Università degli Studi di Pisa
Dipartimento di Statistica e Matematica
Applicata all'Economia

Report n. 224

Missing at random in randomized experiments with imperfect compliance

Andrea Mercatanti

Pisa, Maggio 2002

- Stampato in Proprio – Il lavoro rientra nell'attività svolta ai fini del programma di ricerca denominato "Modelli statistici per l'analisi dei dati di durata", e finanziato con relativo assegno di ricerca.

Via Cosimo Ridolfi, 10 - 56124 PISA – Tel. Segr. Amm. 050 945231 Segr. Stud. 050 945317 Fax 050 945375
Cod. Fisc. 80003670504 - P. IVA 00286820501 - Web http://statmat.ec.unipi.it/ - E-mail: dipstat@ec.unipi.it

Missing at random in randomized experiments with imperfect compliance

Andrea Mercatanti
Dipartimento di Statistica e Matematica Applicata all'Economia
Università di Pisa

Abstract

The paper proposes a method for handling non-responses in a likelihood based analysis of a randomized experiment with imperfect compliance. The study of the complications from which a randomized experiment can suffer, is not only dedicated to a randomized trial analysis. Infact the template of a randomized experiment with imperfect compliance can be used for identify and estimate causal treatment effects also in observational studies. Not only, but some complications like the presence of non-responses in the treatment and/or in the assignment to treatment are more plausible in the context of an observational study than in a randomized trial.

My proposal is based on two conditions, namely in supposing: the non-responses mechanism in the category of "missing at random", and the "distinctness of parameter" condition satisfied. In other words this means that the non-responses mechanism is ignorable only after conditioning on the observable quantities. I present both theoretical and computational results.

1 [1]Introduction

Most of the methodologies proposed in the literature for evaluating causal effects are based on the concept of potential quantities (or counterfactuals).

[1]I would like to thank Andrea Ichino, Guido Imbens, Gilberto Ghilardi, Fabrizia Mealli, as well as seminar participants at the University of Pisa, European University Institute (Fiesole), 40-th Scientific Meeting of the Italian Statistical Society (Firenze), for useful comments and suggestions.

This means that, in order to define causal effects, a comparison between the outcome of a generic individual i assigned to a treatment, and the outcome that it would be observed if the same individual would not be assigned to that treatment is necessary[2] (Rubin, 1974; Holland and Rubin, 1983). Usually, this comparison is made by a difference of potential outcome averages under the different treatments, and over the whole population of units; this is the so called *Average Causal Effect,* A.T.E. (Holland, 1986). A randomized experiment with a binary treatment and imperfect compliance with respect to the assignments needs three basic variables in order to be defined: the outcome, Y_i; the treatment, D_i; and the assignment to treatment, Z_i. The potential outcomes respect to the treatments are defined as: $Y_i(D_i = d)$, with $d \in \{1, 0\}$, and respect to the assignments to treatment as: $Y_i(Z_i = z)$, with $z \in \{1, 0\}$. The two components in any couple of potential quantities, for example $\{Y_i(D_i = 0), Y_i(D_i = 1)\}$, are not contemporarily observable since one excludes necessarily the other, so the concept of potential quantities necessarily implies missing informations.

This paper propose a method for handling missing data in a likelihood-based analysis for a randomized experiment with imperfect compliance. In this context, missing data have not only the meaning of unobserved potential quantities (for example the outcomes of treated individuals if these individuals would not be subjected to the treatment) but have to be meant in a more general sense (like, for example, non-responses in the outcomes of treated individuals). The method proposed is based on the likelihood function introduced by Imbens and Rubin (1997).

The study of the complications from which a randomized experiment can suffer, is not only dedicated to a randomized trial analysis. The template of a randomized experiment with imperfect compliance can be adopted for the identification and estimation of treatment causal effects also in non-experimental situations. Angrist, Imbens and Rubin (1996) show under which set of assumptions, a regression analysis supported by the use instrumental variables identifies causal treatment effects in observational studies. The template is that of a randomized experiment with imperfect compliance in the sense that, in order to identify causal treatment effects, the particular instrumental variable adopted should have the role of a random assigment

[2]The reference to the idea of potential outcomes is not the only way to get a definition of causal effect. See Dawid (1997) for a discussion about this point and the proposal of a different approach based on the Decision Theory.

for which the treatment not necessarily comply. Moreover the complications of the presence of non-responses in the treatment and/or in the assignment to treatment, are more plausible in the context of an observational study that in a randomized trial. In experimental contexts, the treatments and the assignments to treatment have the peculiarities to be "created" by a researcher who randomly assigned a treatment to the units. So the probabilities to have non-response in these two variables are usually remote. Differently, in observational studies the data derived from other kind of sources (for example answers to questionnaires) for which the presence of non-responses is not unusual.

In Section 2 I briefly describe the structure of a likelihood based inference for causal effects in the presence of non-compliance. In Section 3 I present the likelihood function on which the inference can be based in presence of non-responses. In Section 4 I present a computational way for maximizing the likelihood function through the EM algorithm.

2 Theoretical framework

In settings of imperfect compliance with respect to an assigned binary treatment, and basing on the concept of potential quantities, the whole population can be subdivided in four sub-groups characterizing for different compliance behaviors. Units for which $Z_i = 1$ implies $D_i = 1$ and $Z_i = 0$ implies $D_i = 0$ (*compliers*) are induced to take the treatment by the assignment; they can be the most interesting units for example in the evaluation of an encouragement design. Units for which $Z_i = 1$ implies $D_i = 0$ and $Z_i = 0$ implies $D_i = 0$ are called *never-takers* because they never take the treatment, while units for which $Z_i = 1$ implies $D_i = 1$ and $Z_i = 0$ implies $D_i = 1$ are called *always-takers* because they always take the treatment. Finally the units for which $Z_i = 1$ implies $D_i = 0$ and $Z_i = 0$ implies $D_i = 1$ do exactly the opposite of the assignment and were called *defiers* by Balke and Pearl (1993). This four groups in which the population can be subdivided are shown in Table 1; each of them define a particular *compliance-status*.

