Intention-to-Treat Analysis in Clinical Trials with Treatment Discontinuation and Missing Data

Roderick Little

University of Michigan
Outline

I. NRC report (National Research Council 2010; Little, D’Agostino et al. 2012)
   • Review key findings

II. Intention-to-Treat Analysis in Clinical Trials with Treatment Discontinuation and Missing Data (Little and Kang, 2014)
   • A causal perspective on treatment discontinuation
   • Alternative ITT measures
   • When are measures after discontinuation needed/useful?

NJ ASA chapter talk
Outline

I. NRC report (National Research Council 2010; Little, D’Agostino et al. 2012)
  • Review key findings

II. Intention-to-Treat Analysis in Clinical Trials with Treatment Discontinuation and Missing Data (Little and Kang, 2014)
  • A causal perspective on treatment discontinuation
  • Alternative ITT measures
  • When are measures after discontinuation needed/useful?
Panel’s Charge

• To prepare “a report with recommendations that would be useful for USFDA's development of guidance for clinical trials on appropriate study designs and follow-up methods to reduce missing data and appropriate statistical methods to address missing data for analysis of results.”

• Focus is on confirmatory randomized controlled trials of drugs, devices, and biologics
  – With some differences in emphasis, also pertinent for academic and NIH-funded trials, and more generally for various biostatistical investigations
Defining Missing Data

• Missing data are unrecorded values that, if recorded, would be meaningful for analysis. Outcomes that are not defined for some participants are not considered by the panel as missing data
  – Missed clinic visit: yes
  – QOL for individuals who die: no
  – Outcomes if individuals who discontinue drug if they had not discontinued: maybe
Key Take-Home Messages

• Missing data undermines randomization, the lynchpin of inferences in confirmatory trials
• Limiting missing data should be a major consideration when weighing alternative study designs
  – Analysis methods come with unverifiable assumptions, and limiting these assumptions is crucial
• Careful attention to avoiding missing data in trial conduct can greatly limit the scope of the problem
• Analysis methods need to be driven by plausible scientific assumptions
• Sensitivity analyses to assess robustness to alternative analysis models are needed
  – Lack of robust treatment effect from these analyses reinforces the need to limit missing data in trial design and conduct
Design issues

• The estimand: summary outcome measure of interest defined for the population under study
  – a key starting point for the design of a clinical trial
• Alternative choices of estimand may have important implications for trial design and implementation and on the rate of missingness
• Limiting missing data should be a consideration in choice of estimand
• Later consider one estimand that limits missing data – an on-treatment summary
Design to reduce the occurrence of missingness

1. Run-in periods before randomization to identify who can tolerate or respond to the study treatment
2. Flexible-dose (titration) studies
3. Restrict trial to target population for whom treatment is indicated
4. Reduce length of follow-up period
5. Allow rescue medication in the event of poor response
6. Define outcomes that can be ascertained in a high proportion of participants

Benefits of these options need to weighed against costs
Some Trial Conduct Strategies to Reduce Missing Data

• Limit participant burden
  – Reduce the number of visits and assessments
  – Allow a relatively large time window for each follow-up assessment

• Set maximal acceptable rates of missing data, and monitor during the trial

• Provide incentives for investigators and participants to stay in the trial, subject to ethical guidelines

• Continuous update of contact information

• Educate study staff on importance of limiting missing data
Analysis Methods: Principles

1. Missing data: missingness hides a true underlying value that is meaningful for analysis
2. Formulate the analysis for inference about an appropriate and well-defined causal estimand
3. Document, to the degree possible, the reasons for missing data, and incorporate in the analysis
   Some may be MAR, others not
4. Decide on a defensible primary set of assumptions about the missing data mechanism
5. Conduct a statistically valid analysis under the primary missing data assumptions
6. Assess the robustness of the treatment effect inferences by prespecified sensitivity analyses.
Some missing-data analysis methods

- Complete-case analysis
- Single imputation methods, including LOCF, BOCF
- Inverse probability-weighted methods, simple and augmented
- Likelihood – based methods
  - Maximum likelihood, Bayes, Multiple imputation

