Introduction to Clinical Trials

Lecture 8: Data Analysis Approaches

Giorgos Bakoyannis, PhD
Associate Professor
Department of Biostatistics and Health Data Science
Indiana University Indianapolis

Notes by Prof. Constantin T. Yiannoutsos, City University of New York

Section 1

Introduction

Difficulties with data analyses

After design, implementation and monitoring of a clinical trial comes the time to analyze the collected data in order to address the questions that motivated the study in the first place. Invariably, the collected data do not entirely reflect what was anticipated and suffer from a number of "imperfections" that need to be addressed. These, can be coarsely assigned into three categories:

- Protocol non-adherence
- Incomplete or missing observations
- Methodological errors

The problem is how to reconcile these imperfections with the experimental approach to ensure reliable inference. Two approaches have been followed, which have significant downstream implications: A "pragmatic" and an "explanatory" approach.

Pragmatic versus *explanatory* trials
The pragmatic perspective

The "pragmatic" versus "explanatory" approach has a huge impact both on the design and analysis of a clinical trial.

The following is taken from

`http://www.collemergencymed.ac.uk/CEM/Research/technical/guide/prag`

Pragmatic versus *explanatory* trials
The pragmatic perspective

The pragmatic perspective addresses the question of whether a treatment works under real-life conditions and in terms of what is important to the patients.

While, the pragmatic approach might seem eminently logical, it may not be able to determine how or why an intervention works. For this reason, pragmatic studies are useful for making policy decisions of what services should be provided but may give limited insight into why interventions are efficacious or not (we will discuss *efficacy* later in this lecture).

Explanatory studies, on the other hand, address the question of whether a treatment works under ideal conditions or under very selective circumstances.

The explanatory perspective is more concerned with how and why an intervention works and is thus useful for understanding *efficacy* but may have limited value in informing policy decisions for providing a service to the general patient population or in a wide variety of circumstances not considered in the study (for a look at the difference between efficacy and effectiveness see discussion later in this lecture).

Impact of the pragmatic or explanatory point of view
Case study: Pre-operative chemotherapy in early NSCLC

Whether a study is pragmatic or explanatory has a number of important consequences in the implementation of the clinical trial.

Consider a trial where pre-operative chemotherapy (C) is evaluated in the treatment of early-stage non-small-cell lung cancer (NSCLC) for patients going to surgery (S). So, the two approaches are $C + S$ versus S.

Effect of the chosen approach in the case study

The effects of the pragmatic versus explanatory approach will have important consequences in a number of characteristics of the resulting study:

- *Patient selection*

  Under the pragmatic approach the patient cohort should reflect routine practice, so all relevant patients should be studied, with exclusions limited to a small number patients. For explanatory studies patients with co-morbidities or with a doubtful diagnosis will be excluded. So while the explanatory approach will establish the efficacy of the treatment, we will not know whether it will work in a "real-life" setting.

- *Study design*

  The study design might also be radically different. Under the explanatory model, the intervention must be identical in the two groups in all aspects except of the treatment under evaluation. For example, under the explanatory model, surgery might have to be delayed in the S group for the same period as in the chemotherapy group to align the two. Under the pragmatic approach, patients in the S group will have surgery immediately after diagnosis.

Efficacy versus effectiveness

Depending on the perspective chosen, there are consequences of whether the study measures efficacy or effectiveness. These two terms are not equivalent.

Efficacy addresses the question of whether the intervention works under ideal conditions. It is a proof of concept. It answers the question - Can it work?

By contrast, *effectiveness* assesses whether the intervention works in real-life conditions. It answers the question- Does it work?

For further reading see Roland & Torgerson (*BMJ*, 1998) and Haynes (*BMJ*, 1999).

Impact of the pragmatic or explanatory point of view
Chosen analysis

Under the pragmatic model, patients are analyzed as they were assigned to treatment and not as they were ultimately treated (we will take up this issue at length later in this lecture). This is called an *Intention-to-treat* analysis (ITT).

On the other hand, under the explanatory model, patients that changed treatment group are analyzed according to treatment received ("TR" or "as-treated") analysis. Another analysis is one that discards non-adherent patients ("adherers only" or "per-protocol" analysis).

The ITT approach addresses the question of how a treatment intervention will fare when it is administered in general practice. On the other hand, as-treated approaches permit insight into whether a treatment is efficacious but do not address the real-life effectiveness of the treatment in general use.

ITT versus per-protocol approaches

Suppose that, among patients accrued in an single-arm trial, N_E received treatment and, out of those, R_E had a beneficial response. On the other hand, N_1 did not receive the assigned treatment, and R_1 had a beneficial response (usually $R_1 = 0$).

Then, under the ITT (pragmatic) approach,

$$p_{ITT} = \frac{R_E + R_1}{N_E + N_1}$$

while, under the per-protocol approach,

$$p_{PP} = \frac{R_E}{N_E}$$

Thus, unless, $N_1 = 0$ or $R_1/N_1 \geq p_{PP}$ (neither of which is plausible usually) $p_{ITT} < p_{PP}$. In other words, as-treated or per-protocol analyses will exaggerate the effectiveness of an intervention. Seen from the opposite viewpoint, ITT analysis will underestimate the efficacy of the intervention.

