2. From Scientific Setting to Statistical Design

2.1 Refining scientific questions to statistical hypotheses (Chapter 5)
   (a) Refining endpoints and the patient population
   (b) Statistical model for scientific inference
       (i) Choice of probability model
       (ii) Choice of functional
       (iii) Choice of contrast (statistical model)
   (c) Statement of statistical hypotheses

2.2 From statistical hypotheses to scientific decisions (Chapter 7)
   (a) Translation from sample space in to parameter space
   (b) Statistical inference in clinical trials
   (c) Asymptotic inference

2.3 Case Study (PLCO trial; DPP trial)
2.1 Refining scientific questions to statistical hypotheses
(a) Refining endpoints and the patient population

- It’s a t-test (points covered during the discussion):
  - Comments on type of endpoint
  - Comments on type of trial
  - Comments on choice of analysis
  - Issues that must be solved during pre-trial design rather than at post-trial analysis

- Refinements to the endpoint
- Refinements to the patient population
2.1 Refining scientific questions to statistical hypotheses
(a) Refining endpoints and the patient population

- Comments on type of endpoint:
  - Continuous endpoints: mean-based inference leads to the t-test by the CLT *regardless of the underlying distribution.*
  - Binary endpoints: A Chi-square test is a t-test
  - Time-to-event endpoints: special methods are required to address censoring. (Somewhat controversial note: log-rank test is intransitive; consider restricted mean survival.)
  - Longitudinal endpoints: Select the scientifically-relevant aspect of the time trajectory (often leading to a t-test of summary measures - see DPP example, below.)
  - Ordinal endpoints: What is the question?
  - Nominal endpoints: What is the question?
2.1 Refining scientific questions to statistical hypotheses

(a) Refining endpoints and the patient population

- Comments on type of trial:
  - **Single-group trial**: Historic control group is implied. Inference focuses on estimation within the single arm rather than comparing groups.
  - **Two-group trial**: It’s a t-test!
  - **K-group trial**: What is the question? Do we need to control for multiple comparisons? Why isn’t it a t-test?
    - Two or more treatments compared with common control group (all possible comparisons to control?)
    - Multiple dose groups (dose-response trial?)
    - Is interest in all possible 2-way comparisons? Interested in any difference (ANOVA)?
    - 2 × 2 factorial design (looking for interaction? looking for efficiency?)
    - (Note: For really large $K$, we are usually interested in association; i.e., linear regression.)
2.1 Refining scientific questions to statistical hypotheses

(a) Refining endpoints and the patient population

- Comments on choice of analysis:
  - Why isn’t it a t-test? (Incorrect answers)
    - The data are not normally distributed, so we need to use a non-parametric test.
      (INCORRECT)
    - A Wilcoxon test is a test of medians and does not require parametric assumptions.
      (INCORRECT: a Wilcoxon test is not a test of medians; a Wilcoxon test is intransitive.)
    - I need to adjust for confounding.
      (INCORRECT: in RCT randomization removes confounding.)
2.1 Refining scientific questions to statistical hypotheses
(a) Refining endpoints and the patient population

- Comments on choice of analysis:
  - Why isn’t it a t-test? (It is, you just don’t recognize it.)
    - It’s a binary endpoint so we need to use a chi-square test.
      (A chi-square test is a t-test.)
    - I have censored data.
      (It is essentially a t-test only special methods are required to address informative missingness.)
    - I have correlated (clustered) data.
      (Need a t-test that correctly accounts for the clustering.)
    - I need to condition on a stratification variable.
      (Regression methods are a subgroup-specific t-test.)
2.1 Refining scientific questions to statistical hypotheses

(a) Refining endpoints and the patient population

- Comments on choice of analysis:
  - Why isn’t it a t-test? (Please refine your question.)
    - I have more than 2 groups.
      (What is the question; is it simply multiple 2-group t-tests?)
    - I have longitudinal data.
      (What aspect of the longitudinal trajectory is scientifically relevant?)
    - I have ordered groups (e.g., multiple doses)
      (Are you interested in the dose-response relationship?)
      (Are you interested in the dose with the highest response?)
      (Are you interested in the dose with lowest toxicity for adequate response?)
2.1 Refining scientific questions to statistical hypotheses
(a) Refining endpoints and the patient population

