31
Sensitivity Analysis

31.1 Introduction

Even though the assumption of likelihood ignorability encompasses both MAR and the more stringent and often implausible MCAR mechanisms, it is difficult to exclude the option of a more general missingness mechanism. One solution is to fit an MNAR model as proposed by Diggle and Kenward (1994) or Molenberghs, Kenward, and Lesaffre (1997). However, as pointed out by several authors (discussion to Diggle and Kenward 1994, Verbeke and Molenberghs 2000, Chapter 18), one has to be extremely careful with interpreting evidence for or against MNAR using only the data under study. A detailed treatment of the issue is provided in Jansen *et al* (2005).

A sensible compromise between blindly shifting to MNAR models or ignoring them altogether is to make them a component of a sensitivity analysis. It is important to consider the effect on key parameters such as treatment effect. In many instances, a sensitivity analysis can strengthen one's confidence in the MAR model (Molenberghs *et al* 2001, Verbeke *et al* 2001).

Broadly speaking, we could define a sensitivity analysis as one in which several statistical models are considered simultaneously and/or where a statistical model is further scrutinized using specialized tools (such as diagnostic measures). This rather loose and very general definition encompasses a wide variety of useful approaches. The simplest procedure is to fit a selected number of (MNAR) models that are all deemed plausible or one in which a preferred (primary) analysis is supplemented with a num-

ber of variations. The extent to which conclusions (inferences) are stable across such ranges provides an indication about the belief one can put into them. Variations to a basic model can be constructed in different ways. The most obvious strategy is to consider various dependencies of the missing data process on the outcomes and/or covariates. Alternatively, the distributional assumptions of the model can be changed. Thus clearly, several routes to sensitivity analysis are possible and in fact the area is fully in expansion.

Sensitivity analysis can be conducted within the selection model family itself. A perspective is given in Section 31.2. Another promising tool, proposed by Verbeke *et al* (2001), and employed by Thijs, Molenberghs, and Verbeke (2000) and Molenberghs *et al* (2001), is based on local influence (Cook 1986). These authors considered the Diggle and Kenward (1994) model, which is based on a selection model, integrating a linear mixed model for continuous outcomes with logistic regression for dropout.

These ideas have been developed for categorical data as well. Van Steen *et al* (2001) developed a local influence based sensitivity analysis for the MNAR Dale model of Section 29.2. It is presented in Section 31.3. Section 31.4 discusses related ideas for the general Baker, Rosenberger, and DerSimonian (1992) model, introduced in Section 29.3. It is based upon work by Jansen *et al* (2003). Hens *et al* (2005) developed kernel weighted influence measures.

Although classical inferential procedures account for the imprecision resulting from the stochastic component of the model and for finite sampling, less attention is devoted to the uncertainty arising from (unplanned) incompleteness in the data, even though the majority of studies in humans suffer from incomplete follow-up. Molenberghs *et al* (2001) acknowledge both the status of imprecision, due to (finite) random sampling, as well as ignorance, due to incompleteness. Both can be combined into uncertainty (Kenward, Goetghebeur, and Molenberghs 2001). An overview is given in Section 31.5.

Another option is to consider pattern-mixture models as a complement to selection models (Thijs *et al* 2002, Michiels *et al* 2002). The analysis conducted in Section 30.6, along the lines outlined in Sections 30.4 and 30.5, can be viewed as a sensitivity analysis of this type. A perspective is given in Section 31.6.

31.2 Sensitivity Analysis for Selection Models

When data are incomplete, the analysis of the actually observed data is subject to further untestable modeling assumptions. The methodologically simplest case is discussed in Section 27.3, where it is assumed that the missing data are MCAR. However, the MCAR assumption is a strong one and made too often in practice.

When more flexible assumptions, such as MAR or even MNAR, are considered, several choices have to be made. For example, one has to choose between selection and pattern-mixture models, or an alternative framework such as shared-parameter models (Section 26.2.1).

Particularly within the selection modeling framework, there has been an increasing literature on MNAR. At the same time, concern has been growing precisely about the fact that models often rest on strong assumptions and relatively little evidence from the data themselves.

In response to these concerns, there is growing awareness of the need for methods that investigate the sensitivity of the results with respect to the model assumptions. See, for example, Nordheim (1984), Little (1994b), Rubin (1994), Laird (1994), Fitzmaurice, Molenberghs, and Lipsitz (1995), Molenberghs *et al* (1999), Kenward (1998), and Kenward and Molenberghs (1999). Many of these are to be considered useful but *ad hoc* approaches. Whereas such informal sensitivity analyses are an indispensable step in the analysis of incomplete longitudinal data, it is desirable to conduct more formal sensitivity analyses.

At any rate, fitting an MNAR model should be subject to careful scrutiny. The modeler needs to pay attention, not only to the assumed distributional form of the model (Little 1994b, Kenward 1998), but also to the impact one or a few influential subjects may have on the dropout and/or measurement model parameters. Because fitting an MNAR model is feasible by virtue of strong assumptions, such models are likely to pick up a wide variety of influences in the parameters describing the nonrandom part of the dropout mechanism. Hence, a good level of caution is in place. This issue has been studied in detail by Jansen *et al* (2005). These authors not only study the behavior of local influence methods in the presence of a variety of deviations from the posited model, not only in terms of the dropout mechanism, they also study the behavior of the likelihood ratio test statistic, used to test MNAR *versus* MAR. Their conclusion is that such a test is surrounded with both philosophical issues, as well as technical problems. There are philosophical issues because two models, similar or even identical in terms of their fit to the observed data, may produce widely varying predictions of the unobserved data. When unrecognized, this is a problem, as such models cannot be distinguished in terms of statistical arguments only. When the scientific question is, at least in part, in terms of the fit to the unobserved outcomes, it is very difficult to distinguish between such models solely in statistical terms. The technical issues occur because the likelihood ratio test statistic for MNAR *versus* MAR, of the type used in the Diggle and Kenward (1994) and MNAR Dale (Section 29.2) models, exhibits non-standard behavior. This should not come as a surprise, as most of the information on the MNAR parameter(s) comes from distributional assumptions, and not from genuine information in the data. Therefore, classical asymptotics should not be taken for granted. This problem is studied by Jansen *et al* (2005); see also Scharfstein, Rotnitzky, and Robins (1999). By using

a semi-parametric framework, it becomes clear that, under MNAR, semi-parametric assumptions, i.e., moment-based assumptions, are not sufficient to identify model parameters.

31.3 A Local Influence Approach for Ordinal Data with Dropout

Incomplete longitudinal ordinal data can be modeled using a simple logistic regression formulation for the dropout process and using a multivariate Dale model for the response (Molenberghs and Lesaffre 1994, 1999, Molenberghs, Kenward, and Lesaffre, 1997), as described in Section 29.2. To explore the sensitivity of this selection model for repeated ordinal outcomes, Van Steen *et al* (2001), considered a local influence approach.

31.3.1 General Principles

Cook (1986) suggests that more confidence can be put in a model that is relatively stable under small modifications. The best known perturbation schemes are based on case-deletion (Cook and Weisberg 1982) in which the effect is studied of completely removing cases from the analysis. A quite different paradigm is the local influence approach where one investigates how the results of an analysis are changed under small perturbations of the model. In the framework of the linear mixed model Beckman, Nachtsheim, and Cook (1987) used local influence to assess the effect of perturbing the error variances, the random-effects variances and the response vector. In the same context, Lesaffre and Verbeke (1998) have shown that the local influence approach is also useful for the detection of influential subjects in a longitudinal data analysis. Verbeke *et al* (2001) and Verbeke and Molenberghs (2000, Chapter 19) use the same idea to explore the sensitivity of a selection model for repeated continuous outcomes. The principal idea is to explore how small perturbations around MAR, in the direction of MNAR, can have a large impact. These authors have shown that various types of influential subjects can cause a model to apparently be of the MNAR type. This implies that caution should be used before concluding that the model really is MNAR, as many types of influential subjects, different from an MNAR mechanism, can force such a conclusion. This view was confirmed by Jansen *et al* (2005).

Consider the following perturbed dropout model:

$$\text{logit}[p_{id}(H_d, k_d; \boldsymbol{\psi})] = H_d\boldsymbol{\psi} + \omega_i k_d. \tag{31.1}$$

which are the components of the ϕ's in (29.3). We choose to use the individual-level index i, rather than the design-level index r, as the perturbations ω_i are defined at the individual level.

When $\omega_i = 0$ for all i, MAR is obtained. Due to ignorability, no influence on the measurement model parameters is then possible. When small perturbations in a specific ω_i lead to relatively large differences in the model parameters, then this suggests that these subjects may have a large impact on the final analysis. Note that the ω_i are not to be seen as fixed or random subject-specific parameters, but rather as (infinitesimal) perturbations, to which differential geometry will be applied, rather than ordinary parameter estimation.

We first give a general introduction of the local influence methodology as introduced by Cook (1986). In Section 31.3.2, it will be applied to the fluvoxamine study.