Advantages and problems with the two methods

While p_{ITT} does not measure the biological effect of the intervention, there are serious problems with the explanatory analysis (or per-protocol analysis) which leads to the estimation of p_{PP}.

At the same time, p_{PP} does not necessarily estimate a biological effect for two reasons: The effect p_{PP} is influence by adherence to the treatment regimen. And, even if it did measure the biological effect, it degenerates if patients cannot adhere to treatment.

Difficulties of the per-protocol approach

A final technical issue arises when enrolling a patient. The explanatory, per-protocol, analysis, *conditions* on patient adherence. This is of course not known at the time of enrollment, so this approach conditions on a future event.

This is both conceptually wrong and undermines fundamental mathematical foundations of a number of models (e.g., what is called a "predictable process" in survival analysis).

The ITT principle maintains the advantages of treatment randomization.

Analysis of *completers* and the vagaries of missing data

Frequently analysis is performed on patients that have completed a regimen (i.e., have complete data). This analysis is slightly different from a per-protocol analysis as it focuses on data availability and not protocol or treatment adherence. This analytical approach is the default for all statistical software.

However, excluding subjects with missing values may produce seriously biased results if the underlying assumption that the observed data are representative of the missing data does not hold. This is a special case of *selection bias*.

The missing data hierarchy

The levels difficulty of interpreting missing data in a model of response Y, modeled according to a set of predictors (also known as covariates) X, are given below (Little & Rubin, 1987):

- *MCAR*
  Data are missing completely at random. The missingness is not associated with the actual true (but uknown) value and there are no other variables in the data that can predict whether a missing value exists in a particular variable. This is the difference between MCAR and MAR (below).
- *MAR*
  Data are missing at random. The pattern or probability of missingness is associated with other variables in the data *but not with the actual missing value*. This is the difference between MCAR and MNAR (see below).
- *MNAR*
  Data are missing not at random. The chance that the data are missing depends on the actual missing value.

The missing data hierarchy
MCAR

These are data missing completely at random. In clinical trials measuring a response Y based on measurements obtained from the subjects (covariates) X.

MCAR means that missing responses and/or covariate measurements are not dependent upon other responses or covariate measurements.

Examples of MCAR cases

Examples include the following:

- A laboratory specimen is dropped or goes bad
- A subject is run over by a bus
- A survey is lost in the mail

However, if the lab specimen reflects practices at a specific lab (which, in turn, may be associated with the outcome), the survey was lost because the post office is in a certain part of town, or the subject was part of a depression study and may have thrown himself under the bus, then the missing pattern may not be MCAR.

The missing data hierarchy
MAR

- *MAR*
  Missing at random, is the situation where missingness depends on fully observed variables only. Despite its name, there is little randomness about MAR. Examples of MAR missingness are as follows:
  - Sicker patients might be less able to perform some of the tests required by the study protocol
  - Laboratory animals are sacrificed for humane reasons when tumors have grown too large

  In both cases, missingness is MAR because the current (and missing) response can be predicted (and thus imputed) by modeling it on previous responses and/or the covariate measurements already collected on the subjects.

The missing data hierarchy
MNAR or NMAR or non-ignorable missingness

- *MNAR*
  Missing not at random is the hardest case to deal with as data missingness depends on unobserved covariates or missing response observations. The hardest part of MNAR is that, based only on available data, there is no way to determine whether missingess is non-random or whether modeling is appropriate! Examples include:
- Patients are lost to follow-up because they are dead (but vital status is not available because of the death)
- Tumors in laboratory animals are non-palpable because of treatment success so tumor volume is missing (because of the current, and missing, observation not past ones)

Statistics can address some missing data issues

It would be illustrative to go over the mathematical notation of missing data (just so we have some anchors for the components of the statistical analysis).

The complete data is a vector $\mathbf{Y} = (\mathbf{Y}^{\text{obs}}, \mathbf{Y}^{\text{mis}})$. We can also consider a binary (zero/one) matrix $\mathbf{M}$, with elements $m_{ij} = 1$ if observation from the jth subject and ith variable is observed and $m_{ij} = 0$ if it is missing.

Modeling with missing values

What we are trying to model is

$$f(\mathbf{Y}, \mathbf{M}|\mathbf{X}; \theta) = \underbrace{f(\mathbf{Y}^{\text{obs}}, \mathbf{Y}^{\text{mis}}|\mathbf{X}; \theta)}_{\text{Data mechanism}} \underbrace{f(\mathbf{M}|\mathbf{Y}^{\text{obs}}, \mathbf{Y}^{\text{mis}}, \mathbf{X}; \theta)}_{\text{Missingness mechanism}}$$

where $\mathbf{X}$ are all the covariates and θ are parameters (e.g., means, variances, etc.).

So we are jointly modeling the data and the missing mechanism $f(\mathbf{M}|\mathbf{Y}^{\text{obs}}, \mathbf{Y}^{\text{mis}}, \mathbf{X}; \theta)$.