Notes from Chapter 5 (Refining the endpoint):

- Improvements to avoid bias. Examples:
  - Effectiveness of home safety (Southland, NZ)
  - School environment study (San Luis Valley, CO)
  - Blinding (as a general principle)

- Improving precision. Examples:
  - Standardization of measurement procedures (graded exercise tests, blood pressure measurement)
  - Replication of measurements (may increase participant burden)

- Misclassification of outcomes.
  - Central (blinded) adjudication committee.

- Attenuation of effects due to inclusion of non-susceptibles.
2.1 Refining scientific questions to statistical hypotheses

(a) Refining endpoints and the patient population

- Notes from Chapter 5 (Longitudinal endpoints):
  - Note that all endpoints have a reference timeframe, and that the timeframe is an essential consideration. Examples:
    - Sepsis trials (including Daptomycin)
    - PLCO and DPP (see below)

- Notes from Chapter 5 (Risk of missing data):
  - Distinguish between:
    - Lack of compliance (it may be a result of treatment; you should still measure the outcome)
    - Loss to follow-up (minimize through good design; reduce participant burden)
    - Withdrawal of consent (can the subject be retained for key follow-up visits only?)
  - Missing data are bad (there is no foolproof adjustment; the only correct approach is to minimize it)
2.1 Refining scientific questions to statistical hypotheses

(a) Refining endpoints and the patient population

I now want to go through 2 examples to illustrate some of the issues in refining the endpoint and the patient population:

- Example 1: Prostate, Lung, Colorectal, and Ovarian (PLCO) screening trial; NEJM 2009; 360:1310-9
- Example 2: Diabetes Prevention Program (DPP) trial; NEJM 2002; 346:393-403
2.1 Refining scientific questions to statistical hypotheses
(a) Refining endpoints and the patient population

Example (PLCO trial; NEJM 2009; 360:1310-9):

- Setting: PSA screening for prostate cancer is common, but the tradeoff between risks and benefits is not known.
- Public Health Question: Should screening for prostate cancer be part of routine practice?
- Refining the question: What information is needed to recommend for/against routine screening?
2.1 Refining scientific questions to statistical hypotheses
(a) Refining endpoints and the patient population

- Example (PLCO trial; NEJM 2009; 360:1310-9):
  - Potential trial endpoints:
  - Refining the endpoint:
2.1 Refining scientific questions to statistical hypotheses
(a) Refining endpoints and the patient population

- Example (PLCO trial; NEJM 2009; 360:1310-9):
  - Possible study populations:

  - Refining the study population:
2.1 Refining scientific questions to statistical hypotheses

(a) Refining endpoints and the patient population

- Example (PLCO trial; NEJM 2009; 360:1310-9):

  Other considerations:
  - Inclusion of non-susceptibles:
    - What if half of the 76,693 were women?
    - Incidence would be reduced by half.
  - How should the endpoint be ascertained?
    - How is death from prostate cancer determined?
  - Was there complete follow-up?
  - What is the potential effect of drop-in (screening in usual care group) and drop-out (failure to screen in screening group)?
2.1 Refining scientific questions to statistical hypotheses
(a) Refining endpoints and the patient population

- Example (PLCO trial; NEJM 2009; 360:1310-9):
  - **Design**:
    - Groups: Annual screening (PSA and DRE) vs usual care.
    - Randomized: 38,343 to screening, 38,350 to usual care.
    - Primary endpoint: Death from prostate cancer
      \[ \theta_1 = \text{incidence rate with screening} \text{ (deaths per person year)} \]
      \[ \theta_0 = \text{incidence rate with usual} \text{ (deaths per person year)} \]
      \[ \theta = \theta_1 / \theta_0 \]
  - **Primary Results (7-year)**:
    \[ \hat{\theta}_1 = 50 \text{ deaths in 254,295py (2.0 per 10,000py)} \]
    \[ \hat{\theta}_0 = 44 \text{ deaths in 253,317py (1.7 per 10,000py)} \]
    \[ \hat{\theta} = 1.13 \text{ (95% CI: 0.75-1.70)} \]
  - **Cancer diagnosis (7-year)**:
    - Cases diagnosed: 2820 (screened); 2322 (usual care);
    - Rate ratio: 1.22 (95% CI: 1.16 to 1.29)
2.1 Refining scientific questions to statistical hypotheses
(a) Refining longitudinal endpoints