We denote the log-likelihood function corresponding to the model including perturbed dropout model (31.1) by

$$\ell(\boldsymbol{\gamma}|\boldsymbol{\omega}) = \sum_{i=1}^{N} \ell_i(\boldsymbol{\gamma}|\omega_i),$$

in which $\ell_i(\boldsymbol{\gamma}|\omega_i)$ is the contribution of the ith individual to the log-likelihood, and where $\boldsymbol{\gamma} = (\boldsymbol{\theta}, \boldsymbol{\psi})$ is the s-dimensional vector, grouping the parameters of the measurement model and the dropout model, not including the $N \times 1$ vector $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_N)'$ of weights defining the perturbation of the MAR model. It is assumed that $\boldsymbol{\omega}$ belongs to an open subset Ω of $\mathbb{R}^N$. For $\boldsymbol{\omega}$ equal to $\boldsymbol{\omega_0} = (0, 0, \ldots, 0)'$, $\ell(\boldsymbol{\gamma}|\boldsymbol{\omega_0})$ is the log-likelihood function that corresponds to a MAR dropout model.

Let $\widehat{\boldsymbol{\gamma}}$ be the maximum likelihood estimator for $\boldsymbol{\gamma}$, obtained by maximizing $\ell(\boldsymbol{\gamma}|\boldsymbol{\omega_0})$, and let $\widehat{\boldsymbol{\gamma}}_\omega$ denote the maximum likelihood estimator for $\boldsymbol{\gamma}$ under $\ell(\boldsymbol{\gamma}|\boldsymbol{\omega})$. The local influence approach now compares $\widehat{\boldsymbol{\gamma}}_\omega$ with $\widehat{\boldsymbol{\gamma}}$. Similar estimates indicate that the parameter estimates are robust with respect to perturbations of the MAR model in the direction of MNAR. Very different estimates suggest that the estimation procedure is highly sensitive to such perturbations, which suggests that the choice between a random and a non-random dropout model highly affects the results of the analysis. Cook (1986) proposed to measure the distance between $\widehat{\boldsymbol{\gamma}}_\omega$ and $\widehat{\boldsymbol{\gamma}}$ by the so-called likelihood displacement, defined by $LD(\boldsymbol{\omega}) = 2[\ell(\widehat{\boldsymbol{\gamma}}|\boldsymbol{\omega_0}) - \ell(\widehat{\boldsymbol{\gamma}}_\omega|\boldsymbol{\omega_0})]$. This takes into account the variability of $\widehat{\boldsymbol{\gamma}}$. Indeed, $LD(\boldsymbol{\omega})$ will be large if $\ell(\boldsymbol{\gamma}|\boldsymbol{\omega_0})$ is strongly curved at $\widehat{\boldsymbol{\gamma}}$, which means that $\boldsymbol{\gamma}$ is estimated with high precision, and small otherwise. Therefore, a graph of $LD(\boldsymbol{\omega})$ versus $\boldsymbol{\omega}$ contains essential information on the influence of perturbations. It is useful to view this graph as the geometric surface formed by the values of the $N + 1$ dimensional vector $\boldsymbol{\xi}(\boldsymbol{\omega}) = [\boldsymbol{\omega}', LD(\boldsymbol{\omega})]'$ as $\boldsymbol{\omega}$ varies throughout Ω. Because this so-called influence graph can only be depicted when $N = 2$, Cook (1986) proposed to consider local influence, i.e., at the normal curvatures $C_{\boldsymbol{h}}$ of $\boldsymbol{\xi}(\boldsymbol{\omega})$ in $\boldsymbol{\omega}_0$, in the direction of some N dimensional vector

$\boldsymbol{h}$ of unit length. Let $\boldsymbol{\Delta}_i$ be the s dimensional vector defined by

$$\boldsymbol{\Delta}_i = \left. \frac{\partial^2 \ell_i(\boldsymbol{\gamma}|\omega_i)}{\partial \omega_i \partial \boldsymbol{\gamma}} \right|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}},\omega_i=0}$$

and define Δ as the $(s \times N)$ matrix with $\boldsymbol{\Delta}_i$ as its ith column. Further, let $\ddot{L}$ denote the $(s \times s)$ matrix of second-order derivatives of $\ell(\boldsymbol{\gamma}|\boldsymbol{\omega}_0)$ with respect to $\boldsymbol{\gamma}$, also evaluated at $\boldsymbol{\gamma} = \widehat{\boldsymbol{\gamma}}$. Cook (1986) has then shown that $C_{\boldsymbol{h}}$ can be easily calculated by $C_{\boldsymbol{h}} = 2|\boldsymbol{h}'\Delta'\ddot{L}^{-1}\Delta\boldsymbol{h}|$.

Obviously, $C_{\boldsymbol{h}}$ can be calculated for any direction $\boldsymbol{h}$. One evident choice is the vector $\boldsymbol{h_i}$ containing one in the ith position and zero elsewhere, corresponding to the perturbation of the ith weight only. This reflects the influence of allowing the ith subject to drop out non-randomly, whereas the others can only drop out at random. The corresponding local influence measure, denoted by C_i, then becomes $C_i = 2|\boldsymbol{\Delta}_i'\ddot{L}^{-1}\boldsymbol{\Delta}_i|$. Another important direction is the direction $\boldsymbol{h}_{\max}$ of maximal normal curvature $C_{\max}$. It shows how to perturb the MAR model to obtain the largest local changes in the likelihood displacement. It is readily seen that $C_{\max}$ is the largest eigenvalue of $-2\,\Delta'\,\ddot{L}^{-1}\,\Delta$, and that $\boldsymbol{h}_{\max}$ is the corresponding eigenvector.

When a subset $\boldsymbol{\gamma}_1$ of $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2')'$ is of special interest, a similar approach can be used, replacing the log-likelihood by the profile log-likelihood for $\boldsymbol{\gamma}_1$, and the methods discussed above for the full parameter vector directly carry over. Details can be found in Lesaffre and Verbeke (1998), Verbeke *et al* (2001), and Verbeke and Molenberghs (2000, Chapters 11 and 19).

It will be clear from the previous derivations that calculation of local influence measures merely reduces to evaluation of Δ and $\ddot{L}$. In the linear mixed model case, Verbeke *et al* (2001) and Verbeke and Molenberghs (2000) have proposed closed form expressions, with some emphasis on the case of compound symmetry. For the multivariate Dale model, as will be the case for many other non-normal models, this is algebraically very involved and may not yield the same type of insightful expressions. However, when a program is available to fit the full non-random model (3.11), a particularly convenient computational scheme can be used. Indeed, in this case there are usually tools available to obtain a Hessian matrix evaluated in a point of interest (e.g., through EM-aided differentiation, see also page 537). Note that in our situation, it suffices to compute the second derivatives of the likelihood, for each observation separately, after which the subvector $\boldsymbol{\Delta}_i$ pertaining to the $(\boldsymbol{\gamma}, \omega)$-block can be selected.

In practice, the parameter $\boldsymbol{\theta}$ in the measurement model is often of primary interest. Because $\ddot{L}$ is block-diagonal with blocks $\ddot{L}(\boldsymbol{\theta})$ and $\ddot{L}(\boldsymbol{\psi})$, we have that for any unit vector $\boldsymbol{h}$, $C_{\boldsymbol{h}}$ equals $C_{\boldsymbol{h}}(\boldsymbol{\theta}) + C_{\boldsymbol{h}}(\boldsymbol{\psi})$, with

$$C_{\boldsymbol{h}}(\boldsymbol{\theta}) \;=\; -2\boldsymbol{h}' \left[\left. \frac{\partial^2 \ell_{i\omega}}{\partial \boldsymbol{\theta} \partial \omega_i} \right|_{\omega_i=0} \right]' \ddot{L}^{-1}(\boldsymbol{\theta}) \left[\left. \frac{\partial^2 \ell_{i\omega}}{\partial \boldsymbol{\theta} \partial \omega_i} \right|_{\omega_i=0} \right] \boldsymbol{h}$$

$$C_{\boldsymbol{h}}(\boldsymbol{\psi}) = -2\boldsymbol{h}' \left[\left. \frac{\partial^2 \ell_{i\omega}}{\partial \boldsymbol{\psi} \partial \omega_i} \right|_{\omega_i = 0} \right]' \ddot{L}^{-1}(\boldsymbol{\psi}) \left[\left. \frac{\partial^2 \ell_{i\omega}}{\partial \boldsymbol{\psi} \partial \omega_i} \right|_{\omega_i = 0} \right] \boldsymbol{h},$$

evaluated at $\boldsymbol{\gamma} = \widehat{\boldsymbol{\gamma}}$.

31.3.2 Analysis of the Fluvoxamine Data

Van Steen *et al* (2001) applied the local influence ideas of Section 31.3 to the fluvoxamine study introduced in Section 2.4 and analyzed at various instances.

To investigate the sensitivity of inferences reported in Section 29.2.3 with respect to modeling assumptions for the dropout process, the overall C_i, influences $C_i(\boldsymbol{\theta})$ and $C_i(\boldsymbol{\psi})$ for the measurement parameters and dropout parameters, as well as $\boldsymbol{h}_{\max}$ of maximal curvature are displayed in Figure 31.1. Note that the largest C_i are observed for patients #34 and #252 (both having side effects surpassing the therapeutic effect at visit 1 and visit 2), followed by patients #182, #64, #122, #28, #108, #287, #232, #112, and #245, all of whom yield the worst score on side effects at visit 1 and drop out at visit 2. We pay special attention to patient #239, showing side effects interfering significantly with functionality at visit 1, after which dropout occurs.

