Implementation of pattern-mixture models in randomized clinical trials

Implementation of pattern-mixture models in randomized clinical trials

By P. Bunouf[1] and G. Molenberghs[2]

(1) Laboratoires Pierre Fabre, France

(2) I-BioStat, Universiteit Hasselt and KU Leuven, Belgium

Summary. Modern analysis of incomplete longitudinal outcomes involves formulating assumptions about the missingness mechanisms and then using a statistical method that produces valid inferences under this assumption. In this manuscript, we define missingness strategies for analyzing randomized clinical trials (RCTs) based on plausible clinical scenarios. Penalties for drop-out are also introduced in an attempt to balance benefits against risks. Some missingness mechanisms are assumed to be non-future dependent (NFD) which is a subclass of missing-not-at-random. NFD stipulates that missingness depends on the past and the present information, but not on the future. Missingness strategies are implemented in the pattern-mixture modeling (PMM) framework using multiple imputation (MI) and we show how to estimate the marginal treatment effect. Next, we outline that MI can be used to investigate the impact of drop-out strategies in subgroups of interest. Finally, we provide the reader with some points to consider when implementing PMM-MI analyses in confirmatory RCTs. The data set that motivated our investigation comes from a placebo-controlled RCT design to assess the effect on pain of a new compound.

KEY WORDS : incomplete longitudinal outcome ; missing data ; MNAR ; non-future dependence ; pattern-mixture model ; multiple imputation.

1 Introduction

Clinical developers are becoming increasingly aware of the good practices in analyzing incomplete longitudinal outcomes in randomized clinical trials (RCTs). Their analysis involves formulating assumptions about the missingness mechanisms and then using a statistical method that produces valid inferences under this assumption. Consequently, the formulation of missingness assumptions in a transparent and interpretable manner has become a key aspect. A major cause of missingness in RCTs is drop-out. Missingness is at random (MAR) if drop-out occurrence is independent of missing outcome values, conditionally on the observed ones. If the covariates are fully observed, additional dependence on covariates is allowed for too. When MAR fails to hold, missingness is not at random (MNAR). MNAR implies that drop-out occurrence depends on an outside variable not in the model or is related to unobserved outcome values at the drop-out time and possibly afterwards, even when conditioned on available information. The consequence of MNAR is that missing outcome values cannot be reliably predicted using observed measurements (i.e., covariates and outcome values).

Pattern-mixture modeling (PMM) is a framework that can be considered when missingness is MNAR [1]. PMM stratifies the sample of subjects by missingness pattern and formulates distinct models to estimate parameters within each pattern. In RCTs with multiple scheduled visits, time is often modeled as a fixed class effect and patterns are defined based on drop-out at every visit. Whereas parameters are all identified in completers (i.e., subjects who complete the trial until the last visit), some parameters are unidentified in the patterns of drop-out subjects. This can be overcome by setting unidentified parameters equal to functions of the parameters describing the distributions of other patterns.

The so-called identifying restrictions indicate in which patterns information is borrowed. In Little's taxonomy [2], complete-case missing values (CCMV) stipulates that missing information is borrowed from completers. In neighboring-case missing values (NCMV), the closest neighboring pattern is used instead. NCMV implies that any unidentified parameter at a visit is estimated in the pattern of subjects having their last observed outcome value at this visit. Available-case missing values (ACMV) offers a compromise between CCMV and NCMV as all available patterns are used weighted by occurrence of each pattern. ACMV has a particular status since this is the natural counterpart of MAR in the PMM framework. In practice, analysis assuming MAR is often a point of departure for sensitivity analyses assuming MNAR.

Non-future missing values (NFMV) is another identifying restriction that offers appealing perspectives for sensitivity analysis since the user has full freedom to choose the distribution of the first unobserved (or present) outcome value given previous measurements. The appeal of NFMV is that missingness depends on the past and the present, but not on future unobserved outcome values. In other words, missingness is non-future dependent (NFD). This missingness mechanism is a sub-class of MNAR. The correspondence between NFMV and NFD allows the formulation of comprehensible drop-out strategies because mechanisms are directly characterized by the free distributions of present outcome values. In this manuscript, we define several drop-out strategies based on plausible clinical scenarios but penalties for drop-out are also introduced in an attempt to balance benefits against risks. These penalties are intended to reflect the prejudice suffered by subjects.

Drop-out strategies can be implemented in the pattern-mixture modeling (PMM) framework using multiple imputation (MI). MI can be used to allow for MAR as well as MNAR mechanisms. We show how to estimate the marginal treatment effect and results obtained under drop-out strategies are contrasted. We attempt to overcome major drawbacks of well known single-imputation concepts such as baseline-observation-carried-forward (BOCF) and last-observation-carried-forward (LOCF) using appropriate NFD strategies. We also question the choice of the first unobserved visit to characterize the present. Next, it is known that MI offers a transparent way to represent the impact of drop-out strategies [3]. We study this impact in subgroups of drop-out subjects according to the cause of drop-out. Finally, we provide the reader with some points to consider to carry-out PMM-MI analyses in confirmatory RCTs and we describe an implementation using an existing freely-available SAS program [4].

To this end, we have re-analyzed a data set that comes from a placebo-controlled clinical trial to assess the effect on pain of a new compound [5][6]. A continuous visual analogue scale (VAS) was used to assess pain intensity and a binary outcome whose values indicate clinical response or non-response was derived. The next section describes the case study and outlines several points to consider during analysis preparation. Identifying restrictions and drop-out strategies are laid out in Section 3. In Section 4, we describe PMM-MI methods and we explain how to estimate the marginal treatment effect. Results obtained under drop-out strategies are contrasted in Section 5 whereas Section 6 and Section 7 respectively contain the discussion and the concluding remarks. The appendix provides technical information to implement analysis using software.

2 Case study

2.1 Trial design and statistical objectives

The data set was collected in a multi-country European phase-III double-blind clinical trial of which the objective was to compare the effect on pain intensity of a test product (Test) versus placebo (PCB) in subjects suffering from fibromyalgia. Four hundred thirty three (433) subjects were randomized to receive Test versus 447 subjects to receive PCB, that is 880 subjects in total.

After random assignment to either Test or PCB, subjects underwent a four-week escalation dose period followed by three four-week periods under stable dose. All periods ended at the scheduled visits that are visit 1 at the end of dose escalation and monthly visits 2–4 under stable dose. Subjects who discontinued were asked to undergo a specific visit at the discontinuation date, termed the drop-out visit.

The pain intensity level was reported by subjects on a VAS that ranges from 0 (no pain) to 100 (worst pain imaginable) using an electronic diary at each visit. The four values of a continuous longitudinal outcome were derived as the pain intensity level at randomisation (visit 0) minus the ones at visits 1–4. Thus, a positive value indicates a decrease in pain intensity. For the drop-out subjects, the pain intensity level was assessed at the drop-out visit, and the corresponding outcome value was repositioned to the next scheduled visit for analysis. Then, the four values of a longitudinal binary outcome were derived in terms of clinical response or non-response based on a minimum 30% improvement from randomisation on the continuous outcome values. In chronic pain diseases, including fibromyalgia, this gain is regarded as clinically relevant.

The estimation of treatment effect at visit 4 was the primary objective of the trial. The original analyses, as mentioned in the trial protocol, were based on BOCF and LOCF principles to impute missing outcome values. The continuous outcome values were fitted using a covariance analysis model and the binary outcome values were fitted using a logistic regression model. Pain level at randomization was a pre-specified key covariate whereas country was a pre-specified stratum factor. The objective of analyses described in this manuscript is to assess the treatment effects adjusted on the same covariate and stratum factor at visit 4 on the continuous and the binary outcomes under a set of drop-out strategies. Analyses are conducted on all randomized subjects.

2.2 Causes of missingness

Out of the 880 randomized subjects, 77/447 subjects (17.2%) in the PCB group and 126/433 subjects (29.1%) in the Test group dropped out from the trial. In our case study, drop-out from the trial also means drop-out from the follow-up even if treatment intake had stopped earlier. Three subjects who dropped out very early have no assessment after randomization. All the other drop-out subjects have completed their drop-out visit and reported their pain intensity level. There are only two 'intermittent' missing assessments (i.e., missing assessments not due to drop-out). The first occurred at the randomization visit because of a technical problem with the electronic diary whereas the second occurred at visit 3 and was caused by the absence of a subject from home.