- Outcomes are frequently measured over time
- We need to pre-specify the aspect of the time trajectory that is scientifically important
- Notation:
  - Let $Y_{ik}(t)$ denote the outcome in patient $i$ ($i = 1, \ldots, N$), treatment group $k$ ($0 = \text{control}, 1 = \text{new treatment}$), at time $t = 0, 1, 2, \ldots, L$ with $t = 0$ at baseline.
  - Suppose $Y_{ik}(t)$ has expectation $\mu_k(t)$ and variance $\Sigma_k$ with diagonal elements $\sigma^2_{kt}$ and covariances $\sigma_{kst}$. 
2.1 Refining scientific questions to statistical hypotheses
(a) Refining longitudinal endpoints

Types of scientific questions:
- Difference at any time.
- Weighted average across time:

\[ \theta_k = \sum_{t=0}^{L} w_t \mu_k(t) \]

where \( w_t \) denotes the weight at each time.

Examples:

- Final time point: \( w = (0, 0, \ldots, 0, 1) \)
- Change \((L - 0)\): \( w = (-1, 0, \ldots, 0, 1) \)
- Average change: \( w = \left(-1, \frac{1}{L}, \ldots, \frac{1}{L}\right)\)
- Rate of change: \( w = \left(\frac{-L}{2}, \frac{-L}{2} + 1, \frac{-L}{2} + 2, \ldots, \frac{L}{2} - 1, \frac{L}{2}\right)\)
2.1 Refining scientific questions to statistical hypotheses
(a) Refining longitudinal endpoints

Example (Diabetes Prevention Project):

- **Design**: Overweight but healthy individuals randomized to Placebo, Metformin, Lifestyle interventions.
- **Outcomes**: Incidence of type II diabetes
  - Primary: Incidence of type II diabetes
  - Secondary included weight measured at times $t = (0, 6, 12, 18, 24, 30, 36, 42, 48)$ months
- **Data**:
  \[
  Y_{i,0}(t) = \text{weight in } i\text{th subject in placebo group at time } t \\
  Y_{i,1}(t) = \text{weight in } i\text{th subject in lifestyle group at time } t \\
  Y_{i,2}(t) = \text{weight in } i\text{th subject in metformin group at time } t
  \]
- **Hypotheses**:
  - Lifestyle (and metformin): early weight loss with slow regain
  - Placebo: steady (slow) weight gain.
2.1 Refining scientific questions to statistical hypotheses

(a) Refining longitudinal endpoints

• Example (Diabetes Prevention Project):
  ▶ Some potential summary measures:

  Final time point : \( w = (0, 0, \ldots, 0, 1) \)
  Change \((L - 0)\) : \( w = (-1, 0, \ldots, 0, 1) \)
  Average change : \( w = \left( -1, \frac{1}{L}, \ldots, \frac{1}{L} \right) \)
  Rate of change : \( w = \left( \frac{-L}{2}, \frac{-L}{2} + 1, \frac{-L}{2} + 2, \ldots, \frac{L}{2} - 1, \frac{L}{2} \right) \)

  ▶ Scientific relevance?: (Note: weight is surrogate for health consequences)
    ★ Weight is a surrogate for other adverse health effects.
    ★ Permanent weight loss is good.
    ★ Is loss followed by very slow regain good/bad?
    ★ Is weight cycling good/bad?
2.1 Refining scientific questions to statistical hypotheses

(a) Refining longitudinal endpoints

Example (Diabetes Prevention Project):

