) &= c_2 \\ & & m(\{11,12,21,22\}) &= d \end{aligned}$$

leading to the lower and upper functions:

$$\begin{aligned} \text{LOW}(11) &= a_{11} & \text{UP}(11) &= a_{11}+b_1+c_1+d \\ \text{LOW}(12) &= a_{12} & \text{UP}(12) &= a_{12}+b_1+c_2+d \\ \text{LOW}(21) &= a_{21} & \text{UP}(21) &= a_{21}+b_2+c_1+d \\ \text{LOW}(22) &= a_{22} & \text{UP}(22) &= a_{22}+b_2+c_2+d \end{aligned}$$

In addition, the lower and upper functions for the marginal successes are

$$\begin{aligned} \text{LOW}(\{11,12\}) &= a_{11}+a_{12}+b_1 \\ \text{UP}(\{11,12\}) &= a_{11}+a_{12}+b_1+c_1+c_2+d \\ \text{LOW}(\{11,21\}) &= a_{11}+a_{21}+c_1 \\ \text{UP}(\{11,21\}) &= a_{11}+a_{21}+b_1+b_2+c_1+d \end{aligned}$$

while for other responses (such as a concordant pair), they are easy to derive. Let us now specialize to the common case of dropout ($c_1 = c_2 = d = 0$). Above belief and plausibilities then reduce to:

$$\begin{aligned} \text{LOW}(\{11,12\}) = \text{UP}(\{11,12\}) &= a_{11}+a_{12}+b_1 \\ \text{LOW}(\{11,21\}) &= a_{11}+a_{21} \\ \text{UP}(\{11,21\}) &= a_{11}+a_{21}+b_1+b_2 \end{aligned}$$

In other words, there is *no* ignorance about the probabilities at the first measurement occasion, while there is ignorance about the second measurement.

Consider again the setting of Example 4. Let $p$ be the success probability and $q_i$ the probability of being observed, given success ($i = 1$) or failure ($i = 2$).

Table 7: Incomplete and Complete Data of Example 4.

| | Incomplete | | Full | |
|---|---|---|---|---|
| | Success | Failure | Success | Failure |
| Observed | $r$ | $n-r$ | $r$ | $n-r$ |
| Unobserved | $N-n$ | | $R-r$ | $N-n-R+r$ |

In this simple case, when $M_0$ saturates the observed degrees of freedom, its predicted incomplete table equals the observed data. Enumeration of all tables corresponding to this predicted incomplete table can be done in terms of a single parameter. Let the number of successes out of the $N-n$ missing observations be denoted by $R-r$, then the corresponding number of failures is $N-n-R+r$ and we merely have to cycle through all values of $R$ in the interval

$$[\text{LOW}(\{1\}), \text{UP}(\{1\})] = [r, N-n+r].$$

The compatible full data tables are depicted in Table 7.

The second way consist of maximizing an overparameterized observed data likelihood. We will first consider the complete-data likelihood in terms of a generic parameterization. Thereafter, we study two specific reparameterizations in terms of $p$ and $q_i$. The complete-data log-likelihood is

$$\ell^* = r\ln\alpha + (n-r)\ln\beta + (R-r)\ln\gamma_1 + (N-n-R+r)\ln\gamma_2, \tag{5.2}$$

where $\alpha$, $\beta$, $\gamma_1$ and $\gamma_2$ are the four cell probabilities. The maximum likelihood estimator is given by

$$\begin{aligned} \alpha_R &= \frac{r}{N}, & \gamma_{1,R} &= \frac{R-r}{N}, \\ \beta_R &= \frac{n-r}{N}, & \gamma_{2,R} &= \frac{N-n-R+r}{N}. \end{aligned}$$

The subscript $R$ indicates dependence on $R$. While this parameterization is easy to work with, one is usually interested in effects such as success (failure) probabilities, non-response probabilities, etc. Table 8 shows two parameterizations (MAR≡MCAR and MNAR). The complete-data log-likelihood corresponding to $M_0$ can be written as

$$\ell_A^* = R\ln p + (N-R)\ln(1-p) + n\ln q + (N-n)\ln(1-q).$$

Table 8: Two Reparameterizations of the Complete-Data Likelihood.

| **Model $M_0$: MAR≡MCAR** | **Model $M_{\mathbf{sat}}$: MNAR** |
|---|---|
| *Parameterization:* | |
| $\alpha = pq$ | $\alpha = pq_1$ |
| $\beta = (1-p)q$ | $\beta = (1-p)q_2$ |
| $\gamma_1 = p(1-q)$ | $\gamma_1 = p(1-q_1)$ |
| $\gamma_2 = (1-p)(1-q)$ | $\gamma_2 = (1-p)(1-q_2)$ |
| *Solution:* | |
| $p_R = \frac{R}{N}$ | $p_R = \frac{R}{N}$ |
| $q_R = \frac{n}{N}$ | $q_{1R} = \frac{r}{R}$ |
| | $q_{2R} = \frac{n-r}{N-R}$ |
| $p \in \left[\frac{r}{N}, \frac{N-n+r}{N}\right]$ | $p \in \left[\frac{r}{N}, \frac{N-n+r}{N}\right]$ |

The solutions to this likelihood, $p_R$ and $q_R$, are displayed in Table 8, as well as the estimated interval of ignorance for $p$, obtained as the collection of all $p_R$. For this model, the fitted counts do not coincide with the observed counts. Indeed, the four complete data counts are:

| | |
|---|---|
| $\frac{Rn}{N}$ | $\frac{R(N-n)}{N}$ |
| $\frac{(N-R)n}{N}$ | $\frac{(N-R)(N-n)}{N}$ |

and they correspond to the observed counts only if

$$R = \frac{r}{n}N, \tag{5.3}$$

i.e., for the predicted complete table corresponding to MAR. For every value of $R$, the number of missing observations $N-n$ is reproduced exactly, but the number of observed successes and failures are not. Let us now plug this solution into the *observed-data* log-likelihood:

$$\ell = r\ln\alpha + (n-r)\ln\beta + (N-n)\ln(1-\alpha-\beta), \tag{5.4}$$

yielding the following function of $R$:

$$\ell(R) = n\ln n + (N-n)\ln(N-n) - N\ln N + r\ln\left(\frac{R}{N}\right) + (n-r)\ln\left(\frac{N-R}{N}\right),$$

which achieves a maximum, as a function of $R$, for (5.3). Note that $\ell(R)$ measures the fit of the model to the predicted complete table.

For $M_{\text{sat}}$, the complete-data log-likelihood becomes

$$\begin{aligned}\ell_B^* &= R\ln p + (N-R)\ln(1-p) \\ &\quad + r\ln q_1 + (R-r)\ln(1-q_1) + (n-r)\ln q_2 + (N-n-R+r)\ln(1-q_2).\end{aligned}$$

The corresponding solution is displayed in Table 8. The same estimated interval of ignorance for $p$ is obtained. The main distinction is that we now reproduce the predicted complete table and hence also the observed data for each choice of $R$. The observed data likelihood, with the solution substituted, is now constant. The difference in behaviour under both models arises because the enumeration, prescribed by $R \in [r; N-n+r]$, does not distinguish between the saturated $M_{\text{sat}}$ and the identified $M_0$. Let us elaborate on this.

Suppose first that $M^* = M_0$. Fitting $M^*$ to each element of $\mathcal{S}(M_0)$ will yield a non-saturated fit, except for the one predicted complete table predicted by the model. In other words, there will be a range of solutions, $M^*(\mathcal{S}(M_0))$. The alternative procedure consists of fitting $M^* = M_0$ directly which by assumption of identifiability produces a single solution. This implies a general discrepancy between both approaches when $M^*$ is not saturated. Differences remain between both procedures, as long as $M^*$ differs from the saturated model.

In order to remove this discrepancy, the set of predicted complete tables ought to be restricted to the subset, compatible with $M^*$. When $M^* = M_0$, the only predicted complete table allowable is the one originally produced by $M_0$, reducing the sensitivity analysis to a standard analysis. In general, the allowable predicted complete tables should reflect the model $M^*$. This observation then trivially reduces both procedures to a single one.

This suggests treating the possibly overspecified model $M^*$ and its associated likelihood as the fundamental concepts, which will be the subject of the following sections.