Statistics can address some missing data issues
Statistical modeling under different missing patterns

- *MCAR*

$$f(\mathbf{M}|\mathbf{Y},\mathbf{X}) = f(\mathbf{M})$$

in this case the usual *complete case analysis* that discards the observations with missing values is valid.

- *MAR*

$$f(\mathbf{M}|\mathbf{Y},\mathbf{X}) = f(\mathbf{M}|\mathbf{Y}^{\text{obs}},\mathbf{X}))$$

which implies that

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{M}=0,\mathbf{Y}^{\text{obs}},\mathbf{X};\theta) = f(\mathbf{Y}^{\text{mis}}|\mathbf{M}=1,\mathbf{Y}^{\text{obs}},\mathbf{X};\theta)$$

so a correct completer model can be used to impute (fill in) the missing data. In the special case of covariate-dependent missingness the above becomes $f(\mathbf{M}|\mathbf{Y},\mathbf{X}) = f(\mathbf{M}|\mathbf{X})$ and, in this case, the complete case analysis is still valid.

Statistics can address some missing data issues
Statistical modeling under different missing patterns (continued)

- *MNAR*

$$f(\mathbf{M}|\mathbf{Y},\mathbf{X}) = f(\mathbf{M}|\mathbf{Y}^{\text{obs}},\mathbf{Y}^{\text{mis}},\mathbf{X})$$

this implies that we will need to explicitly model $f(\mathbf{Y}^{\text{mis}}|\mathbf{M},\mathbf{X};\theta)$, but there is no data to estimate this model under MNAR!

Example: Simple linear regression with covariate-dependent missingness

Consider the following contrived data set[1]:

Unit	X	Y
1	3.4	5.67
2	3.9	4.81
3	2.6	4.93
4	1.9	6.21
5	2.2	6.83
6	3.3	5.61
7	1.7	5.45
8	2.4	4.94
9	2.8	5.73
10	3.6	.

[1] Taken from http://www.lshtm.ac.uk/msu/missingdata/simple_web/node5.html

Example: Linear regression with covariate-dependent missingness

Complete case analysis

The complete case analysis (regression of Y on X) is given as follows:

```
Call:
lm(formula = y ~ x, data = reg)

Residuals:
    Min      1Q  Median      3Q     Max
-0.7413 -0.4876  0.1951  0.3456  1.0754

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.5601     0.8565   7.660  0.00012 ***
x            -0.3662     0.3085  -1.187  0.27399
---
Signif. codes:  0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1

Residual standard error: 0.6389 on 7 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.1675, Adjusted R-squared:  0.0486
F-statistic: 1.409 on 1 and 7 DF,  p-value: 0.274
```

Example: Linear regression with covariate-dependent missingness

Analysis based on multiple imputation

Based on this model we can follow the multiple-imputation paradigm to impute the missing value. The steps are as follows

(i) Choose the number k of multiple imputations.

(ii) For each $j = 1, \ldots, k$ follow the steps:

1. Simulate $\tilde{\beta} = (\tilde{\beta}_0, \tilde{\beta}_1) \sim N(\hat{\beta}, \hat{\Sigma})$ where $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ and $\hat{\Sigma}$ is the estimated variance matrix of $\hat{\beta}$
2. Simulate the missing value $\tilde{Y}_{10} \sim N(\tilde{\beta}_0 + \tilde{\beta}_1 X_{10}, \hat{\sigma}^2_\epsilon)$
3. Based on the pseudo-complete data estimate $\hat{\beta}^{(j)} = (\hat{\beta}_0^{(j)}, \hat{\beta}_1^{(j)})$

(iii) Estimate the parameter based on multiple-imputation as $\hat{\beta}^{MI} = k^{-1} \sum_{j=1}^{k} \hat{\beta}^{(j)}$.

(iv) Estimate the standard error of $\hat{\beta}^{MI}$ using Rubin's rules for combining the within-imputation and between-imputation variances

Example: The OASIS smoking cessation study

Consider the following table of the OASIS smoking cessation study that compares standard (ST) versus enhanced (ET) counseling interventions[2]:

Table: The OASIS study

Treatment		Month			
		1	3	6	12
ET	Abstinent	0.18	0.09	0.11	0.11
	Smoking	0.83	0.47	0.42	0.34
	Missing	–	0.44	0.45	0.55
ST	Abstinent	0.15	0.09	0.10	0.07
	Smoking	0.85	0.54	0.52	0.52
	Missing	–	0.37	0.38	0.40

[2]Taken from Daniels & Hogan, Missing data in longitudinal studies: Strategies for Bayesian modeling and sensitivity analysis, Chapman & Hall/CRC, 2008.

Example: The OASIS study
Analysis under MCAR

Analysis under the MCAR assumption assumes that all the missing smoking statuses are missing randomly, so the analysis can be done ignoring all missing observations.

If we perform a logistic-regression analysis, based on the subjects with known smoking status (completers), the smoking rate at one year is $p_{ET} = 0.76$ and $p_{ST} = 0.88$ for the enhanced and standard intervention groups respectively.

The odds ratio is OR $= 2.225$.

Alternative analysis

The above analysis is equivalent to an analysis of the following 2×2 table.