The distribution of subjects by drop-out cause by treatment group is given in Table 1.

TABLE 1 – Frequencies (percentages) of subjects by drop-out cause by treatment group

Causes	PCB	Test
Adverse event (AE)	44 (9.8)	96 (22.2)
Serious AE due to treat.	*1 (0.2)*	*3 (0.7)*
AE due to treatment	*30 (6.7)*	*89 (20.6)*
Serious AE not due to treat.	*3 (0.7)*	*0 (_)*
AE not due to treatment	*10 (2.2)*	*4 (0.9)*
Other than AE	33 (7.4)	30 (7.0)
Patients's decision	*11 (2.5)*	*11 (2.5)*
Investigator's decision	*0*	*4 (0.9)*
Therapeutic failure	*18 (4.0)*	*10 (2.3)*
Other reasons	*4 (0.9)*	*5 (1.2)*
Total	*77 (17.2)*	*126 (29.1)*

The main cause of drop-out was adverse event (AE). The higher drop-out rate for Test was caused by a higher occurrence of AEs in this group. Particularly, there are three times more subjects who dropped out for AE due to Test (20.6%) than to PCB (6.7%). Conversely, subjects who dropped out for other causes than AE are similarly distributed in the two groups with 7.4% for PCB and 7.0% for Test. In this category, therapeutic failure concerns twice less subjects under Test (2.3%) than under PCB (4.0%). Further investigations (not shown in Table 1) reveal that the earlier the drop-out, the greater the occurrence of drop-out caused by AE (from 77.8% if drop-out occurs between randomization and visit 1 to 40.6% between visit 3 and visit 4). The distribution of drop-out causes in our case study is classical of many clinical trials in chronic pain.

2.3 Characteristics of patterns

Patterns are defined based on drop-out at every visit. Accordingly, patterns 0–4 respectively consist of subjects with 0 to 4 observed outcome values. Pattern 0 contains the subjects without assessment at randomisation and the three subjects without assessment after randomization. However, pattern 4 doesn't allow distinguishing between drop-out subjects and completers. Indeed, subjects who dropped out after visit 3 have a complete outcome profile because of the repositioning of the drop-out visit to visit 4. To clarify this, we define pattern 5 which contains completers whereas pattern 4 keeps the subjects who dropped out after visit 3. This separation does not contradict the pattern definition given here-above since one may consider that completers actually dropped out at a virtual visit 5 after the end of the trial follow-up.

Figure 1 displays the unadjusted mean profiles per pattern by treatment group.

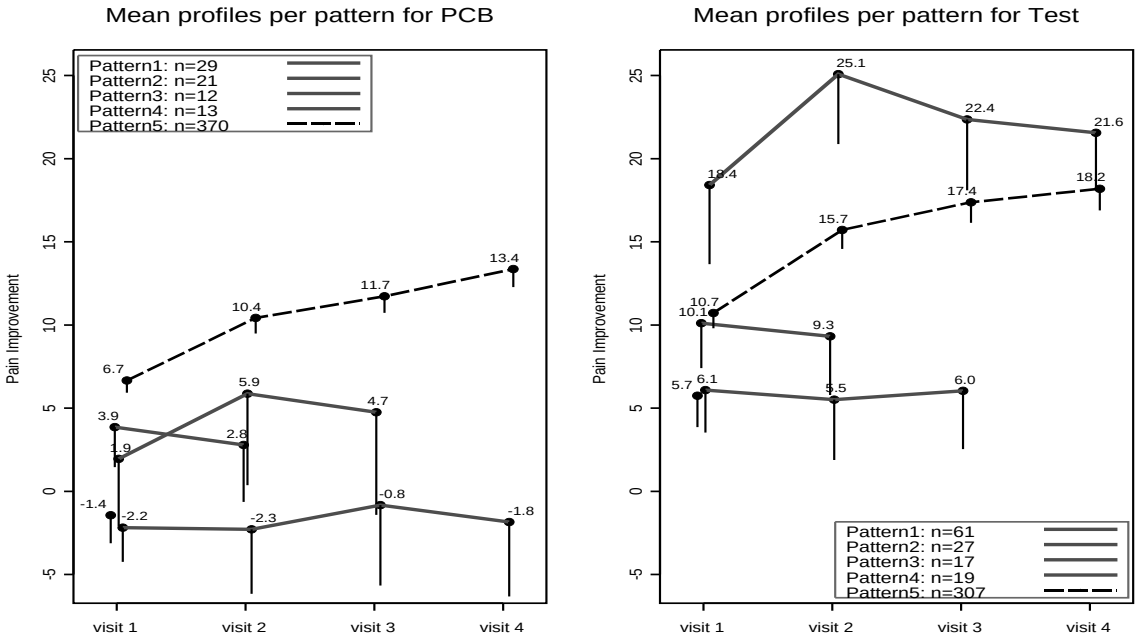

FIGURE 1 – Unadjusted mean profiles and frequencies per pattern by treatment group.

The mean profiles of early drop-out patterns decline immediately whereas the mean profiles of the drop-out subjects who stay longer in the trial show a stagnation and then a decline. Patterns 4 shows a huge difference between treatment groups. This difference is partly explained by the causes of drop-out which are less related to AE, and probably more to efficacy, than in the other drop-out patterns. The patterns of completers exhibit sustained improvements until visit 4. The difference in mean profiles between pattern 4 and pattern 5 supports the decision of separation.

2.4 Pattern sizes by stratum

Pattern sizes by stratum is an important point to consider before implementing a stratified PMM analysis. The reason is that parameter estimation can fail if a minimum number of values are not available. In our case study, pattern 0 is treated separately since this pattern is not used for parameter estimation and is pooled with pattern 1 for imputation (see Section 4.2.2 for further information). The imputation models incorporate the pre-specified stratum factor country and a full group-by-time interaction for fixed effects, and unstructured error covariance matrix. In this setting, parameter estimation would require a minimum of three available outcome values per pattern by country by treatment group. As the trial was conducted in twelve European countries, this pre-requisite is not reached and a tricky pooling of countries must be done.

The trial was conducted in the Czech republic (CZ ; n =55), Denmark (DK ; n =17), Finland (FI ; n =60), France (FR ; n =184), Germany (GE ; n =42), Italy (IT ; n =101), Norway (NO ; n =53), Poland (PL ; n =31), Portugal (PT ; n =16), Romania (RO ; n =59), Spain (ES ; n =111), and Sweden (SE ; n =151). A first pooling of countries into four regions was done based on their geographical proximity. These regions are central Europe (CZ, GE, PL, RO), southern Europe (ES, IT, PT), France (FR) and the Nordic countries (DK, FI, NO, SE). Then, central and southern Europe were pooled into the same region to reach the minimum pattern size. The consequence of this pooling is that the stratum factor country is replaced by the three-class factor region in the imputation model.

The pattern sizes by region by treatment group are provided in Table 2.

TABLE 2 – Pattern sizes by region by treatment group

		central + south	France	nordic countries	All Countries
Patterns 0+1	PCB	1+11	9	1+9	2+29
	Test	1+28	10	1+23	2+61
Pattern 2	PCB	9	4	8	21
	Test	11	5	11	27
Pattern 3	PCB	4	4	4	12
	Test	3	7	7	17
Pattern 4	PCB	4	3	6	13
	Test	11	3	5	19
Pattern 5	PCB	187	70	113	370
	Test	145	69	93	307
All Patterns	PCB	216	90	141	447
	Test	199	94	140	433
	All Groups	415	184	281	880

3 Identifying restrictions and drop-out strategies

3.1 Description of identifying restrictions

The basis of pattern-mixture modeling results from a particular decomposition of the joint distribution of the outcome variable together with the drop-out indicator. The pattern-mixture distribution of the values $y_1, \ldots, y_T$ of complete outcomes is given by :

$$f(y_1, \ldots, y_T) = \sum_{t=1}^{T} \alpha_t f_t(y_1, \ldots, y_T), \tag{1}$$

where α_t denotes the proportion of pattern t and $f_t(y_1, \ldots, y_T)$ stands for $f(y_1, \ldots, y_T|t)$.