## 6 Observed Data Likelihood

In this family of approaches, the likelihood of the observed data will be used rather than the likelihood of the completed data. We will explicitly allow for overspecified parameterizations, in

Table 9: Two Transformations of the Observed-Data Likelihood.

| **Model I** **(MAR)** | **Model II** **(MNAR, $M_{\mathbf{sat}}$)** |
|---|---|
| *Parameterization:* | |
| $\alpha = pq$ | $\alpha = pq_1$ |
| $\beta = (1-p)q$ | $\beta = (1-p)q_2$ |
| $\gamma = 1-q$ | $\gamma = 1 - pq_1 - (1-p)q_2$ |
| *Solution:* | |
| $\hat{p} = \frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}} = \frac{r}{n}$ | $pq_1 = \frac{r}{N}$ |
| $\hat{q} = \hat{\alpha} + \hat{\beta} = \frac{n}{N}$ | $(1-p)q_2 = \frac{n-r}{N}$ |
| | $\frac{r}{q_1} + \frac{n-r}{q_2} = N$ |
| | $p : \left[\frac{r}{N}, \frac{N-n+r}{N}\right]$ |

order to carry out a sensitivity analysis.

There are three distinct approaches. The first one, the *transformation method* considers an identifiable parameterization, which is then transformed, using a one-to-many mapping, into an overspecified parameterization. The second *direct-likelihood method* uses an overspecified parameterization without intermediate steps. The final method treats a number of parameters as fixed and will be referred to as the *sensitivity parameter method*. To exemplify the approaches, the setting of Example 4 will be reconsidered.

## 6.1 Transformation Method

Independent of the parameterization chosen, the observed data log-likelihood can be represented in the form (5.4), where $\alpha$, $\beta$, and $\gamma$ can be seen as moments. We consider two models, of which the parameterization is given in Table 9. The first one is identified, the second one is overparameterized. Denote the corresponding log-likelihoods by $\ell_I$ and $\ell_{II}$ respectively. In both cases,

$$\hat{\alpha} = \frac{r}{N}, \qquad \hat{\beta} = \frac{n-r}{N}.$$

Maximum likelihood estimates for $p$ and $q$ follow immediately under Model I, either by observing that the moments $(\alpha, \beta)$ maps 1–1 onto the pair $(p, q)$ or by directly solving $\ell_I$. The solutions are given in Table 9. The asymptotic variance-covariance matrix for $p$ and $q$ is block-diagonal with well known elements $p(1-p)/n$ and $q(1-q)/N$. Observe that we now obtain only one solution, a strong argument in favour of the current method.

A similar standard derivation is not possible for Model II, since the triplet $(p, q_1, q_2)$ is redundant. This follows directly from Catchpole and Morgan (1997) and Catchpole, Morgan and Freeman (1998) whose theory shows that Model II is rank-deficient and Model I is of full rank. Since Model I is a submodel of Model II and saturates the observed data, so must every solution to $\ell_{II}$, implying the relationships:

$$pq_1 = \frac{r}{N}, \tag{6.5}$$

$$(1-p)q_2 = \frac{n-r}{N}. \tag{6.6}$$

Constraints (6.5) and (6.6) imply

$$\hat{p} = \frac{r}{Nq_1} = 1 - \frac{n-r}{Nq_2}$$

and hence

$$\frac{r}{q_1} + \frac{n-r}{q_2} = N. \tag{6.7}$$

The requirement that $q_1, q_2 \leq 1$ in (6.5) and (6.6) implies the range for $p$ wich was already found with the enumeration method (see Table 8).

## 6.2 Direct-Likelihood Approach

A variation to the observed-likelihood method consists of rewriting the log-likelihood for Model II in terms of $(p, q_1, q_2)$ before maximization:

$$\ell = r\ln(pq_1) + (n-r)\ln[(1-p)q_2] + (N-n)\ln[p(1-q_1) + (1-p)(1-q_2)]. \tag{6.8}$$

The likelihood equations are:

$$\frac{r}{p} - \frac{n-r}{1-p} + \frac{(N-n)(q_2-q_1)}{p(1-q_1)+(1-p)(1-q_2)} = 0, \tag{6.9}$$

$$\frac{r}{q_1} = \frac{(N-n)p}{p(1-q_1)+(1-p)(1-q_2)}, \tag{6.10}$$

$$\frac{n-r}{q_2} = \frac{(N-n)(1-p)}{p(1-q_1)+(1-p)(1-q_2)}. \tag{6.11}$$

Division of (6.11) by (6.10) yields

$$p = \frac{rq_2}{rq_2+(n-r)q_1}. \tag{6.12}$$

Substituting (6.12) into (6.10) produces (6.7), establishing equivalence between both methods.

The equivalence between direct-data likelihood and the transformation method does not hold in general, since the former starts from a saturated model for the observed data and then translates this into relations between parameters of $M^*$. The equivalence is thus restricted to models that saturate the observed data. The direct-likelihood method is the most general one since it allows for any model $M^*$ applied directly to the observed data.

### 6.3 Sensitivity Parameter Approach

As an alternative to fitting the overspecified likelihood, the overspecification can be removed by considering a minimal set of parameters $\boldsymbol{\eta}$, conditional upon which the others, $\boldsymbol{\mu}$, are identified. We term $\boldsymbol{\eta}$ the sensitivity parameter and $\boldsymbol{\mu}$ the estimable parameter. It is important to realize that in general there will not be a unique choice for $\boldsymbol{\eta}$ and hence for $\boldsymbol{\mu}$. Such a technique has been proposed for a specific example by Nordheim (1984). Each value of $\boldsymbol{\eta}$ will produce an estimate $\hat{\boldsymbol{\mu}}(\boldsymbol{\eta})$. The union of these produces the region of ignorance. A natural estimate of the region of uncertainty is the union of confidence regions for each $\hat{\boldsymbol{\mu}}(\boldsymbol{\eta})$. Note that one has to ensure that $\boldsymbol{\eta}$ is within the allowable range. As will be shown below, the choice of sensitivity parameter is non-unique and a proper choice can greatly simplify the treatment. Another issue is whether the parameters of direct scientific interest can overlap with the sensitivity set or not (see White and Goetghebeur 1997). For example, if the scientific question is a sensitivity analysis for treatment effect, then one should consider the implications of including the treatment effect parameters into the sensitivity set. There will be no direct estimate of imprecision available for the sensitivity parameter. Clearly, the particular choice of sensitivity parameter will not affect the estimate of the region of ignorance. However, the region of uncertainty is built from confidence regions which are conditional on a particular value of the sensitivity parameter and hence will typically vary with the choice made.

Table 10: Two Sensitivity Parameterizations for the Observed Data Likelihood.

| **Parameterization III** | **Parameterization IV** |
|---|---|
| *Parameterization:* | |
| $\alpha = pq$ | $\alpha = pq$ |
| $\beta = (1-p)(q+\theta)$ | $\beta = (1-p)q\lambda$ |
| $\gamma = 1 - pq - (1-p)(q+\theta)$ | $\gamma = 1 - pq - (1-p)q\lambda$ |

Let us exemplify this approach by considering two additional parameterizations of (5.4), as presented in Table 10. Models III and IV are based on $\theta$ and $\lambda$ as sensitivity parameters respectively. The maximum likelihood estimators for $p$ and $q$, given a value of the sensitivity parameter, will be subscripted by the sensitivity parameter. With some algebra, Model III is seen to imply $p_\theta = r/(Nq_\theta)$ and $q_\theta$ is found as the valid solution to

$$q_\theta = \frac{1}{2}\left[\frac{n}{N} - \theta \pm \sqrt{\left(\theta - \frac{n}{N}\right)^2 - 4\frac{r}{N}}\right].$$

Calculations are quickly getting cumbersome and therefore, rather than pursuing this approach, we will turn attention to Model IV. Some algebra yields simple expressions for $p_\lambda$ and $q_\lambda$:

$$p_\lambda = \frac{\hat{\alpha}\lambda}{\hat{\beta} + \hat{\alpha}\lambda} = \frac{\lambda r}{n - r(1-\lambda)}, \tag{6.13}$$

$$q_\lambda = \frac{\hat{\beta} + \hat{\alpha}\lambda}{\lambda} = \frac{n - r(1-\lambda)}{N\lambda}. \tag{6.14}$$