- Different summary measures emphasize different aspects of the time trajectory.
- Select a summary measure that is the best indicator of risk for health problems and best reflects the treatment effect on that risk.
- Measure treatment effects as:
  - Change from baseline: $\delta_k(t) = \mu_k(t) - \mu_k(0)$
  - Time trajectory summary: $\theta_k = \sum_{t=1}^{L} w_t \delta_t$
  - Difference between treatment groups: $\theta = \theta_1 - \theta_0$
  - (Note: $\theta < 0$ favors lifestyle.)
Potential time trajectories

Different magnitude of loss (kg)

Months post randomization

θ

-10 -8 -6 -4 -2 0

0 6 12 18 24 30 36 42 48
Potential time trajectories
Potential time trajectories
Potential time trajectories
Potential time trajectories
Potential time trajectories (ordered by “last” contrast)
Potential time trajectories (ordered by “last” contrast)
Potential time trajectories (ordered by “last” contrast)
Potential time trajectories (ordered by “last” contrast)
### Potential time trajectories (ordered by “avg” contrast)

<table>
<thead>
<tr>
<th>-7.1</th>
<th>-6.2</th>
<th>-6.2</th>
<th>-6.2</th>
<th>-6.2</th>
<th>-6.2</th>
<th>-6.2</th>
<th>-6.2</th>
<th>-6.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
</tr>
<tr>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
</tr>
<tr>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
</tr>
<tr>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
<td>-5.3</td>
</tr>
<tr>
<td>-5.3</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
</tr>
<tr>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
</tr>
<tr>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
</tr>
<tr>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
</tr>
<tr>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
</tr>
<tr>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
<td>-4.4</td>
</tr>
</tbody>
</table>
Potential time trajectories (ordered by “avg” contrast)
Potential time trajectories (ordered by "avg" contrast)

2. Scientific Setting to Stat Design ()
2.1 From Science to Stat Hypotheses

14 Feb 2011 32 / 68
Potential time trajectories (ordered by “avg” contrast)
Potential time trajectories (ordered by “slope” contrast)
Potential time trajectories (ordered by “slope” contrast)
Potential time trajectories (ordered by “slope” contrast)
Potential time trajectories (ordered by “slope” contrast)
Potential time trajectories (ordered by variation)
Potential time trajectories (ordered by variation)
Potential time trajectories (ordered by variation)
Potential time trajectories (ordered by variation)
2.1 Refining scientific questions to statistical hypotheses

(a) Refining longitudinal endpoints

- Example (Diabetes Prevention Project):
  - Higher values for the summary measure give more power.
  - Want high power for scientifically important trajectories.
  - Do not want high power for “null” trajectories.
  - In this example the average is best at detecting immediate weight loss followed by long-term maintenance.
  - Other measures are better in other settings; examples:
    - Slope (for slowly evolving processes)
    - Value at a scientifically-meaningful timepoint (for variable trajectories that reach steady state).
2. From Scientific Setting to Statistical Design

2.1 Refining scientific questions to statistical hypotheses (Chapter 5)

(a) Refining endpoints and the patient population
(b) Statistical model for scientific inference
   (i) Choice of probability model
   (ii) Choice of functional
   (iii) Choice of contrast (statistical model)
(c) Statement of statistical hypotheses
2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

- Upon trial completion we estimate $\theta$ and report (scientific inference):
  - Point estimate: $\hat{\theta}$
  - Interval estimate: $(\theta_L, \theta_U)$.
  - p-value

- What are the basic elements of the statistical model that are required for the above scientific inference?
2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

- Common statistical model for continuous data (MLE):
  - Probability model:
    
    \[ Y_{i1} \sim \mathcal{N}(\mu_1, \sigma_1^2) \]
    
    \[ Y_{i0} \sim \mathcal{N}(\mu_0, \sigma_0^2) \]

  - Functional (population mean):
    
    \[ \theta_1 = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2} \left( \frac{x-\mu_1}{\sigma_1} \right)^2} dx \]
    
    \[ \theta_0 = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2} \left( \frac{x-\mu_0}{\sigma_0} \right)^2} dx \]