In our case study, patterns are defined based on drop-out at every visit. More precisely, if the tth outcome value is the last observed one and subject drops out after that, this subject belongs to pattern t, $t = 1, \ldots, T$. In (1), the distribution of the whole population is expressed in terms of a mixture of the distributions of pattern populations. These, in turn can be decomposed as :

$$\begin{aligned} f_t(y_1, \ldots, y_T) &= f_t(y_1, \ldots, y_t) f_t(y_{t+1}, \ldots, y_T|y_1, \ldots, y_t) \\ &= f_t(y_1, \ldots, y_t) \prod_{s=t+1}^{T} f_t(y_s|y_1, \ldots, y_{s-1}). \end{aligned} \tag{2}$$

The first component in (2) is identified from the observed outcome values. The second is a product of conditional pattern distributions, which are unidentified since the values of y_s are unobserved in these patterns. This can be overcome by setting unidentified parameters equal to functions of the parameters describing the distributions of other patterns. The identifying restrictions, informally introduced in Section 1, are used to this effect.

Under CCMV, identification is based on pattern T, the pattern of completers. This can be formalized by :

$$f_t(y_s|y_1, \ldots, y_{s-1}) = f_T(y_s|y_1, \ldots, y_{s-1}), \quad s = t+1, \ldots, T. \tag{3}$$

Under NCMV, the neighboring pattern is used instead :

$$f_t(y_s|y_1, \ldots, y_{s-1}) = f_s(y_s|y_1, \ldots, y_{s-1}), \quad s = t+1, \ldots, T. \tag{4}$$

Identification can also be based on all identified patterns as specified in the formulation :

$$f_t(y_s|y_1, \ldots, y_{s-1}) = \sum_{j=s}^{T} \omega_{sj} f_j(y_s|y_1, \ldots, y_{s-1}), \quad s = t+1, \ldots, T. \tag{5}$$

Every convex set of ω_{sj}'s that sums to 1 provides a valid identification scheme. In [8], the ω_{sj}'s are determined such that (5) corresponds with ACMV.

Alternatively, NFMV offers the appealing characteristic that the distributions of present outcome values, that we denote g_t for the sake of clarity, are left unconstrained [9]. In our analysis, we set the g_t's equal to their f_{t+1} counterparts, in the spirit of NCMV, with a possibility to shift the mean by a value Δ. Formally, this can be expressed by :

$$g_t(y_{t+1}|y_1,\ldots,y_t) = f_{t+1}(y_{t+1}+\Delta|y_1,\ldots,y_t). \tag{6}$$

In our case study, the closest neighboring patterns involved in NCMV and NFMV identifications are all drop-out patterns. Indeed, pattern 5 is never used for identification given that pattern 4 contains complete outcome profiles. Of note, the alternative choice of basing NFMV identification on the pattern of completers, in the spirit of CCMV, would have been inappropriately optimistic. As described in Figure 1, the mean profiles of drop-out patterns show a stagnation and then decline at the drop-out visits, whereas the pattern of completers shows regular improvements in the two treatment groups. Moreover, the large size of patterns 5 would impose an inappropriately low uncertainty to parameter estimates.

The correspondence of NFMV to NFD allows formally relating drop-out occurrence to the g_t's. Through (6), we state that drop-out is caused by, or associated with, a mean decrease by $\triangle$ of present values relative to the values observed in subjects who dropped out at this visit, without other possible causes, or associations, involving future values.

3.2 Description of drop-out strategies

Although MAR is impossible to demonstrate [10], primary analyses in RCTs often relies on this assumption. As laid out in [7], MAR is compatible with *de jure* (or per-protocol) analysis of which an objective is to estimate the treatment effect as if drop-out subject continued the trial under the same conditions until the last scheduled visit. The consequence on PMM analysis under ACMV, which is the counterpart of MAR, is that imputation is based on subject's treatment group. We keep this imputation rule in PMM analyses under CCMV and NCMV, which are often used to study the effect of a certain departure from MAR. These three identifying restrictions define the de jure analyses.

Alternatively, a *de facto* (or intention-to-treat) analysis aims at reflecting the effect of the initially assigned treatment as well as the impact of the treatment withdrawal subsequent to drop-out. In our case study, a de facto analysis can be implemented considering that subjects under PCB continue on PCB after drop-out and subjects under Test switch

to PCB. The consequence on PMM analysis is that imputation is based on PCB in all subjects whatever their treatment group. This approach is referred to as *reference-based imputation* in earlier works [11].

The de facto analyses are defined by the following drop-out strategies, all based on the NFD assumption. We first introduce N-LOCF which is intended to overcome a major drawback of LOCF, i.e., single imputation, by allowing an appropriate degree of uncertainty. Under N-LOCF, the present values are imputed without shifting the distribution mean (i.e., $\triangle = 0$). In our case study, N-LOCF is more conservative than MAR since N-LOCF maintains the pain intensity level in present values even though the mean profiles by treatment group tend to separate over time. Another reason is that N-LOCF imputation is based on PCB whereas ACMV imputation is based on the subject's treatment group. In the same vein, we define N-BOCF which resumes the concept of BOCF single imputation. Under N-BOCF, we set the values of $\triangle$ equal to the individual gains observed from randomization to the last visit.

The following drop-out strategies rely on clinical rationales. Under N-DO5, any drop-out is associated with a penalty on present outcome values in terms of pain intensity which is set to $\triangle = -5$. This value was discussed and approved by clinical experts for its meaningfulness in chronic pain. This also corresponds to minus the expected treatment effect under the alternative in the trial protocol.

The drop-out strategy N-AE5 refines N-DO5 by incorporating information about the cause of drop-out in an attempt to balance benefits against risks. The penalty rule is supposed to reflect the prejudice suffered by subjects because of drop-out. Under N-AE5, missing outcome values are imputed, assuming :

1. NFD($\triangle = -5$) if a subject drops out for AE due to the treatment,
2. NFD($\triangle = -10$) if a subject drops out for serious AE due to the treatment,
3. N-LOCF or NFD($\triangle = 0$) if a subject drops out for AE not due to the treatment,
4. MAR if a subject drops out for other reasons than AE.

In our case study, the lower the value of $\triangle$, the greater the conservatism introduced by NFMV imputation into the analysis because of the greater drop-out rate in the Test group.

The next two drop-out strategies aim at measuring the impact of the penalty value in N-AE5. The value is brought to $\triangle = -10$ in N-AE10 and $\triangle = -15$ in N-AE15 in subjects who dropped out for AE due to the treatment. The other rules, i.e., the doubling of the Δ value if the AE is serious and the zeroing if AE is not due to the treatment, are kept.

In another set of de facto analyses, we question our implementation of NFD strategies for which the present is characterized by the first unobserved visit whereas the distributions of present values are based on the drop-out visits in the closest neighboring patterns. Indeed, the drop-out visits are repositioned to the next scheduled visit for analysis so that the present may be likened to a (near) future. Moreover, the influence of the cause of drop-out, such as an AE, on a patient's assessment is not well characterized. So, we cannot rule out that the outcome value at the drop-out visit combines efficacy and safety. In an attempt to address this, we introduce N-AE5-L, N-AE10-L, and N-AE15-L, which resume the penalty rules applied in N-AE5, N-AE10, and N-AE15, respectively. However, the drop-out visits are removed in the subjects who dropped out for AE and are kept otherwise.

Results of de jure and de facto analyses in the PMM framework will be contrasted to those obtained after BOCF and LOCF single imputations, as well as, in completers and in subjects with complete outcome profiles (i.e., in subjects who drop out after visit 3). To facilitate the comparison, all analyses will be conducted using the same statistical model, which is described in the next section. For BOCF and LOCF analyses specifically, the three subjects of pattern 0 without assessment under treatment are imputed with a 0 value for the continuous outcome and are clinically non-responders. The fourth subject of pattern 0 without assessment at randomization is excluded from analyses.

Table 3 sums up the drop-out strategies defined in this section.