In addition, Figure 31.1 shows some evidence of the fact that influence on measurement model parameters can theoretically only arise from those measurement occasions at which dropout occurs, a fact already observed by Verbeke *et al* (2001). Nevertheless, it should be noted that influence on the measurement model parameters can also arise from complete observations. Indeed, when small perturbations in a specific ω_i lead to relatively large differences in the model parameters, the subject's impact on dropout parameters indirectly influences all functions that include these dropout parameters. An example of such a function is the conditional mean of an unobserved measurement, given the observed measurements and given the fact that the patient belongs to a certain dropout pattern. As a consequence, the corresponding measurement model parameters will *indirectly* be affected as well (Verbeke *et al* 2001).

Influential completers occur in the index plots of C_i, $C_i(\boldsymbol{\psi})$, and of the components of the direction $\boldsymbol{h}_{\max}$ of maximal curvature, but are absent in the index plot for $C_i(\boldsymbol{\theta})$. Focusing on $C_i(\boldsymbol{\theta})$, Figure 31.1 reveals the highest peaks for patients #239 and #128. It appears that the influence of allowing subject #239 to drop out non-randomly, is best visible on the measurement model parameters. Patient #128 has an incomplete sequence, with a relatively mild score for side effects (side effects not interfering with functionality). Hence, the relatively large value for $C_i(\boldsymbol{\theta})$ is somewhat unusual, especially because other index plots do not show evidence of any influential effect, not even globally. One could ask the question whether other, unmeasured factors could have caused this phenomenon.

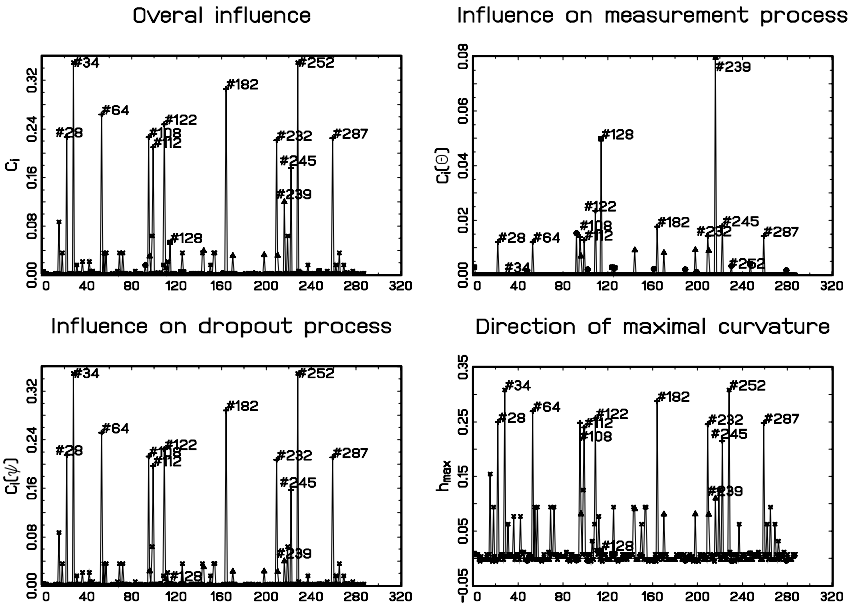

FIGURE 31.1. *Fluvoxamine Trial. Index plots of C_i, $C_i(\boldsymbol{\theta})$, $C_i(\boldsymbol{\psi})$, and of the components of the direction $\boldsymbol{h}_{\max}$ of maximal curvature. The x-axis merely contains sequential indicators. Relevant patient IDs have been added to the plot. Completers (patients with observed responses at visit 1 and visit 2) are indicated with a solid star. A solid circle, a solid square, a solid triangle, or a solid plus is used for subjects whose score on side effects at visit 1 respectively ranges from (1) to (4). Patients with a non-monotone dropout pattern are discarded.*

Before addressing this question, we turn attention to $C_i(\boldsymbol{\psi})$ and $\boldsymbol{h}_{\max}$. To avoid confusion, observe that the scale is different from the one of $C_i(\boldsymbol{\theta})$. The most influential patients appear to be the same as for the overall C_i (#34, #252 and #182, #64, #122, #28, #108, #287, #232, #112, #245). The same patients are also shown in the index plot for $\boldsymbol{h}_{\max}$.

Observe that in all plots, 'layers' of influential cases may be distinguished. The higher the layers, the more they seem to be associated with particular response levels. For instance, in Figure 31.1, patients #34 and #252 give rise to components of $\boldsymbol{h}_{\max}$ that are larger than 0.3. Patients #182, #64, #122, #28, #108, #287, #232, #112 and #245 (corresponding to the influential patients in the previous paragraph) refer to $\boldsymbol{h}_{\max}$ components that are all smaller than 0.3 but larger than 0.2. The layer formation is not clear though, and recalling the particular behavior of patient #128, one is led to believe that another distorting factor is involved, blurring the picture. Therefore, we investigate the effect of covariates on the ability to interpret influence plots.

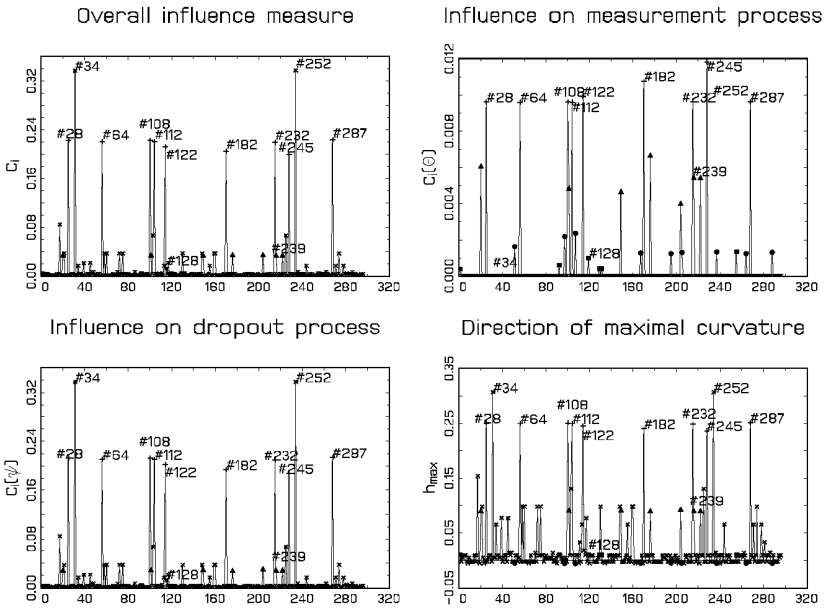

FIGURE 31.2. *Fluvoxamine Trial. Index plots of C_i, $C_i(\boldsymbol{\theta})$, $C_i(\boldsymbol{\psi})$, and of the components of the direction $\boldsymbol{h}_{\max}$ of maximal curvature, where 'age' is considered as the sole covariate in the Dale model. The x-axis contains sequential indicators. Completers are indicated with a solid star. A solid circle, a solid square, a solid triangle, or a solid plus is used for subjects whose score at visit 1 on side effects respectively ranges from (1) to (4).*

To this end, we consider two additional models. The first one includes 'sex' as the only covariate in the measurement model, the second one uses 'age' as the only covariate. These models perform worse than the model including both 'age' and 'sex,' augmented with 'duration' and 'severity,' but they are merely intended for illustrative purposes. The resulting influence plots are enlightening. Figure 31.2 shows the index plots when 'age' is included as only covariate, Figure 31.3 displays the corresponding pictures in case 'sex' is the only source of covariate information. In both cases, much smaller values are obtained for $C_i(\boldsymbol{\theta})$. The high peaks for patients #239 and #128 have disappeared. Patients #122, #245, and #182 also show up in Figure 31.2 with the highest peaks for $C_i(\boldsymbol{\theta})$, although hard to distinguish from the peaks for patients #287, #232, #28, #108, #64, and #112. The variability observed in $C_i(\boldsymbol{\theta})$ values also appears in Figure 31.3. However, in this case, it seems to be caused by the fact that patients #108, #182, #287, and #232 have $C_i(\boldsymbol{\theta})$ equal to about 0.0116 compared to ap-

proximately 0.0097 for patients #28, #245, #64, #122, and #112. This layer effect may be explained by the binary character of 'sex' as opposed to 'age,' the latter of which entered the model as a continuous variable. Also note that patients #108, #182, #287, and #232 are all male, whereas patients #28, #245, #64, #122, and #112 are all female. All these patients drop out at visit 2 and showed side effects surpassing therapeutic effect at visit 1. In Figures 31.2 and 31.3, the same patient group (i.e., patients #34, #252, #287, #108, #28, #112, #64, #232, #122, #182, and #245) is distinguished as globally influential, with highest C_i values for #34 and #252. The layering effect is again the most explicit when 'sex' is considered as only covariate (Figure 31.3). Influential patients for $C_i(\boldsymbol{\psi})$ and $\boldsymbol{h}_{\max}$ appear to be the same as before, where 'sex' and 'age' were both considered in the pool of covariates, with the exception of subject #239 whose corresponding component in $\boldsymbol{h}_{\max}$ is now less than 0.1000. The distribution over potential values becomes more discrete when 'age' is considered to be the only covariate in the multivariate Dale model. Changing 'age' for 'sex' causes the distribution to be even more discrete and therefore the layer effect more explicit.

