Introduction to Clinical trials

Instructor:
Constantin Yiannoutsos, Ph.D.
Professor of Biostatistics
Indiana University
School of R.M. Fairbanks School of Public Health, Department of Biostatistics

Analysis of data from clinical trials

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 effective or not (we will discuss *effectiveness* later in this lecture).

Pragmatic versus *explanatory* trials
The Explanatory perspective

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 allign 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 oen 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 SA 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 $R_1/N_1 \geq p_{ITT}$ (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 that p_{PP} measures is confounded with 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.

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
Covariate-dependent missingness

An intermediate situation is missing data dependent only on values of the covariates X but not on Y. This is called covariate-dependent missingness by Little (1995, *JASA*). In this case, missing data are dependent only on the covariates X but not on the outcomes Y. Examples include:

- Cultural beliefs affect follow-up and retention in a clinical trial but are unrelated to the outcome Y
- Treatment-related dropout that is constant within each treatment group and unrelated to the outcome Y can be considered MCAR

However, if the cultural beliefs reflect groups that have different overall response to therapy (or are associated with certain risk factors, such as a higher triple-negative breast cancer rates among African-American women), or dropout within a treatment group is related to lack of efficacy, the missing data will not be covariate-dependent MCAR.

The missing data hierarchy
MAR

- *MAR*
  Missing at random, is the situation where missingness depends on both the observed covariates X and on the observed response Y. 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 j^{th} subject and i^{th} variable is observed and $m_{ij} = 0$ if it is missing.

Modeling with missing values

What we are trying to model is

$$f[\mathbf{Y}, \mathbf{M}|\mathbf{X}, \mathbf{\Theta}] = \underbrace{f[(\mathbf{Y}^{\text{obs}}, \mathbf{Y}^{\text{mis}})|\mathbf{X}, \mathbf{\Theta}]}_{\text{Data mechanism}} \underbrace{f[\mathbf{M}|(\mathbf{Y}^{\text{obs}}, \mathbf{Y}^{\text{mis}}), \mathbf{X}, \mathbf{\Theta}]}_{\text{Missing data mechanism}}$$

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

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

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

- *MCAR*

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

so this implies that $f[(\mathbf{Y}^{\text{obs}},\mathbf{Y}^{\text{mis}})|\mathbf{X},\boldsymbol{\Theta}] = f[\mathbf{Y}^{\text{obs}}|\mathbf{X},\boldsymbol{\Theta}]$ and the usual completer analysis can be carried out. In the case of covariate-dependent missingness the above becomes $f(\mathbf{M}|\mathbf{Y},\mathbf{X}) = f(\mathbf{M}|\mathbf{X})$ leading to the same analysis (i.e., excluding subjects with any missing data) as above.

- *MAR*

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

which implies that $f[\mathbf{Y}^{\text{mis}}|\mathbf{M},\mathbf{X},\boldsymbol{\Theta}] = f[\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{X},\boldsymbol{\Theta}]$, so a correct completer model can be used to impute (fill in) the missing data.

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 $f[\mathbf{Y}^{\text{mis}}|\mathbf{M},\mathbf{X},\boldsymbol{\Theta}]$, but there is nothing in the collected data that can give us any comfort about how good our model is!

Example: Simple linear regression with covariate-dependent missingness

Consider the following contrived data set*:

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	.

* Taken from http://www.lshtm.ac.uk/msu/missingdata/simple_web/node5.html

Example: Linear regression with covariate-dependent missingness

Completer analysis

The completer 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 with imputed missing values

The above model suggests a way to perform an analysis on the complete data set by imputing (filling in) a value for the missing observation Y.