	Group				
Smoking status	ET		ST		Total
Abstinent	16	(24%)	11	(12%)	27
Smoking	51	(76%)	78	(88%)	129
Total	67		89		156

The 95% CI of the odds-ratio is (0.96, 5.18). So the OR is not statistically significant at the 95% level (since the value OR=1 is included in the confidence interval). In other words, this analysis fails to show any difference in the effectiveness of the interventions in groups ET and ST.

Example: The OASIS study
Analysis under MAR

Under the MAR assumption, it is assumed that the missing smoking statuses are missing randomly *within each intervention group*.

In other words, it is assumed that, within each intervention group, the smoking rate among subjects with missing smoking status is the same as the rate determined from subjects with known smoking status.

Example: The OASIS study
Analysis under MAR (continued)

An analysis under MAR would *impute* (fill in) the missing statuses as smoking in 76% and 88% of the missing subjects in group ET and ST respectively and as non-smoking in 24% and 12%.

This is like analyzing the following table:

	Group				
Smoking status	ET		ST		Total
Abstinent	36	(24%)	18	(12%)	54
Smoking	113	(76%)	131	(88%)	244
Total	149		149		298

Example: The OASIS study
Analysis under MAR (continued)

The odds ratio is $\widehat{\text{OR}} = (36 \times 131)/(18 \times 113) = 2.31$, with an approximate 95% CI

$$\left(e^{\log(2.31)-1.96\sqrt{1/36+1/113+1/18+1/131}}, e^{\log(2.31)+1.96\sqrt{1/36+1/113+1/18+1/131}}\right)$$

$$(1.24, 4.29)$$

This analysis shows that the smoking rate in the standard is significantly higher than the enhanced group (or, equivalently, that the enhanced intervention is more effective).

Example: The OASIS study
Analysis under MAR (continued)

The previous analysis has a significant drawback. It does not take into account the fact that 55% of the missing observations in the ET group and 40% of the missing observations in the ST group were deterministically imputed.

Thus, the estimate of the variability of the odds ratio will be underestimated.

Example: The OASIS study
Analysis under MAR (continued)

To overcome this, we can perform simulations of the data where each missing value is imputed with some degree of error in a random fashion.

In the following we show the analysis of the smoking cessation trial where a uniform random number $U \in (0, 1)$ was generated and the missing observations in the ET group were replaced with one (smokers) if $U < 0.76$ and zero otherwise.

Similarly, the missing observations in the ST group were replaced with one if $U < 0.88$ and zero otherwise.

Example: The OASIS study
Analysis under MAR: Simulations of the OR

In the following figure I present a histogram of 1,000 replications of the imputed data and the resulting values of the OR.

Descriptive statistics are given in the next slide.

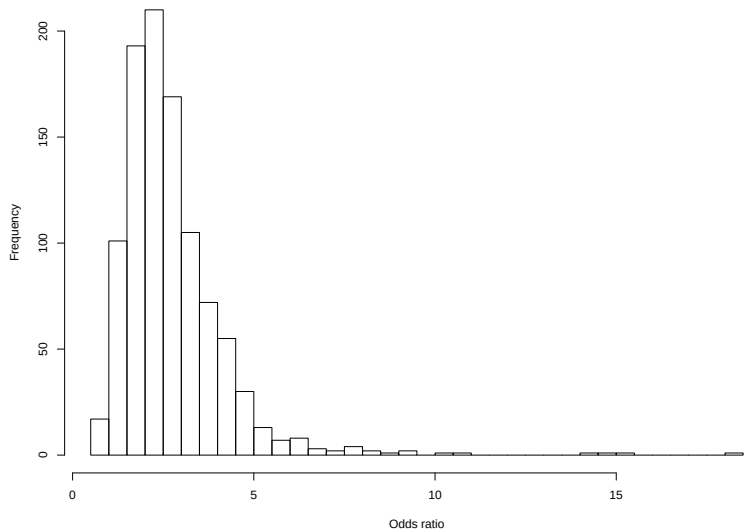

Figure: Histogram of odds ratios from 1,000 simulated data sets

Example: The OASIS study
Analysis under MAR: Simulations of the OR

The median of the simulated odds ratio is 2.45[3].

A way to obtain the two-sided 95% confidence interval for the odds ratio is to determine the 2.5th and 97.5th percentiles.

```
> quantile(sampleOR[,4])
       0%        25%        50%        75%       100%
 0.6060277  1.8567031  2.4109671  3.1720622 12.4419048
> quantile(sampleOR[,4], probs=c(0.025, 0.975))
    2.5%    97.5%
1.087558 6.065612
```

This interval is (1.09, 6.07), suggesting that, under the MAR assumption, the enhanced intervention is significantly more effective than the standard intervention.

[3]Note: Since the distribution of the odds ratio is skewed to the right the median, rather than the mean, should be used as a measure of central tendency

Example: The OASIS study
Analysis under MNAR

While the MCAR assumption seems completely far-fetched for this example, there are serious suspicions that the MAR hypothesis is a stretch at best. It is commonly accepted that subjects that smoke tend to drop out more readily from a program. This means that the smoking rates among subjects with missing smoking status will be higher (potentially significantly so) compared to the observed rates in each intervention group.