In an attempt to improve insight into the driving forces present in the set of data, which may explain possible deviations from a random dropout process, we exclude patients #34 and #252 from the data set and apply the same measurement model as in the beginning of Section 5 (thus including the covariates 'age,' 'sex,' 'duration,' and 'severity'). Provided MAR is the correct alternative hypothesis and provided the parametric form for the MAR process is correct (again, no covariates were included), there seems to be even less evidence for MAR; the likelihood ratio test statistic comparing MCAR with MAR equals G^2=0.94, based on 1 degree of freedom ($p = 0.333$). Note that now borderline evidence for MNAR is observed, since a comparison between the non-random and random dropout model generates a likelihood ratio test statistic of G^2= 3.74 with 1 degree of freedom ($p = 0.053$). Hence, the suggested local influence approach bridges the gap between the random and the non-random model: some of the mechanisms that cannot be explained by the random model and are captured by the non-random model, the latter resting on untestable assumptions, can be attributed to the observations for patients #34 and #252.

Repeating the previous analysis on a reduced data set, where patient #239 is excluded instead of patients #34 and #252, we find no evidence for MAR against MCAR ($G^2 = 0.01$, $p = 0.913$). After investigating the likelihood ratio test statistic for comparing the non-random with the random dropout model ($G^2 = 2.13$, $p = 0.145$), we may conclude that the MCAR assumption is fairly plausible. It is not surprising that conclusions remain similar. Indeed, although patient #239 appeared to be most influential patient with respect to the measurement model parameters, it should be noted that (i) the value for $C_i(\boldsymbol{\theta})$ is "only" 0.079 (further investigation is required to define some critical value above which $C_i(\boldsymbol{\theta})$ can be said to be

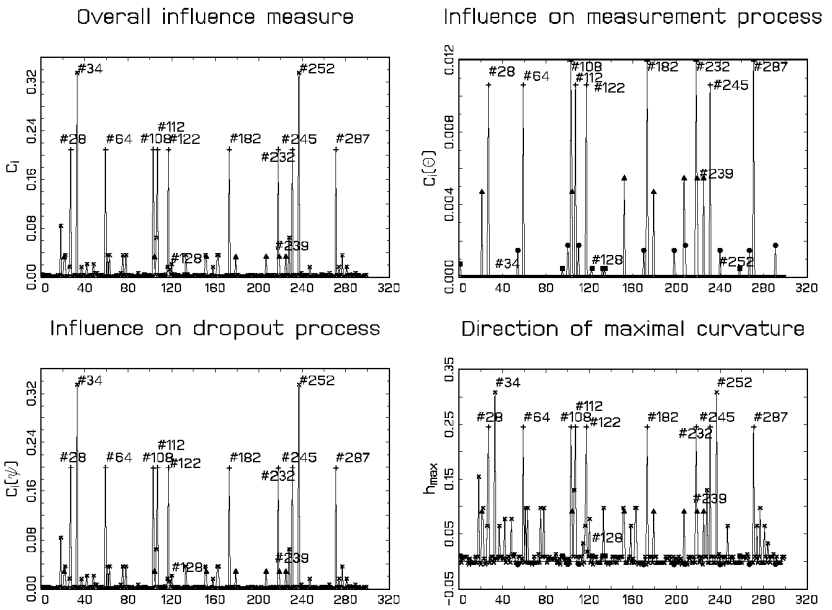

FIGURE 31.3. *Fluvoxamine Trial. Index plots of C_i, $C_i(\boldsymbol{\theta})$, $C_i(\boldsymbol{\psi})$, and of the components of the direction $\boldsymbol{h}_{\max}$ of maximal curvature, where 'sex' is considered as only covariate in the Dale model. The x-axis contains sequential indicators. Completers are indicated with a solid star. A solid circle, a solid square, a solid triangle, or a solid plus is used for non-completers whose score at visit 1 on side effects respectively ranges from (1) to (4).*

statistically significantly large) and that (ii) patient #239 did not appear to be influential overall.

31.4 A Local Influence Approach for Incomplete Binary Data

31.4.1 General Principles

For multivariate and longitudinal binary data, subject to non-monotone missingness, one can focus on the model proposed by Baker, Rosenberger, and DerSimonian (1992). They considered a log-linear type of model for two possibly binary outcomes, subject to non-monotone missingness. Jansen *et al* (2003) reformulated the model such that its membership of the selection model family is unambiguously clear. Next, they extended the original

model to accommodate for, possibly continuous, covariates, turning the model into a regression tool for several categorical outcomes. Further, a parameterization was proposed that avoids the risk of invalid solutions. In other words, all combinations of the natural parameters produce probabilities between 0 and 1. The model is introduced in Section 29.3.

As a consequence of these extensions, the closed-form solutions of Baker, Rosenberger, and DerSimonian (1992) no longer apply. Of course, given the focus on continuous covariates, the derivation of closed-form solutions should not be of primary concern. Finally, Jansen *et al* (2003) coupled a local influence approach with the model strategy, to assess which observations have a strong impact on the comparison of two nested models within the BRD family.

Jansen *et al* (2005) consider perturbations of a given BRD model in the direction of a model with one more parameter in which the original model is nested, implying that perturbations lie along the edges of Figure 29.2: for each of the nested pairs in Figure 29.2, the simpler of the two models equates two parameters from the more complex one. For example, BRD4 includes $\beta_{.k}$, $(k = 1, 2)$, whereas in BRD1 only $\beta_{..}$ is included. For the influence analysis, ω_i is then included as a contrast between two such parameters; for the perturbation of BRD1 in the direction of BRD4, one considers $\beta_{..}$ and $\beta_{..} + \omega_i$. Such an ω_i is not a subject-specific parameter, but rather an infinitesimal perturbation. The vector of all ω_i's defines the direction in which such a perturbation is considered. Clearly, other perturbation schemes are possible as well, or one could consider a different route of sensitivity analysis altogether. Ideally, several could be considered within an integrated sensitivity analysis. Note that our influence analysis focuses on the missingness model, rather than on the measurement model parameters. This may be seen as slightly odd, because often, scientific interest focuses on the measurement model parameters. However, it has been documented (discussion to Diggle and Kenward 1994, Kenward 1998, Verbeke *et al* 2001) that the missingness model parameters are often the most sensitive ones to take up all kinds of misspecification and influential features. These may then, in turn, impact conclusions coming from the measurement model parameters (e.g., time evolution) or combinations from both (e.g., covariate effects for certain groups of responders).

31.4.2 Analysis of the Fluvoxamine Data

We will now apply the local influence ideas, outlined in the previous section, to the BRD models in order to contradict or strengthen the conclusions of Section 29.3.1. Whereas all comparisons along the edges of Figure 29.2 are possible, we propose to primarily focus on the comparison of BRD1 with BRD4 (Figure 31.4), as the first one was the most adequate model when no duration effect is included and when duration is included in both parts of the model, whereas the second one was the model of choice when duration

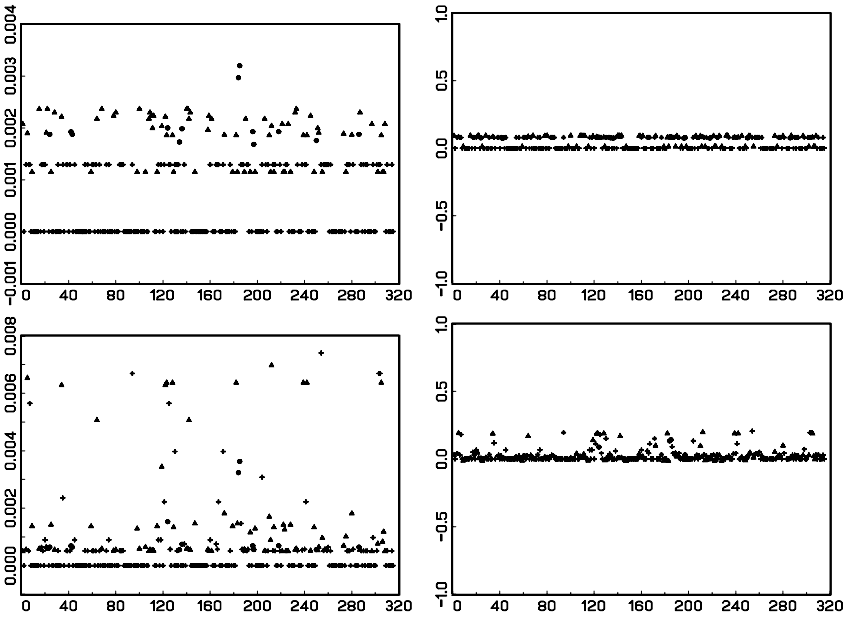

FIGURE 31.4. *Fluvoxamine Trial. Index plots of C_i (left panels) and of the components of the direction $\boldsymbol{h}_{\max}$ of maximal curvature (right panels) for comparison BRD1–4, without (top panels) or with (bottom panels) duration as a covariate in the missingness models.*

is included in the measurement model only. In addition, we will consider the comparisons BRD4–7 (Figure 31.5) and BRD4–8 (plot not shown), the supermodels of BRD4. The symbols used in these figures are the following: +: both observations are available, (1,1) type; black triangle: only the first observation is available, (1,0) type; black square: only the second observation is available, (0,1) type; •: both measurements are missing, (0,0) type.