Using the delta method, an asymptotic variance-covariance matrix of $p_\lambda$ and $q_\lambda$ is seen to be built from:

$$\widehat{\mathrm{Var}}(\widehat{(p_\lambda)}) = \frac{p_\lambda(1-p_\lambda)}{N\lambda q_\lambda}\left\{1 + \frac{1-\lambda}{\lambda}(1-p_\lambda)[1 - p_\lambda q_\lambda(1-\lambda)]\right\}, \tag{6.15}$$

$$\widehat{\mathrm{Cov}}((\widehat{p_\lambda}, \widehat{q_\lambda})) = -\frac{1}{N}p_\lambda(1-p_\lambda)\frac{1-\lambda}{\lambda}q_\lambda,$$

$$\widehat{\mathrm{Var}}(\widehat{(q_\lambda)}) = \frac{q_\lambda(1-q_\lambda)}{N}\left\{1 + \frac{1-p_\lambda}{1-q_\lambda}\frac{1-\lambda}{\lambda}\right\}.$$

Note that the parameter estimates are asymptotically correlated, except when $\lambda = 1$, i.e., under the

Table 11: Limiting Cases for the Sensitivity Parameter Analysis.

| Estimator | $\lambda$ | $\lambda = \frac{n-r}{N-r}$ | $\lambda = 1$ | $\lambda = \frac{N-(n-r)}{r}$ |
|---|---|---|---|---|
| $p_\lambda$ | $\frac{\lambda r}{n-r(1-\lambda)}$ | $\frac{r}{N}$ | $\frac{r}{N}$ | $\frac{N-n+r}{N}$ |
| $q_\lambda$ | $\frac{n-r(1-\lambda)}{N\lambda}$ | 1 | $\frac{n}{N}$ | $\frac{r}{N-(n-r)}$ |
| $q_\lambda\lambda$ | $\frac{n-r(1-\lambda)}{N}$ | $\frac{n-r}{N-r}$ | $\frac{n}{N}$ | 1 |
| $\frac{p_\lambda}{1-p_\lambda}$ | $\lambda\frac{r}{n-r}$ | $\frac{r}{N-r}$ | $\frac{r}{n-r}$ | $\frac{N-(n-r)}{n-r}$ |

MAR assumption, or under boundary values ($p_\lambda = 0, 1; q_\lambda = 0$). This is in line with the ignorable nature of the MAR model (Rubin, 1976).

We need to determine the set of allowable values for $\lambda$ by requiring $0 \leq p_\lambda, q_\lambda, \lambda q_\lambda \leq 1$. These six inequalities reduce to

$$\lambda \in \left[\frac{n-r}{N-r}, \frac{N-(n-r)}{r}\right].$$

Clearly, $\lambda = 1$ is always valid.

Table 11 presents estimates for limiting cases. The interval of ignorance for the success probability is thus seen to be as in Table 8. It is interesting to observe that the interval of ignorance for the success odds is linear in the sensitivity parameter and equal to

$$\text{odds}(p) : \left[\frac{r}{N-r}, \frac{N-n+r}{n-r}\right].$$

For the success probability, the variance of $p_\lambda$ is given by (6.15). For the success odds, we obtain:

$$\widehat{\text{Var}}(\widehat{\text{odds}}(p_\lambda)) = \frac{1}{N\lambda q_\lambda}\frac{p_\lambda}{1-p_\lambda}\left\{1 + \frac{1-\lambda}{\lambda}(1-p_\lambda)[1-p_\lambda q_\lambda(1-\lambda)]\right\}$$

and for the success logit:

$$\widehat{\text{Var}}(\widehat{\text{logit}}(p_\lambda)) = \frac{1}{N\lambda q_\lambda}\frac{1}{p_\lambda(1-p_\lambda)}\left\{1 + \frac{1-\lambda}{\lambda}(1-p_\lambda)[1-p_\lambda q_\lambda(1-\lambda)]\right\}.$$

For each $\lambda$, a confidence interval $C_\lambda$ can be constructed for every point within the allowable range of $\lambda$. The union of the $C_\lambda$ is the *interval of uncertainty*, for either $p_\lambda$, its odds, or its logit. For a set of data with $N = 50, n = 40, r = 20$, these intervals are depicted graphically in Figure 1.

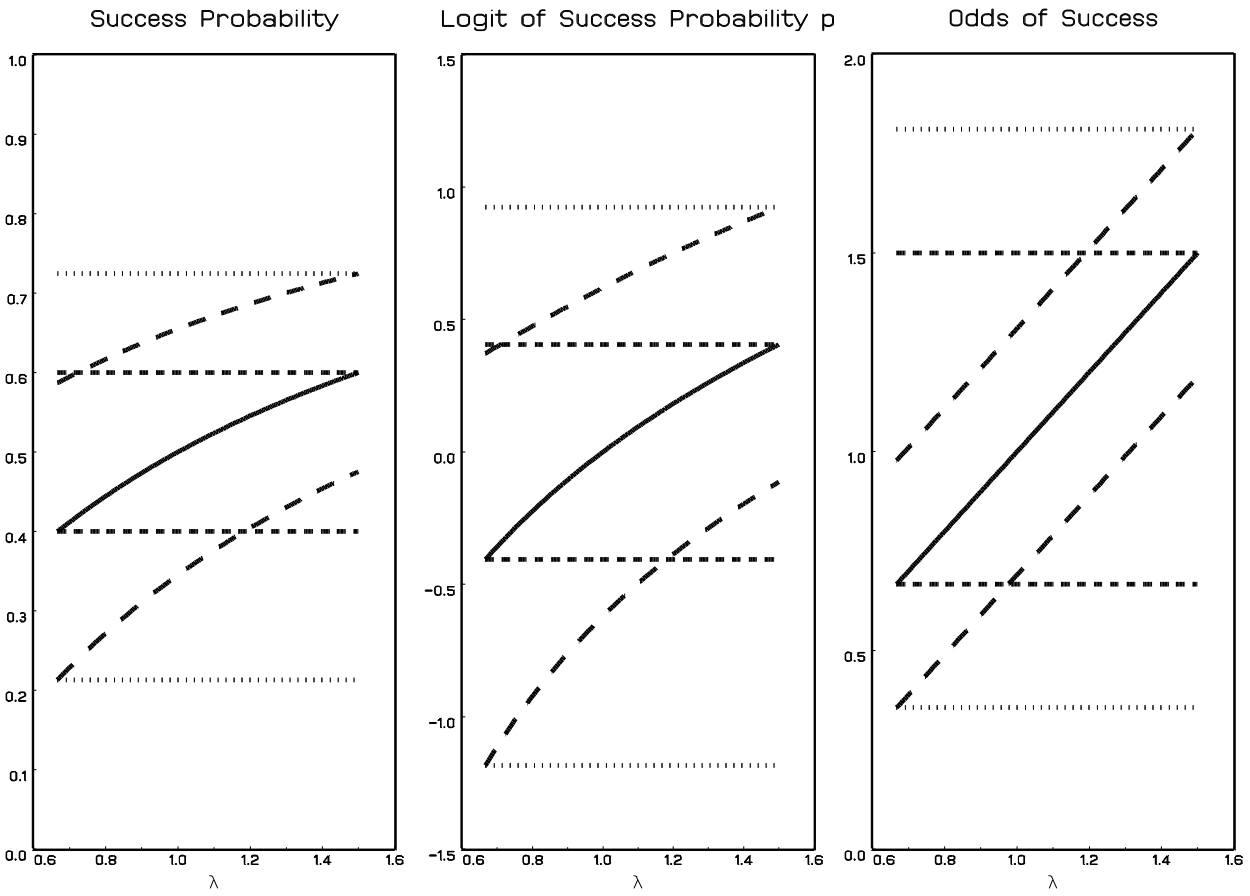


Figure 1: *Graphical Representation of Interval of Ignorance and Interval of Uncertainty For a Single Binary Outcome With Supplemental Count.*

# 7 Comparison of Methods

The simple examples in our development have shown that all methods yield the same results for models that saturate the complete-data. In contrast, non-saturated methods produce differences. The MAR model, applied to the $2 \times 1$ table, yields a range of solutions with the enumeration method, but only one solution with the observed-data procedures. The transformation method is only designed for models that saturate the observed-data likelihood.

Both procedures are completely equivalent for a saturated model $M^*$, but not otherwise. This results holds in general since, under a saturated model, the solution to the complete-data likelihood is in 1–1 relation with the corresponding table. Thus, each solution to the observed

data likelihood obeys the same mapping. Therefore, under a saturated $M^*$, the observed-likelihood is equivalent to enumeration. From the counterexample it is clear that this result is confined to saturated models.