Since it is not known how well any model represents reality, a reasonable approach is to consider a number of models and see how they affect the results. This is called *sensitivity* analysis.

Example: The OASIS study
Sensitivity analysis

In the OASIS study, one conservative approach would be to consider all subjects with missing smoking status are smokers. Another is the approach the investigators followed. They consulted with four experts about the likely probability that subjects with missing smoking status are smokers.

While the analysis chosen is beyond the scope of this lecture, the resulting probabilities from the five approaches are given in the following Table[4]:

	Model/Expert					
Intervention	MAR	Conservative	1	2	3	4
ET	0.78	0.89	0.83	0.87	0.87	0.83
ST	0.88	0.93	0.90	0.91	0.90	0.90
OR	2.31	1.51	1.54	1.61	1.41	1.47

[4]Note that we have considerably simplified the analysis. For more details refer to Daniels & Hogan (2008).

Note that we expressed the odds ratio as ST versus ET.

Example: The OASIS study
Sensitivity analysis (continued)

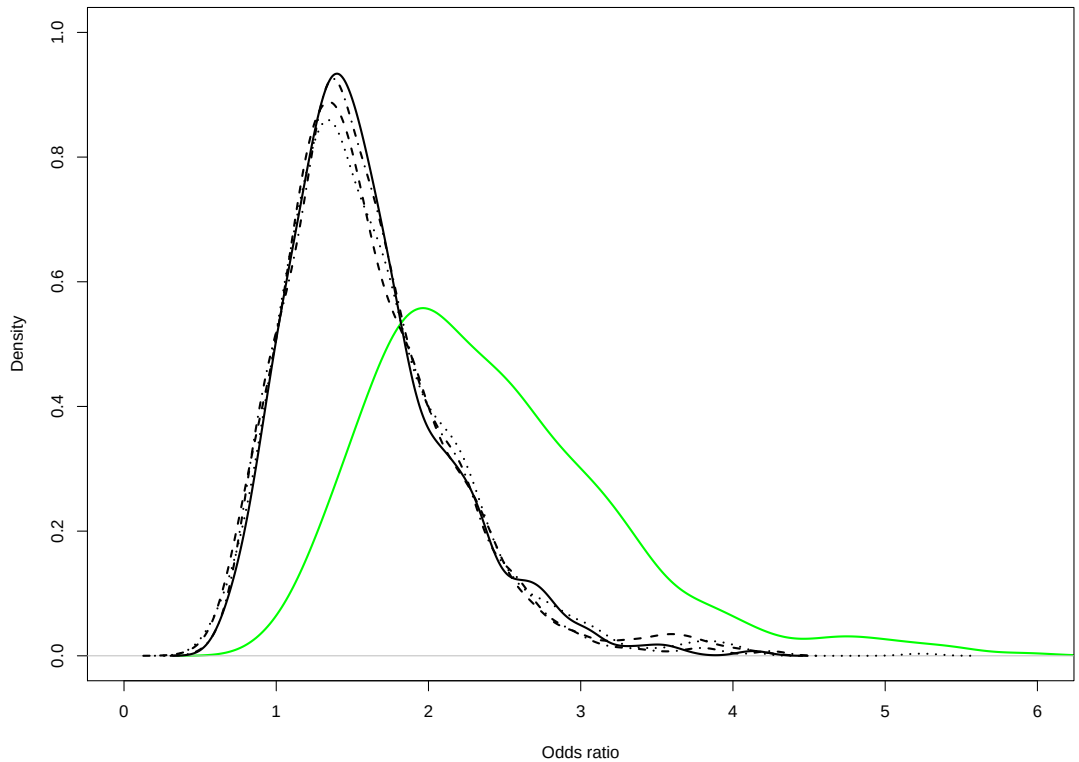

Figure: Monte Carlo simulations under MAR (green) and the four experts (grey).

Implications of the analysis

The analysis shown in Figure 2 represents a histogram of 1,000 simulated datasets with a smoother run through it.

It is obvious that the MAR assumption is the most optimistic viewpoint and that most likely the treatment effect is, at best, minor, and certainly not statistically significant at the 5% level.

Section 2

Causality and non-compliance

Causality

Randomization allows us to make causal statements regarding the effect of treatment on the response of interest.

In this section we will give a more formal definition of cause and effect. To do this we use what are called *counterfactual* random variables.

As usual, we consider an overall population of individuals that we are interested in and assume that the participants in a clinical trial represent a random sample from this population. Within this clinical trial we will compare an experimental treatment (e.g., treatment 1) to a standard treatment or placebo (treatment 0).

Definition of counterfactual random variables

We define the counterfactual random variable Y_1^* to denote the response. This may be a binary or continuous outcome that a randomly selected individual would have if, possibly contrary to fact, they received treatment 1 (the experimental treatment).

Similarly, we define the counterfactual random variable Y_0^* to denote the response that a randomly selected individual would have if, possibly contrary to fact, that individual received treatment 0 (standard treatment or placebo).

We imagine that both random variables Y_0^* and Y_1^* exist, even though in actuality it would be impossible to observe both responses on the same individual (from which the term *counterfactual* or "contrary to fact" emanates).

