- The overall rationale and features of a trial are described in a trial protocol

Table 1.7 *Protocol outline.*

1. Background of the study
2. Objectives
   (a) Primary question and response variable
   (b) Secondary question and response variable
   (c) Subgroups hypotheses
   (d) Adverse effects
3. Design of the study
   (a) Study population
      i. Inclusion criteria
      ii. Exclusion criteria
   (b) Sample size assumptions and estimates
   (c) Enrollment of participants
      i. Informed consent
      ii. Assessment of eligibility
      iii. Baseline examination
      iv. Intervention allocation (e.g., randomization method)
   (d) Intervention
      i. Description and schedule
      ii. Measures of compliance

- The overall rationale and features of a trial are described in a trial protocol (contd)

(e) Follow-up visit and description and schedule
(f) Ascertainment of response variables
  i. Training
  ii. Data collection
  iii. Quality control
(g) Data analysis
  i. Interim monitoring
  ii. Final analysis
(h) Termination policy

4. Organization
(a) Participating investigators
(b) Statistical unit or data coordinating center
  i. Laboratories and other special units
  ii. Clinical center(s)
(c) Study administration
  i. Steering committees and subcommittees
  ii. Data monitoring committee
  iii. Funding organization

Appendix
Definitions of eligibility criteria
Definitions of response variables

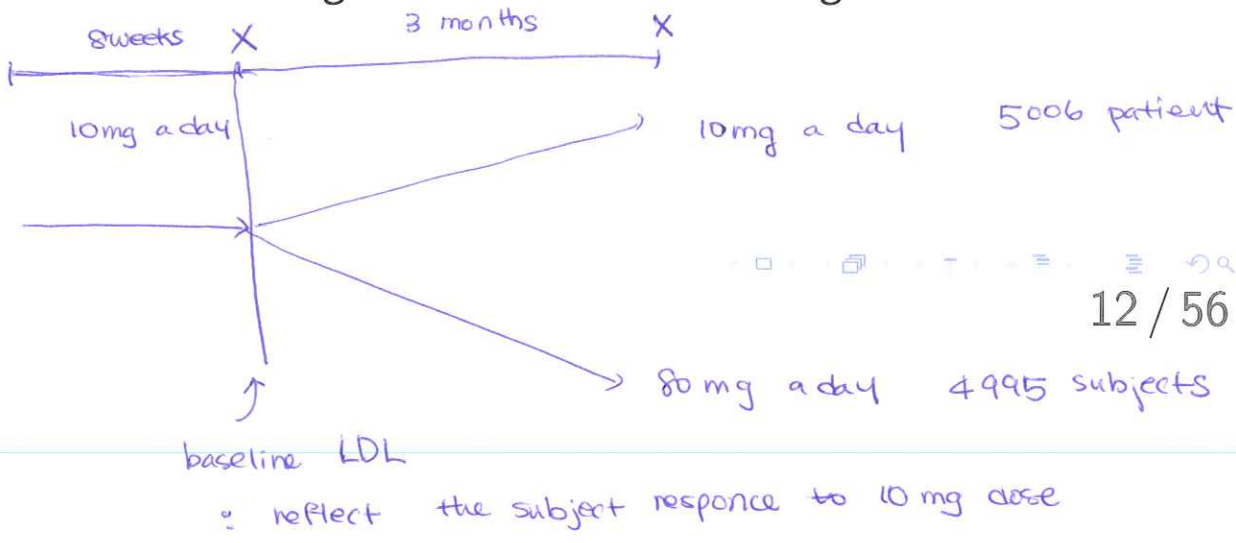
† Goal of a clinical trial: To establish the safety and effectiveness of the intervention in the target population.

e.g.: "Is the intervention X safe and effective in individuals with (or at risk for) disease Y?"