TABLE 3 – Description of drop-out strategies

Type of analysis	Drop-out strategy	Missingness mechanisms	Type of imputation	Parameter estimation	Data set
_	BOCF	Unknown	Single	_	All visits
_	LOCF	Unknown	Single	_	All visits
_	Complete	_	_	_	Complete profiles
_	Completers	_	_	_	Completers
de jure	NCMV	Unknown	Multiple	By group	All visits
de jure	ACMV	MAR	Multiple	By group	All visits
de jure	CCMV	Unknown	Multiple	By group	All visits
de facto	N-BOCF	NFD	Multiple	PCB	All visits
de facto	N-LOCF	NFD	Multiple	PCB	All visits
de facto	N-DO5	NFD	Multiple	PCB	All visits
de facto	N-AE5	NFD/MAR	Multiple	PCB	All visits
de facto	N-AE5-L	NFD/MAR	Multiple	PCB	Scheduled visits
de facto	N-AE10	NFD/MAR	Multiple	PCB	All visits
de facto	N-AE10-L	NFD/MAR	Multiple	PCB	Scheduled visits
de facto	N-AE15	NFD/MAR	Multiple	PCB	All visits
de facto	N-AE15-L	NFD/MAR	Multiple	PCB	Scheduled visits

4 Implementation of PMM analysis using MI

4.1 General features

The PMM analyses are implemented following the standard MI approach, as described in [12], which includes pattern parameters estimation, missing values imputation, and pooled analysis. Some other basic features follow general recommendations for MI analysis of continuous and derived binary outcomes which are stated, justified, and exemplified in [6]. These are :

1. MI in the original scale followed by analysis of the desired derived outcome is a more informed strategy than direct analysis of the derived outcome ;
2. Analysis at the desired timepoint provides valid inferences if all the effects are properly fitted by imputation models ;
3. MI offers a transparent way to represent the impact of drop-out strategies.

In our case study, missing outcome values are imputed on the continuous scale in accordance with recommendation 1. Then, we derive the values of the binary outcome in terms of clinical response or non-response, by subject and by imputation. In this way, all MI analyses are based on the same complete data sets for the continuous outcome.

Analysis of the outcome values at visit 4 is based on the original statistical models, as mentioned in the trial protocol. The continuous outcome is analyzed using a covariance analysis model whereas the binary outcome is analyzed using a logistic regression model. The pain level at randomization is a covariate and the three-class stratum factor region a stratum factor. According to recommendation 2, the analysis of the outcome values at visit 4 provides valid treatment-effect inferences, like any longitudinal model if the covariate and class-factors are properly fitted by the imputation models. The imputation models used are mixed models for repeated measures (MMRM) with the factor region and the full baseline-by-visit and group-by-visit interactions for the fixed effects and unstructured error covariance matrix.

We implement recommendation 3 using summary outcome values by subject directly obtained from the complete data sets. The summary continuous values are simply the means by subject over imputations. For the summary binary values, the clinical responses or non-responses are first derived by subject and by imputation. The subject is declared as clinical responder if at least half of these binary values correspond to clinical responses.

Next, we re-use the covariance analysis and the logistic regression models here-above to analyze the summary outcome values in the subgroups of subjects who dropped out for AE and those who dropped out for other reasons. We also provide raw frequencies of clinical responders based on the summary binary values, for illustrative purposes.

An analysis of the summary outcome values ignores between-imputation variance unlike pooled analysis in Rubin's method. The relative increase of variance due to missing data, introduced by Rubin [12], depends on the ratio of the between-imputation to the total variance. In our case study, the maximum value obtained for this criterion was *0.17*. This indicates that the magnitude of ignored variance information is moderate. It also supports the use of the summary outcome values to illustrate the impact of drop-out strategies.

4.2 Estimation of the marginal treatment effect

In this section, we describe a procedure in three stages to estimate the marginal treatment effect in a PMM-MI analysis. It is important to recall that the original Rubin's method provides inferences that are conditional on patterns by construction. To analyze the marginal treatment effect, the pattern-specific effects must be combined into a pattern-average effect. Some aspects of the procedure are further detailed in [13]. A complement of information to implement analysis using a SAS program which combine R functionalities is available in the Appendix.

4.2.1 Pattern parameter estimation

Distinct models are formulated within each pattern. Let us denote by $\mathbf{Y}_i = (y_{i,1}, \ldots, y_{i,T})$ the complete outcome vector in the i*th* subject of pattern t and $\mathbf{Y}_{i,obs} = (y_{i,1}, \ldots, y_{i,t})$ its observed part. The MMRMs per pattern can be expressed as :

$$\mathbf{Y}_{i,obs} = X_i \boldsymbol{\beta}_t + \boldsymbol{\epsilon}_i, \tag{7}$$

where $\boldsymbol{\epsilon}_i \sim \mathrm{N}(\mathbf{0}, \Sigma_t)$, Σ_t is unstructured, and the $\boldsymbol{\epsilon}_i$'s are independent. The matrix X_i contains the known subject covariates whereas $\boldsymbol{\beta}_t$ contains the unknown fixed effects. This first stage is aimed at estimating the pattern parameters $\boldsymbol{\beta}_t$ and Σ_t, whose estimators are denoted $\widehat{\boldsymbol{\beta}}_t$ and $\widehat{\Sigma}_t$ in what follows.

4.2.2 Imputation

Missing values imputation is conducted sequentially by value. We describe here-below how to obtain a run of M imputed values of $y_{i,t+1}$. Multiple imputation of $y_{i,t+2}, \ldots, y_{i,T}$

follows the same process by considering the previous imputed values as observed ones. The identifying restriction chosen determines which patterns are used for imputation. Whereas imputations under CCMV (3) and NCMV (4) use unique patterns, imputations under ACMV and NFMV (5) is based on several patterns.

In our illustration, we suppose that $y_{i,t+1}$ is imputed from pattern r ($t+1 \leq r \leq T$). Let us introduce $\boldsymbol{\mu}_{i,r}$ the mean of $\mathbf{Y}_i$, which is $\boldsymbol{\mu}_{i,r} = X_i \boldsymbol{\beta}_r$. Based on appropriate parts of $\boldsymbol{\mu}_{i,r}$ and Σ_r, we further define the distributions of the components of $\mathbf{Y}_i$, which are $\mathbf{Y}_{i,obs} \sim \mathrm{N}(\boldsymbol{\mu}_{i,r,1}, \Sigma_{r,11})$ and $y_{i,t+1} \sim \mathrm{N}(\mu_{i,r,2}, \Sigma_{r,22})$. Their covariances are denoted $\Sigma_{r,12}$ and $\Sigma_{r,21}$. Using 2|1 as notation for $y_{i,t+1}|y_{i,1}, \ldots, y_{i,t}$, the conditional pattern distribution of $y_{i,t+1}$ given $y_{i,1}, \ldots, y_{i,t}$ is described by :

$$f_r(y_{i,t+1}|y_{i,1}, \ldots, y_{i,t}) \sim N(\mu_{i,r,2|1}, \Sigma_{r,2|1}),$$

where

$$\begin{aligned} \mu_{i,r,2|1} = &\quad \mu_{i,r,2} + \Sigma_{r,21}[\Sigma_{r,11}]^{-1}(\mathbf{Y}_{i,obs} - \boldsymbol{\mu}_{i,r,1}), \\ \Sigma_{r,2|1} = &\quad \Sigma_{r,22} - \Sigma_{r,21}[\Sigma_{r,11}]^{-1}\Sigma_{r,12}. \end{aligned} \tag{8}$$

Uncertainty pertaining to the pattern parameters $\boldsymbol{\beta}_r$ and Σ_r is incorporated through Bayesian distributions. On the basis of non-informative Jeffreys' priors in the Gaussian context, the values of $\widehat{\boldsymbol{\beta}}_r^{(m)}$ and $\widehat{\Sigma}_r^{(m)}$, $m = 1, \ldots, M$, are randomly drawn from their respective posterior predictive distributions. After the derivation of $\widehat{\boldsymbol{\mu}}_{i,r}^{(m)}$, the imputed values of $y_{i,t+1}$ are drawn from the conditional pattern distributions, which are expressed by :

$$f_r^{(m)}(y_{i,t+1}|y_{i,1}, \ldots, y_{i,t}) \sim N(\widehat{\boldsymbol{\mu}}_{i,r,2|1}^{(m)}, \widehat{\Sigma}_{r,2|1}^{(m)}), \qquad m = 1, \ldots, M. \tag{9}$$

This stage requires the covariate and the outcome value at visit 1 to be available. In pattern 0, these missing values are multiply imputed to begin with. Missing covariates are imputed from the conditional bivariate normal distribution of both variables given the first outcome value. Missing first outcome values are imputed from the conditional bivariate normal distribution of both variables given the covariate in pattern 1 under NCMV, the whole sample under ACMV, and pattern 5 under CCMV. Next, the only intermittent missing value at visit 3 is multiply imputed from the conditional multivariate normal distribution of all outcome values given the observed ones in pattern 5, which is the subject's pattern.