  - Contrast: \( \theta = \theta_1 - \theta_0 \).
2.1 Refining scientific questions to statistical hypotheses
(b) Statistical model for scientific inference

- Common statistical model for continuous data (moment estimator):
  - Probability model:
    \[
    Y_{i1} \sim F_1 \quad \text{with} \quad E(Y_{i1}) = \mu_1, \quad \text{var}(Y_{i1}) = \sigma^2_1 \\
    Y_{i0} \sim F_0 \quad \text{with} \quad E(Y_{i0}) = \mu_0, \quad \text{var}(Y_{i0}) = \sigma^2_0
    \]

  By CLT:
  \[
  \hat{\theta}_1 = \overline{Y}_1 \sim \mathcal{N} \left( \mu_1, \frac{\sigma^2_1}{N} \right) \\
  \hat{\theta}_0 = \overline{Y}_0 \sim \mathcal{N} \left( \mu_0, \frac{\sigma^2_0}{N} \right)
  \]
2.1 Refining scientific questions to statistical hypotheses
(b) Statistical model for scientific inference

- Common statistical model for continuous data (moment estimator):
  - Functional (population mean):
    - Contrast: $\theta = \theta_1 - \theta_0$. 

\[
\begin{align*}
\theta_1 &= \int_{-\infty}^{\infty} x \frac{\sqrt{N}}{\sqrt{2\pi}\sigma_1^2} e^{-\frac{N}{2} \left( \frac{x-\mu_1}{\sigma_1} \right)^2} \, dx \\
\theta_0 &= \int_{-\infty}^{\infty} x \frac{\sqrt{N}}{\sqrt{2\pi}\sigma_0^2} e^{-\frac{N}{2} \left( \frac{x-\mu_0}{\sigma_0} \right)^2} \, dx
\end{align*}
\]
2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

- Common statistical model for binary data (MLE):
  - Probability model:

\[
Y_{i1} \sim \mathcal{B}(1, p_1) \\
Y_{i0} \sim \mathcal{B}(1, p_0)
\]

- Functional (population mean):

\[
\theta_1 = \sum_{k=1}^{N} \frac{k}{N} \frac{N!}{k!(N-k)!} p_1^k (1 - p_1)^{N-k} \\
\theta_0 = \sum_{k=1}^{N} \frac{k}{N} \frac{N!}{k!(N-k)!} p_0^k (1 - p_0)^{N-k}
\]

- Contrasts: \( \theta = \theta_1 - \theta_0 \)
  \( \theta = \frac{\theta_1}{\theta_0} \)
2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

- Common statistical model for binary data (MLE):
  - Probability model:
    
    \[ Y_{i1} \sim \mathcal{B}(1, p_1) \]
    \[ Y_{i0} \sim \mathcal{B}(1, p_0) \]

  - Functional (population odds):
    
    \[ \theta_1 = \frac{p_1}{1 - p_1} \]
    \[ \theta_0 = \frac{p_0}{1 - p_0} \]

  - Contrast: \( \theta = \frac{\theta_1}{\theta_0} \)
2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

- Common statistical model for censored time-to-event data:
  - Probability model ($Y_{ik}$ denotes time to event):
    
    \[
    \begin{align*}
    Y_{i1} & \sim F_1(y) \\
    Y_{i0} & \sim F_0(y)
    \end{align*}
    \]

    with $1 - F_1(y) = [1 - F_0(y)]^{e^\theta}$.

    (Note: $F_k(y)$ denotes the probability cdf; i.e., $F_k(y) = P(Y_{ik} \leq y)$.)

  - This is an example of a “semi-parametric model”.
2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

- Common statistical model for censored time-to-event data:
  - The functional and contrast are implied in the probability model (proportional hazards model).
  