Tab.1

		$Z_i = 0$	
		$D_i = 0$	$D_i = 1$
$Z_i = 1$	$Z_i = 0$	*never-taker*	*defier*
	$D_i = 1$	*complier*	*always-taker*

The starting point for handling the problem of missing data in a randomized experiment with imperfect compliance can be the likelihood function introduced by Imbens and Rubin (1997), and used in some empirical works, for example by Little and Yau (1997), or by Hirano, Imbens, Rubin, Zhou (1998). Let's indicate: with $\underline{Y_i}$ the couple of potential outcomes under the two different assignments to treatment, $\{Y_i(Z_i = 1), Y_i(Z_i = 0)\}$; and with $\underline{D_i}$ the couple of potential treatments received under the two different assignments to treatment, $\{D_i(Z_i = 1), D_i(Z_i = 0)\}$. Given the concept of potential quantities, only one value for each of the two couples $\underline{D_i}$ and $\underline{Y_i}$ is observable at individual level. These observable values will be indicate with $\underline{D_{obs,i}}$ and $\underline{Y_{obs,i}}$, differently from the non-observed values that will be indicate with $\underline{D_{mis,i}}$ and $\underline{Y_{mis,i}}$. Let's indicate $\underline{\mathbf{D}}$ to be the $n \times 2$ matrix (where n is the dimension of the sample), generated by the vertical concatenation of the n couples $\underline{D_i}$; $\underline{\mathbf{Y}}$ to be the $n \times 2$ matrix created by the vertical concatenation of the n couples $\underline{Y_i}$; $\underline{\underline{\mathbf{Y}}}$ to be the $n \times 4$ matrix created by the horizontal concatenation of the two matrices $\underline{\mathbf{D}}$ and $\underline{\mathbf{Y}}$, and finally $\underline{\underline{\mathbf{Y}}}_{\mathbf{obs}}$ and $\underline{\underline{\mathbf{Y}}}_{\mathbf{mis}}$ to be the observed and non-observed quantities in $\underline{\underline{\mathbf{Y}}}$.

Given the assumptions of *S.U.T.V.A. (Stable Unit Treatment Value Assumption)* by which the potential quantities for each unit are unrelated to the treatment status of other units, and "*Random assignment to treatment*" by which the probability to be assigned to treatment is the same for each individual [for these two conditions see: Angrist, Imbens and Rubin, (1996)] the likelihood function for the complete data $\underline{\underline{\mathbf{Y}}}$ can be written, as outlined in Imbens and Rubin (1997):

$$L(\boldsymbol{\theta}|\underline{\underline{\mathbf{Y}}}) \propto f(\underline{\underline{\mathbf{Y}}}; \boldsymbol{\theta}) = \Pi_{i=1}^{N} f(\underline{d_i}, \underline{y_i}; \boldsymbol{\theta}), \qquad (1)$$

where $f(\underline{d_i}, \underline{y_i}; \boldsymbol{\theta})$ is the probability or density function of $(\underline{d_i}, \underline{y_i}) = (\underline{d_{obs,i}}, \underline{d_{mis,i}}, \underline{y_{obs,i}}, \underline{y_{mis,i}})$ for a given value of the parametrical vector $\boldsymbol{\theta}$.

The likelihood function based on the observed quantities is proportional to the integral of (1) respect to the non-observed part of $\underline{\underline{\mathbf{Y}}}$, [Rubin (1978); Imbens and Rubin (1997)]:

$$L\left(\boldsymbol{\theta}|\underline{\underline{\mathbf{Y}}}_{\mathbf{obs}}\right) \propto f(\underline{\underline{\mathbf{Y}}}_{\mathbf{obs}}; \boldsymbol{\theta}) = \int\int \left[\prod_{i=1}^{N} f\left(\underline{d_i}, \underline{y_i}; \boldsymbol{\theta}\right)\right] d\underline{\underline{\mathbf{Y}}}_{\mathbf{mis}} = \prod_{i=1}^{N} \int\int f\left(\underline{d_i}, \underline{y_i}; \boldsymbol{\theta}\right) d\underline{y_{mis,i}}\, d\underline{d_{mis,i}} \,.$$

The resolution of the integral produces:

$$L\left(\boldsymbol{\theta}|\underline{\underline{\mathbf{Y_{obs}}}}\right) \propto f(\underline{\underline{\mathbf{Y_{obs}}}};\boldsymbol{\theta}) = \Pi_{i\in\varsigma(1,0)}(\omega_a g^i_{a0}+\omega_d g^i_{d0})\times\Pi_{i\in\varsigma(0,1)}(\omega_n g^i_{n1}+\omega_d g^i_{d1})\times$$

$$\times\Pi_{i\in\varsigma(1,1)}(\omega_a g^i_{a1}+\omega_c g^i_{c1})\times\Pi_{i\in\varsigma(0,0)}(\omega_n g^i_{n0}+\omega_c g^i_{c0}), \qquad (2)$$

where: the parametrical vector is

$$\boldsymbol{\theta} = \left(\omega_a,\omega_n,\omega_c,\omega_d,\boldsymbol{\eta}_{a0},\boldsymbol{\eta}_{a1},\boldsymbol{\eta}_{n0},\boldsymbol{\eta}_{n1},\boldsymbol{\eta}_{c0},\boldsymbol{\eta}_{c1},\boldsymbol{\eta}_{d0},\boldsymbol{\eta}_{d1}\right),$$

and $\varsigma(d,z)$ is the set of units having treatment $\underline{D_{obs,i}} = d$, and assigned to $\underline{Z_{obs,i}} = z$. Each of the four parameters ω_t represents the probabilities of an individual of being in the t group, where $t = c$ (*complier*), n (*never-taker*), a (*always-taker*), d (*defier*); the function $g^i_{tz} = g_{tz}(y_i)$ (dependent on the parametrical vector $\boldsymbol{\eta}_{tz}$) is the outcome distribution for an individual in the t group and assigned to the treatment z, with $z = 0,1$.