† Example 2.1: CD page 31

- Low density lipoprotein (LDL): a particle containing fatty acids that circulates in the blood and is produced in the liver.
- High levels of LDL are associated with increased risk of cardiovascular disease which leads to heart attacks, strokes, and other adverse consequences.
- A class of drugs known as *statins* have been shown to lower LDL levels in the blood and reduce the incidence of certain adverse cardiovascular events.

- The TNT ("Treating to New Targets") trial was a double-blind randomized trial conducted.
- **Study Question***: To determine if lowering LDL beyond the commonly accepted levels using a high daily dose (80*mg*) of atorvastatin results in fewer major cardiovascular events than that achieved with a lower daily dose (10*mg*) of atorvastain.
- Prior to randomization, all subjects meeting inclusion criteria* received 10*mg* of atorvastatin for eight weeks, after which subjects meeting the screening criteria were randomized either continue with the 10*mg* dose or receive the 80*mg* dose.



- A population of interest:
  - ⋆⋆ The study population is precisely defined by the eligibility criteria for a trial. The criteria are clearly specified in the study protocol
  - ⋆⋆ The eligibility criteria are critical because they have an impact on the generalizability of the result and the ease of subject recruitment.
    - less restrictive criteria: more generalizable results, possibly easier recruitment, may need a larger sample size
    - more restrictive criteria: less generalizable and more difficult recruitment.
  - ⋆⋆ "Healthy volunteer effect": these individuals are often at lower risk than the general population of subjects.

⇒ can result in "under-powered" trial

- Study questions
  - ⋆⋆ The questions of interest generally – direct treatment comparison

    e.g.: Is there a greater reduction in three month serum LDL in subjects with elevated baseline LDL taking $80mg$ atorvastatin than in subjects taking $10mg$ atorvastatin?

    $\Rightarrow$ hypothesis testing

    $\Rightarrow$ generally a one-to-one correspondence between the question and the outcome.

  - ⋆⋆ Multiple questions of interest in a trial: primary, secondary

    * Secondary questions are generally other important clinical questions that either add support to the primary question or are of interest in their own right.

⋆⋆ Multiple questions of interest in a trial: primary, secondary

e.g.: TNT trial

* Primary outcome: the composite of coronary heart disease death, nonfatal myocardial infarction, resuscitated cardiac arrest, and fatal and nonfatal stroke.

* Secondary outcome: all-cause morality

* The observed incidence rate for the primary outcome was 20% lower in the high dose arm than in the low dose arm

* The rates of all-cause morality were virtually identical.

e.g.: Cancer trial

* Primary outcome: disease free survival

* Secondary outcome: total mortality

↺ Clinical outcomes: outcomes that reflect the survival or symptomatic status of the subject.

↺ Once the population and the questions of interest have been defined, the investigators must precisely specify the outcome measures.

e.g.: *Primary question*: the effect of the treatment on all-cause mortality

⇒ *Outcome*: the dichotomous variable indicating whether or not death has occurred, and possibly the date or time of death.

↺ Surrogate outcomes: responses that have been shown to be associated with clinical outcomes but do not in themselves directly reflect survival or the clinical status of the subject.

e.g.: blood pressure, cholesterol, and tumor size

- ⋆⋆ Ascertained more quickly and more sensitive to the direct effect of treatment
- ⋆⋆ Require fewer subjects and follow-up time ⇒ trials can be conducted more quickly and efficiently than trails using clinical outcomes.
- ⋆⋆ Ideal for early phase trials (not intended to definitively establish efficacy)
- ⋆⋆ For more, read Chapter 2.4

Table 1: Elicited Utilities, $u_{\text{cycle}}(y)$.

Efficacy Scores	Toxicity Severity Level		
	Mild	Moderate	Severe
PD	25	10	0
SD	70	50	25
PR/CR	100	80	50

Figure 1: The decision problem consists of an alternating sequence of actions d_c and responses (Y_c, Z_c).