The overall C_i are considered, as well as the components of the direction of maximal curvature $\boldsymbol{h}_{\max}$. The top right panel in Figure 31.4 shows essentially no structure, whereas in the top left there are two important observations. First, a clear layering effect is present, consistent with the analysis in Section 31.3.2. Again, this is not surprising, as there are quite a number of discrete features to the model: the responses and the missingness patterns. On the other hand, the continuous covariate duration is included in the measurement model. In this case, mainly the missingness patterns are noticeable, although the top layer shows a good deal of variability. These layers are reminiscent of a pattern-mixture structure (Little 1995) even though the model is of a selection nature.

Two views can be taken. Either, focus can be on two observations, #184 and #185, that stand out. These subjects have no measurements at all for

TABLE 31.1. *Fluvoxamine Trial. Negative log-likelihood values for three additional sets of analysis. I: #184 and #185 removed, no covariates; II: #184 and #185 removed, duration as covariate in the measurement model; III: all observations in the (0,0) group removed, duration as covariate in the measurement model.*

Set	BRD1	BRD2	BRD3	BRD4	BRD5
I	559.59	558.18	558.70	558.18	558.97
II	543.65	541.87	542.16	540.35	542.43
III	496.19	494.33	495.26	492.53	495.53
Set	BRD6	BRD7	BRD8	BRD9	
I	557.59	557.32	557.59	557.32	
II	540.61	538.53	538.81	540.34	
III	493.71	491.67	491.95	493.43	

side effects. Alternatively, the entire pattern without follow up measurements can be given further consideration. We will return to this issue later in this section. This phenomenon is in contrast to the analyses made by Verbeke *et al* (2001) and Molenberghs *et al* (2001) who found that the influential observations are invariably completers. In this case, the situation is different since the "empty" observations are explicitly modeled in the BRD models. Therefore, assumptions about the perturbations in the direction of such observations have an impact on the values such an individual *would have had* had the measurements been made; hence a strong sensitivity. This is an illustration of the fact that studying influence by means of perturbations in the missingness model may lead to important conclusions regarding the measurement model parameters. Indeed, the measurement model conclusions depend, not only on the observations actually made, but also on the expectation of the missing measurements. In an MNAR model, such expectations depend on the missingness model as well, since they are made *conditional on an observation being missing*. A high level of sensitivity means that the expectations of the missing outcomes and the resulting measurement model parameters strongly depend on the missingness model (Verbeke *et al* 2001). As stated earlier, the only continuous characteristics of the observations are the levels for duration. These are 38 and 41, respectively, the largest values within the group without observations and the 91st and 92nd percentile values within the entire sample. Thus, the conclusions are driven by a very high value of duration.

Let us now turn to the bottom panels of Figure 31.4. The right hand panel still shows little or no structure. On the left hand side, the layering has been blurred due to the occurrence of duration as a continuous feature into the missingness model. The fact that no sets of observations stand

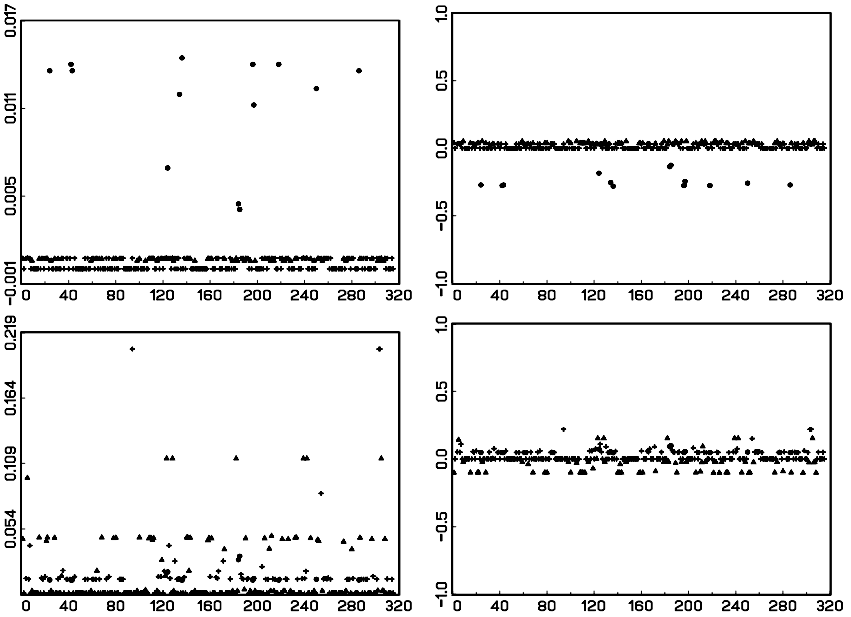

FIGURE 31.5. *Fluvoxamine Trial. Index plots of C_i (left panels) and of the components of the direction $\boldsymbol{h}_{\max}$ of maximal curvature (right panels) for comparison BRD4–7, without (top panels) or with (bottom panels) duration as a covariate in the missingness models.*

out as such, confirms the impression that a good fit has been obtained by including duration in both parts of the model.

Let us now turn to Figure 31.5. A qualitative difference with Figure 31.4 (top left panels) is that now the entire group with no follow-up measurements shows more influential than all other subjects. In this case, $\boldsymbol{h}_{\max}$ displays the same group of subjects with no follow-up. However, all of this disappears when one turns to the bottom panels, again underscoring the importance of duration in the missingness model.

The consequence of these findings is that, as soon as duration is included in the missingness model, a reasonable amount of confidence can be put into the conclusions so obtained. Nevertheless, based on the comparison BRD1–4, it seems wise to further study the effect of subjects #184 and #185, as well as from the group without follow up measurements. To this effect, three additional analyses are considered (Table 31.1): two sets pertain to removal of subjects #184 and #185: without (I) and with (II) duration as a covariate in the measurement model. Note that we do not consider removal in case duration is included in the missingness model because, in this case, these two subjects did not show up as locally influential. Finally, removing all subjects without follow-up measurements and using duration as covariate in the measurement model is reported as family III.

In analysis I, BRD1 is still the preferred model; in II, evidence still points towards BRD4, although slightly less extreme than before: likelihood ratio test statistics for BRD1–4, BRD4–7, and BRD4–8 are 6.60, 3.64, and 3.08, respectively, compared to 7.10, 2.10, and 1.52 obtained initially. However, while the two subjects deleted in I and II cannot explain the apparent non-random missingness, the same conclusions are reached when all subject in pattern (0,0) are deleted (analysis III), as then a few likelihood ratios are above the significance threshold (7.17, attained for BRD3–7 and for BRD5–8; and 7.32 for BRD1–4). Thus, removing these subjects does not change the conclusions about the non-random nature of the data. This is useful supplemental information. Indeed, it is confirmed that the largest impact on the conclusion regarding the nature of the missingness mechanism, is coming from the inclusion of the covariate duration, and neither from isolated individuals, nor from a specific missingness pattern (those without measurements). A nice side effect of this conclusion is that the selected analysis encompasses all subjects and therefore avoids the need of subject deletion, which, if at all possible, should be avoided in statistical analysis.

These analyses can be seen as a useful component of a sensitivity analysis. Given the intrinsic problems with incomplete data models, one can never be completely sure the nature of the missingness mechanism is as posited in the model of choice and therefore several sensitivity assessments simultaneously and/or substantive knowledge have to be considered. When a number of possible causes for the observed non-randomness are found, one might ideally add substantive arguments as to their relative plausibility.

Subjects in an influence graph are displayed without a particular order. Several alternatives are possible, each with pros and cons. For example, one could order the subjects by covariate level, but this method cannot be considered when there are several covariates. Alternatively, the subjects could be ordered by C_i or h_i level, but then different orderings would exist on different plots.

31.5 Interval of Ignorance

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. This is usually done by means of precision or interval estimates. Less attention is devoted to the uncertainty arising from (unplanned) incompleteness in the data. Through the choice of an identifiable model for MNAR missingness, one narrows the possible data generating mechanisms to the point where inference only suffers from imprecision. Some proposals have been made for assessing the sensitivity of these model assumptions; many are based on fitting several plausible but competing models.

Molenberghs, Kenward, and Goetghebeur (2001) and Kenward, Goetghebeur, and Molenberghs (2001) showed an approach that identifies and incorporates both sources of uncertainty in inference: imprecision due to finite sampling, and ignorance due to incompleteness. A simple sensitivity analysis considers a finite set of plausible models. This idea can be taken one step further, by considering more degrees of freedom than the data support. This produces sets of estimates, termed *region of ignorance*, and sets of confidence region, combined into so-called *regions of uncertainty*.

We focus on the model proposed by Baker, Rosenberger, and DerSimonian (1992) and used before in Sections 29.3 and 31.4. Two of the main advantages of this family are ease of computation in general, and the existence of a closed-form solution for several of its members, at least in the initial formulation. Molenberghs, Kenward, and Goetghebeur (2001) used this family of models in a reanalysis of the Slovenian plebiscite data of Rubin, Stern, and Vehovar (1995).