4.2.3 Pooled analysis

The outcome values at visit 4 are analysed per pattern and by imputation. Let us denote the MT-dimensional vector of treatment-effect estimates by $\boldsymbol{\delta}$=($\hat{\delta}_1^{(1)}$,...,$\hat{\delta}_1^{(M)}$,...,$\hat{\delta}_T^{(1)}$,...,$\hat{\delta}_T^{(M)}$), where $\hat{\delta}_t^{(m)}$ is the estimate in pattern t and imputation m.

We first derive the vector of pooled estimates per pattern which is noted $\boldsymbol{\delta}^*$=(δ_1^*,...,δ_T^*). According to Rubin's rule, the δ_t^*'s are the means of the $\hat{\delta}_t^{(m)}$'s per pattern and the $T\times T$-dimensional total covariance matrix of $\boldsymbol{\delta}^*$ is given by :

$$V^* = W^* + \left(\frac{M+1}{M}\right) B^*,$$

where W^* is the within-imputation covariance matrix and B^* is the between-imputation covariance matrix. W^* is a diagonal matrix whose coefficients are the means of the δ_t^*'s standard errors and B^* is the correlation matrix of $\boldsymbol{\delta}^*$.

Then, the pattern-specific information is combined with respect to pattern probabilities. To this end, we use Rubin's rule a second time to derive the pattern-average parameters which are denoted $\delta^\dagger$, $W^\dagger$, and $B^\dagger$, respectiely in what follows. Let us define the vector $\boldsymbol{\Pi}$ that contains the overall pattern probabilities regardless of treatment groups. We estimate the marginal treatment effect using $\boldsymbol{\delta}^\dagger = \boldsymbol{\Pi}'\boldsymbol{\delta}^*$. The total variance is given by :

$$V^\dagger = W^\dagger + \left(\frac{T+1}{T}\right) B^\dagger,$$

where $W^\dagger$=$\boldsymbol{\Pi}'W^*\boldsymbol{\Pi}$+$\boldsymbol{\delta}^{*'}\mathrm{Var}(\boldsymbol{\Pi})\boldsymbol{\delta}^*$ is the within-pattern variance and $B^\dagger$=$\boldsymbol{\Pi}'B^*\boldsymbol{\Pi}$ the between-pattern variance. These formulas were derived and justified in [13]. Then, the treatment effect is tested using a Student t-statistic with the freedom degree $(T-1)*\{1+W^\dagger/[(1+1/T)*B^\dagger)]^2\}$ given in Rubin [12].

It is important to realize that the formulaes described here-above assume that pattern probabilities are equal in the two groups. Whereas this assumption is necessary, and acceptable, to derive an analytic approximation of the total variance, the treatment effect can be estimated by dispensing with it. This is all the more necessary in our case study where Figure 1 yields the pattern probabilities vectors $\widehat{\boldsymbol{\Pi}}'_P$=(.065, .047, .027, .029, .831) for PCB and $\widehat{\boldsymbol{\Pi}}'_T$ = (.141, .063, .039, .044, .712) for Test and the classical χ^2 to test homogeneity between groups provides p<0.01. To tackle this, it is quite possible to use the treatment-effect estimate provided by direct pooled analysis which is the average of the estimates by imputation regardless of patterns. This estimate is conditional on patterns but consistent. So, our implementation of PMM-MI analysis combines the estimate obtained from direct pooled analysis and the total variance of the pattern-average effect.

5 Results

5.1 Continuous outcome analysis

Figure 2 displays treatment-effect inferences obtained with the continuous outcome at visit 4 under the drop-out strategies specified in Table 3. We provide the differences between treatment groups with their 95% confidence limits and p-values. Results by treatment group are expressed in terms of adjusted means with standard errors.

As expected, owing to the greater drop-out rate in the Test group, the treatment effect estimated after LOCF imputation [4.30 (1.41) p=0.002] lies between that obtained after BOCF imputation [2.85 (1.37) p=0.038] and in subjects with complete outcome [5.87 (1.61) p<.0001]. Relative to BOCF, LOCF increases the means by treatment group by 0.1 for PCB and 1.6 for Test, suggesting that drop-out subjects under Test benefit from the treatment. Next, the treatment effect estimated in completers [5.14 (1.66) p=0.002] is lower than that obtained in subjects with complete outcome. This results was expected since the set of subjects with complete outcome is composed of the subjects of pattern 4 and the subjects of pattern 5 (i.e., the completers). Moreover, Figure 1 exhibits a huge difference between treatment groups in pattern 4.

Among PMM analyses, the per-protocol analyses exhibit moderate divergences. As expected, CCMV which uses the pattern of completers for imputation produces larger means by treatment group (11.9 for PCB and 16.9 for Test) than ACMV which uses all available patterns (11.7 for PCB and 16.4 for Test) and NCMV which uses the closest neighboring patterns (10.7 for PCB and 15.6 for Test). CCMV also produces larger treatment effect [5.03 (1.63) p=0.002] than ACMV [4.78 (1.63) p=0.004] and NCMV [4.85 (1.67) p=0.004].

The de facto analyses exhibit greater divergences. The treatment effect estimated under N-LOCF [3.92 (1.72) p=0.024] is related to that obtained with the single-imputation counterpart LOCF whereas N-BOCF [1.52 (1.89) p=0.424] provides a lower estimate than that obtained with BOCF. The allowance for appropriate uncertainty in the NFD strategies increases standard errors and p-values. Unlike BOCF, the treatment effect under N-BOCF is not statistically significant.

Because of the greater drop-out rate in the Test group, the penalty $\Delta = -5$ applied to drop-out subjects decreases the treatment effect under N-DO5 [3.21 (1.78) p=0.073] by 0.71 relative to N-LOCF which involves $\Delta = 0$. The inferences produced by N-AE5 [3.13 (1.74) p=0.0073] are very close to that of N-DO5 although the penalty is zeroed if subjects drop out for other reasons than AE. Indeed, these subjects are similarly distributed in

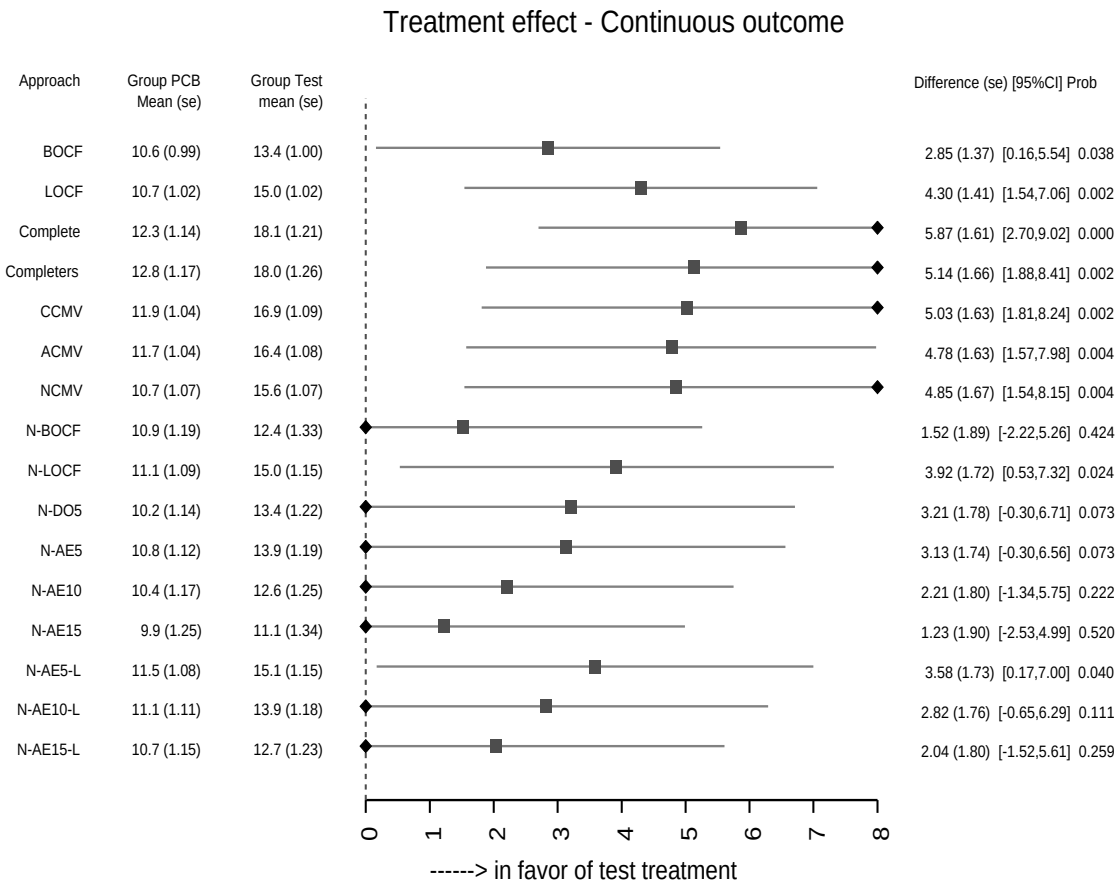

FIGURE 2 – Treatment-effect estimates obtained with the continuous outcome at visit 4.