In conclusion, the enumeration method and similarly the transformation method, are easy to apply but confined to saturated models and therefore hard to generalize to more complicated data settings. The direct-likelihood and sensitivity parameter methods are more promising. Both produce exactly the same regions of ignorance. This is clear by observing that the sensitivity parameter method extends profile likelihood (Welsh 1996, p. 97).

# 8 Analysis of Psychiatric Study

## 8.1 Monotone Patterns

In Section 3 five models were introduced for a $2 \times 2$ contingency table with supplemental margin (Table 5). The first three were fitted to the monotone cases of the psychiatric study. Since these models were identified classical methods could be used. We will now fit overspecified Models 4 and 5 to the same data for the side-effects outcome.

We will focus on three quantities of interest: the first marginal probability, the second marginal probability, and the (log) odds ratio describing the association between the first and the second outcome. Results for all five models are given in Table 12. Models 1 and 2 are both ignorable and hence all measurement model quantities are independent of the choice between MAR and MCAR. Since Models 1–3 are identified, the interval of ignorance collapses to a classical point estimate and the corresponding interval of uncertainty is an ordinary confidence interval.

For Model 4, there is one sensitivity parameter, which we choose to be $\gamma$ (measuring the extent of non-randomness). When $\gamma = 0$ MAR Model 2 is recovered. The value of $\gamma$ which corresponds to $\beta = 0$ yields protective Model 3. Since there is only one sensitivity parameter, a graphical representation (Figure 2) is straightforward. The heavy lines represent the point estimates conditional on the sensitivity parameter, together with their bounds (interval of ignorance). The thin lines are the corresponding delta method confidence intervals, together with their bounds (interval of uncertainty). Since among the monotone cases the first measurement is always obtained, there is no ignorance about the first marginal probability and hence the interval of ignorance for

Table 12: Interval of Ignorance (II) and Interval of Uncertainty (IU) for Marginal Probabilities and (Log) Odds Ratio for Monotone Patterns of Side-Effects Data (Psychiatric Study).

| Parameter | | Model 1/2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| First Margin | II | 0.43 | 0.43 | 0.43 | 0.43 |
| | IU | [0.37;0.48] | [0.37;0.48] | [0.37;0.48] | [0.37;0.48] |
| Second Margin | II | 0.64 | 0.59 | [0.49;0.74] | [0.49;0.74] |
| | IU | [0.58;0.70] | [0.53;0.65] | [0.43;0.79] | [0.43;0.79] |
| Log Odds Ratio | II | 2.06 | 2.06 | [1.52;2.08] | [0.41;2.84] |
| | IU | [1.37;2.74] | [1.39;2.72] | [1.03;2.76] | [0.0013;2.84] |
| Odds Ratio | II | 7.81 | 7.81 | [4.57;7.98] | [1.50;17.04] |
| | IU | [3.95;15.44] | [4.00;15.24] | [2.79;15.74] | [1.0013;32.89] |

this quantity is still a point. This is not true for the other two quantities.

Often, fitting a couple of identifiable models (e.g., Model 2 and Model 3) is seen as a sensitivity analysis. This example shows this can be misleading. Both models differ by about 0.05 in the second marginal probability, but the II of Model 4 shows the range is about 0.25 ! Similarly, Models 2 and 3 yield virtually the same result for the odds ratio, but the II of Model 4 shows there is still some variation possible.

The impact of fitting an overspecified but, at the complete-data level, non-saturated model is seen by contrasting Model 4 with the fully saturated Model 5. As expected, both models coincide for the first marginal probability. It turns out that there is little difference in II and IU for the second marginal probability (no difference seen to the precision reported). In contrast, the length of the II for the log odds ratio is now about 5 times longer. The lower limit of the IU (Model 5) is very close to zero, whereas its Model 4 counterpart shows clear evidence for a strong positive association between both outcomes.

By construction, the data do not provide evidence for choosing between Models 4 and 5. Both are overspecified at the observed data level and both encompass Models 2 and 3. Model 5 is saturated at the observed data level as well and therefore the limits derived from it are absolute. The reduced width of the intervals produced under Model 4 are entirely due to the model assumption

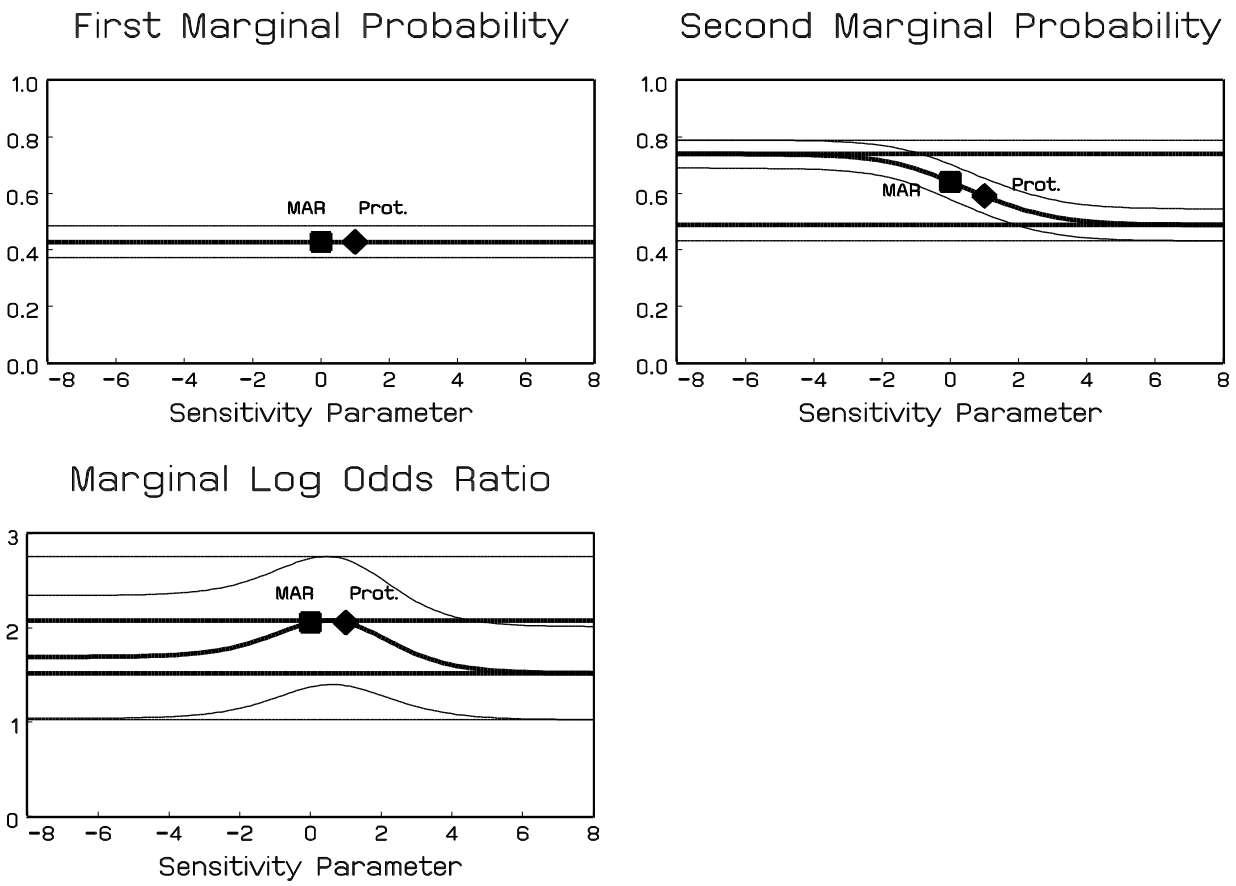


Figure 2: *Graphical Representation of Intervals of Ignorance and Intervals of Uncertainty for Mononotone Patterns of Psychiatric Study (Side Effects).*

that the dropout probability depends on both outcomes through their main effects only. Such an assumption has to be the subject of careful reflection and, while it can be avoided in this simple data setting, it cannot in more complex categorical data settings where covariates are recorded or with continuous response variables.