31.5.1 General Principle

It is useful to distinguish between two types of statistical uncertainty. The first, statistical imprecision, is due to finite sampling. The second source of uncertainty, due to incompleteness, will be called statistical ignorance. Statistical imprecision is classically quantified by means of estimators (standard error and variance, confidence regions, etc.) and properties of estimators (consistency, asymptotic distribution, efficiency, etc.). To quantify statistical ignorance, it is useful to distinguish between complete and observed data.

For the BRD model, the 16 complete-data degrees of freedom and the 9 observed-data degrees of freedom are represented in Table 29.1. A sample from this table produces empirical proportions representing the π's with error. This imprecision disappears as the sample size tends to ∞. What remains is ignorance regarding the redistribution of all except the first four πs over the missing outcome value. This leaves ignorance regarding any probability in which at least one of the first or second indices is equal to 0 and hence regarding any derived parameter of scientific interest. For such a parameter, θ say, a region of possible values that is consistent with Table 29.1 is called a *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 θ consistent with the observed data provides an estimated region of ignorance. The $100(1-\alpha)\%$ *region of uncertainty* is a larger region in the spirit of a confidence region, designed to capture the combined effects of imprecision and ignorance. Various ways of constructing regions of ignorance and regions of uncertainty are conceivable.

In standard statistical practice, ignorance is hidden in the consideration of a single identified model, such as models BRD1–BRD9. Among those,

BRD6–BRD9 are said to saturate the degrees of freedom. To be precise, they saturate the *observed data* degrees of freedom. A model that would saturate the *complete data* degrees of freedom, would need 15 rather than 8 parameters. From a (classical) observed data perspective, such a model would be overspecified, as would be any model with 9 or more parameters. Note that it is possible to construct an overspecified model with degrees of freedom less than those in an identifiable saturated model at the observed level.

We construct three such overspecified models. Write the missingness part of the model as (29.8). We will consider two models (Models 10 and 11) with a single sensitivity parameter, while Model 12 will include two sensitivity parameters. Model 10 is defined as (α_k, β_{jk}) with

$$\beta_{jk} = \beta_0 + \beta_j + \beta_k, \tag{31.2}$$

an additive decomposition for missingness on the independence question.

Similarly, Model 11, (α_{jk}, β_j), uses

$$\alpha_{jk} = \alpha_0 + \alpha_j + \alpha_k, \tag{31.3}$$

an additive decomposition of the missingness parameter on the attendance question.

Finally, we define Model 12, $(\alpha_{jk}, \beta_{jk})$, as a combination of both (31.2) and (31.3).

We will now outline the general principle behind considering such overspecified models and then focus on the sensitivity parameter approach.

We start from the classical approach of fitting a single identifiable model M_0 to incomplete data (e.g., a particular BRD model). Maximum likelihood estimation produces a parameter estimate $\widehat{\boldsymbol{\pi}}$ along with measures of imprecision (estimated standard errors). From $\widehat{\boldsymbol{\pi}}$ four predicted contingency tables can be derived as in Table 29.1.

The fitted complete tables collapse back to fitted values for the incomplete Table 29.1. 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, as we want to focus on ignorance, we will assume the fit is acceptable. In what follows, models with poor fit (or boundary solutions) will be dropped.

One can now range through *all possible* complete tables, which collapse back to the M_0 predicted incomplete table. One could call such tables 'M_0 compatible' and we denote the set by $\mathcal{S}(M_0)$. The general principle is that to each table in $\mathcal{S}(M_0)$ an extended model M^* will be fitted. This implies that each table produces an estimated parameter vector and a confidence region. The union of those are termed *region of ignorance* and *region of uncertainty*, respectively. For scalar parameters the terms interval of ignorance (II) and interval of uncertainty (IU) will be used.

Apart from explicitly constructing the (real-valued) set of complete tables, one can proceed in an alternative way. This is done by fitting the model M^* directly to the observed data. This implies that the general principle translates to fitting an overspecified model to the observed data, which will produce a *range* of parameters maximizing the observed data likelihood. This range is then the region of ignorance. If this route is followed, there are technically several ways to find the region. One method is described next.

31.5.2 Sensitivity Parameter Approach

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. Such a technique has been proposed for specific examples by Nordheim (1984) and Vach and Blettner (1995). Foster and Smith (1998) expand on this idea and by referring to Baker and Laird (1988) and to Rubin, Stern, and Vehovar (1995), they suggest imposing a prior distribution on a range. Each value of $\boldsymbol{\eta}$ will produce an estimate $\hat{\boldsymbol{\mu}}(\boldsymbol{\eta})$. The union of these yields the region of ignorance. It is important to realize that in general there will not be a unique choice for $\boldsymbol{\eta}$ and hence for $\boldsymbol{\mu}$. Changing the partitioning will produce the same region for $\boldsymbol{\theta} = (\boldsymbol{\eta}', \boldsymbol{\mu}')'$. Models 10 and 11 have a single sensitivity parameter. We chose $\eta = \beta_k$ and $\eta = \alpha_k$ from (31.2) and (31.3), respectively. In Model 12, both these parameters $\boldsymbol{\eta} = (\beta_k, \alpha_k)'$ are treated as sensitivity parameters. In practice, an easy computation scheme is to consider a grid in the sensitivity parameter space, at each value of which the estimable parameter is maximized.

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. Because 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 (White and Goetghebeur 1998). 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 in 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 that are conditional on a particular value of the sensitivity parameter and hence will typically vary with the choice made.

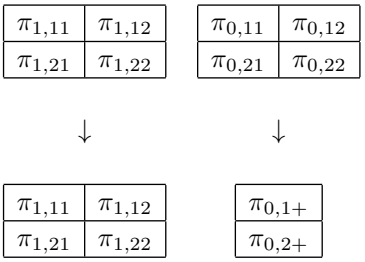

FIGURE 31.6. *Monotone Patterns. Theoretical distribution over complete and observed cells. (Monotone patterns).*

31.5.3 Models for Monotone Patterns and a Bernoulli Experiment

To further illustrate the II ideas, let us focus on the relatively simple setting of two binary outcomes, of which the first one is always observed but the second one is missing for some subjects. This setting is depicted in Figure 31.6. Decompose the cell probabilities as

$$\pi_{r,ij} = p_{ij} q_{r|ij}, \tag{31.4}$$

where p_{ij} parameterizes the measurement process and $q_{r|ij}$ describes the non-response (or dropout) mechanism. In what follows, we will leave p_{ij} unconstrained and consider various forms for $q_{r|ij}$, as listed in Table 31.2. In this setting, there are 7 complete-data degrees of freedom, d.f.(comp)= 7 and 5 observed-data degrees of freedom, d.f.(obs)= 5.

Model M_{sat} (Model 5) has 3 measurement parameters and 4 dropout parameters and saturates d.f.(comp). However, there are only 5 observed degrees of freedom, rendering this model overspecified when fitted to the observed data.

Three models are identified. Conventional restrictions result from assuming an MCAR or MAR model (Models 1 and 2, respectively). Another identified model lets dropout depend on the potentially unobserved second measurement, but not on the first one (Michiels and Molenberghs 1997). Brown (1990) who proposed this estimator for normally distributed endpoints, used the term *protective* estimator because it can be fitted without explicitly addressing the missingness model. We refer to it in Table 31.2 as MNAR 0. Models 2 and 3 both saturate d.f.(obs), and hence cannot be distinguished from each other purely on statistical grounds, in terms of the observed data. In Model 4, dropout is allowed to depend on both measurements but not on their interaction. As a consequence, it overspecifies d.f.(obs) and underspecifies d.f.(comp).

Before turning to setting (31.6), let us illustrate the ideas outlined in Section 31.5.1 by means of the simple setting of a Bernoulli experiment with N trials, where r denotes the number of observed successes, $n - r$

TABLE 31.2. *Monotone Patterns. Dropout models corresponding to the setting of Figure 31.6.*

Dropout models				
Model	$q_{r\|ij}$	Par.	Obs. d.f.	Comp. d.f.
1. MCAR	q_r	4	Non-sat.	Non-sat.
2. MAR	$q_{r\|i}$	5	Sat.	Non-sat.
3. MNAR 0	$q_{r\|j}$	5	Sat.	Non-sat.
4. MNAR I	$\text{logit}(q_{r\|ij}) = \alpha + \beta_i + \gamma_j$	6	Oversp.	Non-sat.
5. M_{sat}	$\text{logit}(q_{r\|ij}) = \alpha + \beta_i + \gamma_j + \delta_{ij}$	7	Oversp.	Sat.

the number of observed failures, and $N - n$ the number of unclassified subjects. Independent of the parameterization chosen, the observed data log-likelihood can be represented in the form

$$\ell = r \ln \alpha + (n - r) \ln \beta + (N - n) \ln(1 - \alpha - \beta), \tag{31.5}$$

with α the probability of an observed success and β the probability of an observed failure. It is sometimes useful to denote $\gamma = 1 - \alpha - \beta$. We consider two models, of which the parameterization is given in Table 31.3. The first one is identified, the second one is overparameterized. Here, p is the probability of a success (whether observed or not), q_1 (q_2) is the probability of being observed given a success (failure), and λ is the odds for being observed for failures versus successes. For Model I, the latter is assumed to be unity. Denote the corresponding log-likelihoods by ℓ_I and ℓ_{II} respectively. In both cases,

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

Maximum likelihood estimates for p and q follow immediately under Model I, either by observing that the moments (α, β) map 1–1 onto the pair (p, q) or by directly solving ℓ_I. The solutions are given in Table 31.3. 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 favor of the current model.

