Supplementary materials to

A structural mean model to allow for non-compliance in a randomized trial comparing two active treatments

KRISTA FISCHER
MRC Biostatistics Unit, Cambridge, UK
Krista.Fischer@ut.ee

ELS GOETGHEBEUR
Ghent University, Dept. Applied Mathematics and Computer Science, Ghent, Belgium

BERNARD VRIJENS
Pharmionic Research Center, Visé, Belgium

IAN R. WHITE
MRC Biostatistics Unit, Cambridge, UK

A Interpretation and estimation of the treatment contrast when the two separate parameters are not identifiable

When the condition in section 4.1 is not fulfilled, i.e. when $\mathrm{E}(C_i^B|\mathbf{X}_i) = k\mathrm{E}(C_i^A|\mathbf{X}_i)$, the estimable contrast $\delta = \psi^A - k\psi^B$ deserves more discussion.

For the subpopulation $\mathbf{X}_i = \mathbf{x}$ with $\mathrm{E}(C_i^A|\mathbf{X}_i = \mathbf{x}) = \bar{c}_x$, $\delta\bar{c}_x$ reflects the expected difference in response to assigned treatment A versus B (i.e. the ITT effect at a given $\mathbf{x}$-level). As an example, suppose all women are fully compliant to A and on average 50% compliant to B, while all men are on average 50% compliant to A and 25% compliant to B. Adjusting for gender only, $k = 0.5$ and the effect of assigning A instead of B would be δ for women and 0.5δ for men.

For additional causal meaning one needs additional assumptions. For

instance, if $k = 1$ and we assume the model of form 2.3 (with γ's linear as in equation 4.1), δ can be interpreted as the difference in average treatment effects for the subgroup $C_i^A = C_i^B = 1$. Similarly, δc can be interpreted as the difference in the subgroup $C_i^A = C_i^B = c$.

Estimator 3.4 can estimate δ as long as the inverse $\{\mathbf{G}' P_X \mathbf{Z}\}^{-1}$ exists in the sample. Although the two distinct parameter estimates may have a large variance, the contrast $\hat{\delta}$ often has reasonable precision (as demonstrated in our simulation study in section B). The parameter k can be estimated as the ratio of average compliance summaries on the two arms.

Alternatively, define $H_i = Y_i - \delta R_i^A C_i^A$ with $\mathrm{E}(H_i|\mathbf{X}_i) = \mathrm{E}(Y_i^B|\mathbf{X}_i)$. The causal estimand then compares outcomes for individuals at given levels of treatment A with their average response if assigned to B, regardless of potential compliance with treatment B. The estimating equations for $\hat{\delta}$, derived from equation 2.5, now become equivalent to those for the SMM analysis of a placebo-controlled trial (Goetghebeur and Lapp, 1997; Fischer-Lapp and Goetghebeur, 1999), with B the reference treatment ("placebo"). The function q_{opt} in 2.5 is then directly estimable from a regression of Y_i^B on $\mathbf{X}_i$ and estimating g_{opt} involves regressing C_i^A on $\mathbf{X}_i$. The contrast δ is now estimated directly without need to estimate k beforehand (although the latter may be useful for interpretation purposes).

When instead A is taken as the reference, one estimates $\delta^* = \psi^B - \frac{1}{k}\psi^A$ with correspondingly reversed interpretation. The two choices of reference give 2 different (although asymptotically equivalent) estimators. The choice of reference is naturally driven by what is considered the standard treatment.

Finally, if one is particularly interested in the actual parameters ψ^A and ψ^B even though only δ is estimable, one may resort to a sensitivity analysis as the implied equality $\psi^B = (\psi^A - \delta)/k$ leads to possible values of ψ^B for a range of realistic values of ψ^A, given estimates of k and δ.