### 8.2 Non-Monotone Patterns

Let us now turn attention to the more general but more complicated problem of modelling all patterns in Table 1. The discrepancy between full data and observed data is larger, since the full data comprise 15 degrees of freedom, while there are only 8 observed degrees of freedom. An interesting class of models has been proposed by Baker, Rosenberger, and DerSimonian (1992). It

is based on log-linear models for the four-way classification of both outcomes, together with their respective missingness indicators. Denote the counts by $Y_{r_1r_2jk}$ where $r_1, r_2 = 0, 1$ indicate whether a measurement is missing or taken at occasions 1 and 2 respectively, and $j, k = 1, 2$ indicate the response categories for both outcomes. The models are written as:

$$\begin{aligned} E(Y_{11jk}) &= m_{jk}, & E(Y_{01jk}) &= m_{jk}\alpha_{jk}, \\ E(Y_{10jk}) &= m_{jk}\beta_{jk}, & E(Y_{00jk}) &= m_{jk}\alpha_{jk}\beta_{jk}\gamma, \end{aligned}$$

with $m_{jk} = Y_{++++}\pi_{11jk}$, and

$$\alpha_{jk} = \frac{q_{01|jk}}{q_{11|jk}}, \qquad \beta_{jk} = \frac{q_{10|jk}}{q_{11|jk}}, \qquad \gamma = \frac{q_{11|jk}q_{00|jk}}{q_{10|jk}q_{01|jk}}.$$

The subscripts are missing from $\gamma$ since Baker *et al* have shown that this quantity is independent of $j$ and $k$. These authors consider nine identifiable models, based on setting $\alpha_{jk}$ and $\beta_{jk}$ constant in one or more indices:

$$\begin{array}{lllllllll} \text{BRD1} & : & (\alpha, \beta) & \text{BRD4} & : & (\alpha, \beta_k) & \text{BRD7} & : & (\alpha_k, \beta_k) \\ \text{BRD2} & : & (\alpha, \beta_j) & \text{BRD5} & : & (\alpha_j, \beta) & \text{BRD8} & : & (\alpha_j, \beta_k) \\ \text{BRD3} & : & (\alpha_k, \beta) & \text{BRD6} & : & (\alpha_j, \beta_j) & \text{BRD9} & : & (\alpha_k, \beta_j). \end{array}$$

Interpretation is straightforward. For example, BRD1 is MCAR, in BRD4 missingness in the first variable is constant, while missingness in the second variable depends on its value. Two of the main advantages of this family is ease of computation in general, and the existence of closed-form solution for several of its members (BRD2–BRD9). Molenberghs, Goetghebeur, and Lipsitz (1997) used these models in an informal sensitivity analysis of the psychiatric study. The fit of the models with an interior solution is given in Table 13 and the fitted counts are displayed in Table 14.

We consider a slightly different but equivalent parameterization

$$\pi_{r_1r_1jk} = p_{jk}\frac{\exp[\beta_{jk}(1-r_2) + \alpha_{jk}(1-r_1) + \gamma(1-r_1)(1-r_2)]}{1 + \exp(\beta_{jk}) + \exp(\alpha_{jk}) + \exp(\beta_{jk} + \alpha_{jk} + \gamma)}, \tag{8.16}$$

which contains the marginal success probabilities $p_{jk}$ and forces the missingness probabilities to obey their range restrictions.

Table 13 reveals that Models BRD1–BRD9 show little variation in the marginal probabilities and in the measure of association. Considered as an informal sensitivity analysis, this could be seen as evidence for the robustness of these measures. However, we have seen in Section 8.1 that such a conclusion could be deceptive. To this end, we consider an overspecified model, analogous

Table 13: Model Fit For Side Effects (d.f.: degrees of freedom; $G^2$: likelihood ratio test statistic, $P$ value, estimates and confidence limits for marginal probabilities and marginal (log) odds ratio.) For Model 10, II and IU are presented instead.

| Model | d.f. | $G^2$ | $P$ | Marg. Prob. | | Odds Ratio | |
|---|---|---|---|---|---|---|---|
| | | | | First | Second | Orig. | Log |
| BRD1 | 6 | 4.5 | 0.104 | 0.43[0.37;0.49] | 0.64[0.58;0.71] | 7.80[3.94;15.42] | 2.06[1.37;2.74] |
| BRD2 | 7 | 1.7 | 0.192 | 0.43[0.37;0.48] | 0.64[0.58;0.70] | 7.81[3.95;15.44] | 2.06[1.37;2.74] |
| BRD3 | 7 | 2.8 | 0.097 | 0.44[0.38;0.49] | 0.66[0.60;0.72] | 7.81[3.95;15.44] | 2.06[1.37;2.74] |
| BRD4 | 7 | 1.7 | 0.192 | 0.43[0.37;0.48] | 0.58[0.49;0.68] | 7.81[3.95;15.44] | 2.06[1.37;2.74] |
| BRD7 | 8 | 0.0 | - | 0.44[0.38;0.49] | 0.61[0.53;0.69] | 7.81[3.95;15.44] | 2.06[1.37;2.74] |
| BRD9 | 8 | 0.0 | - | 0.43[0.38;0.49] | 0.66[0.60;0.72] | 7.63[3.86;15.10] | 2.03[1.35;2.71] |
| Model 10:II | 9 | 0.0 | - | [0.425;0.429] | [0.47;0.75] | [4.40;7.96] | [1.48;2.07] |
| Model 10:IU | 9 | 0.0 | - | [0.37;0.49] | [0.41;0.80] | [2.69;15.69] | [0.99;2.75] |

Table 14: Complete Data Counts For Models Fitted to Side Effects Data.

| | (1,1) | | (1,0) | | (0,1) | | (0,0) | | (+,+) | |
|---|---|---|---|---|---|---|---|---|---|---|
| BRD1 | 84.00 | 12.12 | 28.13 | 4.06 | 0.74 | 0.11 | 5.26 | 0.76 | 118.13 | 17.05 |
| | 60.21 | 67.67 | 20.16 | 22.66 | 0.53 | 0.60 | 3.77 | 4.23 | 84.67 | 95.16 |
| BRD2 | 89.42 | 12.89 | 22.73 | 3.27 | 0.80 | 0.12 | 4.24 | 0.61 | 117.19 | 16.89 |
| | 57.27 | 64.42 | 23.06 | 25.94 | 0.51 | 0.58 | 4.30 | 4.82 | 85.14 | 95.76 |
| BRD3 | 83.67 | 12.22 | 28.02 | 4.09 | 1.17 | 0.00 | 8.16 | 0.00 | 121.01 | 16.31 |
| | 59.85 | 68.25 | 20.04 | 22.85 | 0.83 | 0.00 | 5.84 | 0.00 | 86.57 | 91.11 |
| BRD4 | 89.42 | 12.89 | 18.58 | 7.42 | 0.80 | 0.12 | 3.47 | 1.39 | 112.27 | 21.82 |
| | 57.27 | 64.42 | 11.90 | 37.10 | 0.51 | 0.58 | 2.22 | 6.93 | 71.90 | 109.03 |
| BRD7 | 89.00 | 13.00 | 18.58 | 7.42 | 1.22 | 0.00 | 8.53 | 0.00 | 117.33 | 20.42 |
| | 57.00 | 65.00 | 11.90 | 37.10 | 0.78 | 0.00 | 5.47 | 0.00 | 75.15 | 102.10 |
| BRD9 | 89.00 | 13.00 | 22.69 | 3.31 | 1.22 | 0.00 | 6.97 | 0.00 | 119.87 | 16.31 |
| | 57.00 | 65.00 | 22.89 | 26.11 | 0.78 | 0.00 | 7.03 | 0.00 | 87.71 | 91.11 |

to Model 4 in Table 5. The choice can be motivated by observing that both BRD7 and BRD9 yield an interior solution and differ only in the $\beta$-model. Therefore, Model 10 will be defined as $(\alpha_k, \beta_{jk})$