3 The likelihood function with non-responses

In this section the specification of the likelihood function for a randomized experiments with imperfect compliance and with the possibilty to have non-responses in the variables Y_i, D_i, Z_i, is introduced. The problem of non-responses is, in general, particularly important and its study is justified because of the difficulties in having complete datasets. As it was shown in the previous section non-responses are not the only sources of missing informations in randomized experiments with imperfect compliance; we saw, in fact, that missing data in the form of unobserved potential outcomes (that is the outcome and the treatment under the treatment not assigned: $\underline{Y_{mis,i}}$ and $\underline{D_{mis,i}}$) are always present in this context.

An analysis based on (2) can be performed only using datasets without non-responses, consequently the units having at least a non-response in the variables Y_i, D_i, or Z_i, should be necessarily dropped out of the analysis. But this procedure is not easily justified, because in this way it would be necessary to satisfy the assumption called "*missing completely at random*", by which the probability of non-response for every variable is the same for all the units. Indeed, only in this case deleting the units presenting at least a non-response in the variables Y_i, D_i, or Z_i, does not influence a likelihood-based analysis (Rubin, 1976; Little and Rubin, 1987; Schafer, 1997).

Less restrictive conditions can be imposed assuming: the *"missing at random" (MAR),* and the *"distinctness of parameter" (DOP)* conditions (Rubin, 1976; Little and Rubin, 1987; Schafer, 1997). In order to explain these two conditions, let $\mathbf{R}$ indicates a $n \times p$ matrix of indicators, namely a matrix of binary variables assuming value 1 if the corresponding element of a generic $n \times p$ dataset $\mathbf{Y}$ is observed (a response), and 0 otherwise (a non-response). Consequently the dataset $\mathbf{Y}$ can be decomposed in two parts: the observed part $\mathbf{Y}_{obs}$, and the unobserved part $\mathbf{Y}_{mis}$. Let's introduce a model for $\mathbf{R}$, that depends on both $\mathbf{Y}$ and a parameters vector $\boldsymbol{\varepsilon}$: $f(\mathbf{R}|\mathbf{Y};\boldsymbol{\varepsilon})$, and let's call it the *"non-response mechanism"*. The *MAR* assumption states that the probability or density function of $\mathbf{R}$ is independent on the unobserved part of the dataset, $\mathbf{Y}_{mis}$:

$$f(\mathbf{R}|\mathbf{Y};\boldsymbol{\varepsilon}) = f(\mathbf{R}|\mathbf{Y}_{obs},\mathbf{Y}_{mis};\boldsymbol{\varepsilon}) = f(\mathbf{R}|\mathbf{Y}_{obs};\boldsymbol{\varepsilon}). \tag{3}$$

Consequently, by introducing the assumption of independence between individuals behaviors, we can state that the probability of a certain value for $\mathbf{r}_i$ (i row of $\mathbf{R}$, $i = 1,..,n$) is the same for all the individuals having the same value of the row vector $\mathbf{y}_{i,obs}$ (observed part of the i row of $\mathbf{Y}$, $i = 1,..,n$). To understand the importance of (3) we should consider the further condition of *DOP* that with *MAR* allows us to simplify the inference. The *DOP* condition (Little e Rubin, 1987; Rubin, 1987; Schafer 1997), states that the parameters $\boldsymbol{\theta}$ of the model generating the data, $f(\mathbf{Y};\boldsymbol{\theta})$, and the parameters $\boldsymbol{\varepsilon}$ of the non-response mechanism have to be distinct. This means that from, a frequentist point of view, the joint parameter space of $(\boldsymbol{\theta},\boldsymbol{\varepsilon})$ must be the Cartesian cross-product of the marginal parameter spaces for $\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}$, and from a Bayesian viewpoint that the prior distribution of $\boldsymbol{\theta}$ must be independent on that of $\boldsymbol{\varepsilon}$. This is a reasonable condition in a lot of practical situations.

Given the two conditions of *MAR* and *DOP*, it is possible to show that in a likelihood-based analysis, the consideration of the non-response mechanism can be avoided. In fact, the joint distribution of the observed quantities (namely the observed part of the dataset, $\mathbf{Y}_{obs}$, and the $\mathbf{R}$ matrix) is:

$$f(\mathbf{R},\mathbf{Y}_{obs};\boldsymbol{\theta},\boldsymbol{\varepsilon}) = \int f(\mathbf{R},\mathbf{Y};\boldsymbol{\theta},\boldsymbol{\varepsilon})\, d\mathbf{Y}_{mis} = \int f(\mathbf{R}|\mathbf{Y};\boldsymbol{\varepsilon})\, f(\mathbf{Y};\boldsymbol{\theta})\, d\mathbf{Y}_{mis},$$

that under *MAR*, becomes:

$$f(\mathbf{R},\mathbf{Y}_{obs};\boldsymbol{\theta},\boldsymbol{\varepsilon}) = f(\mathbf{R}|\mathbf{Y}_{obs};\boldsymbol{\varepsilon}) \int f(\mathbf{Y};\boldsymbol{\theta})\, d\mathbf{Y}_{mis} = f(\mathbf{R}|\mathbf{Y}_{obs};\boldsymbol{\varepsilon}) \cdot f(\mathbf{Y}_{obs};\boldsymbol{\theta}),$$

Then under the further condition of *DOP*, a likelihood-based inference for $\boldsymbol{\theta}$ is not dependent on $\boldsymbol{\varepsilon}$ or on $f(\mathbf{R}|\mathbf{Y}_{obs};\boldsymbol{\varepsilon})$. It is then possible to avoid the consideration of the non-response mechanism, and base the inference only on the likelihood function based on the observed part of the dataset, $\mathbf{Y}_{obs}$:

$$L(\boldsymbol{\theta}|\mathbf{y}_{obs}) \propto f(\mathbf{Y}_{obs};\boldsymbol{\theta}) = \int f(\mathbf{Y};\boldsymbol{\theta})\, d\mathbf{Y}_{mis}.$$