Formal definition of the causal effect of treatment

Definition (Causal treatment effect)

At the individual level, the (random) *causal* treatment effect is

$$Y_1^* - Y_0^*$$

The mean causal effect of treatment

Clearly, if we knew the response of an individual to both treatments, then we would choose whichever treatment gave the better response.

Of course, this is not possible at the individual level but perhaps we can look at this question at the population level. That is, we will estimate the causal treatment response by the population mean causal effect

$$\Delta = E(Y_1^* - Y_0^*) = E(Y_1^*) - E(Y_0^*)$$

If Δ is positive, then, on average, the response on treatment 1 will be better than on treatment 0.

Note that, at the individual level, this does not necessarily imply that any specific individual will be guaranteed to benefit from the treatment found to be superior based on Δ but, on average, the population as a whole will benefit.

Observable data

The data that we actually observe from a clinical trial are summarized by (Y_i, A_i, X_i), with $i = 1, \cdots, n$, where, for a randomly selected individual i,

- $A_i \in \{0, 1\}$ denotes the treatment assignment (to the new treatment or the standard treatment or placebo respectively)
- Y_i denotes the response
- X_i denotes any additional characteristics, collected on the individual prior to treatment assignment (baseline characteristics)

We will refer to these as the observable random variables.

Note: We distinguish between the *observed response* Y_i for the i-th individual and the *counterfactual responses* Y_{1i}^* and Y_{0i}^*.

The *consistency* assumption

We make the reasonable assumption that $Y_i = Y_{1i}^*$ if $A_i = 1$ and that $Y_i = Y_{0i}^*$ if $A_i = 0$. In other words, we assume that the observed treatment response is equal to the counterfactual treatment response if the individual were assigned the same treatment as the one we observe to be assigned to the individual. This is the *consistency* assumption.

The consistency assumption in causal inference is defined as

$$Y_i = Y_{1i}^* I(A_i = 1) + Y_{0i}^* I(A_i = 0)$$

where $I(\cdot)$ denotes the indicator function of an event and it equals 1 when the event is true and 0 otherwise.

Association versus causation

Traditional statistical methods allow us to estimate *associational* relationships. For example, we can use regression models that allow us to estimate relationships such as $E(Y|A, X)$. These models explore the association of the mean outcome Y to the assigned treatment A and other measured characteristics (prognostic factors or covariates) X.

These associational relationships are not the causal relationships that are the parameters of interest.

However, associational statements are more easily assessed. So the question of estimating causal effects is modified as:

"Under what conditions or assumptions can we estimate causal parameters such as Δ, from observable data?"

Randomization

This is where randomization plays a key role. Since treatment is randomly assigned to the patient in a randomized study, treatment assignment is independent of any pre-treatment characteristics of the individual.

Consequently, we make the following assumption:

In randomized clinical trials,

$$A_i \text{ is independent of } (Y_{1i}^*, Y_{0i}^*, X_i)$$

That is, randomization severs any association between how an individual would have responded if given treatment 1 and how he/she would have responded if given treatment 0 and the treatment he/she was randomized to.

Remark

It is important to note that the assumption of independence between the treatment assignment A_i and the counterfactual response of individual i, (i.e., Y_{1i}^* or Y_{0i}^*), is not the same as saying that A_i is independent of Y_i (the observed response).

Since $Y_i = Y_{1i}^* I(A_i = 1) + Y_{0i}^* I(A_i = 0)$, Y_i is a function both of counterfactual responses and the treatment assignment and, as such, will not be independent of A_i.

In fact, if treatment is effective, as one hopes, then we would expect (and want) Y_i to depend on A_i.

Mean causal treatment effect

Another assumption is that $0 < P(A = 1) < 1$, also knonw as the *positivity assumption* (i.e., each treatment has a non-zero probability).

Based on the consistency, independence, and positivity assumptions we have that

$$\begin{aligned} E(Y|A=1) - E(Y|A=0) &= E(Y_1^* A + Y_0^*(1-A)|A=1) \\ &\quad -E(Y_1^* A + Y_0^*(1-A)|A=0) \\ &= E(Y_1^*|A=1) - E(Y_0^*|A=0) \\ &= E(Y_1^*) - E(Y_0^*) \\ &= \Delta \end{aligned}$$

Now we have an expression for the causal parameter Δ in terms of quantities that can be observed!

Estimation of the mean causal treatment effect

To estimate Δ it suffices to estimate $E(Y|A=1)$ and $E(Y|A=0)$.

These can be estimated by

$$\bar{Y}_1 = \frac{\sum_{i=1}^{n} Y_i A_i}{\sum_{i=1}^{n} A_i}$$

and

$$\bar{Y}_0 = \frac{\sum_{i=1}^{n} Y_i (1 - A_i)}{\sum_{i=1}^{n} (1 - A_i)}$$

respectively.

Thus, an consistent estimator of the causal treatment effect Δ can be derived from a randomized study using

$$\hat{\Delta} = \bar{Y}_1 - \bar{Y}_0$$

Non-compliance

The arguments outlined above assume that patients take the treatment to which they are randomized. In most clinical trials however, this is rarely the case. This is called *non-compliance*.