B Additional details of the SMM analysis of the trial comparing two anti-depressants

B.1 Selection of baseline covariates

Eight baseline predictors that were significant at the 0.01 level in multivariate regression models for either of the two compliance measures or the outcome (in the whole dataset or in either arm separately) were selected:

- ADCQ questions. Each question was formulated as a statement, where respondents had to indicate their level of agreement on a 4-point scale. The following statements were significantly associated with subsequent treatment adherence and/or the final Hamilton score of the patient:
  - "Antidepressants can change my personality"
  - "When I am more depressed, I can take more of the prescribed dose"
  - "Skipping certain days prevents the body from becoming resistant against or used to the anti-depressant"
  - "Skipping certain days prevents the body from becoming dependent on the anti-depressant"
  - "I am satisfied with the explanations my physician gave me about my depression"
- Other covariates:
  - Compliance at the end of the run-in period (binary indicator of taking a pill a day before Visit 2)
  - WHO well-being score: sum of scores 2 and 3 at Visit 2 (a score from 0 to 5, indicating level of agreement with the statements: "I feel calm and peaceful" and "I am full of energy")
  - Hamilton score at Visit 2 (baseline)

Hamilton Score at Visit 2 was the strongest predictor of the final Hamilton score, whereas run-in compliance was the strongest predictor of compliance summaries.

The resulting predictions of the two expected compliance summaries have a low empirical correlation between trial arms (Pearson's coefficient of correlation being -0.13, 95%CI -0.36 to 0.11, for $C^A(2,8)$ and $C^B(2,8)$), hence the identifiability conditions of a SMM with one parameter per arm are fulfilled. However, the two compliance variables $C(2,8)$ and $C^w(8)$ depend on $\mathbf{X}$ in a similar fashion, so we cannot expect to distinguish the effects of these two variables when estimating all 4 parameters.

B.2 Missing value handling

We used multiple imputation for the missing Hamilton scores, (Rubin, 1987) assuming the data are missing at random (MAR). With approximate normality of the Hamilton scores, we imputed missing values from their estimated conditional distribution, given observed Hamilton scores, baseline covariates and compliance summaries. This was repeated 30 times. Each SMM was fitted on the 30 imputed datasets and the resulting parameter estimates and estimated variance-covariance matrices were combined using Rubin's rules (Rubin, 1987), to produce estimates of the SMM parameters and their standard errors (R packages `norm` and `mitools` were used, `http://www.r-project.org/`).

The null hypothesis that all parameters in the model are 0 can in complete data sets be tested by a Wald test, assuming approximate normality of the parameter estimates. With multiple imputation of the missing data, the Wald test is replaced by an F-test, with degrees of freedom given by Reiter (2007).

C A simulation study

To study the properties of the SMM estimator, we generated datasets according to the following rules:

1. First, 3 "known" and 1 "unknown" independent baseline characteristics, X_1, X_2, X_3 and U were generated, each having a $U(-0.5; 0.5)$ distribution (uniform from -0.5 to 0.5, chosen to ensure that generated compliance measures and outcomes lay within realistic limits).

2. The two compliance summaries C^A and C^B, representing percentages of prescribed drug taken by a patient while randomized to treatment A or B, respectively, were generated to satisfy:

$$\begin{aligned} C^A &= c_x^A + 0.2U + 0.3\varepsilon_c^A \text{ with } c_x^A = 0.6 + 0.1X_1 + 0.4X_2 \\ C^B &= c_x^B + 0.2U + 0.3\varepsilon_c^B, \end{aligned}$$

   where ε_c^A and ε_c^B were independent $U(-0.5; 0.5)$ error terms and c_x^B was specified in five ways:

   (a) $c_x^B = 0.6 + 0.3X_1 + 0.2X_2$. This corresponds to differential prediction of C^A and C^B, with $\text{Cor}[\text{E}(C^A|\mathbf{X}), \text{E}(C^B|\mathbf{X})] \approx 0.75$.

   (b) $c_x^B = 0.6 + 0.2X_1 + 0.3X_2$. Now $\text{Cor}[\text{E}(C^A|\mathbf{X}), \text{E}(C^B|\mathbf{X})] \approx 0.95$.

   (c) $c_x^B = 0.15 + c_x^A$. Arm B patients have on average a 15% higher compliance, a difference which is not associated with observed baseline characteristics.

   (d) $c_x^B = 1.25c_x^A$. Now compliance on arm B tends to be 1.25 times that on arm A. Note that mean difference in compliance summaries is 0.15, as in (c).