Preferred methods
Masked MNAR for double-blind trials

\[ Y = \text{outcome (with missing values)} \]
\[ S = \text{outcome-related side effects} \]
\[ X = \text{baseline covariates} \]
\[ T = \text{treatment indicators} \]
\[ R = 1 \text{ if } Y \text{ observed, 0 if } Y \text{ is missing} \]

\[ \text{MAR : Pr}(R = 1 \mid T, X, S, Y) = \text{Pr}(R = 1 \mid T, X, S) \]
\[ \text{MMNAR : Pr}(R = 1 \mid T, X, S, Y) = \text{Pr}(R = 1 \mid X, S, Y) \]

Kang and Little (2014 Clin Trials, forthcoming) discusses use of MMNAR assumption to identify models
Sensitivity Analysis

• Parameters of MNAR models cannot be reliably estimated – identifiability requires structural assumptions that are often questionable
• Varying certain parameters in a sensitivity analysis is the preferred approach
• In many (not all) situations, it would be reasonable to choose an MAR primary model, and look at MNAR models via a sensitivity analysis to assess plausible deviations from MAR
• Xiang Sun presents such a sensitivity analysis in her talk at this workshop
Outline

I. NRC report (National Research Council 2010; Little, D’Agostino et al. 2012)
   • Review key findings

II. Intention-to-Treat Analysis in Clinical Trials with Treatment Discontinuation and Missing Data (Little and Kang, 2014)
   • A causal perspective on treatment discontinuation
   • Alternative ITT measures
   • When are measures after discontinuation needed/useful?
Missing outcome data

• Individuals do not discontinue assigned treatment, but outcome values are missing
  • E.g. administrative censoring, missed clinical visits
• This is like a standard missing data problem
• Usual methods and concepts apply
  • MAR vs MNAR models
  • Likelihood or Augmented IPW methods
  • Sensitivity analysis
Treatment discontinuation

• Individuals discontinue assigned treatment for reasons potentially associated with that treatment
  – Side effects, ineffectiveness, advice of physician

• Outcome data may continue to be recorded
  – But might be outcomes under a treatment different from the treatment assigned

• Can view treatment discontinuation as a form of noncompliance, and invoke the causal literature on noncompliance
Complier-average causal effect (CACE)

- The CACE measures the treatment effect in the subgroup of principal compliers -- individuals who would comply under either treatment (e.g. Angrist, Imbens and Rubin 1996, Little, Long and Lin 2006)
- Special case of principal stratification (Frangakis and Rubin 2002)
- Need to “impute” compliance under treatment(s) not assigned
- Similarly we can define “Completer-average causal effect”, and apply instrumental variable methods
- My focus here on “Intention to treat” estimands, which apply to the whole population that is randomized
Example: Insulin Trial

- Eli Lilly study of a new oral anti-hyperglycemic medication for patients with type 2 diabetes (T2DM), compared to the standard therapies based on injected insulin Glargine.
- Randomized, parallel-group study of individuals experiencing lack of control of glucose levels, as measured by the HbA1c laboratory test (low is good).
- 3 treatment arms:
  - New (n = 221)
  - Glargine (IG, n = 213)
  - Combined (new and IG) arm (n = 115)
Lilly T2DM Study

- Measures of HbA1c were obtained at baseline and at weeks 1, 2, 4, 6, 8, 10, 12 and 24.
- Primary analysis in protocol is by ITT, change in HbA1c from baseline to 24 weeks, including all randomized patients with a baseline and at least one follow-up measure after baseline.
- Rescue medications were allowed for IG and New treatments, and HbA1C values while on rescue medications were included.
- Thus, “treatment” was in fact a “treatment protocol” which includes any effects of the rescue medications.
- No measures after subsequent treatment discontinuation: missing data imputed by last observation carried forward.
Example. Lilly T2DM Trial: Discontinuation and Missing Data in Study Groups by Reason

