**Appendices for Handling Intercurrent Events Through Hypothetical Strategy in Delayed-Start Designs**

Yue Wei1\*, Wen Li2, Fang Liu3, G. Frank Liu3, Richard Entsuah3

1University of Pittsburgh, 130 De Soto Street, Pittsburgh, PA 15261 USA

2Vaccine Clinical Research & Development, Pfizer, Inc., Collegeville, PA, 19426, USA

3Merck & Co., Inc., 2000 Galloping Hill Road, Kenilworth, NJ 07033 USA

**Appendix I – Derivation of Jump to Reference Data for the Simulations**

Two versions of jump-to-reference method are considered:

* **J2R(V1):** the missing value is imputed from the control profile before 24 months. After that, a pseudo-control profile is generated using the average increment per 6 months from 6-24 months change from baseline (CFB) changes.
* **J2R(V2):** the missing value is imputed from the delayed-start cohort.

Note that, the two versions of J2R assume a same reference profile before 24 months. The following repeated-measurement model is used as the basis of the jump-to-reference method:

 (A.1)

where is the response of the *j*th patient at the time *t*; m denotes the months after randomization i.e. 6, 12, 18, 24, 30, 36, 42, 48; and is the indicator for the treatment groups and takes on the value of 1 for the early-start cohort and the value of 0 for the delayed-start cohort. In addition, we have:

 the mean CFB from 3 months to baseline (0 month) of the delayed-start cohort;

the mean CFB from 3 months to baseline of the early-start cohort, as compared to the delayed-start cohort;

: the difference between the mean CFB at *m* months and the mean CFB at 3 months of the delayed-start cohort;

 the difference between the mean CFB at *m* months and the mean CFB at 3 months of the early-start cohort, as compared to the delayed-start cohort;

the mean CFB change for every unit change in baseline at 3 months;

 difference between the mean CFB change for every unit change in baseline at *m* months and that at 3 months

For the hypothesis of the overall trial duration (i.e. from baseline to 48 months), the mean CFB of the early-start cohort is lower than that of the delayed-start cohort. Denote the mean CFB profile of the early-start cohort and the delayed-start cohort at time *t* as and .

For J2R(V1), we firstly construct the pseudo-control profile after 24 months:

where is the average of the baseline values.

The mean CFB for the pseudo-control profile at 48 months is:

Note that,

Next, we construct the test statistics. To simplify the derivation, within each simulation, the baseline values are centralized, so that the average value of baseline is 0. Therefore, we have

Let denote *non-missing* patients at months in the early-start cohort, and denote *non-missing* patients at months in delayed-start cohort. We have and , where and are the probability of being present at months for each treatment cohort. Assume and are independent.

With missing values in both cohorts imputed from the pseudo-control profile:

where

The parameter estimates and their variance-covariance matrix can be obtained from the repeated-measurement model (A.1).

For J2R(V2), with missing values in both cohorts imputed from the delayed-start cohort:

where

**Appendix II – Correlation Matrix Assumed in the Simulation**

In our simulation, the correlation matrix of longitudinal ADAS-Cog scores at baseline, 3-month, 6-month, 12-month, 18-month, 24-month, 30-month, 36-month, 42-month, and 48-month is in the following matrix.

**Appendix III – A Setting with More Missing Before 24 Months**

For the simulation setting described in Section 4.2, , are the fixed as constant across time points so that the generated proportions of missing data per visit are evenly distributed. For the simulation setting below, we adjust the setting of the cohort with 30% missing (via or ), such that 20% missing comes from the first 24 months and 10% missing comes is after 24 months.

* MAR1: missing at random with , and , are selected to have about 20% and 30% (i.e. about 20% missing from the first 24 months and 10% missing is after 24 months) missing data in the early-start cohort and delayed-start cohort at 48 months, respectively;
* MAR2: missing at random with , and , are selected to have about 30% (i.e. about 20% missing from the first 24 months and 10% missing is after 24 months) and 20% missing data in the early-start cohort and delayed-start cohort at 48 months, respectively;
* MNAR1: missing not at random with  , and , are selected to have about 20% and 30% (i.e. about 20% missing from the first 24 months and 10% missing is after 24 months) missing data in the early-start cohort and delayed-start cohort at 48 months, respectively;
* MNAR2: missing not at random with  , and ,  are selected to have about 30% (i.e. about 20% missing from the first 24 months and 10% missing is after 24 months) and 20% missing data in the early-start cohort and delayed-start cohort at 48 months, respectively.

The results are summarized in Table A.1 and Table A.2 below.