   (e) $c_x^B = c_x^A$.

3. The potential treatment-free response Y^0 was generated to satisfy

$$Y^0 = 5 + 2X_2 + 3X_3 + 2U + \varepsilon^0,$$

   with ε^0 being an independent normal $N(0, 0.7)$ variate. Note that the baseline covariate X_1 predicts compliance but not outcome, whereas covariate X_3 predicts just the outcome.

4. The (potential) responses to treatment A and B were generated to satisfy

$$\begin{aligned} Y^A &= Y^0 + 7C^A + \varepsilon^A, \\ Y^B &= Y^0 + 9C^B + \varepsilon^B, \end{aligned}$$

with ε^A and ε^B being independent $N(0, 0.7)$ error terms (results were unchanged when ε^A and ε^B were correlated or even equal).

5. The randomized assignment $R^A = 1 - R^B$ was an independent binomial distribution $Bin(1, 0.5)$. The "observed" variables were defined as

$$Y = Y^A R^A + Y^B R^B; \;\; C^A_{obs} = C^A R^A; \;\; C^B_{obs} = C^B R^B.$$

6. For scenario (f) in Table 1, the data was generated exactly as under scenario (a), but now deleting the values of Y where $M = 1$, with M generated to satisfy: $logit[P(M = 1|\mathbf{X})] = 0.5X_3 - X_2 - 1.3$. ($logit(p) = p/(1-p)$). This led to about 30% of the values of Y being missing (as in the depression trial) and satisfying the MAR assumption.

Simulations were carried out with total sample sizes $n = 100, 400$ and 2000 and for all scenarios (a)–(f). For each simulated dataset under the scenarios (a)–(e), the parameters ψ^A and ψ^B were estimated together with their variance-covariance matrix, using the algorithm described in Section 3. Based on these estimates, also the difference $\psi^A - \psi^B$ and its standard error were found.

Under the scenario (f) we used a multiple imputation procedure as in the depression analysis (Section B.2). Specifically, we imputed each missing outcome 30 times by sampling from their estimated conditional distribution, given the three baseline covariates, randomization indicator and the compliance summary. The SMM was then fitted on each of the 30 imputed datasets and the resulting parameter estimates and estimated variance-covariance matrices were combined using Rubin's rules (Rubin, 1987).

The results are shown in Table 1. Since results for the second parameter ψ^B were similar to those for ψ^A (due to equal error variances in the data-generating models for compliance and outcomes on both arms) they are omitted.

As expected, the estimator behaves quite well when expected compliance measures, given baseline characteristics, are different with correlation less than 1 (scenarios a and b). When the correlation reaches 0.95, the mean squared error increases (as compared to the correlation of 0.75) and there is an indication of a small finite sample bias with small sample sizes. However, the magnitude of such bias is about 10 times smaller than the standard deviation of the estimates. The estimated standard error of the estimates approximates well the empirical standard deviation found from simulations. The resulting 95% confidence intervals (assuming asymptotic normality of the parameter estimates) have coverage probability close to 0.95.

Scenarios (c), (d) and (e) correspond to cases where the correlation between the two expected compliance summaries is 1. When the two summaries differ by an additive constant (scenario c), the estimate is still reasonably precise for sample size of 2000 and might also work for $n = 400$. More imprecise estimates are obtained when the two summaries differ by a multiplicative constant or are equal (scenarios d and e): in these cases one is unlikely to identify the two distinct parameters. Overly large standard errors lead to conservative confidence intervals.

The difference between causal parameters ψ^A and ψ^B is always estimated with much better precision than the two distinct parameters. However, the precision of the estimated difference is considerably decreased under scenarios (c) and (d), especially when the sample size is small.

Under scenario (e), when the two expected compliance summaries are equal, one can estimate the difference $\delta = \psi^A - \psi^B$ directly using the SMM methodology for placebo-controlled trials, as described in Section A. Simulation results of the direct method compare favorably with the estimates from the 2-stage procedure in Table 1. Estimated mean squared errors are very close, but the standard errors are less conservative, observed coverage of the 95% confidence interval being 93%, 95% and 95% for sample sizes 100, 400 and 2000, respectively.