In other words this means that the non-response mechanism is ignorable only after conditioning on the observable quantities. This assumption is different from the assumptions on which other recent methods proposed in the literature for dealing with the problem of missing data in experiments with non-compliance are based, see Frangakis and Rubin (1999); Baker (2000), Little and Yau (2001). In particular Frangakis and Rubin (1999) propose a method for handling the problem of non-responses on the outcome, working under the usual set of assumptions for the LATE identification, in absence of always-takers. Their procedure is based on two conditions: the *"compound exclusion restriction"* stating that assignment have no effect on both the potential outcomes and the response behaviors, and the *"latent ignorabilty"* of the non-response mechanism. The satisfaction of these two last assumptions is different from supposing the *MAR* and *DOP* conditions satisfied. The *"compound exclusion restriction"* and the *"latent ignorabilty"* make the missing data mechanism ignorable only after conditioning on the compliance status (never-takers, treated compliers, untreated compliers) that can be not always observed. The satisfaction of *MAR* and *DOP* conditions is not in general less restrictive than the satisfaction of the *"latent ignorabilty"* and the *"compound exclusion restriction"*,the choice depends essentially on the context of the study and has to be evaluated case by case. In any case the Frangakis and Rubin method does not consider the presence of missing data in the treatments and/or in the assigments to treatment.

In a randomized experiment with imperfect compliance and missing data the likelihood function based on $\mathbf{Y}_{obs}$, under *MAR*, *DOP*, and the usual *S.U.T.V.A.* and *"Random assignment to treatment"* conditions (for these two last assumptions, see Angrist, Imbens and Rubin; 1996) can be obtained by integrating (2) with respect to the non-responses:

$$L(\boldsymbol{\theta}|\mathbf{Y}_{obs}) \propto \int\int\int f(\underline{\mathbf{Y}_{\mathbf{obs}}};\boldsymbol{\theta})\, dz_{mis,i}\; dd_{mis,i}\; dy_{mis,i} =$$

$$= \prod_{i=1}^{N} \int\int\int\int\int f\left(\underline{d_i}, \underline{y_i}; \boldsymbol{\theta}\right) d\underline{y_{mis,i}}\ d\underline{d_{mis,i}}\ dz_{mis,i}\ dd_{mis,i}\ dy_{mis,i}\ , \qquad (4)$$

where: $\underline{y_{mis,i}}$ and $\underline{d_{mis,i}}$ are the unobserved parts of $\underline{y_i}$ and $\underline{d_i}$ (due to the concept of potential quantities), like in the (2); $z_{mis,i}$, $d_{mis,i}$ and $y_{mis,i}$ are the non-responses on the potentially observed part of $\underline{y_i}$ and $\underline{d_i}$. The parametrical vector $\boldsymbol{\theta}$ is:

$$\boldsymbol{\theta} = \left(\omega_a, \omega_n, \omega_c, \omega_d, \boldsymbol{\eta}_{a0}, \boldsymbol{\eta}_{a1}, \boldsymbol{\eta}_{n0}, \boldsymbol{\eta}_{n1}, \boldsymbol{\eta}_{c0}, \boldsymbol{\eta}_{c1}, \boldsymbol{\eta}_{d0}, \boldsymbol{\eta}_{d1}, \pi_z\right).$$

The only difference with respect to the vector $\boldsymbol{\theta}$ in (2), is the presence of the parameter representing the probability to be assigned to the treatment: $\pi_z = P(Z_i = 1)$. This parameter could be omitted in (2) because not influential on a likelihood-based analysis, given that the *"Random assignment to treatment"* condition was supposed to be satisfied, and so:

$$f(\underline{\underline{\mathbf{Y}}}, \mathbf{Z}) = f(\underline{\underline{\mathbf{Y}}})\, f(\mathbf{Z}),$$

where $\mathbf{Z}$ is the n component column vector of assignments to treatment. But in the (4), given the possibility to have non-responses in the assignment to treatment and in order to avoid wasting of informations, the probability to be assigned, π_z, has not to be drop out of the likelihood function.

The likelihood function (2), factors in four terms, and every term includes the likelihood for the units in the set $\varsigma(\underline{D_{obs,i}} = d,\ \underline{Z_{obs,i}} = z)$. The resolution of the integrals in the (4) produces a similar factorization, but in eighteen terms. Any of these terms refers to the sets obtainable by intersecting the four sets $\varsigma(\underline{D_{obs,i}} = d,\ \underline{Z_{obs,i}} = z)$:

$$\{\varsigma(0,0), \varsigma(0,1), \varsigma(1,1), \varsigma(1,0)\}$$

with the eight sets $\tau(\underline{ry_{obs,i}}, \underline{rd_{obs,i}}, \underline{rz_{obs,i}})$, where $\underline{ry_{obs,i}}$, $\underline{rd_{obs,i}}$ and $\underline{rz_{obs,i}}$ are indicators assuming value 0 in case of non-response and value 1 in case of response on $\underline{y_{obs,i}}$, $\underline{d_{obs,i}}$ and $\underline{z_{obs,i}}$ respectevely:

$$\{\tau(1,1,1), \tau(0,1,1), \tau(1,0,1), \tau(1,1,0), \tau(0,0,1), \tau(0,1,0), \tau(1,0,0), \tau(0,0,0)\}.$$

All the possible intersections between $\varsigma(\underline{D_{obs,i}} = d,\ \underline{Z_{obs,i}} = z)$ and $\tau(\underline{ry_{obs,i}}, \underline{rd_{obs,i}}, \underline{rz_{obs,i}})$ produce $(4 \times 8) = 32$ sets $\upsilon(\underline{D_{obs,i}} = d, \underline{Z_{obs,i}} = z, \underline{ry_{obs,i}}, \underline{rd_{obs,i}}, \underline{rz_{obs,i}})$.