<table>
<thead>
<tr>
<th>Type</th>
<th>Combined</th>
<th>IG</th>
<th>New</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 completed</td>
<td>25</td>
<td>48</td>
<td>42</td>
</tr>
<tr>
<td>1 subject</td>
<td>10</td>
<td>12</td>
<td>33</td>
</tr>
<tr>
<td>2 physician</td>
<td>5</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>3 Protocol viol</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>4 Adverse event</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>5 Death</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6 Sponsor terminated</td>
<td>68</td>
<td>146</td>
<td>119</td>
</tr>
<tr>
<td>7 Lost follow-up</td>
<td>3</td>
<td>10</td>
<td>13</td>
</tr>
</tbody>
</table>

1-4: treatment discontinuation + missing data
6-7: missing data
5: not missing data, since dead people do not have Hba1c levels
Example. Lilly T2DM Trial: Discontinuation and Missing Data in Study Groups by Reason

<table>
<thead>
<tr>
<th>Type</th>
<th>Combi ned</th>
<th>IG no rescue</th>
<th>IG rescue</th>
<th>New no rescue</th>
<th>New rescue</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 completed</td>
<td>25</td>
<td>47</td>
<td>1</td>
<td>24</td>
<td>18</td>
</tr>
<tr>
<td>1 subject</td>
<td>10</td>
<td>12</td>
<td>0</td>
<td>30</td>
<td>3</td>
</tr>
<tr>
<td>2 physician</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>3 Protocol viol</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4 Adverse event</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>5 Death</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6 Sponsor terminated</td>
<td>68</td>
<td>137</td>
<td>9</td>
<td>94</td>
<td>25</td>
</tr>
<tr>
<td>7 Lost follow-up</td>
<td>3</td>
<td>10</td>
<td>0</td>
<td>13</td>
<td>0</td>
</tr>
</tbody>
</table>

Rescue for “IG” is minor, rescue for “New” is sizeable: should rescue values be included?
Treatment Discontinuation: ITT options

• If (as here) the primary estimand involves measures after discontinuation, methods need to in effect impute (explicitly or implicitly, e.g. by weighting) these measures

• Imputation method needs to be appropriate for the estimand, which involves assumptions about treatments after discontinuation

• The standard ITT approach measures the effect of randomization to treatment $Y(t_{\text{true}})$, where $t_{\text{true}}$ is set of actual treatments received after discontinuation
  – measured if follow-up measures are obtained after treatment
  – includes effects of any treatments between discontinuation and end of study (so these should be specified in the protocol)
  – But are we interested in including effects of these in the analysis?
Treatment Discontinuation: ITT options

We could conceive of other estimands, for the ITT population:

• Estimand under Assigned Treatment $Y(t_a)$: estimand if discontinuers had continued to take the assigned treatment

• Estimand under Control Treatment $Y(t_c)$: estimand if discontinuers had taken control (or reference) treatment
  – Latter two are counterfactual and need to be imputed (see Little and Yau 1996, Ratitch et al., 2013)
  – In particular, $Y(t_c)$ is estimated by methods that impute by jumping to reference after discontinuation. Whether these alternatives make sense varies according to context

• In most protocols, the method of imputation is described without stating the estimand
Imputation for Lilly T2DM Example

- The Lilly protocol specified LOCF imputation
- Which of the outcomes $Y(t_{true})$, $Y(t_a)$ or $Y(t_c)$, is being imputed by LOCF? Here, as in many trials, this is left unspecified and unclear. This clouds the nature of the treatment
- All of these estimands are problematic in the new treatment arm:
  - $Y(t_{true})$ and $Y(t_c)$, both include the effects of Glargine administered after drop-out, not the new treatment
  - $Y(t_a)$ is counter-factual, and there are no data to impute this since, for safety of participants, HbA1c levels need to be brought under control
- Suggest that there is a better alternative: on-treatment summaries
On-treatment summaries