The direct method enables estimation of the contrast $\delta = \psi^A - k\psi^B$ under scenario (d). With $\psi^A = 7$, $\psi^B = 9$ and $k = 1.25$, the true value of δ is -4.25. In our simulated datasets the precision of $\hat{\delta}$ was very close to

the precision under scenario (e). However, if the direct method is mistakenly used under scenario (c), which in practice may be hard to distinguish from (d), the estimated quantity would lack any meaningful interpretation.

In practice it would usually be safe to obtain the distinct point estimates first and then, based on their estimated variance-covariance matrix, the contrasts of interest. Only if there is sufficient certainty that the two expected compliance summaries are proportional (or equal), one could consider using the direct method to estimate the contrast δ.

Under scenario (f), missing data leads to somewhat increased MSE of the estimates using multiple imputation (as expected), but no biases are observed and coverage probabilities are close to the complete data scenario. As the sample size 100 is close to the sample size in the depression trial, this adds confidence in the validity of the data analysis results in Section 5.3, if the MAR assumption is valid.

We also compared the estimated SMM difference $\hat{\psi}^A - \hat{\psi}^B$ with the ITT difference $\hat{E}(Y^A - Y^B|\mathbf{X})$.

First note that the two approaches test different hypotheses: while the SMM estimate can be interpreted as the difference in treatment efficacies for full compliers, the ITT analysis estimates and tests the effect of treatment A assignment compared to B, capturing at the same time both compliance and efficacy differences (without enabling distinction between the two aspects). When the average compliance levels are the same for treatments A and B, as under scenarios (a), (b), (e) and (f), the ITT null hypothesis can be true only when the SMM null hypothesis is and vice versa. So it is of interest to compare the power of the two methods to detect a departure from the null.

The results indicate that under scenarios (a), (b) and (f), the SMM approach may lead to better power, compared to the ITT approach, to detect a significant treatment difference, with the power being comparable for the two approaches under scenario (e), where $E(C^A|\mathbf{X}) = E(C^B|\mathbf{X})$. Under scenarios (c) and (d) the ITT and SMM parameters and corresponding estimates of the difference between the two treatments have opposite directions, as Treatment B has higher average compliance level, but lower average efficacy for full compliers. Although in all these settings, ITT approach provides a

Table 1: *Simulation results on estimating the parameter ψ^A, the difference $\psi^A - \psi^B$ and the ITT difference $E(Y^A - Y^B|\mathbf{X})$, under 6 scenarios (a)–(f) and with three different sample sizes n (number of simulations is 1000 for each case): mean and standard error (SE) of the estimates over all simulated datasets, coverage of the asymptotic 95% Confidence Interval, power to detect a significant difference at 5% significance level. The true values are $\psi^A = 7$ and $\psi^B = 6$.*

parameter:		SMM: $\hat{\psi}^A$			SMM: $\hat{\psi}^A - \hat{\psi}^B$				ITT: E$(Y^A - Y^B|\mathbf{X})$		
scenario	n	mean	SE	coverage	mean	SE	coverage	power	mean	SE	power
(a)	100	7.16	3.15	0.96	1.01	0.40	0.96	0.68	0.60	0.31	0.50
	400	7.05	1.44	0.94	1.00	0.19	0.96	1.00	0.59	0.15	0.98
	2000	7.02	0.61	0.95	1.00	0.09	0.94	1.00	0.60	0.07	1.00
(b)	100	7.87	6.40	0.98	1.01	0.43	0.96	0.61	0.61	0.31	0.54
	400	7.17	2.92	0.97	1.00	0.19	0.97	1.00	0.60	0.15	0.98
	2000	7.00	1.24	0.96	1.00	0.08	0.96	1.00	0.60	0.07	1.00
(c)	100	8.72	8.37	0.99	1.33	1.71	0.98	0.20	-0.31	0.31	0.20
	400	7.59	5.27	0.98	1.12	1.03	0.98	0.32	-0.30	0.15	0.52
	2000	7.06	2.24	0.96	1.01	0.44	0.96	0.66	-0.30	0.07	1.00
(d)	100	9.89	11.13	0.98	1.58	2.27	0.98	0.14	-0.31	0.30	0.18
	400	10.20	10.36	0.98	1.65	2.07	0.98	0.15	-0.30	0.15	0.51
	2000	9.88	12.16	0.99	1.57	2.45	0.99	0.14	-0.30	0.06	1.00
(e)	100	9.92	9.98	0.98	1.01	0.47	0.98	0.52	0.61	0.30	0.52
	400	9.88	10.57	0.99	0.99	0.27	0.98	0.90	0.60	0.15	0.98
	2000	9.94	9.15	0.98	1.00	0.10	0.99	0.99	0.60	0.06	1.00
(f)	100	7.28	4.37	0.98	1.01	0.53	0.95	0.48	0.60	0.38	0.42
	400	7.09	1.91	0.96	0.99	0.24	0.96	0.98	0.59	0.18	0.92
	2000	7.06	0.79	0.96	0.99	0.11	0.96	1.00	0.60	0.08	1.00