There is almost always some form of noncompliance from the intended treatment regimen. Some reasons for non-compliance are:

- A refusal by the patient to start or continue the assigned treatment, due to side effects or a belief that the treatment is ineffective
- A failure to comply with detailed instructions, such as drug dose, or to attend examinations when requested to do so
- A change of treatment imposed by the physician for clinical reasons, such as adverse effects or deterioration of the patient's health
- An administrative error. In its most extreme form, this may be the implementation of the wrong treatment.

Analytical strategies in the face of non-compliance

Some strategies that have been proposed include the following:

- **Intent-to-Treat Analysis** (ITT; As randomized)
  Everyone is included in the analysis and the comparison of treatments is based on the difference of the average response between the randomized groups ignoring the fact that some patients were non-compliant.
- **As-treated analysis**
  This type of analysis follows the general idea that only patients who fully complied with their assigned treatment regimen are to be compared and all non-compliant patients are excluded from the analysis.

The ITT principle and the dogma of clinical trials

The intent-to-treat (ITT) analysis principal complies (no pun intended) with the central dogma in clinical trial research:

Exclusions based on post-randomization considerations, such as noncompliance, are not allowed for the primary analysis.

This is because exclusion of patients from the analysis may result in bias in the treatment comparisons.

Example: The Clofibrate study

To illustrate some of the difficulties that can result from non-compliance, we consider the results from a study conducted by the Coronary Drug Project (New England Journal of Medicine, 1980).

This was a double-blind placebo-controlled trial comparing Clofibrate to Placebo. The following table shows the results from the ITT analysis:

Table: Intent-to-Treat Analysis

	Treatment	
	Clofibrate	Placebo
	$N = 1065$	$N = 2695$
5-year mortality	0.18	0.19

Effect of non-compliance

Table: 5-year mortality by treatment adherence

Adherence (% of capsules taken)	Clofibrate 5-year mortality	Clofibrate Number of patients	Placebo 5-year mortality	Placebo Number of patients
Poor ($< 80\%$)	0.25	357	0.28	882
Good ($> 80\%$)	0.15	708	0.15	1813

It is clear from these data that compliant patients are prognostically different from non-compliant patients. Therefore, the as-treated approach may lead to severe biases because it cannot separate the prognostic effect of noncompliance from the prognostic effect of treatment.

By contrast, the intent-to-treat analysis does not suffer from this type of bias. At the same time, when some patients do not comply with the treatment, an ITT analysis would diminish the effect of a treatment.

A simple trial design

Consider a randomized study where patients are randomized with equal probability to active drug (treatment 1) or placebo (control) (treatment 0).

Response is dichotomous. The main goal of the clinical trial is to estimate the difference in the probability of response between active drug and placebo

For simplicity, we assume that every patient either takes their assigned treatment or not (partial compliance is not considered) and their compliance can be assessed by a simple assay.

We also consider that the patients assigned to placebo do not have access to the study drug and that compliance cannot be determined for these patients.

Counterfactual and observable random variables

The problem above can be conceptualized as follows:

Let the counterfactual random variables Y_1^* and Y_0^* denote the response (1=response, 0=non-response) of a randomly selected individual if they received treatment 1 or 0 respectively.

Also let C denote the counterfactual random variable corresponding to whether or not a randomly selected individual complies or not $C = (1, 0)$. This is a *counterfactual* random variable because we do not know the compliance status for patients randomized to placebo.

Counterfactual random variables (continued)

Denote by $\theta = P(C = 1)$ the population probability of complying with the assigned treatment, while $\pi_1^{\mathrm{COM}} = P(Y_1^* = 1|C = 1)$ and $\pi_1^{\mathrm{NC}} = P(Y_1^* = 1|C = 0)$ are the probability of response among those who comply or do not comply if given active drug respectively.

Also, denote by $\pi_0^{\mathrm{COM}} = P(Y_0^* = 1|C = 1)$ and $\pi_0^{\mathrm{NC}} = P(Y_0^* = 1|C = 0)$ the probability of response among those who comply or do not comply if given active placebo.

As it is not reasonable to assume that Y_1^* and Y_0^* are independent of C, so we would not expect $\pi_1^{\mathrm{COM}} = \pi_1^{\mathrm{NC}}$ or $\pi_0^{\mathrm{COM}} = \pi_0^{\mathrm{NC}}$.

Estimating the causal treatment effect

Using some simple probability calculations we get that

$$\begin{aligned} E(Y_1^*) &= P(Y_1^* = 1) \\ &= P(Y_1^* = 1|C = 1)P(C = 1) + P(Y_1^* = 1|C = 0)P(C = 0) \\ &= \pi_1^{\mathrm{COM}}\theta + \pi_1^{\mathrm{NC}}(1-\theta) = \pi_1 \end{aligned}$$

and, similarly,

$$E(Y_0^*) = P(Y_0^* = 1) = \pi_0^{\mathrm{COM}}\theta + \pi_0^{\mathrm{NC}}(1-\theta) = \pi_0$$

Therefore, the average causal treatment effect equals

$$\Delta = E(Y_1^*) - E(Y_0^*) = \pi_1 - \pi_0 = \Delta^{\mathrm{COM}}\theta + \Delta^{\mathrm{NC}}(1-\theta)$$

where $\Delta^{\mathrm{COM}} = \pi_1^{\mathrm{COM}} - \pi_0^{\mathrm{COM}}$ and $\Delta^{\mathrm{NC}} = \pi_1^{\mathrm{NC}} - \pi_0^{\mathrm{NC}}$.