But given that the units out of the set $\tau(1,1,1)$ or $\tau(0,1,1)$ have at least one non-response in the treatment or in the assignment to treatment, the number of possible sets reduce to eighteen. For example the units in the set $\tau(1,0,1)$ have a non-response in the treatment, then it is not possible to specify the four possible sets created by intersecting $\tau(1,0,1)$ with $\{\varsigma(0,0),\varsigma(0,1),\varsigma(1,1),\varsigma(1,0)\}$:

$$\{\upsilon(0,0,1,0,1),\upsilon(0,1,1,0,1),\upsilon(1,1,1,0,1),\upsilon(1,0,1,0,1)\}. \tag{5}$$

The question can be resolved by defining $\tau(.,z,1,0,1)=\tau(0,z,1,0,1)\cup\tau(1,z,1,0,1)$ be the units in set $\tau(1,0,1)$ and for which $\underline{Z_{obs,i}}=z$. Analogous arguments hold for the units in $\tau(1,1,0),\tau(0,0,1),\tau(0,1,0),\tau(1,0,0),\tau(0,0,0)$, and produce respectevely the sets: $\upsilon(d,.,1,1,0),\upsilon(.,z,0,0,1),\upsilon(d,.,0,1,0),\upsilon(.,.,1,0,0)$, and $\upsilon(.,.,0,0,0)$. Now, let's consider the units in the set $\tau(.,z,1,0,1)$; the probability or density function for a unit in this set is obtainable by integrating $f(\underline{d_i},\underline{y_i};\boldsymbol{\theta})$ with respect to the unobserved quantities:

$$\begin{cases} \int\int\int f(\underline{d_i},\underline{y_i};\boldsymbol{\theta})\, d\underline{y_{mis,i}} d\underline{d_{mis,i}} dd_{mis,i} & \text{if } z=0 \\ \int\int\int f(\underline{d_i},\underline{y_i};\boldsymbol{\theta})\, d\underline{y_{mis,i}} d\underline{d_{mis,i}} dd_{mis,i} & \text{if } z=1 \end{cases}$$

From (4), the integrations with respect to $\underline{y_{mis,i}}$ and $\underline{d_{mis,i}}$ produce:

$$\begin{cases} (1-\pi_z)\int\left\{I_{(d=1)}(\omega_a g^i_{a0}+\omega_d g^i_{d0})+I_{(d=0)}(\omega_c g^i_{c0}+\omega_n g^i_{n0})\right\} dd_{mis,i} & \text{if } z=0 \\ \pi_z\int\left\{I_{(d=1)}(\omega_a g^i_{a1}+\omega_c g^i_{c1})+I_{(d=0)}(\omega_n g^i_{n1}+\omega_d g^i_{d1})\right\} dd_{mis,i} & \text{if } z=1 \end{cases}$$

the further integration respect to $d_{mis,i}$ eliminates the indicators $I_{(d=1)}$ and $I_{(d=0)}$:

$$\begin{cases} (1-\pi_z)(\omega_a g^i_{a0}+\omega_d g^i_{d0}+\omega_c g^i_{c0}+\omega_n g^i_{n0}) & \text{if } z=0 \\ \pi_z(\omega_a g^i_{a1}+\omega_c g^i_{c1}+\omega_n g^i_{n1}+\omega_d g^i_{d1}) & \text{if } z=1 \end{cases}$$

In other words, the units in the set $\tau(.,z,1,0,1)$ are a mixture of compliers, defiers, always-takers and never-takers assigned to the treatment z.

Analogous arguments hold for units in $\upsilon(d,.,1,1,0),\upsilon(.,z,0,0,1),\upsilon(d,.,0,1,0),\upsilon(.,.,1,0,0)$, and $\upsilon(.,.,0,0,0)$. Then, the resolution of the multiple integrations in (4) produces the likelihood function based on the observed quantities that factors in eigth terms:

$$L(\boldsymbol{\theta}|\mathbf{Y}_{obs})=\Pi_{d=1,0}\Pi_{z=1,0}\Pi_{i\in\upsilon(d,z,1,1,1)} f_{d,z,1,1,1}\times\Pi_{d=1,0}\Pi_{z=1,0}\Pi_{i\in\upsilon(d,z,0,1,1)} f_{d,z,0,1,1}\times$$

$$\times\Pi_{z=1,0}\Pi_{i\in v(.,z,1,0,1)}\, f_{.,z,1,0,1}\times\Pi_{z=1,0}\Pi_{i\in v(.,z,0,0,1)}\, f_{d,z,0,0,1}\times\Pi_{d=1,0}\Pi_{i\in v(d,.,1,1,0)}\, f_{d,.,1,1,0}$$

$$\times\Pi_{d=1,0}\Pi_{i\in v(d,.,0,1,0)}\, f_{d,.,0,1,0}\times\Pi_{i\in\varsigma(.,.,1,0,0)}\, f_{.,.,1,0,0}\times\Pi_{i\in v(.,.,0,0,0)}\, f_{.,.,0,0,0}. \quad (6)$$

Table 2 presents the specification of the probability, or density, functions in the (6):

Tab.2

$f_{d,z,1,1,1} = I_{(d=1,z=0)}\,(1-\pi_z)\,(\omega_a g^i_{a0}+\omega_d g^i_{d0}) + I_{(d=0,z=1)}\pi_z(\omega_n g^i_{n1}+\omega_d g^i_{d1})+$ $+I_{(d=1,z=1)}\pi_z(\omega_a g^i_{a1}+\omega_c g^i_{c1}) + I_{(d=0,z=0)}\,(1-\pi_z)\,(\omega_n g^i_{n0}+\omega_c g^i_{c0})$
$f_{d,z,0,1,1} = I_{(d=1,z=0)}\,(1-\pi_z)\,(\omega_a+\omega_d) + I_{(d=0,z=1)}\pi_z(\omega_n+\omega_d)+$ $+I_{(d=1,z=1)}\pi_z(\omega_a+\omega_c) + I_{(d=0,z=0)}\,(1-\pi_z)\,(\omega_n+\omega_c)$
$f_{.,z,1,0,1} = I_{(z=0)}\,(1-\pi_z)\,(\omega_a g^i_{a0}+\omega_d g^i_{d0}+\omega_n g^i_{n0}+\omega_c g^i_{c0})+$ $+I_{(z=1)}\pi_z(\omega_a g^i_{a1}+\omega_d g^i_{d1}+\omega_n g^i_{n1}+\omega_c g^i_{c1})$
$f_{d,z,0,0,1} = I_{(z=0)}\,(1-\pi_z) + I_{(z=1)}\pi_z$
$f_{d,.,1,1,0} = I_{(d=0)}(\omega_d g^i_{d1}+\omega_n g^i_{n0}+\omega_n g^i_{n1}+\omega_c g^i_{c0}) + I_{(d=1)}(\omega_d g^i_{d0}+\omega_a g^i_{a0}+\omega_a g^i_{a1}+\omega_c g^i_{c1})$
$f_{d,.,0,1,0} = I_{(d=0)}(\omega_d+\omega_n+\omega_c) + I_{(d=1)}(\omega_d+\omega_a+\omega_c)$
$f_{.,.,1,0,0} = g^i_{d1}+g^i_{d0}+g^i_{n0}+g^i_{n1}+g^i_{a0}+g^i_{a1}+g^i_{c0}+g^i_{c1}$
$f_{.,.,0,0,0} = 1$