the two groups (7.4% for PCB and 7.0% for Test based on Table 1). Consequently, the means by treatment group increase similarly. As a global rule in the N-AEx strategy class, the increase of penalty value makes the treatment-effect estimate decreasing, as observed under N-AE10 [2.21 (1.80) p=0.222] and N-AE15 [1.23 (1.90) p=0.520], since twice more subjects dropped out for AE in the Test group compared to PCB.

In the N-AEx-L strategy class, the removal of the last visit if drop-out is caused by an AE increases the means by treatment group relative to the corresponding N-AEx counterparts. This impact was expected since Figure 1 exhibits mean outcome profiles in the drop-out patterns, which decline at the last visit. The removal of the last visit also increases the separation between treatment groups under N-AE5-L [3.58 (1.73) p=0.040], N-AE10-L [2.82 (1.76) p=0.111], and N-AE15-L [2.04 (1.80) p=0.259] relative to the respective N-AEx counterparts.

The search of a penalty value applicable to drop-out subjects at which conclusions change from being favorable for Test to being unfavorable was suggested by authors in different contexts [14]. In our case study, we have conducted such investigation under the

strategies N-AEx and N-AEx-L. The limit values which zero the treatment effect (i.e., no difference between treatment groups) are $\Delta = -21$ under N-AEx and $\Delta = -30$ under N-AEx-L. In the same vein, the limit values for statistical non-significance (i.e., p=0.05) are $\Delta = -3$ under N-AEx and $\Delta = -6$ under N-AEx-L. These limit values can be interpreted as the maximum price to pay for drop-out so that Test remains beneficial to subjects in the trial. If the penalty is below these limit values, the risks for subjects do not balance the benefits of Test. Of note, several authors have suggested to introduce conservatism by deliberately penalizing subjects of the Test group, such as in the δ-method (see, e.g., [15]). In our analysis, penalties are applied to drop-out subjects whatever their treatment groups. Ultimately, this allows a sensible representation of true treatment effects, undermining a main trial objective [16].

5.2 Binary outcome analysis

Figure 3 displays treatment-effect estimates obtained with the binary outcome at visit 4 under the drop-out strategies specified in Table 3. We provide odds-ratios with their 95% confidence limits and p-values. Results by treatment group are described with frequencies and percentages of clinical responders based on the summary binary values.

The inferences obtained after BOCF [1.37 (0.15) p=0.029] and LOCF [1.53 (0.14) p=0.003] imputations confirm the major trends observed with the continuous outcome. The greater odds-ratio obtained after LOCF imputation relative to BOCF is caused by 16 additional clinical responders for Test against 5 for PCB. Despite the change of outcome scale, the p-values obtained after BOCF and LOCF imputations, as well as in subjects with complete outcome [1.74 (0.16) p<0.001] and in completers [1.63 (0.16) p=0.002], are surprisingly identical to those obtained with the continuous outcome.

The per-protocol analyses now exhibit a moderate divergence between CCMV [1.63 (0.18) p=0.006], ACMV [1.58 (0.18) p=0.011], and NCMV [1.46 (0.20) p=0.061]. Particularly, the treatment effect under NCMV is not statistically significant. Among de facto analyses, the treatment effect estimated under N-LOCF [1.52 (0.19) p=0.027] is still greater than that estimated under N-BOCF [1.39 (0.21) p=0.117]. The difference is caused by 13 additional clinical responders for Test against 3 for PCB. As with the continuous outcome, N-DO5 [1.48 (0.21) p=0.059] and N-AE5 [1.47 (0.19) p=0.047] provide similar treatment-effect estimates although only N-AE5 is statistically significant. Of note, both strategies produce treatment effects on the edge of the statistical significance.

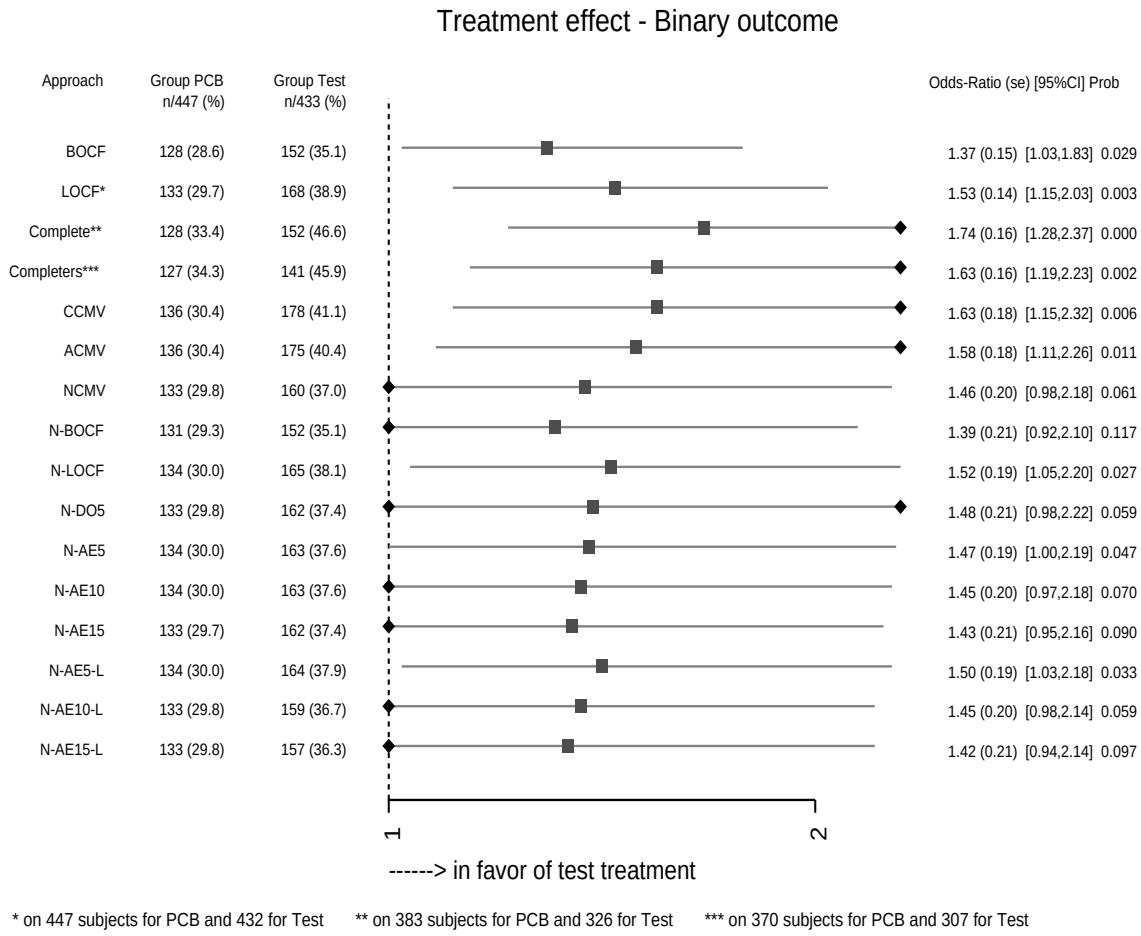

FIGURE 3 – Treatment-effect estimates obtained with the binary outcome at visit 4.