A similar standard derivation is not possible for Model II, as the triplet (p, q_1, q_2) or, equivalently, the triplet (p, q, λ), 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. Because Model I is a submodel of Model II and saturates the observed data, so must every solution to ℓ_{II}, implying the relationships:

$$pq_1 = \frac{r}{N}, \qquad (1 - p)q_2 = \frac{n - r}{N}. \tag{31.6}$$

TABLE 31.3. *Bernoulli Experiment. Two transformations of the observed-data likelihood.*

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

Constraints (31.6) imply

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

and hence

$$\frac{r}{q_1} + \frac{n-r}{q_2} = N. \tag{31.7}$$

The requirement that $q_1, q_2 \leq 1$ in (31.6) implies a range for p:

$$p \in \left[\frac{r}{N}, \frac{N-n+r}{N}\right]. \tag{31.8}$$

Such overspecification of the likelihood can be managed in a more general way using the method outlined in Section 31.5.2. It is not always the case that the range for $\boldsymbol{\eta}$ will be an entire line or real space and hence specific measures may be needed to ensure that $\boldsymbol{\eta}$ is within its allowable range. As the choice of sensitivity parameter is non-unique, a proper choice can greatly simplify the treatment. It will be seen in what follows that the choice of λ as in Table 31.3 is an efficient one from a computational point of view. In contrast, the choice $\theta = q_2 - q_1$ would lead to cumbersome computations and will not be pursued. Of course, what is understood by a proper choice will depend on the context.

TABLE 31.4. *Bernoulli Experiment. Limiting cases for the sensitivity parameter analysis.*

Estimator	λ	$\lambda = \frac{n-r}{N-r}$	$\lambda = 1$	$\lambda = \frac{N-(n-r)}{r}$
p_λ	$\frac{\lambda r}{n-r(1-\lambda)}$	$\frac{r}{N}$	$\frac{r}{n}$	$\frac{N-n+r}{N}$
q_λ	$\frac{n-r(1-\lambda)}{N\lambda}$	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
$\frac{p_\lambda}{1-p_\lambda}$	$\lambda\frac{r}{n-r}$	$\frac{r}{N-r}$	$\frac{r}{n-r}$	$\frac{N-(n-r)}{n-r}$

For example, the sensitivity parameter can be chosen from the nuisance parameters, rather than from the parameters of direct scientific interest. Whether the latter parameters can overlap with the sensitivity set or not is itself an issue (White and Goetghebeur 1998). 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 in the sensitivity set. There will be no direct estimate of imprecision available for the sensitivity parameter. Alternatively, if, given a certain choice of sensitivity parameter, the resulting profile likelihood has a simple form (analogous to the Box-Cox transformation, where conditioning on the transformation parameter produces essentially a normal likelihood), then such a parameter is an obvious candidate.

Given our choice of sensitivity parameter λ, simple algebra yields estimates for p and q (subscripted by λ to indicate dependence on the sensitivity parameter):

$$p_\lambda = \frac{\widehat{\alpha}\lambda}{\widehat{\beta}+\widehat{\alpha}\lambda} = \frac{\lambda r}{n-r(1-\lambda)}, \tag{31.9}$$

$$q_\lambda = \frac{\widehat{\beta}+\widehat{\alpha}\lambda}{\lambda} = \frac{n-r(1-\lambda)}{N\lambda}. \tag{31.10}$$

Using the delta method, an asymptotic variance-covariance matrix of p_λ and q_λ is seen to be built from:

$$\widehat{\mathrm{Var}}(p_\lambda) = \frac{p_\lambda(1-p_\lambda)}{N\lambda q_\lambda} \times\left\{1+\frac{1-\lambda}{\lambda}(1-p_\lambda)[1-p_\lambda q_\lambda(1-\lambda)]\right\}, \tag{31.11}$$

$$\widehat{\mathrm{Cov}}(p_\lambda, q_\lambda) = -\frac{1}{N}p_\lambda(1-p_\lambda)\frac{1-\lambda}{\lambda}q_\lambda, \tag{31.12}$$

TABLE 31.5. *Fluvoxamine Trial. The first subtable represents the complete observations. Subjects with only the first outcome, only the last outcome, or no outcome at all reported are presented in subtables 2, 3, and 4, respectively.*

Side effects:

	time 2				
time 1	yes	no			
yes	89	13	26	2 0	14
no	57	65	49		

Therapeutic effect:

	time 2				
time 1	no	yes			
no	11	1	7	0 2	14
yes	124	88	68		

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

Note that the parameter estimates are asymptotically correlated, except when $\lambda = 1$, i.e., under the 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 λ by requiring $0 \leq p_\lambda, q_\lambda, \lambda q_\lambda \leq 1$. These six inequalities reduce to

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

Table 31.4 presents estimates for limiting cases. The interval of ignorance for the success probability is thus seen to be as in (31.8). It is interesting to observe that the success odds estimator is linear in the sensitivity parameter; the resulting interval of ignorance equals

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

For the success probability, the variance of p_λ is given by (31.11). For the success odds, we obtain:

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

TABLE 31.6. *Fluvoxamine Trial. Identifiable models, fitted to monotone patterns.*

	(1,1)		(1,0)		-2 logl
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 (MNAR 0)	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 (MNAR 0, Unconstr.)	124.0	88.0	80.5	-12.5	
Therapeutic effect	11.6	1.0	6.4	0.0	385.8
Model 3 (MNAR 0, Constr.)	123.4	88.0	68.5	0.0	

and for the success logit:

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

For each λ, a confidence interval C_λ can be constructed for every point within the allowable range of λ. The union of the C_λ is the *interval of uncertainty*, for either p, its odds, or its logit.

31.5.4 Analysis of the Fluvoxamine Data

We focus on the setting of Table 29.1. A version for the fluvoxamine study, based on the first and last follow-up measurements, is given in Table 31.5. There are two patients with a non-monotone pattern of follow-up, whereas 14 subjects have no follow-up data at all. This enables us to treat these data both from the monotone non-response or dropout perspective, as well as from the more complicated but more general non-monotone point of view. We will first the identifiable models and then switch to sensitivity analysis.

31.5.4.1 Identified Models

We first consider the monotone patterns and the corresponding models of Table 31.2. Table 31.6 shows the predicted complete tables for Models 1, 2, and 3. The effect of ignorance is clearly seen by comparing the MAR and

TABLE 31.7. *Fluvoxamine Trial. Marginal probabilities and (log) odds ratio for monotone patterns of side-effects data. Models 1–3: point estimate and 95% confidence interval; Models 4–5: interval of ignorance (II) and interval of uncertainty (IU); these models are defined in Section 31.5.3.*

Parameter		Model 1/2	Model 3	Model 4	Model 5
First Marg.	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 Marg.	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 O.R.	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]
O.R.	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]

protective models: they provide a substantially different prediction for the partially observed table, while producing the same deviance. In addition, the protective model produces a boundary solution, or even an invalid solution if predicted proportions are not constrained to lie within the unit interval, for therapeutic effect.

We now interpret these results in terms of possible quantities of interest, for instance the first and second marginal probability of side effects and the odds ratio, capturing the association between both measurements (Table 31.7). Models 4 and 5 will be discussed in the sensitivity analysis. Models 1 and 2 are both ignorable and hence all measurement model quantities are independent of the choice between MAR and MCAR.

The quantities in Tables 31.6 and 31.7 differ in one important way. The former quantities are calculated conditional on the dropout pattern; the latter follow directly from the marginal measurement probabilities p_{ij}, which are common to all three models while the dropout probabilities $q_{r|ij}$ depend on the model. As a consequence, while MAR and MCAR are equivalent for the quantities in Table 31.7, this does not carry over to the predicted cell counts in Table 31.6. Further, the stability of the estimates in Table 31.7 (at least for Models 1–3) is in marked contrast to the variation among the predicted cell counts in Table 31.6. These considerations suggest that stability may be restricted to *certain* functions of parameters in *certain* sets of data.

We now introduce the non-monotone patterns into the analysis and fit the nine identifiable BRD models of Table 29.2. The fitted counts of the models with an interior solution are given in Table 31.8 and the marginal quantities of interest are displayed in Table 31.9. Note that a subgroup of

TABLE 31.8. *Fluvoxamine Trial. 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

models produces invalid solutions without appropriate constraints, such as automatically imposed by form (29.8) for the dropout model.

In spite of the fact that we are now looking at a larger class of models, the results are comparable with those obtained for the monotone patterns. Table 31.9 reveals that Models BRD1–9 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. We will revisit this conclusion following a more formal sensitivity analysis and deduce that it is strongly misleading.

31.5.4.2 Intervals of Ignorance

Turning to the overspecified models, let us consider the monotone patterns first. In addition to the three identifiable models from Table 31.2, we now fit overspecified Models 4 and 5 to the same data. Results for these additional models are also given in Table 31.7.

For Model 4, there is one sensitivity parameter, which we choose to be γ (measuring the extent of non-randomness). When $\gamma = 0$ the MAR Model 2 is recovered. The value of γ which corresponds to $q_{r|ij} = q_{r|j}$ in Table 31.2 yields the protective Model 3. Because there is only one sensitivity parameter, a graphical representation (Figure 31.7) is straightforward. Because 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 this quantity is still a point. This is not true for the other two quantities.