valid comparison of assignment effects of treatments A and B, a SMM analysis, complemented by a comparison of average compliance levels, would give more insight into the nature of action of the two treatments.

D Extensions

D.1 Allowing for contamination

Contamination means that some patients randomized to treatment A actually get treatment B (instead of A or in addition to A) and/or vice versa. Let us assume now that treatment dosage summaries $C_i^{A(A)}$, $C_i^{B(A)}$, $C_i^{B(B)}$ and $C_i^{A(B)}$ are available, with $C_i^{J(K)}$ summarizing the amount of treatment J

received while being randomized to treatment K. One may assume now:

$$\mathrm{E}[Y_i^A - \psi^{A(A)} C_i^{A(A)} - \psi^{B(A)} C_i^{B(A)} | \mathbf{X}_i] = \mathrm{E}[Y_i^B - \psi^{B(B)} C_i^{B(B)} - \psi^{A(B)} C_i^{A(B)} | \mathbf{X}_i]. \tag{D.1}$$

This is again a special case of a SMM with multivariate compliance summaries. One often assumes the effect of received treatment to be independent of randomised group, i.e. $\psi^{A(B)} = \psi^{A(A)}$ and $\psi^{B(A)} = \psi^{B(B)}$. In this case the model can be rewritten as

$$\mathrm{E}[Y_i^A - \psi^{A(A)} D_i^A | \mathbf{X}_i] = \mathrm{E}[Y_i^B - \psi^{B(B)} D_i^B | \mathbf{X}_i],$$

with $D_i^A = C_i^{A(A)} - C_i^{A(B)}$ and $D_i^B = C_i^{B(B)} - C_i^{B(A)}$, interpreted as the excess dose of the treatment (A or B) received when assigned to the corresponding treatment arm. Although D_i^A and D_i^B are not directly observed, the randomization assumption allows for estimation of $\mathrm{E}(D_i^A | \mathbf{X}_i)$ and $\mathrm{E}(D_i^B | \mathbf{X}_i)$. To gain in causal interpretation, one might in D.1 additionally condition on the entire vector of potential compliance and contamination summaries $\mathbf{C}_i = (C_i^{A(A)}, C_i^{B(A)}, C_i^{B(B)}, C_i^{A(B)})$, as in equation 2.3.

This methodology can also be used for estimation of the causal effect of treatment in placebo-controlled trials with noncompliance and contamination. If treatment B represents a placebo without any dose effect, one can simplify (D.1) by taking $\psi^{B(B)} = \psi^{B(A)} = 0$. Allowing $\psi^{A(B)}$ and $\psi^{A(A)}$ to be different would make sense in trials where contamination occurs in a somewhat different way to compliance with the active treatment (e.g. later start of the therapy, different product with the same active compound used, contaminators being unblinded). If, however, $\psi^{A(B)} = \psi^{A(A)}$ is assumed, the model becomes equivalent to a SMM for placebo-controlled trials, described by Goetghebeur and Lapp (1997) and Fischer-Lapp and Goetghebeur (1999), with D_i^A used instead of the compliance summary.