Whereas treatment-effect inferences obtained so far with the binary outcome remain coherent with the continuous case, the N-AEx strategies exhibit a divergence. The increase of the Δ value only implies a slight decrease of the odds-ratio as observed under N-AE10 [1.45 (0.20) p=0.070] and N-AE15 [1.43 (0.21) p=0.090] relative to N-AE5. However, this impact combined with a slight increase of standard error makes the p-values obtained under N-AE10 and N-AE15 non statistically significant, unlike under N-AE5. The slightness of the impact on treatment effect is caused by a floor effect on the numbers of clinical responders. From N-AE5 to N-AE15, these numbers vary from 134 to 133 for PCB and from 163 to 162 for Test. A second consequence of the floor effect is the lack of divergences between N-AE5-L [1.50 (0.19) p=0.033], N-AE10-L [1.45 (0.20) p=0.059], and N-AE15-L [1.42 (0.21) p=0.097], and with their respective N-AE counterparts. A third consequence is that the treatment-effect magnitude remains in favor of Test (i.e., odds-ratio>1) under the N-AEx and N-AEx-L strategies, whatever the penalty value. As for the rest, the limit values for statistical non-significance are $\Delta = -6$ under N-AE and $\Delta = -8$ under N-AEx-L.

5.3 Subgroup analysis

Table 4 provides the treatment-effect estimates obtained with the continuous and the binary outcomes at visit 4 in the subgroups of subjects who dropped out for AE (DOAE) and other reasons than AE (DOnotAE) under the drop-out strategies specified in Table 3.

TABLE 4 – Treatment-effect estimates at visit 4 in the subgroups of subjects who dropped out for AE and other reasons than AE († indicates a treatment effect in favor of PCB).

Subgroups	Drop-out	Mean		Diff (se)	p	Freq (%)		OR (se)	p
	strategies	PCB	Test			PCB	Test		
DOAE	BOCF	-.10	2.2	2.3 (1.4)	0.098	0/44 (0.0)	4/96 (4.2)	5.23 (1.37)	0.228
	LOCF	5.0	7.6	2.6 (2.9)	0.372	5/44 (11.4)	18/96 (18.7)	1.9 (0.56)	0.257
	CCMV	9.1	13.2	4.0 (2.7)	0.130	7/44 (15.9)	26/96 (27.1)	1.74 (0.47)	0.234
	ACMV	7.9	11.4	3.5 (2.5)	0.172	7/44 (15.9)	24/96 (25.0)	1.57 (0.47)	0.335
	NCMV	.45	8.5	8.1 (2.1)	<.001	5/44 (11.4)	10/96 (10.4)	0.85 (0.56)	0.768 †
	N-BOCF	-1.4	-2.4	-1.0 (2.8)	0.722	3/44 (6.8)	4/96 (4.2)	0.59 (0.72)	0.460 †
	N-LOCF	3.4	6.2	2.8 (2.7)	0.308	6/44 (13.6)	15/96 (15.6)	1.10 (0.51)	0.854
	N-DO5	-1.9	.36	2.3 (2.8)	0.418	5/44 (11.4)	12/96 (12.5)	1.02 (0.55)	0.966
	N-AE5	-.64	.62	1.3 (2.9)	0.667	5/44 (11.4)	12/96 (12.5)	1.02 (0.55)	0.966
	N-AE10	-4.9	-5.3	-.38 (3.2)	0.907 †	5/44 (11.4)	11/96 (11.5)	0.93 (0.55)	0.903 †
	N-AE15	-9.6	-12.0	-2.1 (3.7)	0.576 †	4/44 (9.1)	11/96 (11.5)	1.19 (0.59)	0.770
	N-AE5-L	5.7	5.2	-.47 (2.1)	0.822 †	5/44 (11.4)	13/96 (13.5)	1.07 (0.54)	0.902
	N-AE10-L	1.8	.07	-1.7 (2.2)	0.431 †	4/44 (9.1)	8/96 (8.3)	0.83 (0.61)	0.755 †
	N-AE15-L	-2.2	-5.1	-2.9 (2.3)	0.206 †	4/44 (9.1)	6/96 (6.3)	0.63 (0.63)	0.467 †
DOnotAE	BOCF	-1.0	6.7	7.7 (2.7)	0.006	1/33 (3.0)	7/30 (23.3)	9.0 (0.95)	0.021
	LOCF	-5.7	12.3	18.1 (3.6)	<.001	1/33 (3.0)	9/30 (30.0)	17.7 (1.02)	0.005
	CCMV	-2.2	15.8	18.0 (2.6)	<.001	2/33 (6.0)	11/30 (36.7)	8.71 (0.78)	0.006
	ACMV	-3.3	14.8	18.1 (3.3)	<.001	2/33 (6.0)	11/30 (36.7)	8.71 (0.78)	0.006
	NCMV	-5.8	12.0	17.8 (2.3)	<.001	1/33 (3.0)	9/30 (30.0)	12.7 (0.96)	0.008
	N-BOCF	-.89	.54	1.43 (4.4)	0.747	1/33 (3.0)	7/30 (23.3)	7.10 (0.92)	0.034
	N-LOCF	-5.1	11.1	16.2 (3.5)	<.001	1/33 (3.0)	9/30 (30.0)	12.7 (0.96)	0.008
	N-DO5	-9.5	7.25	16.7 (3.7)	<.001	1/33 (3.0)	9/30 (30.0)	12.7 (0.96)	0.008
	N-AEx	-3.3	14.3	17.7 (3.5)	<.001	2/33 (6.0)	10/30 (33.3)	7.62 (0.79)	0.010
	N-AEx-L	-2.8	14.6	17.4 (3.5)	<.001	2/33 (6.0)	10/30 (33.3)	7.62 (0.79)	0.010

An overall look at results in the subgroup DOAE reveals that only NCMV on the continuous scale yields a statistically significant effect of Test. In this subgroup, the other drop-out strategies produce treatment effects that are either moderate or counterbalanced by penalties for drop-out in the N-AEx and N-AEx-L strategies. The subgroup DOnotAE exhibits opposite results since all the drop-out strategies, except N-BOCF on the continuous scale, yield statistically significant effects of Test. In this subgroup, the continuous outcome means by treatment group are all positive for Test and negative for PCB, whereas the rates

of clinical responders are always greater for Test. These results indicate that there is no evidence of treatment effect in the subjects who dropped out for AE whereas the subjects who dropped out for other reasons unambiguously benefited from Test.

Further attention is needed to interpret the results after BOCF imputation, which exhibit non-null continuous outcome means by treatment group and some clinical responders although BOCF assigns a 0 value and a non-clinical response if visit 4 is missing. In fact, all subjects in Table 4 dropped out but the subjects of pattern 4 dropped out after visit 3 and have their drop-out visit repositioned to visit 4 for analysis. So, the non-null continuous outcome values and the clinical responses come from this pattern.

Results obtained under NCMV in the subgroup DOAE also deserve explanations since the effect of Test is statistically significant with the continuous outcome [8.1 (2.1) $p<.001$] whereas analysis on the binary scale yields a slight advantage in favor of PCB [OR=0.85 (0.56) p=0.768]. Relative to CCMV, the means by treatment group decrease by -8.6 for PCB and -4.7 for Test. Such impact was expected from Figure 1, which exhibits a marked divergence between the drop-out patterns 1–4 used in NCMV and pattern 5 of completers used in CCMV. In parallel, the numbers of clinical responders decrease by $-2/44$ (-4.5%) for PCB and $-15/96$ (-15.6%) for Test. This contradictory result is only caused by a scale effect. The substantial gain for PCB with the continuous outcome corresponds to a few additional clinical responders because the threshold of 30% improvement is not reached. These results in the subgroup DOAE explain the slight divergence observed in analysis under NCMV in the whole sample. The treatment effect is statistically significant on the continuous scale and not on the binary scale whereas results obtained with both outcomes are perfectly in line under CCMV and ACMV.

Under the N-AEx and N-AEx-L strategies, the impact of the penalty applied to subjects who dropped out for AE is unsurprisingly quasi-linear on the continuous scale in the subgroup DOAE. This trend is not observed with the binary outcome because of the floor effect. In the subgroup DOnotAE, the penalty has no impact since subjects are imputed assuming MAR. Of note, the results obtained under ACMV and N-AEx correspond in the PCB group with 2/33 clinical responders and differ in the Test group with respectively 11/33 and 10/33 clinical responders. The reason is that ACMV imputation is based on the treatment group whereas N-AEx imputation is based on PCB only.