  Let $S_k(y) = 1 - F_k(y)$ be the survival function:

  $\begin{align*}
  S_1(y) &= [S_0(y)]^e^\theta \\
  \log[S_1(y)] &= e^\theta \log[S_0(y)] \\
  \frac{d}{dy} \log[S_1(y)] &= \frac{d}{dy} e^\theta \log[S_0(y)] \\
  \lambda_1(y) &= e^\theta \lambda_0(y) \\
  e^\theta &= \frac{\lambda_1(y)}{\lambda_0(y)}
  \end{align*}$

  where $\lambda_k(y)$ denotes the hazard of an event at time $y$. 
2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

- Elements of the statistical model for scientific inference:
  2. *Functional*: A summary measure on the probability model.
2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

- **Probability Model:**
  - Note: with the exception of binary data, the true probability distribution for the data is unknown; i.e., $Y_{i1} \sim F_1(y)$.
  - Common representations:
    - **Parametric models:** Common forms include:
      - **Normal:** $Y_{ik} \sim \mathcal{N}(\mu_k, \sigma_k^2)$:
        
        $E(Y_{ik}) = \mu_k; \; \text{var}(Y_{ik}) = \sigma_k^2$
      
      - **Binomial:** $Y_{ik} \sim \mathcal{B}(1, p_k)$:
        
        $E(Y_{ik}) = p_k; \; \text{var}(Y_{ik}) = p_k(1 - p_k)$
      
      - **Poisson:** $Y_{ik} \sim \mathcal{P}(\lambda_k)$:
        
        $E(Y_{ik}) = \lambda_k; \; \text{var}(Y_{ik}) = \lambda_k$
      
      - **log-Normal:** $\log(Y_{ik}) \sim \mathcal{N}(\mu_k, \sigma_k^2)$:
        
        $E(Y_{ik}) = e^{\mu_k+\sigma^2/2}; \; \text{var}(Y_{ik}) = e^{2\mu_k+\sigma^2}(e^{\sigma^2} - 1)$
      
    - **Others:** Exponential, Weibull
2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

- Probability Model:
  - Common representations:
    - Semi-parametric probability models: Common forms include:
      - Proportional Hazards: \( S(y) = S_0(y)^{e^{X\theta}} \)
      - Shift family: \( F(y) = F_0(y + \theta) \)
    - Non-parametric (infinite-parameter) probability models:
      \[
      Y_{i1} \sim F_1(y) \\
      Y_{i0} \sim F_0(y)
      \]
2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

- Functional: Characteristic of the probability model that best captures the effects of treatment
  - Common choices:

\[
\begin{align*}
\text{Mean:} & \quad \theta_{k}^{(\text{mean})} = \int y f_{k}(y) dy \\
\text{Geometric mean:} & \quad \theta_{k}^{(\text{geo})} = \exp \left[ \int \ln(y) g(y) dy \right] \\
\text{Restricted mean:} & \quad \theta_{k}^{(\text{Ravg})} = \int_{0}^{T} y f_{k}(y) dy \\
\text{Median:} & \quad \theta_{k}^{(\text{med})} \text{ s.t. } \int_{-\infty}^{\theta_{k}^{(\text{med})}} f_{k}(y) dy = 0.5
\end{align*}
\]
2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

- Functional: Characteristic of the probability model that best captures the effects of treatment
  - Common choices:

    Proportion above $W$:
    \[ \theta_{k}^{(prop)} = F_{k}(W) \]

    Odds above threshold:
    \[ \theta_{k}^{(odds)} = \frac{\theta_{k}^{(prop)}}{1 - \theta_{k}^{(prop)}} \]

    cdf:
    \[ \theta_{k}^{(cdf)} = F_{k}(y) \]

    Hazard:
    \[ \theta_{k}^{(haz)} = \frac{f_{k}(y)}{S_{k}(y)} \]
2.1 Refining scientific questions to statistical hypotheses
(b) Statistical model for scientific inference

- Contrast: Comparison of groups
  - Common choices:
    - Difference: $\theta = \theta_1 - \theta_0$
    - Ratio: $\theta = \frac{\theta_1}{\theta_0}$
    - Wilcoxon: $\theta = Pr(Y_{i1} > Y_{i0})$
    - cdf Difference: $\theta = \sup_y [F_1(y) - F_0(y)]$
2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

- Example:
  12-month weight change (e.g., DPP)