Estimation of the probabilities π_1^{COM}, π_0^{COM}

How can we estimate the probabilities π_1^{COM}, π_0^{COM} using the data we observe from a randomized clinical trial when there is noncompliance?

- θ (the overall compliance rate)

$$P(C=1|A=1) = P(C=1) = \theta$$

- π_0 (the overall response rate)

$$P(Y=1|A=0) = P(Y_0^*=1|A=0) = P(Y_0^*=1) = \pi_0$$

- π_0^{NC} (the response rate in the control arm under non-compliance)

$$\begin{aligned} P(Y=1|A=1,C=0) &= P(Y_0^*=1|A=1,C=0) \\ &= P(Y_0^*=1|C=0) = \pi_0^{\mathrm{NC}} \end{aligned}$$

- π_1^{COM} (the response rate in the active arm under compliance)

$$\begin{aligned} P(Y=1|A=1,C=1) &= P(Y_1^*=1|A=1,C=1) \\ &= P(Y_1^*=1|C=1) = \pi_1^{\mathrm{COM}} \end{aligned}$$

- π_0^{COM} (the response rate in the control arm under compliance)

$$\pi_0^{\mathrm{COM}} = \frac{\pi_0 - \pi_0^{\mathrm{NC}}(1-\theta)}{\theta}$$

Estimation of the treatment effect under Intent-to-treat

We are estimating

$$\Delta_{\mathrm{ITT}} = P(Y = 1|A = 1) - P(Y = 1|A = 0)$$

Again, by the assumptions made and some probability calculations we get

$$P(Y = 1|A = 1) = \pi_1^{\mathrm{COM}}\theta + \pi_1^{\mathrm{NC}}(1 - \theta)$$

Similarly,

$$P(Y = 1|A = 0) = \pi_0^{\mathrm{COM}}\theta + \pi_0^{\mathrm{NC}}(1 - \theta)$$

Thus, Δ_{ITT} is

$$\begin{aligned} \Delta_{\mathrm{ITT}} &= (\pi_1^{\mathrm{COM}} - \pi_0^{\mathrm{COM}})\theta + (\pi_1^{\mathrm{NC}} - \pi_0^{\mathrm{NC}})(1 - \theta) \\ &= \theta\Delta^{\mathrm{COM}} + (1 - \theta)\Delta^{\mathrm{NC}} \end{aligned}$$

This, unless the compliance rate is 100%, will be less (and perhaps much less) than the causal treatment effect under full compliance.

Remarks

Recall that

$$\Delta^{\mathrm{COM}} = P(Y_1^* = 1|C = 1) - P(Y_0^* = 1|C = 1) = E(Y_1^* - Y_0^*)$$

is the difference in the mean counterfactual responses between the two treatment arms among patients that would comply with treatment.

As such, Δ^{COM}, some argue, is the causal parameter of greatest interest since it quantifies the benefit among patients who will comply with the new treatment.

However, we are in fact able to estimate Δ^{COM} since we can estimate the parameter θ, the overall compliance rate if offered the new treatment by

$$\Delta^{\mathrm{COM}} = \frac{\Delta_{\mathrm{ITT}} - (1-\theta)\Delta^{\mathrm{NC}}}{\theta}$$

Since all the quantities are easily estimated from the data of a clinical trial, this means we can estimate the causal parameter Δ^{COM}.

More remarks on the ITT principle

If the null hypothesis of no treatment effect is true; namely

$$H_0 : \ \Delta^{\mathrm{COM}} = \Delta^{\mathrm{NC}} = \Delta = 0$$

the intent-to-treat analysis gives an unbiased estimator of treatment difference (under H_0) and can be used to compute a valid test of the null hypothesis

However, the above results make it clear that the ITT analysis will yield an estimator which diminishes a causal treatment effect under compliance.

As-treated analysis

In one version of an as-treated analysis we compare the response rate of patients randomized to active drug who comply to all patients randomized to receive the control. That is, we compute

$$\Delta_{\mathrm{AT}} = E(Y|A=1, C=1) - E(Y|A=0)$$

After some algebra we get that

$$\Delta_{\mathrm{AT}} = \Delta + (1-\theta)(\pi_0^{\mathrm{COM}} - \pi_0^{\mathrm{NC}})$$

where Δ is the average causal treatment effect.

This makes clear that when there is noncompliance, (i.e., when $\theta < 1$), the as-treated analysis will yield an unbiased estimate of the average causal treatment effect only if $\pi_0^{\mathrm{COM}} = \pi_0^{\mathrm{NC}}$.

Since this assumption is not generally true, the as-treated analysis can result in biased estimation even under the null hypothesis.