D.2 Relaxing the exclusion restriction

The exclusion restriction assumes that for noncompliers the outcome does not depend on the randomized arm:

$$\mathrm{E}(Y_i^A - Y_i^0 | \mathbf{C}_i^A \equiv 0, \mathbf{X}_i) = \mathrm{E}(Y_i^B - Y_i^0 | \mathbf{C}_i^B \equiv 0, \mathbf{X}_i) = 0.$$

While this is a natural assumption in a double-blind trial, in other settings one may need to allow for a direct effect of being randomized to a certain arm, regardless of compliance. Examples include open-label trials, but also cases where $\mathbf{C}_i^A$ and $\mathbf{C}_i^B$ do not capture all aspects of actually received treatment so that even when $\mathbf{C}_i^A \equiv \mathbf{C}_i^B \equiv 0$, there are some components of therapy still received by patients that can differ between arms. It is known that wrongly assuming the exclusion restriction would seriously bias compliance-adjusted analysis (Hirano *and others*, 2000).

The absence of the exclusion restriction does not preclude the use of SMM methodology. Relaxing the restriction while assuming linear SMM's and conditioning on univariate compliance summaries C_i^A and C_i^B jointly (as in equation 2.3) would lead to:

$$\mathrm{E}[Y_i^A - \psi_0^A - \psi_1^A C_i^A | \mathbf{X}_i, C_i^A, C_i^B] = \mathrm{E}[Y_i^B - \psi_0^B - \psi_1^B C_i^B | \mathbf{X}_i, C_i^A, C_i^B]. \quad \text{(D.2)}$$

As the two distinct constants ψ_0^A and ψ_0^B would not be identifiable, one may be able to estimate the three parameters from

$$\mathrm{E}[Y_i^A - \psi_0 - \psi_1^A C_i^A | \mathbf{X}_i] = \mathrm{E}[Y_i^B - \psi_1^B C_i^B | \mathbf{X}_i], \quad \text{(D.3)}$$

where $\psi_0 = \psi_0^A - \psi_0^B$. Now ψ_0 is interpreted as the effect of treatment A assignment instead of B for the subset of patients with $C_i^A = C_i^B = 0$. If, however, one is not prepared to assume 2.3, but D.3 is derived from 2.1 by relaxing the exclusion restriction in both equations, such interpretation is not necessarily valid.

The same identifiability and estimation issues apply as for multivariate compliance summaries, with $\mathbf{C}_i^A = (1, C_i^A)$. So we can identify the parameters of interest, provided the matrix $\mathrm{E}(\mathbf{C}|\mathbf{X})$ with ith row $[1, \mathrm{E}(C_i^A|\mathbf{X}_i), \mathrm{E}(C_i^B|\mathbf{X}_i)]$ is of full column rank. Thus in particular the predicted compliance summaries $\mathrm{E}(C_i^A|\mathbf{X}_i)$ and $\mathrm{E}(C_i^B|\mathbf{X}_i)$ should have correlation < 1, which is only possible when there are at least 2 baseline covariates available that predict compliance.

Regardless of whether 2.1 or 2.3 is assumed, if ψ_0 proves to be non-zero, this would be an evidence for a differential dose-independent effect of assignment on the two arms. However, without an untreated reference group

one would not be able to tell whether such direct effect of the assignment is present on one or both arms. Similarly, the SMM methodology is unable to detect such effects if they are similar for both arms ($\psi_0 = 0$).

References

Fischer-Lapp, K. and Goetghebeur, E. (1999). Practical properties of some structural mean analyses of the effect of compliance in randomized trials. *Controlled Clinical Trials* **20**, 531–546.

Goetghebeur, E. and Lapp, K. (1997). The effect of treatment compliance in a placebo-controlled trial: Regression with unpaired data. *Journal of the Royal Statistical Society (C)* **46**, 351–364.

Hirano, K., Imbens, G. W., Rubin, D. B. and Zhou, X.-H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics* **1**, 69–88.

Reiter, J. (2007). Small-sample degrees of freedom for multi-component significance tests with multiple imputation for missing data. *Biometrika* **94**, 502–508.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley and Sons.