4 Computation of the maximum likelihood estimands by the EM algorithm

If the interest of the analysis is in maximizing the likelihood function based on the observed quantities, $L(\boldsymbol{\theta}|\mathbf{y}_{obs})$, the calculations can be performed by usual iterative methods. In particular, in order to exploit the particular structure of the likelihood function induced by the *MAR* and *DOP* assumptions, the maximization of the likelihood function $L(\boldsymbol{\theta}|\mathbf{y}_{obs})$ can be easly obtained using the EM algorithm (Dempster, Laird, Rubin 1977; Tanner 1996). This is a computational method that can be used as a part of a Bayesian or likelihood analysis and whose goal is locating the mode or modes of the posterior or likelihood function. It is a deterministic algorithm (it does not require the input of a stream of random numbers), and it works by running two steps iteratively. The first step, the E-step, works by imputing values to latent or

unobserved data, and the second-step, the M-step, works by maximizing the likelihood function based on the augmented dataset. The phylosopy is then in augmenting the dataset with latent data in order to produce a likelihood function easier to maximize respect to the original likelihood function. This particular way for locating the maximum of a likelihood function, is particulary adapte for making inference in the presence of missing data, without dropping out observations. In fact, in presence of missing data, the only way to write down a likelihood function whitout dropping observations with missing values, is by imputing values to the missing data. In these situations, the EM algorithm works (at the t iteration):

- in the E-step by calculating the expected value of the unobserved part of the dataset, $\mathbf{Y}_{mis}$, given the observed quantities, $\mathbf{Y}_{obs}$, and given the value of the parameter vector at the previous time $\boldsymbol{\theta}^{(t-1)}$:

$$\mathbf{Y}_{mis}^{(t)} = E\left(\mathbf{Y}_{mis}|\mathbf{Y}_{obs};\boldsymbol{\theta}^{(t-1)}\right);$$

- and in the M-step, by maximizing the likelihood of $\boldsymbol{\theta}$ given the complete data, that is the observed part of the dataset augmented by the unobserved quantities $\mathbf{Y}_{mis}^{(t)}$:

$$\boldsymbol{\theta}^{(t)} = \max_{\boldsymbol{\theta}} L\left(\boldsymbol{\theta}|\mathbf{Y}_{mis}^{(t)},\mathbf{Y}_{obs}\right).$$

The evaluation of the conditional expected value of $\mathbf{Y}_{mis}$ in the E-step is easy if MAR is true. This assumption in fact allows to define $f\left(\mathbf{Y}_{mis}|\mathbf{Y}_{obs};\boldsymbol{\theta}^{(t-1)}\right)$ without the consideration of the non-response mechanism, and then simply by:

$$f\left(\mathbf{Y}_{mis}|\mathbf{Y}_{obs};\boldsymbol{\theta}^{(t-1)}\right) = \frac{f\left(\mathbf{Y}_{mis},\mathbf{Y}_{obs};\boldsymbol{\theta}^{(t-1)}\right)}{f\left(\mathbf{Y}_{obs};\boldsymbol{\theta}^{(t-1)}\right)}.$$

Indeed under the (3), and for every value of $\boldsymbol{\varepsilon}$:

$$f\left(\mathbf{Y}_{mis}|\mathbf{Y}_{obs};\boldsymbol{\theta}^{(t-1)}\right) = \frac{f\left(\mathbf{R}|\mathbf{Y}_{obs};\boldsymbol{\varepsilon}\right)\ f\left(\mathbf{Y}_{mis},\mathbf{Y}_{obs};\boldsymbol{\theta}^{(t-1)}\right)}{\int f\left(\mathbf{R}|\mathbf{Y}_{obs};\boldsymbol{\varepsilon}\right)\ f\left(\mathbf{Y}_{mis},\mathbf{Y}_{obs};\boldsymbol{\theta}^{(t-1)}\right) d\mathbf{Y}_{mis}} =$$

$$= \frac{f\left(\mathbf{R}|\mathbf{Y}_{obs};\varepsilon\right)\ f\left(\mathbf{Y}_{mis},\mathbf{Y}_{obs};\boldsymbol{\theta}^{(t-1)}\right)}{f\left(\mathbf{R}|\mathbf{Y}_{obs};\varepsilon\right)\ \int\ f\left(\mathbf{Y}_{mis},\mathbf{Y}_{obs};\boldsymbol{\theta}^{(t-1)}\right)d\mathbf{Y}_{mis}} = \frac{f\left(\mathbf{Y}_{mis},\mathbf{Y}_{obs};\boldsymbol{\theta}^{(t-1)}\right)}{\int f\left(\mathbf{Y}_{mis},\mathbf{Y}_{obs};\boldsymbol{\theta}^{(t-1)}\right)d\mathbf{Y}_{mis}} =$$

$$= \frac{f\left(\mathbf{Y}_{mis},\mathbf{Y}_{obs};\boldsymbol{\theta}^{(t-1)}\right)}{f\left(\mathbf{Y}_{obs};\boldsymbol{\theta}^{(t-1)}\right)}.$$