<table>
<thead>
<tr>
<th></th>
<th>Placebo</th>
<th>Lifestyle</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>995</td>
<td>995</td>
</tr>
<tr>
<td>Mean (kg)</td>
<td>-0.42</td>
<td>-6.84</td>
</tr>
<tr>
<td>Std Dev (s)</td>
<td>4.78</td>
<td>7.04</td>
</tr>
</tbody>
</table>
2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

- Example: 12 month weight change (e.g., DPP)
  - Scientific Inference (model A):

<table>
<thead>
<tr>
<th>Prob Model:</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional:</td>
<td>Mean</td>
</tr>
<tr>
<td>Contrast:</td>
<td>Difference</td>
</tr>
</tbody>
</table>

\[ Y_{ik} \sim \mathcal{N}(\mu_k, \sigma_k^2) \]
\[ \theta_k = \mu_k \]
\[ \theta = \theta_1 - \theta_0 \]

**Results:**
- Point estimate
  \[ \hat{\theta} = \hat{\theta}_1 - \hat{\theta}_0 \]
  -6.42
- Standard error
  \[ SE^2 = \frac{s_1^2}{N} + \frac{s_0^2}{N} \]
  0.270
- Interval estimate
  \[ \hat{\theta} \pm 1.96 \times SE \]
  (-6.94, -5.89)
- P-value
  \[ Z = \hat{\theta}/SE \]
  \(< < 0.00001\)
2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

Example: 12 month weight change (e.g., DPP)

Scientific Inference (model B):

\[ Y_{ik} \sim \mathcal{N}(\mu_k, \sigma_k^2) \]
\[ \theta_k = \mu_k \]
\[ \theta = \theta_1 / \theta_0 \]

Results:

Point estimate \[ \hat{\theta} = \hat{\theta}_1 / \hat{\theta}_0 \]

16.273

Standard error* \[ SE^2 = \frac{s_1^2}{N\hat{\theta}_1^2} + \frac{s_0^2}{N\hat{\theta}_0^2} \]

0.362

Interval estimate \[ \hat{\theta} \pm 1.96 \times SE \]

(8.00, 33.11)

P-value \[ Z = \hat{\theta} / SE \]

<< 0.00001

*SE for log(\(\hat{\theta}\))
2.1 Refining scientific questions to statistical hypotheses
(b) Statistical model for scientific inference

- Example: 12 month weight change (e.g., DPP)
  - Scientific Inference (model C):

  \[
  \text{Prob Model: Non-parametric} \quad Y_{ik} \sim F_k(y)
  \]
  \[
  \text{Functional: Mean} \quad \theta_k = \mu_k
  \]
  \[
  \text{Contrast: Difference} \quad \theta = \theta_1 - \theta_0
  \]
  \[
  \text{Results:}
  \]
  \[
  \text{Point estimate} \quad \hat{\theta} = \hat{\theta}_1 - \hat{\theta}_0 \quad -6.42
  \]
  \[
  \text{Standard error} \quad SE^2 = \frac{s_1^2}{N} + \frac{s_0^2}{N} \quad 0.270
  \]
  \[
  \text{Interval estimate} \quad \hat{\theta} \pm 1.96 \times SE \quad (-6.94, -5.89)
  \]
  \[
  \text{P-value} \quad Z = \hat{\theta}/SE \quad << 0.00001
  \]
2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

Example: 12 month weight change (e.g., DPP)

Scientific Inference (model D):

- **Prob Model:** Non-parametric
- **Functional:** Mean
- **Contrast:** Ratio

Results:

- Point estimate: $\hat{\theta} = \hat{\theta}_1 / \hat{\theta}_0 = 16.273$
- Standard error*: $SE^2 = \frac{s_1^2}{N\hat{\theta}_1^2} + \frac{s_0^2}{N\hat{\theta}_0^2} = 0.362$
- Interval estimate: $exp[log(\hat{\theta}) \pm 1.96 \times SE] = (8.00, 33.11)$
- P-value: $Z = \hat{\theta} / SE << 0.00001$

*SE for log($\hat{\theta}$)
2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