This value is

$$\hat{Y} = 6.56 + (-0.366)(3.6) = 5.24$$

Example: Linear regression with covariate-dependent missingness

Pluging in $\hat{Y} = 5.24$ for the missing observation and repeating the analysis results in the following output:

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.74134 -0.44626  0.09756  0.32374  1.07543

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.5601     0.7641   8.585 2.62e-05 ***
x            -0.3662     0.2663  -1.375    0.206
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.5977 on 8 degrees of freedom
Multiple R-squared:  0.1911,Adjusted R-squared:  0.09002
F-statistic:  1.89 on 1 and 8 DF,  p-value: 0.2064
```

Example: Linear regression with covariate-dependent missingness

Accounting for the variability

We see that the estimate for the regression slope $\hat{\beta}$ has remained the same. We also see that the standard error of the estimation has been reduced from s.e.$\hat{\beta} = 0.309$ to 0.266. This is because the imputed value of Y did not account for the fact that it is an estimate of the true value of Y.

To deal with this issue we can simulate the data by imputing for the missing Y an observation

$$\hat{Y} = 6.56 + (-0.366)(3.6) = 5.24 + \epsilon$$

where $\epsilon \sim N(0, 0.408)$ where 0.408 is the mean square error given above in the ANOVA table.

Example: Linear regression with covariate-dependent missingness
Accounting for the variability (continued)

In the following figure and list of descriptive statistics, I have simulated the data set 1,000 times and obtained an empirical distribution of the true value of β with a more reasonable standard error.

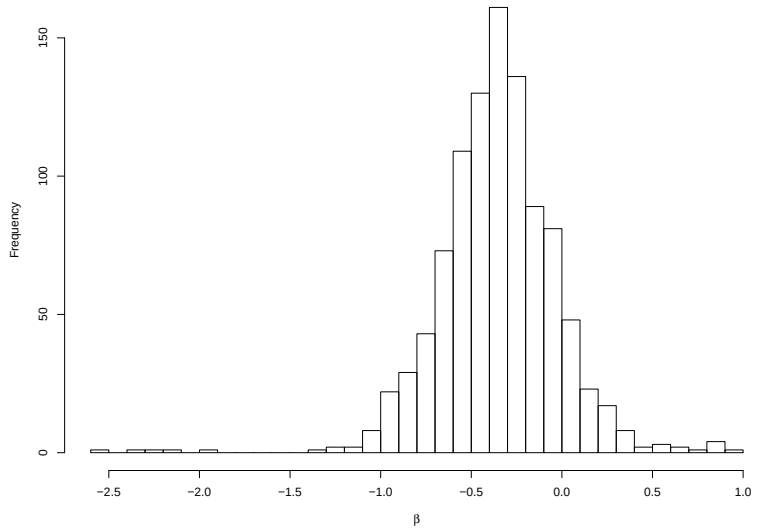

Figure 1: Histogram of 1,000 simulated data sets

Estimated coefficients

The descriptive statistics of the generated estimates are as follows:

```
> apply(sample.beta,2,mean)
          it    betaHat0    betaHat1
500.5000000   6.5304405  -0.3557927
> apply(sample.beta,2,median)
          it    betaHat0    betaHat1
500.5000000   6.5044536  -0.3453537
>
> # The standard error is the sample SD
> apply(sample.beta,2,sd)
          it    betaHat0    betaHat1
288.8194361   0.9448023   0.3313254
```

This means that the average estimate of $\hat{\beta} = -0.356$ with an attendant standard error s.e.$(\hat{\beta}) = 0.331$, which is much more realistic than the one produced by the naïve imputation procedure described earlier.

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*:

Table 1: 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

* Taken from Daniels & Hogan, Misisng 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.

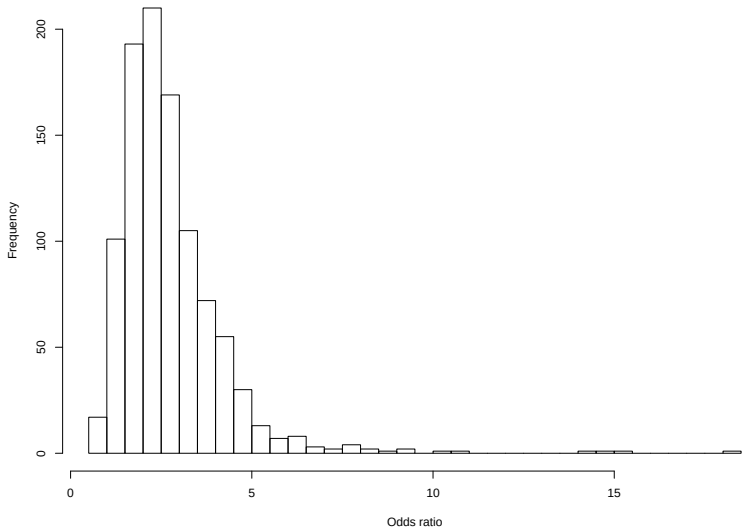

Figure 2: Histogram of odds ratios from 1,000 simulated data sets

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

The median of the simulated odds ratio is 2.45*.

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.

---

*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*:

	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

* 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)

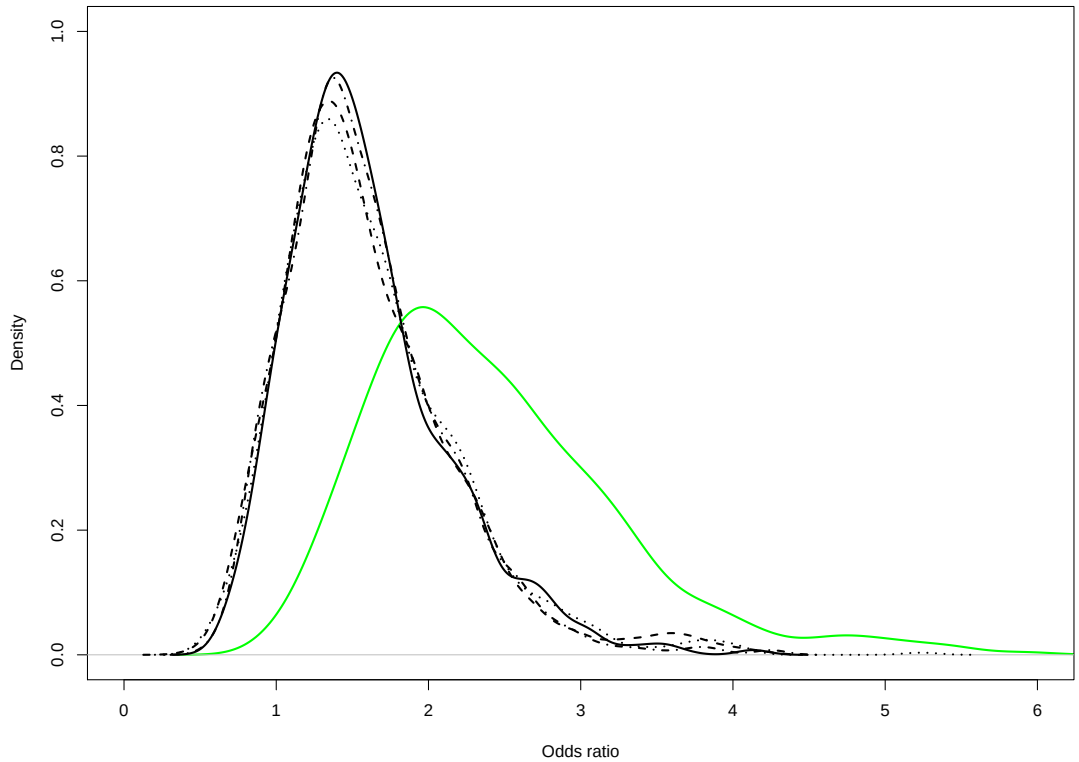

Figure 3: Monte Carlo simulations under MAR (green) and the four experts (grey).

Implications of the analysis

The analysis shown in Figure 3 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.