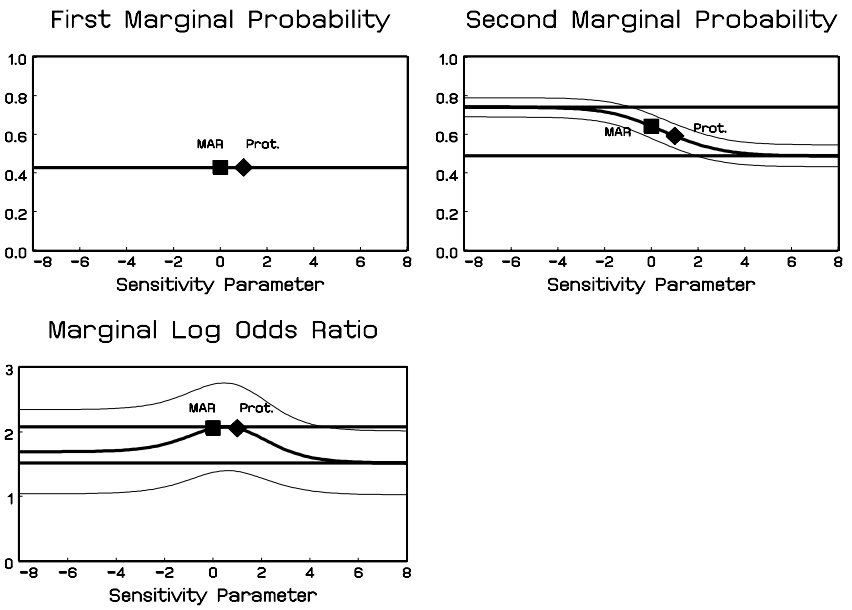

FIGURE 31.7. *Fluvoxamine Trial. Graphical representation of intervals of ignorance and intervals of uncertainty for monotone patterns of psychiatric study (side effects). The bold curve graphs the point estimates conditional on the sensitivity parameter. The bold horizontal lines project the interval of ignorance on the vertical axes. The extremes of the thin lines correspond to the interval of uncertainty. The MAR and protective point estimates have been added to the figure.*

Commonly, fitting a pair of identifiable models (e.g., Models 2 and 3) is regarded as a sensitivity analysis. This example shows how misleading this can be. 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 that this proximity is fortuitous.

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. The sensitivity parameter for Model 4 is γ_1 in Table 31.2. For Model 5, the two sensitivity parameters are γ_1 and δ_{11} (all other γ and δ parameters need to be set to zero for classical identifiability purposes). As expected, both models coincide for the first marginal probability. It turns out that their respective intervals of ignorance and uncertainty for the second marginal probability exhibit considerable overlap. In contrast, the length of the II for the log odds ratio is now about 5 times longer. The Model 5 lower limit of the IU is very close to zero, whereas its Model 4 counterpart shows clear evidence for a strong positive association between both outcomes.

TABLE 31.9. *Fluvoxamine Trial. Model fit for side effects (par: number of model parameters; G^2: likelihood ratio test statistic for model fit, corresponding p-value, estimates and 95% confidence limits for marginal probabilities and marginal (log) odds ratio.) For Model 10 (31.2), intervals of ignorance and uncertainty are presented instead.*

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

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 not model-based. The reduced width of the intervals produced under Model 4 are entirely due to the unverifiable model assumption that the dropout probability depends on both outcomes through their main effects only and *not* on the interaction between both outcomes. If this assumption is deemed implausible, it can easily be avoided by including an extra degree of freedom. However, in more complicated settings, such as when covariates are included or with continuous responses, assumptions are unavoidable in the interest of model parsimony. Now including the non-monotone patterns, any model within the BRD family with more than 8 parameters is non-identifiable. To simplify the sensitivity analysis, let us consider a slightly different but

equivalent parameterization

$$\pi_{r_1 r_1, ij} = p_{ij} \frac{\exp[\beta^*_{ij}(1-r_2) + \alpha^*_{ij}(1-r_1) + \gamma^*(1-r_1)(1-r_2)]}{1 + \exp(\beta^*_{ij}) + \exp(\alpha^*_{ij}) + \exp(\beta^*_{ij} + \alpha^*_{ij} + \gamma^*)}, \quad (31.13)$$

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

Although Models BRD1–9 have shown stability in the estimates of the marginal parameters of interest, it has been revealed in the monotone context, that such a conclusion could be deceptive. To study this further, we consider an overspecified model, analogous to Model 4 in Table 31.2. The choice can be motivated by observing that both BRD7 and BRD9 yield an interior solution and differ only in the β-model. Therefore, Model 10, defined by (31.2), will be fitted. Because one parameter is redundant, we propose using β_j as the sensitivity parameter. Although the II, obtained in this way, is acceptable, the IU shows aberrant behavior (plot not shown), toward larger values of the sensitivity parameter, leading to very wide IUs. This problem is entirely due to the zero count in pattern (0,1) (see Table 31.5), as can be seen by adding 0.5 to this zero count. The results are presented in Figure 31.8. The resulting II and IU are presented in Table 31.9, and they are very similar to the results for Model 4, as displayed in Table 31.7. Due to the non-monotone patterns, there is a (very small) ignorance in the first marginal probability as well. Once again, it is seen that fitting identifiable models only may be misleading because, for example, the log odds ratio shows much more variability than seen among Models BRD1–9.

31.6 Sensitivity Analysis and Pattern-mixture Models

Pattern-mixture models (Chapter 30) can be of use in the context of sensitivity analysis. Given there are several, quite distinct, strategies to formulate such models (Section 30.2), one can consider one strategy as a sensitivity analysis for another one. For example, the sensitivity of simple, identified models can be checked using identifying restrictions (Section 30.3). Also, a set of identifying restrictions can be considered, rather than a single one, by way of sensitivity analysis. Thijs *et al* (2002) and Molenberghs *et al* (2004) discuss strategies for fitting pattern-mixture models.

Obviously, one can formulate selection models for one's primary analysis, and then fit pattern-mixture models to assess sensitivity. This was done in Sections 30.4 and 30.5. Michiels *et al* (2002) followed this route.

Molenberghs, Michiels, and Kenward (1998) formulated models that combine aspects of both selection models and pattern-mixture models, and used pseudo-likelihood ideas to fit such models.

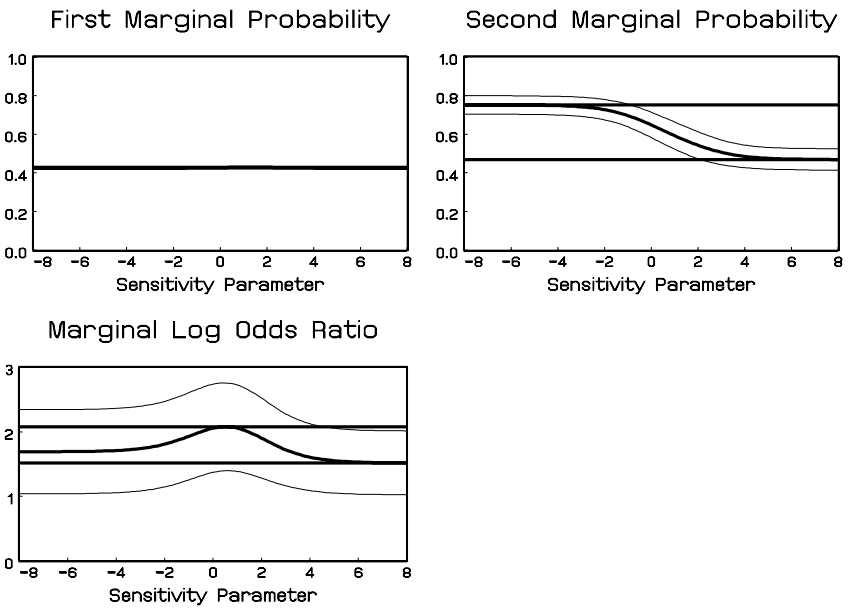

FIGURE 31.8. *Fluvoxamine Trial. Graphical representation of intervals of ignorance and intervals of uncertainty for monotone patterns (side effects). A value of 0.5 is added to the zero count in pattern (1,0). The bold curve graphs the point estimates conditional on the sensitivity parameter. The bold horizontal lines project the interval of ignorance on the vertical axes. The extremes of the thin lines correspond to the interval of uncertainty.*

31.7 Concluding Remarks

When fitting models to incomplete (longitudinal) data, especially of the MNAR type but also of the MAR and MCAR types, it is important to assess the sensitivity of the conclusions to unverifiable model assumptions. Generally, a sensitivity analysis can be conducted within different frameworks, and there are times where the setting will determine which framework is the more appropriate one (for example Bayesian or frequentist), in conjunction with technical and computational considerations. Draper (1995) has considered ways of dealing with model uncertainty in the very natural Bayesian framework. We have focused on local influence methods, the interval of ignorance, and the use of pattern-mixture models. Although these methods are useful, it ought to be clear they are by no means the only routes to sensitivity analysis. This field, in the context of incomplete data, is still in full development and more work will undoubtedly emerge in times to come.