Again in the M-step, by the *DOP* assumption, the maximization of $L\left(\boldsymbol{\theta}|\mathbf{Y}_{mis}^{(t)},\mathbf{Y}_{obs}\right)$ doesn't require to take into account the non-response mechanism. Indeed:

$$f\left(\mathbf{R},\mathbf{Y}_{mis}^{(t)},\mathbf{Y}_{obs};\boldsymbol{\theta},\varepsilon\right) = f\left(\mathbf{R}|\mathbf{Y}_{mis}^{(t)},\mathbf{Y}_{obs};\varepsilon\right)\cdot f\left(\mathbf{Y}_{mis}^{(t)},\mathbf{Y}_{obs};\boldsymbol{\theta}\right).$$

I now outline the general structure of the EM algorithm for a randomized experiment with impefect compliance by first considering the E-step. This step is dedicated to evaluate the expected value of the unobserved quantities given the observed quantities and a current value of the vector $\boldsymbol{\theta}$. The unobserved quantities can be: the missing outcomes and/or the compliance status. In particular the compliance status can be unobserved because of two reasons. One reasons is due to the concept of potential outcomes, infact the compliance status of a unit having treatment $D_i = d$ and assignment $Z_i = z$, is exactly determined only by knowing the value of the treatment D_i under the alternative assignment $Z_i = |1-z|$, and this can represents an unobserved counterfactual situation. For example the compliance status of an unit for which $D_i = 1$ and $Z_i = 1$ is unobserved, this unit could be a complier or an always-taker. The other reason is due to the presence of non-responses on the treatment and/or on the assignment. For example, in absence of defiers the compliance status of a unit for which $D_i = 0$ and $Z_i = 1$ is observed, the unit is surely a never-taker; but if the treatment is missing the compliance status is unobserved, infact that unit could be a complier, a never-taker, or an always-taker.

Tables 3 and 4 presents the inputs for calculating the expected values of the unobserved quantities given the observed quantities and a current value of the vector $\boldsymbol{\theta}$. Table 3 considers the compliance status, that can be represented by a four-component indicator $t = c$ (*complier*), n (*never-taker*), a (*always-taker*), d (*defier*). The conditional probability of subject i, in the set

$\tau(r\underline{y}_{obs,i}, r\underline{d}_{obs,i}, r\underline{z}_{obs,i})$, being type t given the observed data and a current value of the vector $\boldsymbol{\theta}$, is obtainable by a ratio of two quantities. The numerator of this ratio is the corresponding Table 3 entry and the denominator is the corresponding row total. The expected value of the compliance status for a subject i, is then represented by a four-component indicator of conditional probabilities.

Tab.3

	$\underline{d}_{obs,i}$	$\underline{z}_{obs,i}$	$t=c$	$t=n$	$t=a$	$t=d$
$\tau(1,1,1)$	0	0	$\omega_c g^i_{c0}$	$\omega_n g^i_{n0}$	0	0
	0	1	0	$\omega_n g^i_{n1}$	0	$\omega_d g^i_{d1}$
	1	0	0	0	$\omega_a g^i_{a0}$	$\omega_d g^i_{d0}$
	1	1	$\omega_c g^i_{c1}$	0	$\omega_a g^i_{a1}$	0
$\tau(0,1,1)$	0	0	ω_c	ω_n	0	0
	0	1	0	ω_n	0	ω_d
	1	0	0	0	ω_a	ω_d
	1	1	ω_c	0	ω_a	0
$\tau(1,0,1)$	.	0	$\omega_c g^i_{c0}$	$\omega_n g^i_{n0}$	$\omega_a g^i_{a0}$	$\omega_d g^i_{d0}$
	.	1	$\omega_c g^i_{c1}$	$\omega_n g^i_{n1}$	$\omega_a g^i_{a1}$	$\omega_d g^i_{d1}$
$\tau(0,0,1)$	.	0	ω_c	ω_n	ω_a	ω_d
	.	1	ω_c	ω_n	ω_a	ω_d
$\tau(1,1,0)$	0	.	$(1-\pi_z)\, g^i_{c0}$	$(1-\pi_z)\, g^i_{n0}+$ $+\pi_z g^i_{n1}$	0	$\pi_z g^i_{d1}$
	1	.	$\pi_z g^i_{c1}$	0	$(1-\pi_z)\, g^i_{a0}+$ $+\pi_z g^i_{a1}$	$(1-\pi_z)\, g^i_{d0}$
$\tau(0,1,0)$	0	.	$(1-\pi_z)\,\omega_c$	$(1-\pi_z)\,\omega_n$	0	$\pi_z \omega_d$
	1	.	$\pi_z \omega_c$	0	$\pi_z \omega_a$	$(1-\pi_z)\,\omega_d$
$\tau(1,0,0)$	.	.	$(1-\pi_z)\, g^i_{c0}+$ $+\pi_z g^i_{c1}$	$(1-\pi_z)\, g^i_{n0}+$ $+\pi_z g^i_{n1}$	$(1-\pi_z)\, g^i_{a0}+$ $+\pi_z g^i_{a1}$	$(1-\pi_z)\, g^i_{d0}+$ $+\pi_z g^i_{d1}$
$\tau(0,0,0)$	.	.	ω_c	ω_n	ω_a	ω_d

For what concern the computation of the conditional expected values for the unobserved outcomes, Table 4 presents the conditional distributions of the outcome given the observed data, for an individual i in the set

$$\tau(r\underline{y}_{obs,i} = 0, r\underline{d}_{obs,i}, r\underline{z}_{obs,i}) : f^i(y|\mathbf{Y}_{obs}).$$