6 Discussion

Planning PMM-MI analysis requires to question and address all the aspects of implementation. Some characteristics of patterns like the pattern sizes (e.g., majority completer versus spread over many patterns) and the proximity between patterns (e.g., are there major jumps in mean and/or variance structure from one pattern to the neighboring one) should be cautiously examined as they provide a valuable information on the impact of drop-out strategies. In some cases, the use of the pattern of completers for imputation contributes to unduly over-estimate the treatment effect and imposes inappropriately low uncertainty. In our case study, imputation in de facto analyses is based on the drop-out patterns in the PCB group. These patterns have homogeneous sizes and their mean profiles exhibit stable trends over time. The consequence is that, outside the effect of the free distribution of present values in NFD, drop-out strategies have a moderate impact on treatment-effect estimate while some amount of uncertainty is allowed.

Other aspects of implementation are specific to NFD strategies. In practice, the present often corresponds to the first unobserved visit. We also define the distributions of present values relative to the drop-out visits in the closest neighboring patterns. Therefore, it is important to assess the impact of these options in sensitivity analyses. This is all the more necessary that the drop-out visits are repositioned to the next scheduled ones for analysis and we cannot rule out that the assessments at these visits combine other aspects than efficacy. In our case study, the exclusion of the drop-out visits when drop-out is caused by an AE from analysis implies a moderate increase of treatment-effect estimate. This impact was expected from the description of patterns which exhibits mean profiles which decline at the drop-out visits. Based on this clinical argument and the absence of strong divergence between analyses, we considered that drop-out visits do not bias efficacy evaluation and, on the contrary, provide valuable information for treatment-effect estimation.

7 Concluding remarks

As already stated and exemplified in [17], PMM-MI methods permit relevant and accessible assumptions on drop-out mechanisms. Among them, NFD offers an appealing setting to formalize beliefs in analyzing RCTs. Drop-out strategies can easily be tailored according to plausible clinical scenarii. We show that NFD implementation via PMM-MI methods offers powerful solutions to tackle major drawbacks of well known single-imputation concepts

such as BOCF and LOCF. We also show how the ratio of risks to benefits can be investigated with the introduction of penalties for drop-out. Last, we highlight the importance of a thorough investigation of the impact of drop-out strategies. To this end, MI can be used to analyze subgroups of interest.

Our implementation of PMM-MI methods allows valid inferences of the marginal treatment effect, which is compatible with analysis of confirmatory RCTs. Patterns are defined based on every visit and outcome values are fitted using MMRMs with full group-by-visit interaction. However, any candidate method implies the possibility to plan all the aspects of analysis before breaking the blind. In this manuscript, we highlight several points to consider that can be addressed during analysis preparation. We also show that software implementation is very feasible using a freely available existing program. Accordingly, there is now no reason not to consider NFD assumption and PMM-MI methods for primary and sensitivity analyses in confirmatory RCTs.

References

[1] Molenberghs G, Kenward MG. *Missing Data in Clinical Studies*, Wiley & Sons, 2007.

[2] Little RJA. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 1993 ; **88** : 125-134

[3] National Academy of Sciences. *Prevention and Treatment of Missing Data in Clinical Studies*. National Academies Press ; 2010.

[4] Bunouf P, Molenberghs G, Grouin J-M, Thijs H. A SAS program combining R functionalities to implement pattern-mixture models. *J. STAT. SOFTW.* 2015 ; **00** : 000-000.

[5] Bunouf P, Grouin J-M, Molenberghs G, Koch G. Analysing intensive longitudinal data after summarization at landmarks : an application to daily pain evaluation in a clinical trial. *J. R. Statist. Soc. A* 2012 ; **175**, Part 2 : 513-534.

[6] Bunouf P, Grouin J-M, Molenberghs G. Analysis of an incomplete binary outcome derived from frequently recorded longitudinal continuous data : application to daily pain evaluation. *Statist. Med.* 2012 ; **31** : 1554-1571.

[7] Carpenter JR, Kenward MG. Missing data in clinical trials – a practical guide. *Birmingham : National Health Service Co-ordinating Centre for Research Methodology* ; 2008.

[8] Molenberghs G, and Michiels B, Kenward MG, Diggle PJ. Missing data mechanisms and pattern-mixture Models. *Statistica Neerlandica* 1998 ; **52** : 153-161.

[9] Kenward M, Molenberghs G, and Thijs H. Pattern-mixture models with proper time dependence. *Biometrika* 2003 ; **90** : 53-71.

[10] Molenberghs G, Beunckens C, Sotto C, Kenward MG. Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of the Royal Statistical Society, Series B* 2008 ; **70** : 371–388.

[11] Mallinkrodt CH. *Preventing and Treating Missing Data in Longitudinal Clinical Trials : A Practical Guide*, New York, NY : Camridge University Press ; 2013.

[12] Rubin DB. *Multiple Imputation for nonresponse in Surveys*, Wiley & Sons ; 1987.

[13] Thijs H, Molenberghs G, Michiels B, Verbeke G, Curran D. Strategies to fit pattern-Mixture Models. *Biostatistics* 2002 ; **3** : 245-265.

[14] Yan X, Lee S, Li N. Missing data handling methods in medical device clinical trials. *Journal of Biopharmaceutical Statistics* 2009 ; **19** :1085-1098.

[15] Ratitch R, O'Kelly M, Tosiello R. Missing data in clinical trials : from clinical assumptions to statistical analysis using pattern mixture models. *Pharmacentical Statistics* 2013 ; **12** : 337-347.

[16] Gallo P, Chuang-Stein C. A note on missing data in non-inferiority trials. *Drug Information Journal* 2009 ; **43** : 469-474.

[17] Carpenter JR, Roger, J, and Kenward, MG. Analysis of longitudinal trials with missing data : A framework for relevant, accessible assumptions, and inference via multiple imputation. *Journal of Biopharmaceutical Statistics* 2013 ; **23** : 1352-1371.

Appendix : Implementation using software

This appendix provides the reader with additional information to implement marginal treatment-effect estimation in PMM-MI analysis. We base implementation on an existing SAS program which combines R functionalities, available at the [4] publication web-site. This program provides pooled treatment-effect estimates under NCMV, ACMV, NCMV, and NFMV using the three-stage Rubin's method. Minimal knowledges in SAS and R are needed to implement additional functionalities.

In the existing program, parameters estimation and pooled analysis are performed in SAS whereas MI is entirely carried out under R. A first limitation to implement

our case study is that MI is conducted under the same identifying restrictions in all subjects. The reason is that the program structure does not allow for individual imputation schemes as MI parameters, including the specification of identifying restriction, are entered in the SAS program and are then exported to R to perform MI in all subjects. To correct this, the loop for subjects in the MI procedure must be transfered from the R script into the SAS program. Of note, this transfert causes much more calls of R from SAS when running the program since the R script is now nested in the loop.

Next, the existing program does not impute missing covariates and first outcome values. This functionality must be inserted in the R script using a syntax similar to that used for imputation of the other missing values. Another modification in the R script concerns the numberings of patterns and last visits which match by default. In our case study, subjects with complete outcome profiles may pertain to patterns 4 and 5 although visit 4 is the last visit in both patterns. Information on how to decouple the patterns and the last visit numbers is available in Section 7.4 in [4].

To conclude, analysis of the marginal treatment effect requires to replace the syntax to perform the direct pooled analysis in the SAS program by syntax to perform pooled analyses per pattern and then average treatment effects accross patterns. Using the notation in Section 4.2.3, we suggest the following algorithm :

1. Use SAS Proc MIXED or SAS Proc LOGISTIC to fit outcome values at visit 4 per pattern by imputation and estimate $\boldsymbol{\delta}$ and standard errors ;
2. Use SAS Proc FREQ to estimate the pattern probabilities in $\boldsymbol{\Pi}$;
3. Use SAS Proc MIANALYZE to estimate $\boldsymbol{\delta}^*$ and W^* which contain pooled treatment effects and within-imputation standard errors per pattern ;
4. Use SAS Proc CORR to estimate the between-imputation covariance matrix B^* ;
5. Use R or SAS to calculate the pattern-averaged parameters $W^\dagger$, $B^\dagger$, $V^\dagger$ and produce inferential results.