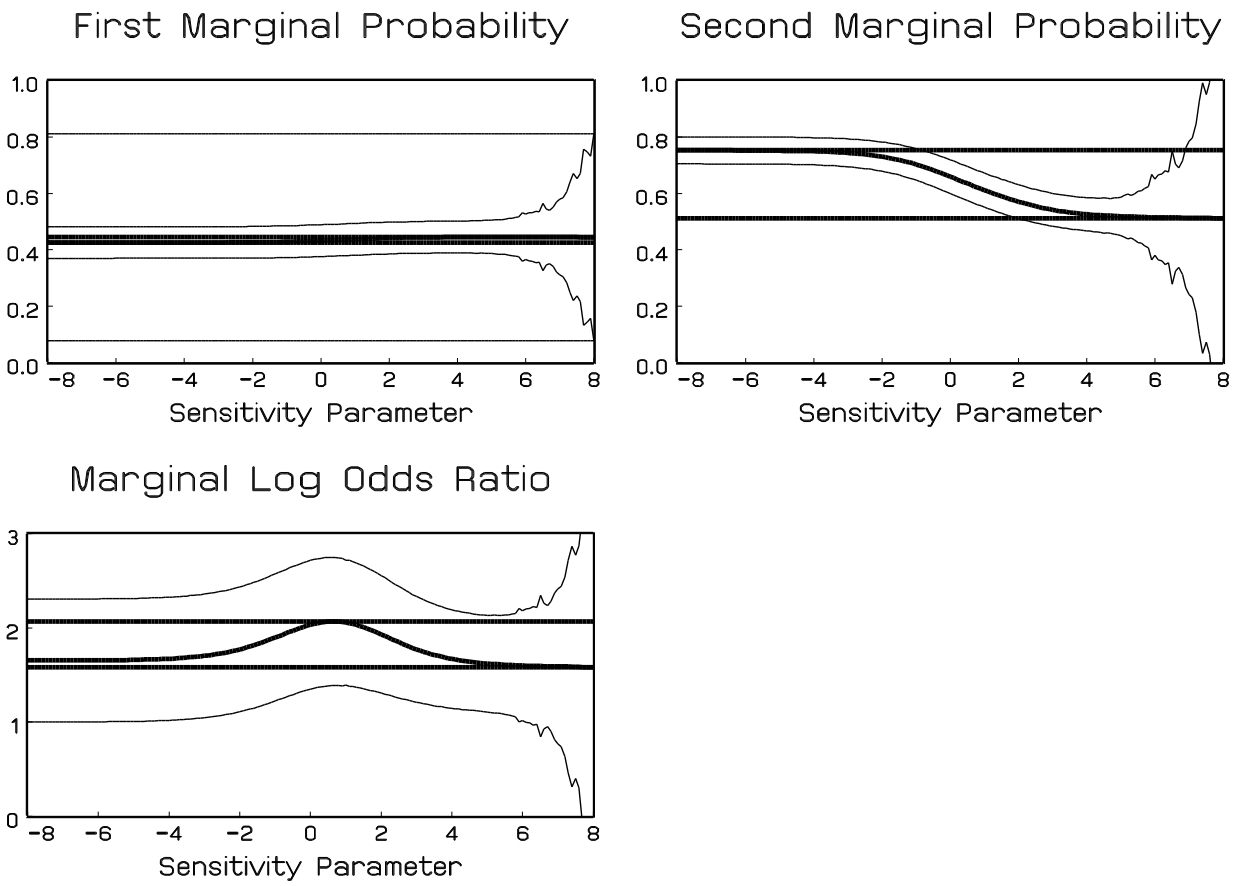


Figure 3: *Graphical Representation of Intervals of Ignorance and Intervals of Uncertainty for Non-Mononotone Patterns of Psychiatric Study (Side Effects).*

with

$$\beta_{jk} = \beta_0 + \beta_j + \beta_k. \tag{8.17}$$

Since one parameter is redundant, we propose to use $\beta_k$ as the sensitivity parameter. The resulting II and IU are plotted in Figure 3. While the II seem acceptable and in fact close to their counterparts in Figure 2, there is clearly aberrant behaviour in the IU towards larger values of the sensitivity parameter, leading to very wide IU's. This problem is entirely due to the zero count in pattern (0,1) (see Table 1), as can be seen by adding 0.5 to this zero count. The results are presented in Figure 4. The resulting II and IU are presented in Table 13, and they are very similar to the results for Model 4, as displayed in Table 12. Due to the non-monotone patterns, there is a (very small)

Figure 4: *Graphical Representation of Intervals of Ignorance and Intervals of Uncertainty for Mononotone Patterns of Psychiatric Study (Side Effects). A Value of 0.5 is Added to the Zero Count in Pattern (1,0).*

ignorance in the first marginal probability as well. Once again, it is seen that fitting identifiable models only may be misleading since, for example, the log odds ratio shows much more variability than seen among Models BRD1–BRD9.

## 9 Analysis of the Slovenian Plebiscite

Rubin, Stern, and Vehovar (1995) conducted several analyses of the data. Their main emphasis was on determining the proportion $\theta$ of the population that would attend the plebiscite and vote for independence. The three other combinations of both binary outcomes would be treated as voting "no". Their estimates are reproduced in Table 15. The conservative method is the ratio of the

Table 15: Estimates of the proportion $\theta$ attending the plebiscite and voting for independence, as presented in Rubin, Stern, and Vehovar (1995).

| Estimation method | $\theta$ | NO via nonattendance |
|---|---|---|
| Conservative | 0.694 | 0.192 |
| Complete cases | 0.928 | 0.020 |
| Available cases | 0.929 | 0.021 |
| MAR (2 questions) | 0.892 | 0.042 |
| MAR (3 questions) | 0.883 | 0.043 |
| Non-ignorable | 0.782 | 0.122 |
| Plebiscite | 0.885 | 0.065 |

(yes, yes) answers to the (attendance, independence) pair and the total sample, i.e., 1439/2076. This is the most pessimistic scenario. At the opposite end of the spectrum, we can add the most optimistic estimate that replaces the numerator by all who are no definite "no":

$$\frac{1439 + 159 + 144 + 136}{2076} = \frac{1878}{2076} = 0.904.$$

Both estimates together yield the range $II = [0.694; 0.904]$. The complete case estimate is based on the subjects answering all three questions and the available case estimate is based on the subjects answering the two questions of interest here. It is noteworthy that both estimates fall outside the interval of ignorance and should be disregarded.

There are two MAR models, the first one based on two questions only, the second one using all three. The non-ignorable model is based on the assumption that missingness on a question depends on the answer to that question but not on the other questions. The authors argue that the MAR results are very close to the true response, unlike the non-ignorable model, and suggest that the MAR assumption is generally more plausible in carefully designed surveys.

The result of fitting the models of Baker, Rosenberger, and DerSimonian (1992) is presented in Table 16. Observe that BRD1, being MAR, is equivalent to MAR (2 questions) in Table 15. Model BRD2 produces an estimate for $\theta$ which is extremely close to the results of the plebiscite. It assumes that missingness on the independence question depends on the attendence question.

Table 16: Estimates of the proportion $\theta$ (confidence interval) attending the plebiscite and voting for independence, following from fitting the Baker, Rosenberger, and DerSimonian models (1992).

| Model | d.f. | loglik | $\theta$ |
|---|---|---|---|
| BRD1 | 6 | -2503.06 | 0.891[0.877;0.906] |
| BRD2 | 7 | -2476.38 | 0.884[0.868;0.899] |
| BRD3 | 7 | -2471.59 | 0.881[0.865;0.896] |
| BRD4 | 7 | -2476.38 | 0.779[0.702;0.857] |
| BRD5 | 7 | -2471.59 | 0.848[0.814;0.882] |
| BRD6 | 8 | -2440.67 | 0.822[0.792;0.850] |
| BRD7 | 8 | -2440.67 | 0.774[0.719;0.828] |
| BRD8 | 8 | -2440.67 | 0.753[0.691;0.815] |
| BRD9 | 8 | -2440.67 | 0.866[0.849;0.884] |
| Model 10 | 9 | -2440.67 | [0.762;0.893][0.744;0.907] |
| Model 11 | 9 | -2440.67 | [0.766;0.883][0.715;0.920] |
| Model 12 | 10 | -2440.67 | [0.694;0.904] |

Note that BRD8 assumes that missingness on either question depends on the question itself and therefore is very similar to the non-ignorable model of Rubin, Stern, and Vehovar (1995).

Next, we present three estimated II. Model 10 in Table 16 is based on 1 sensitivity parameter, as in (8.17). Similarly, Model 11 uses

$$\alpha_{jk} = \alpha_0 + \alpha_j + \alpha_k, \tag{9.18}$$

while Model 12 combines both (8.17) and (9.18). The estimated II for Models 10 and 11 are very similar and the true plebiscite value is marginal within these II. Note that Model 11, and hence also Model 12, does contain a number of boundary solutions, which can be seen as evidence against these models.