Computation of the conditional expected value of the unobserved outcome is obtainable by:

$$E^i(Y|\mathbf{Y}_{obs}) = \int y\ f^i(y|\mathbf{Y}_{obs})\, dy.$$

Tab.4

	$d_{obs,i}$	$z_{obs,i}$	$f^i(y\|\mathbf{Y}_{obs})$
$\tau(0,1,1)$	0	0	$(g^i_{c0}\omega_c + g^i_{n0}\omega_n)/(\omega_c + \omega_n)$
	0	1	$(g^i_{n1}\omega_n + g^i_{d1}\omega_d)/(\omega_n + \omega_d)$
	1	0	$(g^i_{a0}\omega_a + g^i_{d0}\omega_d)/(\omega_a + \omega_d)$
	1	1	$(g^i_{c1}\omega_c + g^i_{a1}\omega_a)/(\omega_c + \omega_a)$
$\tau(0,0,1)$	.	0	$(g^i_{c0}\omega_c + g^i_{n0}\omega_n + g^i_{a0}\omega_a + g^i_{d0}\omega_d)$
	.	1	$(g^i_{c1}\omega_c + g^i_{n1}\omega_n + g^i_{a1}\omega_a + g^i_{d1}\omega_d)$
$\tau(0,1,0)$	0	.	$(1-\pi_z)\, g^i_{c0}\omega_c + [(1-\pi_z)\, g^i_{n0} + \pi_z g^i_{n1}]\omega_n + \pi_z g^i_{d1}\omega_d /$ $/[(1-\pi_z)\,(\omega_c + \omega_n) + \pi_z(\omega_n + \omega_d)]$
	1	.	$(1-\pi_z)\, g^i_{d0}\omega_d + [(1-\pi_z)\, g^i_{a0} + \pi_z g^i_{a1}]\omega_a + \pi_z g^i_{c1}\omega_c /$ $/[(1-\pi_z)\,(\omega_a + \omega_d) + \pi_z(\omega_c + \omega_a)]$
$\tau(0,0,0)$	.	.	$[(1-\pi_z)\, g^i_{c0} + \pi_z g^i_{c1}]\omega_c + [(1-\pi_z)\, g^i_{d0} + \pi_z g^i_{d1}]\omega_d +$ $+[(1-\pi_z)\, g^i_{n0} + \pi_z g^i_{n1}]\omega_n + [(1-\pi_z)\, g^i_{a0} + \pi_z g^i_{a1}]\omega_a$

The M-step is dedicated to the maximization of the likelihood function based on the augmented dataset, that is based on the dataset created by the union of observed and imputed data. Units with missing outcome, will have an imputed value for the outcome equal to $E^i(Y|\mathbf{Y}_{obs})$. The maximization can be performed by a weighted maximum likelihood procedure, where subjects are differently classified in the different compliance groups, t, with weights equal to the conditional probabilities of being in t, calculated in the E-step.

5 Conclusions

In this article I have proposed a likelihood based method for handling the problem of non-responses in a randomized experiment with imperfect compliance. This method relyes on the assumption that the non-response mechanism is ignorable only after conditioning on the observable quantities. This is different, but not in general less restrictive, from the assumptions on which other methods proposed recently in the literature are based, see Frangakis and Rubin (1999), Baker (2000), Little and Yau (2001). Respect to these my

proposal is adapte for dealing with the problem of missing data not only in the outcome but also in the treatment and in the assigment to treatment.

Other than theoretical results, a computational way for maximizing the likelihood function by exploiting the particular strutcure of the function induced by the satisfaction of the MAR and DOP was proposed.

References

[1] Angrist J.D., G.W. Imbens, D.B.Rubin (1996); *Identification of causal effect using instrumental variables*; J.A.S.A., Vol.91, No.434, 444-455.

[2] Baker S.G. (2000); *Analyzing a randomized cancer prevention trial with a missing binary outcome, an auxiliary variable, and all-or-none compliance;* J.A.S.A., Vol.95, No. 449, 43-50.

[3] Dawid A.P. (1997); *Causal inference without counterfactuals*; Dep. of Statistical Science, University College London, Research report N.188.

[4] Dempster A.P., N. Laird, D.B. Rubin (1977); *Maximum likelihood estimation from incomplete data using the EM algorithm*; Journal of the Royal Statistical Society, Ser.B, Vol.39, 1-38.

[5] Frangakis C.E., D.B. Rubin (1999); *Addressing complications of intention to treatment analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes*; Biometrika, Vol.86, 365-379.

[6] Hirano K., G.W. Imbens, D.B. Rubin, X. Zhou (1998); *Estimating the effect of an influenza vaccine in an encouragement design*; Working paper, Dep. of Economics, U.C.L.A.

[7] Holland P.W. (1986); *Statistics and casual inference*; J.A.S.A., Vol. 81, 945-960.

[8] Holland P.W., D.B. Rubin (1983); *On Lord's Paradox*; in Principals of modern psychological measurement, eds. H.Wainer, S.Messick, Hillsdale, NJ:Lawrence Erlbaum.

[9] Imbens G.W., J.D. Angrist (1994); *Identification and estimation of local average treatment effects*; Econometrica, Vol.62, No.2.

[10] Imbens G.W., D.B. Rubin (1997); *Bayesian inference for causal effects in randomized experiments with non-compliance*; The Annals of Statistics, Vol.25, No.1.

[11] Little R.J.A., D.B. Rubin (1987): *Statistical analysis with missing data*; J. Wiley and Sons, New York.

[12] Little R., L. Yau (1997); *Statistical techniques for analyzing data from prevention trials: treatment of no-shows using Rubin's causal model*; Working paper, Dep. of Biostatistics, University of Michigan.

[13] Little R.J., L.Yau (2001): *Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed;* J.A.S.A., Vol. 96, 1232-1244.

[14] Rubin D.B. (1974); *Estimating causal effects of treatments in randomized and nonrandomized studies*; Journal of educational psychology, Vol.66, 688-701.

[15] Rubin D.B. (1976); *Inference and missing data*; Biometrika, Vol.63, No.3.

[16] Rubin D.B. (1978); *Bayesian inference for causal effects: the role of randomization;* The Annals of Statistics, Vol.6, 34-58.

[17] Rubin D.B. (1987); *Multiple imputation for nonresponse in surveys*; J.Wiley and Sons, New York.

[18] Schafer J.L. (1997); *Analysis of incomplete multivariate data*; Chapman & Hall.

[19] Tanner M.A. (1996); *Tools for statistical inference*; Springer.