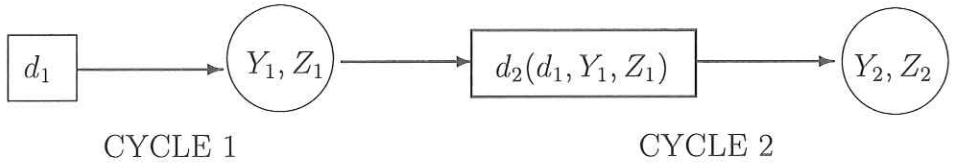
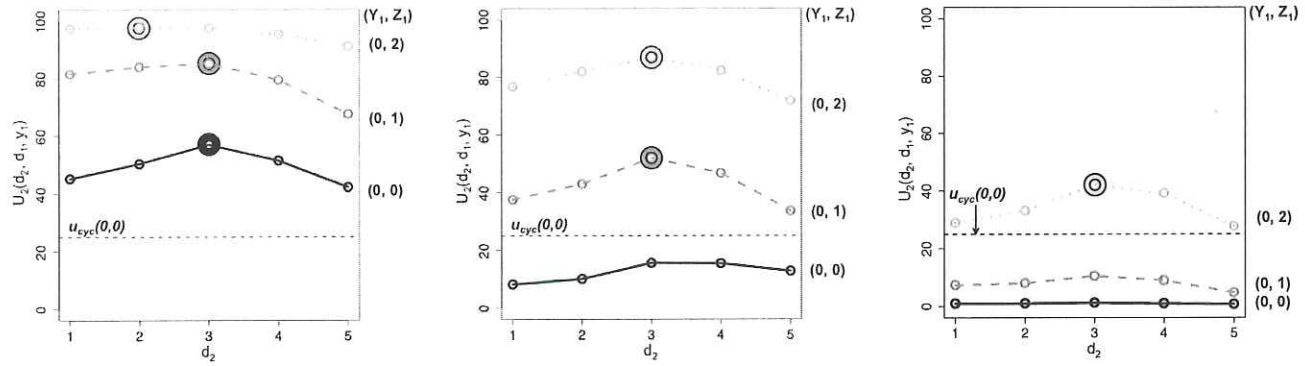


Figure 2: The true expected cycle 2 utilities of taking d_2 given $\boldsymbol{y}_1 = (Y_1, Z_1)$, $U_2(d_2, d_1, \boldsymbol{y}_1)$ with $d_1 = 3$ for scenario 3. $d_2^\star$ is marked with a **bold** circle for each $\boldsymbol{y}_1$ given $d_1 = 3$ if the corresponding expected utility is greater than $u_{\text{cycle}}(0, 0)$. If not, $d_2^\star = NT$ and none of d_2 is marked with a **bold** circle.



(a) $Y_1 = 0$ (b) $Y_1 = 1$ (c) $Y_1 = 2$

↻ Composite outcomes: outcome that is obtained by combining two or more distinct responses into a single outcome

- ⋆⋆ The full effect of treatment cannot be captured meaningfully by a single outcome and a hierarchy of responses is required.

  e.g.: For subjects with a respiratory disorder such as emphysema, both survival and pulmonary function are of interest and a composite outcome could be constructed using 6 months survival and 6 month pulmonary function among the survivors.

  How to combine? an outcome score can be constructed in which subjects who die are assigned a low value, otherwise the pulmonary function score is used.

- ⋆⋆ If treatment is expected to provide benefit for several different responses, it may be possible to increase power by combining them into a single outcome.

TRT 1

non fatal outcome (bad) \ survival	Y	N	
Y	30	0	30
N	70	1	71
			101

TRT 2

non fatal outcome \ survival	Y	N	
Y	3	0	3
N	27	70	97

- Outcome or Response Measures:
  - ⋆⋆ Binary

    e.g.: the presence or absence of a condition
  - ⋆⋆ Ordered categorical

    e.g.: severity of heart failure ranging from I (no symptoms) to IV (severe symptoms)
  - ⋆⋆ Continuous

    e.g.: blood pressure and blood chemistries
  - ⋆⋆ Failure time

- Revisit the TNT trial.
  - LDL levels is only one simple outcome among a number of possible outcomes as the effect of the drug.
  - LDL levels are a *surrogate* outcome and its effect is only meaningful if it results in a decrease in the risk of adverse outcomes associated with the progression of cardiovascular disease.
  - For simplicity, let's consider only the effect of treatment on LDL levels rather than clinical events.