Another quantity which Rubin, Stern, and Vehovar (1995) reported is the proportion of NO's via nonattendance (see Table 15). Observe that all estimates are way below the plebiscite value. We can gain some insight in this phenomenon by plotting the region of ignorance for both $\theta$ and the proportion of NO's via nonattendance. Since Models 10 and 11 are based on a single sensitivity parameter, the regions of ignorance are curves, while a planar region is obtained for

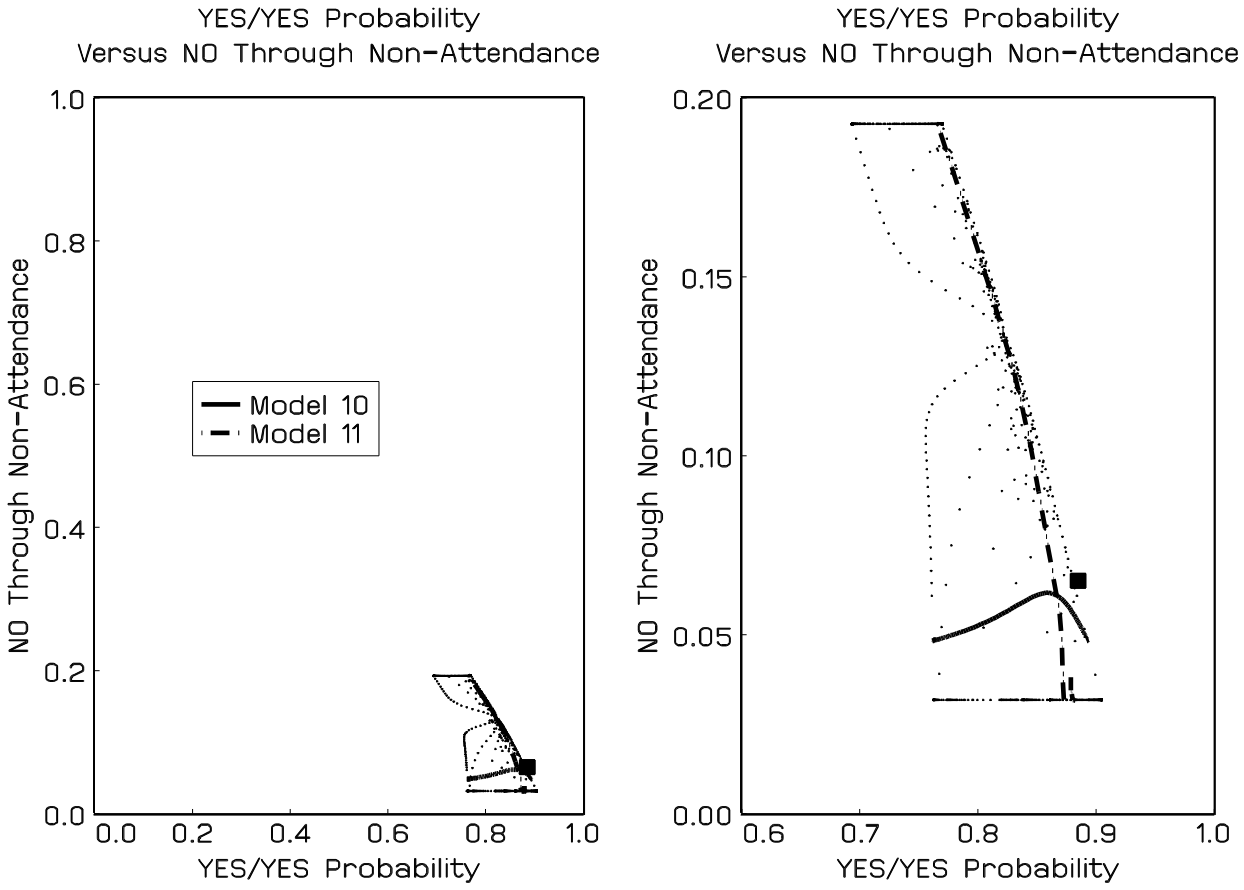


Figure 5: *Graphical Representation of Regions of Ignorance for the Slovenian Plebiscite. Proportion of YES Votes Versus Proportion of NO via Nonattendance. Models 10, 11, and 12.*

Model 12. A graphical sketch is given in Figure 5, where Models 10 and 11 are represented by curves and Model 12 by the result of sampled points from the bivariate sensitivity parameter. Model 10 incorporates relatively little ignorance about NO via nonattendance, but at the same time fails to include the plebiscite value. Models 11 and 12 on the contrary allow a relatively large II for this quantity. A black square marks the plebiscite values for both quantities. It is clear that the plebiscite result is *outside* of the range produced by Model 12. Note that a saturated model would incorporate 5 extra sensitivity parameters !

## 10 Validity

Given a family of probability models, for the complete data distribution, we have identified the theoretical region of ignorance, $\mathcal{R}$, implied by the observed data structure for one or more parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ of interest. Several methods were proposed to estimate such region from the data. For a frequentist, it is valuable to study the sampling properties of the stochastic region to enable formal evaluation and contrasting of estimators. Hence, we must define expected values and convergence of stochastic regions. To cover precision we will involve measures of overlap and distance between such regions.

For a single key parameter of interest, the region of ignorance may be an interval and hence determined by 2 points. One could define expectation and convergence for the interval through expectation and convergence of the characterizing 2-dimensional vector. Similar arguments apply to other regularly shaped regions such as rectangles and ellipsoids.

For a more general discussion of validity of an estimated region of ignorance, we propose the following definitions of weak and strong consistency.

**Definition 1 (Weak Consistency of an Estimated Region of Ignorance)** *The estimated region of ignorance $R_n$ for $\boldsymbol{\theta}$ is* weakly consistent *if it converges in probability to the true region of ignorance $\mathcal{R}$ :*

$$\forall \epsilon > 0, \delta > 0, \exists\, n_0 \, s.t. \; \forall n > n_0 : \Pr(||R_n - \mathcal{R}|| > \delta) < \epsilon,$$

where the chosen distance measure between $R_n$ and $\mathcal{R}$ is based on a norm function $||.||$ with maximum value 1 and minimum value 0 in the $k$-dimensional space.

**Definition 2 (Strong Consistency of an Estimated Region of Ignorance)** *The estimated region of ignorance $R_n$ for $\boldsymbol{\theta}$ is* strongly consistent *if*

$$\Pr(||R_n - \mathcal{R}|| \to 0 \; as \; n \to \infty) = 1,$$

*with $\mathcal{R}$ the true region of ignorance.*

Following the laws of large numbers, in the i.i.d. case for the multinomial table, it suffices

to substitute observed proportions for cell probabilities in the boundary expression for $\mathcal{R}$ to obtain a strongly consistent estimator.

An estimated interval of uncertainty, $U_n$, is wider than $R_n$ by construction and aims to cover the true region with 95% probability. We propose two definitions which coincide in the case of a classic point estimator and reduce then to the classical notion of coverage for a $(1-\alpha)100\%$ confidence interval.

**Definition 3 (Weak Coverage Probability of the Region of Uncertainty)**

$$E\left\{\frac{||U_n \cap \mathcal{R}||}{||\mathcal{R}||}\right\}$$

Hence we get the expected proportion of volume of $\mathcal{R}$ covered by $R_n$.

**Definition 4 (Strong Coverage Probability of the Region of Uncertainty)**

$$\Pr(\mathcal{R} \subset U_n)$$

Consider Example 3 with $\theta = \Pr(Y_2 = 1)$. The theoretical region of ignorance is then the interval $[\pi_{1+1}, \pi_{1+1} + \pi_{0++}]$. A strongly consistent estimator is provided by $[\widehat{\pi_{1+1}}, \widehat{\pi_{1+1}} + \widehat{\pi}_{0++}]$.

A simple interval acknowledging imprecision has bounds

$$\left[\widehat{\pi_{1+1}} - c_{1-\alpha/2}\sqrt{\hat{\mathrm{Var}}(\widehat{\pi_{1+1}})}; \widehat{\pi_{1+1}} + \widehat{\pi}_{0++} + c_{1-\alpha/2}\sqrt{\hat{\mathrm{Var}}(\widehat{\pi_{1+1}} + \hat{\pi}_{0++})}\right]$$

where the critical point $c_{1-\alpha/2}$ is derived from the binomial or asymptotically normal distribution for the observed corresponding combination of proportions.