Table A.1. Rejection rates (in percentage) for hypothesis testing based on 2000 simulations

|  |  |  |  |
| --- | --- | --- | --- |
| Missing Setup | Methods for Handling IEs | Power (%) | Type I Error (%) |
| Scenario 1: DM Effect | Scenario 2: Symptomatic Effect | Scenario 3: No Effect |
| Reject H2 (80%)\* | Reject H2 (2.5%)\* | Reject H2 (2.5%)\* |
| MAR1 | MAL | 63.9 | 0.6 | 0.3 |
| J2R(V1) | 96.0 | 13.8 | 7.7 |
| J2R(V2) | 84.5 | 2.6 | 2.4 |
| MMRM | 84.4 | 2.6 | 2.4 |
| ETRANK (Median) | 77.0 | 8.2 | 2.8 |
| ETRANK (Max) | 86.5 | 40.8 | 5.0 |
| MAR2 | MAL | 96.8 | 12.3 | 10.8 |
| J2R(V1) | 54.2 | 0.9 | 1.0 |
| J2R(V2) | 84.2 | 3.2 | 2.6 |
| MMRM | 84.0 | 3.2 | 2.6 |
| ETRANK (Median) | 61.5 | 3.6 | 0.9 |
| ETRANK (Max) | 77.2 | 38.1 | 1.9 |
| MNAR1 | MAL | 56.7 | 0.2 | 0.2 |
| J2R(V1) | 90.8 | 3.4 | 1.3 |
| J2R(V2) | 74.0 | 0.7 | 0.7 |
| MMRM | 73.8 | 0.7 | 0.7 |
| ETRANK (Median) | 70.8 | 4.5 | 1.7 |
| ETRANK (Max) | 80.6 | 30.1 | 2.6 |
| MNAR2 | MAL | 99.1 | 23.4 | 22.6 |
| J2R(V1) | 82.7 | 3.4 | 5.5 |
| J2R(V2) | 97.4 | 11.9 | 11.1 |
| MMRM | 97.4 | 11.8 | 11.0 |
| ETRANK (Median) | 72.3 | 7.3 | 1.8 |
| ETRANK (Max) | 86.4 | 53.3 | 3.6 |
| \*Rejection rate based on a two-sample t-test (one-sided) when there is no missing data. |

Table A.2. Bias and coverage of 95% CI based on 2000 simulations

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Missing Setup | Methods for Handling IEs | Scenario 1: DM effect | Scenario 2: Symptomatic Effect | Scenario 3: No Effect |
| Bias | 95% CI Coverage (%) | Bias | 95% CI Coverage (%) | Bias | 95% CI Coverage (%) |
| MAR1 | MAL | 0.49 | 59.6 | 0.21 | 88.3 | 0.22 | 88.3 |
| J2R(V1) | 0.00 | 91.1 | -0.21 | 87.0 | -0.14 | 91.2 |
| J2R(V2) | 0.24 | 87.4 | 0.01 | 94.8 | 0.00 | 94.8 |
| MMRM | 0.00 | 94.9 | 0.00 | 94.9 | 0.00 | 94.8 |
| MAR2 | MAL | 0.08 | 93.9 | -0.21 | 88.4 | -0.21 | 88.6 |
| J2R(V1) | 0.60 | 39.8 | 0.24 | 83.1 | 0.12 | 91.2 |
| J2R(V2) | 0.35 | 74.0 | 0.00 | 94.6 | -0.01 | 94.9 |
| MMRM | 0.00 | 94.8 | 0.00 | 94.7 | 0.00 | 94.9 |
| MNAR1 | MAL | 0.60 | 38.3 | 0.31 | 79.7 | 0.31 | 80.1 |
| J2R(V1) | 0.23 | 83.3 | 0.02 | 92.7 | 0.11 | 92.3 |
| J2R(V2) | 0.43 | 66.9 | 0.20 | 89.0 | 0.20 | 89.1 |
| MMRM | 0.25 | 89.2 | 0.25 | 89.2 | 0.25 | 89.2 |
| MNAR2 | MAL | -0.03 | 94.4 | -0.32 | 77.9 | -0.32 | 77.3 |
| J2R(V1) | 0.41 | 58.7 | 0.05 | 90.5 | -0.05 | 92.2 |
| J2R(V2) | 0.17 | 87.8 | -0.18 | 88.6 | -0.18 | 88.5 |
| MMRM | -0.26 | 88.6 | -0.26 | 88.5 | -0.26 | 88.6 |
| Note that, the ETRANK methods are not presented in this table, since it is designed for hypothesis testing and no estimation is obtained through the ETRANK methods.Note that, the definition of for *u*=1,2,3 and *v*=1,2, refers to Section 4.1. |