- The p-value is extremely small. $\Rightarrow$ the $80mg$ dose results in highly statically significant additional LDL lowering relative to the $10mg$ dose.

Table 2.1 *Baseline and three-month LDL levels in TNT for subjects with values at both time points.*

Dose	N	Mean LDL (mg/dL) Baseline	3-Month
$80mg$	4868	97.3	72.5
$10mg$	4898	97.6	99.0
Difference		0.3	26.5

p-value for difference < 0.0001

↺ Suppose that we conclude that the observed mean difference of 26.5 *mg* is clinically relevant in spite of a possible increase in the risk of serious adverse side-effects with the higher dose might offset the potential benefit.

↺ Let's think a bit further. The validity of the claim relies on several factors;

- A new patient is comparable to the typical subject enrolled in the trial.
- An unbiased estimate of the treatment effect can be obtained from the data.
- The model used for estimation adequately describes the treatment effect.

Is a new patient comparable to the typical subject enrolled in the trial??

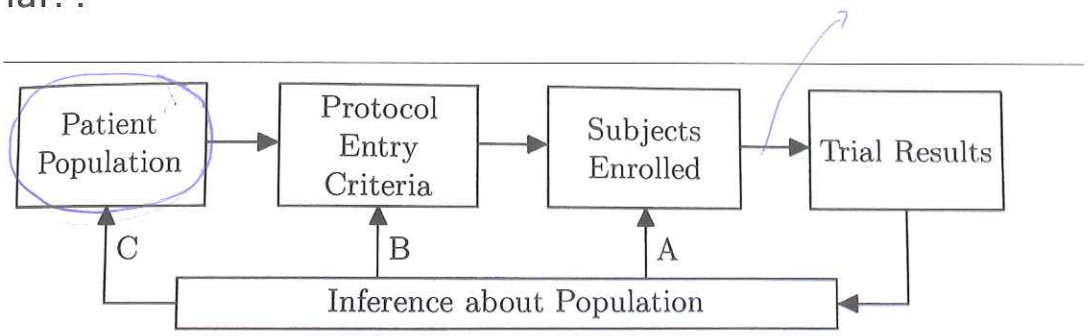


Figure 2.1 *Populations for which clinical trial inference is conducted.*

- Subjects in clinical trials are generally seen more frequently in the clinic and subject to more extensive diagnostic tests than other patients. $\Rightarrow$ maybe not representative of patient population.
- Estimates of the benefits of a new intervention obtained in a clinical trial may bear little resemblance to what will be observed in general practice.

↺ Does the observed difference in outcomes reflect the actual effect of treatment?

- 235 subjects had missing values of LDL at the three-month visit.
  - ∗ Likely that the reason that these data are missing is related to the condition of the subjects.
- 246 subjects discontinues their assigned study medication before the three-month visit (received only a fraction of their assigned medication).
  - ∗ may not reflect the full effect of treatment.
- In most trials the treatment effect under full adherence can never be obtained ⇒ statistical methods to account for missing data are needed.

↺ Is the model adequate to estimate the effect? How is the outcome of three-month LDL meaningfully defined?

- 8 subjects in the $80mg$ group and 7 subjects in the $10mg$ group died before the three month visit.
- What if there is an association between treatment, the outcome of interest (LDL) and mortality?
- Restricting the model to only survivors may cast doubt on whether observed associations are causal or simply artifacts of more complex interrelationships between the outcome, treatment, and morality.
- Cannot be demonstrated on purely statistical grounds, but *requires informed clinical judgment.*

- Subgroup questions: another category of secondary questions

e.g.: effects of treatment in specific baseline subgroups of subjects, demographic factors, disease categories, or known risk factors.