For this interval, weak coverage is expected to be larger than $(1-\alpha)100\%$ and strong coverage to be smaller. Indeed, at the 95% level we find 0.9617 and 0.9199 respectively. Through back calculation one can find the critical point that gives the exact weak or strong coverage that is desired. For the weak coverage we find 0.9506 for $c_{1-\alpha/2} = 1.5$. A graphical representation of the simulation process is given in Figure 6.

## 11 Discussion

In this paper we have defined the concept of *ignorance* and incorporated this into a frequentist framework by combining it with the familiar idea of statistical imprecision, producing a measure of

Figure 6: *The twenty first simulations where the estimated interval of ignorance is represented as a bold line segment, and the interval of uncertainty is shown as a dashed line segment. The vertical lines show theoretical interval of ignorance plus the true* $P(Y_2 = 1) = 0.5$.

*uncertainty*. As an extension of the concept of confidence, uncertainty is expressed as an interval for scalar unknowns (parameters) and a region for vectors. These reduce to conventional confidence intervals and regions when it is assumed that there is no ignorance about the statistical model underlying the data. In the special case of simple categorical data settings with missing values, we have seen that intervals of ignorance can be constructed in a relatively straightforward way, and these reflect in a natural way ignorance about underlying relationships involving unobserved data. The construction of the intervals of uncertainty in the examples are seen to convey useful

information about the problems concerned, providing information not previously appreciated. In particular, we see that earlier conclusions about the selection and behaviour of classes of models for the Slovenian Plebiscite are not strictly justified.

We can approach the calculation of the interval of ignorance in several ways, but it is seen that a (possibly) overspecified model and associated likelihood are the more natural concepts to use. This approach will be important, if not essential when we extend the simple categorical setting to more complex problems with continuous and finely stratified covariates and, very importantly, to continuous responses. For these the complete table approach is not suitable because a missing continuous value should be replaced by an unbounded set, such as the whole, or half, real line. In these settings careful consideration needs to be given to the family of models within which we are assumed to be ignorant.

Formal tools to assess validity of the new concepts are clearly needed. We suggested consistency definitions for the region of ignorance and coverage for the region of uncertainty. They extend familiar concepts used when there is no ignorance and they might provide a reasonable starting point for further exploration of the notions introduced.

## References

Baker, S.G. and Laird, N.M. (1988) Regression analysis for categorical variables with outcome subject to nonignorable nonresponse. *Journal of the American Statistical Association*, **83**, 62–69.

Baker, S.G., Rosenberger, W.F., and DerSimonian, R. (1992) Closed-form estimates for missing counts in two-way contingency tables. *Statistics in Medicine*, **11**, 643–657.

Bickel, P.J. and Doksum, K.A. (1977) *Mathematical Statistics*. Englewood Cliffs: Prentice Hall.

Brown, C.H. (1990) Protecting against nonrandomly missing data in longitudinal studies. *Biometrics*, **46**, 143–155.

Catchpole, E.A. and Morgan, B.J.T. (1997) Detecting parameter redundancy. *Biometrika*, **84**, 187–196.

Catchpole, E.A., Morgan, B.J.T., and Freeman, S.N. (1998) Estimation in parameter-redundant models. *Biometrika*, **85**, 000–000.

Clayton, D. and Hills, M. (1993) *Statistical Models in Epidemiology.* Oxford: Oxford University Press.

Copas, J. and Eguchi, S. (1997) Sensitivity approximations for selectivity bias in observational data analysis. *Submitted for publication.*

Dale, J.R. (1986) Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics*, **42**, 909–917.

Diggle, P.D. and Kenward, M.G. (1994) Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics*, **43**, 49–93.

Draper, D. (1995) Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society, Series B*, **57**, 45–97.

Glynn, Laird and Rubin (1986) Selection Modelling versus mixture modelling with nonignorable nonresponse. In *Drawing Inferences from Self-Selected Samples*, Ed. H. Wainer, pp. 115–142. New York: Springer Verlag.

Heitjan, D.F. and Rubin, D.B. (1991) Ignorability and coasre data. *The Annals of Statistics*, **19**, 2244–2253.

Kenward, M.G. (1998) The effect of untestable assumptions for data with nonrandom dropout. *Statistics in Medicine*, 00, 000–000.

Laird, N.M. (1988) Missing data in longitudinal studies. *Statistics in Medicine*, **7**, 305–315.

Lang, J.B. and Agresti, A. (1994) Simultaneously modelling joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association*, **89**, 625–632.

Lesaffre, E., Molenberghs, G., and Dewulf, L. (1996) Effect of dropouts in a longitudinal study: an application of a repeated ordinal model. *Statistics in Medicine*, **15**, 1123–1141.

Liang, K.-Y., Zeger, S.L., and Qaqish, B. (1992) Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society, Series B*, **54**, 3–40.

Lipsitz, S.R., Laird, N.M., and Harrington, D.P. (1991) Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika*, **78**, 153–160.

Little, R.J.A. and Rubin, D.B. (1987) *Statistical Analysis with Missing Data*. New York: Wiley.

Little, R.J.A. (1993) Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, **88**, 125–134.

Little, R.J.A. (1994) A class of pattern-mixture models for normal incomplete data. *Biometrika*, **81**, 471–483.

Little, R.J.A. (1995) modelling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, **90**, 1112–1121.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models. Second Edition*. London: Chapman and Hall.

Michiels, B. and Molenberghs, G. (1997) Protective estimation of longitudinal categorical data with nonrandom dropout. *Communications in Statistics*, **26**, 65–94.

Molenberghs, G. and Goetghebeur, E. (1997) Simple fitting algorithms for incomplete categorical data. *Journal of the Royal Statistical Society, Series B*, **59**, 401–414.

Molenberghs, G., Goetghebeur, E., and Lipsitz, S.R. (1998) Non-random missingness in categorical data: strengths and limitations. *Submitted*.

Molenberghs, G., Kenward, M.G., and Lesaffre, E. (1997) The analysis of longitudinal ordinal data with nonrandom dropout. *Biometrika*, **84**, 33–44.

Molenberghs, G. and Lesaffre, E. (1994) Marginal modelling of correlated ordinal data using an n-way Plackett distribution. *Journal of the American Statistical Association*, **89**, 633–644.

Molenberghs, G. and Lesaffre, E. (1997) Marginal modelling of multivariate categorical data. *Submitted for publication*.

Molenberghs, G., Michiels, B., Kenward, M.G., and Diggle, P.J. (1997) Missing data mechanisms and pattern-mixture models. *to appear in Statistica Neerlandica*.

Murray, G.D. and Findlay, J.G. (1988) Correcting for the bias caused by drop-outs in hypertension trials. *Statistics in Medicine*, **7**, 941–946.

Plackett, R.L. (1965) A class of bivariate distributions. *Journal of the American Statistical Association*, **60**, 516–522.

Rubin, D.B. (1976) Inference and missing data. *Biometrika*, **63**, 581–592.

Rubin, D.B. (1977) Formalizing Subjective Notions About the Effect of Nonrespondents in Sample Surveys. *Journal of the American Statistical Association*, **72**, 538–543.

Rubin, D.B. (1987) *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Rubin, D.B. and Schenker, N. (1986) Multiple imputation for interval estimation from random samples. *Journal of the American Statistical Association*, **81**, 366–374.

Rubin, D.B., Stern, H.S., and Vehovar, V. (1995) Handling "Don't Know" Survey Responses: The Case of the Slovenian Plebiscite. *Journal of the American Statistical Association*, **90**, 822–828.

Smith, D.M., Robertson, B., and Diggle, P.J. (1996) *Object-oriented Software for the Analysis of Longitudinal Data in S*. Technical Report MA 96/192. Department of Mathematics and Statistics, University of Lancaster, LA1 4YF, United Kingdom.

Verbeke, G., Lesaffre, E., Molenberghs, G., Thijs, H., and Kenward, M.G. (1997). Sensitivity analysis for non-random dropout: a local influence approach. *Submitted for publication.*

Verbeke, G. and Molenberghs, G. (1997). *Linear Mixed Models In Practice: A SAS Oriented Approach*, Lecture Notes in Statistics 126. New York: Springer-Verlag.

Welsh, A.H. (1996) *Aspects of Statistical Inference*. New York: Wiley.