- A measure of the effectiveness of a treatment that only uses information while individuals are on the assigned treatment. Examples:
  - Dropout as failure. Define a binary measure for success or failure, and treat discontinuers as failures
  - Area under curve (measured relative to baseline value) while on treatment. Dropout is penalized in that area is restricted to time while on assigned treatment.
  - Change from baseline to min(dropout, end of study). This estimate is the same change from baseline to end with LOCF for dropouts, but it avoids (unreasonable) assumption of no change after dropout
  - Impute zero change for dropouts. Estimate same as BOCF.
On-treatment summary for Lilly study

• The protocol specified LOCF imputation, equivalent to the on-treatment summary: change in HbA1c levels between baseline and min(dropout, end of study)

• A problem with this outcome is that it does not penalize dropout prior to end of trial (though it does reflect tendency for higher HbA1c values at time of discontinuation)

• On the other hand BOCF corresponds to imputing zero change, which voids any benefit of treatment before dropout.

• An alternative on-treatment summary measure is

• $P =$ proportion of 24 weeks where individual was on treatment and HbA1c levels were under control

• This penalizes early discontinuation appropriately

• Other more quantitative summaries might be developed
Lilly T2DM Study: administrative censoring

- Analysis of $P$ still requires imputation for cases incomplete because of administrative censoring. A (better?) alternative to LOCF is:
  - (a) Impute discontinuation indicator and (if 1) time of discontinuation for censored cases, given their history up to censoring time
  - (b) Impute $P$ given the discontinuation time imputed in (a)
- Repeat (a) and (b) and apply MI combining rules to propagate imputation uncertainty
- Since the administrative censoring is plausibly missing at random – unrelated to individual outcome measures – this is a defensible approach, and a sensitivity analysis for deviations from MAR seems unnecessary.
Three ITT ANCOVA Analyses of Diabetes Data

<table>
<thead>
<tr>
<th>Regressor</th>
<th>A. Outcome = Change from Baseline to Week 52</th>
<th>B. Outcome = Transformed Proportion of 52 weeks when on Treatment and HbA1c≤7.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Types of Missing Data Treated by LOCF Imputation</td>
<td>B1. Admin Censoring or Loss to Follow Up Treated by MI</td>
</tr>
<tr>
<td>Intercept</td>
<td>Estimate (95% CI)</td>
<td>P-Value</td>
</tr>
<tr>
<td></td>
<td>-0.50 (-0.83,-0.17)</td>
<td>0.003</td>
</tr>
<tr>
<td>“Inhaled”-“IG”</td>
<td>-0.18 (-0.38,0.03)</td>
<td>0.090</td>
</tr>
<tr>
<td>“Inhaled+IG”-“IG”</td>
<td>-0.38 (-0.63,0.14)</td>
<td>0.002</td>
</tr>
<tr>
<td>Taking Insulin Secretagogue</td>
<td>-0.08 (-0.28,-0.13)</td>
<td>0.847</td>
</tr>
<tr>
<td>Baseline (centered to 0)</td>
<td>-0.47 (-0.57,-0.37)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Country (DF=10)</td>
<td>---</td>
<td>0.585</td>
</tr>
</tbody>
</table>
Reasons to Collect Data After Discontinuation

• Monitor side effects of treatments prior to discontinuation

• Primary estimand involves data after discontinuation – otherwise these data require imputation (implicit or explicit)
Reasons to Collect Data After Discontinuation

• If the usual ITT estimand, involving $Y(t_{\text{true}})$, is appropriate, then it is better to follow up dropouts rather than impute values.

• For other ITT estimands, the value of measure is less clear…

• Follow up is not needed for an on-treatment summary measure (though follow-up might still be important to monitor side effects)
  – If sponsor and regulator can agree on a suitable on-treatment summary, the result might make both parties happy!
Summary

• Follow up after discontinuation allows side effects after discontinuation to be monitored

• It is also indicated if the measures after discontinuation are needed for the primary outcome and relevant to the treatment under study – which is more likely if treatments after discontinuation are specified in the protocol as part of a treatment regimen

• Defining an on-treatment summary to measure of treatment effect deserves more consideration –
  – does not require follow-up measures discontinuation
  – limits the amount of missing data.
References


Thanks to Eli Lilly for providing example data.