  - ⋆⋆ Subgroups may be used to confirm a result from a previous trial.
  - ⋆⋆ Subgroup analysis can also be used to generate new hypotheses.

- The TNT population is divided into approximately equal sized groups (quartiles) by baseline LDL.

Table 2.2 *Differences in three-month LDL levels in TNT as a function of baseline LDL. Note that the standard errors are sufficiently small that the overall trends are highly statistically significant.*

Baseline	Mean Difference in Three-month LDL (mg/dL)	% change in LDL 3-Month
LDL $\leq$ 85	23.2	30.7
85 $<$ LDL $\leq$ 97	25.1	27.4
97 $<$ LDL $\leq$ 110	27.1	26.2
110 $<$ LDL	29.4	24.6

- The differences in LDL between treatment groups vary as a function of baseline value.
- Interaction between change in LDL and the baseline.
- The expected difference in mean response between groups in a particular trial depends on the distribution of subjects enrolled in the trial

- Safety questions.
  - ⋆⋆ To fully evaluate the benefit to risk ratio, it is equally important that the safety profile should be established.
  - ⋆⋆ Generally obtained from a broad range of standard assessments including laboratory tests and reports of adverse events.

    e.g.: Adverse events: death, life-threatening, hospitalization (initial or prolonged), disability, congenital anomaly or requires intervention to prevent permanent impairment or damage.
  - ⋆⋆ While the efficacy questions can be limited, the safety questions must be comprehensive.

$$\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = \sigma\sqrt{\frac{1}{n_1}+\frac{1}{N-n_1}}$$

An important feature in clinical trial design is the allocation of subjects to treatments and is typically done through the process of **randomization** (DC Chapter 5).

- In observational studies, treatments are assigned based on a patient's diagnosis $\Rightarrow$ selection bias.
- Selection bias? Bias affecting the interventions that a patient may receive or which individuals are entered into the study.
  - Propensity score: the probability of treatment assignment conditional on observed baseline characteristics.

Randomized controlled trials are the "gold standard" of medical research.

Randomization serves as the basis for inference in the "frequentist approach, whereas the basis for inference in the Bayesian approach is subjective probabilities, which does not require randomization.

↺ The role of randomization: protect against sources of bias due to confounding with known and, especially, unknown risk factors.

- Randomization ensures independence between assigned treatment and outcome (no selection bias).

  $\Rightarrow$ allow us to attribute observed differences between treatment groups not attributable to chance to the causal effect of the treatment

- Randomization produces comparable groups with regard to measured and unmeasured risk factors.

  $\Rightarrow$ make the comparison between treatments more credible.

↺ Usually the total sample size is predetermined.

↺ Some allocation schemes make only limited use of randomization.

- Some allocation schemes make only limited use of randomization.
- There are differences in the Bayesian and frequentist views of randomization .
  - *Frequentist:* randomization serves as the basis for inference
  - *Bayesian:* the basis for inference is subjective probability, which does not require randomization.

  "In research situations usually approached by Decision Theory, it is only considered one researcher who collects a sample and makes a decision based on it. It can be shown that randomization of the sample does not improve the utility of the obtained results."

  from In Defense of Randomization: A Subjectivist Bayesian Approach by Bonassi et al.

↻ Fixed randomization procedures: randomization with some restrictions.

- ⋆⋆ Restrictions are made to avoid imbalance (i.e., inequality of treatment group sizes).
- ⋆⋆ Reasoning: Increased imbalance results in decreased power for statistical test.
- ⋆⋆ Concern about selection bias due to the restrictions.

  e.g. random allocation rule, complete randomization, and permuted-block randomization.

- In specific, *permuted-block randomization*
  - ∗ Suppose two treatments. We have blocks, each of size m (must be even).
  - ∗ Within each block $m/2$ subjects are assigned to each treatment.

Say, m = 4

1, 1, 2, 2

1 2 2 1

2 2 1 1