- Example: 12 month weight change (e.g., DPP)
  - Scientific Inference (model E):

  \[
  \text{Prob Model: Normal} \quad Y_{ik} \sim \mathcal{N}(\mu_k, \sigma^2_k) \\
  \text{Functional: Median} \quad F_k(\theta_k) = 0.5 \\
  \text{Contrast: Difference} \quad \theta = \theta_1 - \theta_0 \\
  \]

  \[
  \begin{align*}
  \text{Results:} & & \\
  \text{Point estimate} & \hat{\theta} = \hat{\theta}_1 - \hat{\theta}_0 & \text{-6} \\
  \text{Standard error} & \sqrt{\frac{1}{4f_1^2(\hat{\mu}_1)N} + \frac{1}{4f_0^2(\hat{\mu}_0)N}} & 0.0863 \\
  \text{Interval estimate} & \hat{\theta} \pm 1.96 \times SE & (-6.17,-5.83) \\
  \text{P-value} & Z = \hat{\theta}/SE & << 0.00001
  \end{align*}
  \]


2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

- Example: 12 month weight change (e.g., DPP)
  - Scientific Inference (model F):
    - **Prob Model:** Normal
    - **Functional:** Median
    - **Contrast:** Ratio

\[
Y_{ik} \sim \mathcal{N}(\mu_k, \sigma_k^2)
\]
\[
\theta_k : F_k(\theta_k) = 0.5
\]
\[
\theta = \theta_1 / \theta_0
\]

**Results:**
- Point estimate: \( \hat{\theta} = \hat{\theta}_1 / \hat{\theta}_0 \)
  - \(-6/0 = -\infty\)
- Standard error:
  \[
  SE^2 = \frac{1}{\hat{\sigma}_1^2 \text{Var}(\hat{\mu}_1) N} + \frac{1}{\hat{\sigma}_0^2 \text{Var}(\hat{\mu}_0) N}
  \]
- Interval estimate:
  \[
  \exp[\log(\hat{\theta}) \pm 1.96 \times SE]
  \]
- P-value:
  \[
  Z = \hat{\theta} / SE
  \]

*SE for log(\(\hat{\theta}\))
**Not estimable
2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

- Example: 12 month weight change (e.g., DPP)
  - Scientific Inference (model G):

<table>
<thead>
<tr>
<th>Prob Model:</th>
<th>Semi-parametric</th>
<th>$S_1(y) = S_0(y)e^{\theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional:</td>
<td>Hazard</td>
<td>$\theta_k : \frac{f_k(y)}{S_k(y)}$</td>
</tr>
<tr>
<td>Contrast:</td>
<td>Ratio</td>
<td>$\theta = \theta_1 / \theta_0$</td>
</tr>
</tbody>
</table>

**Results:**
- Point estimate: $\hat{\theta} = \hat{\theta}_1 / \hat{\theta}_0$ = 2.54
- Standard error*: $SE^2 = \frac{1}{4} 2N$ = 0.0448
- Interval estimate: $exp[log(\hat{\theta}) \pm 1.96 \times SE]$ = 2.33-2.77
- P-value: $Z = \hat{\theta} / SE$ = $<< 0.00001$

*SE for log($\hat{\theta}$)
**Not estimable
2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

- Note: A test of hazards may reject even if the means are the same: (HR = 2.54; CI = (2.33-2.77)):
2.1 Refining scientific questions to statistical hypotheses

(b) Statistical model for scientific inference

- Note: A test of hazards may reject even if the means are the same:
  \( (HR = 0.77; CI = (0.70-0.84)) \):
2.1 Refining scientific questions to statistical hypotheses

(c) Statement of Statistical Hypotheses

- One sided (greater-than) test
  
  \[ H_0 : \theta \leq \theta_0 \]
  \[ H_+ : \theta \geq \theta_+ \]

- One sided (less-than) test
  
  \[ H_0 : \theta \geq \theta_0 \]
  \[ H_- : \theta \leq \theta_- \]

- Two sided test
  
  \[ H_- : \theta \leq \theta_- \]
  \[ H_0 : \theta = \theta_0 \]
  \[ H_+ : \theta \geq \theta_+ \]