Background and Motivation

• Unlike randomized control trials (RCTs), observational studies and patient registries typically address objectives rather than test specific hypotheses.
  > Nevertheless, estimation of sample size is an important part of the planning process.
  > A minimum sample size is required to ensure:
    - adequate exploration of the objectives
    - ensure sufficient generalisability

• Sample size estimation for observational studies is often complex
  > Subgroup analyses and modelling are to be expected in observational studies
    - methods require more assumptions and larger sample sizes.

• Analysis follows design
**Landscape**
(not necessarily unique to observational studies)

- Studies with an observational unit other than the patient
  - Patient-years
  - Site/provider
- Single-cohort studies
  - Outcome comparisons against historical comparator
  - Outcome comparisons where patients serve as their own control (i.e., historical control, paired responses)
  - No comparison
- Multiple comparison adjustments to support:
  - Comparisons between multiple study sites or multiple patient types
  - Comparisons using multiple interim/intermediate analyses ("refreshes of the data")

**Objective**

To explore sample size estimations performed for a variety of observational studies with an array of challenges and objectives:

- Missing Data
- Selection Bias
- Unknown Enrollment Distributions
- Time-to-Event
- Precision around Point Estimate
- Need for Generalisability of Results
Sample Size Estimation in Anticipation of Missing Data, Selection Bias and Unknown Enrollment Distributions

Eric Gemmen, MA

Missing Data

- Missing data is the norm in observational studies.
- Covariate-adjusted analyses use complete data only
- Example:
  > Required sample size = 300
  > Expect 5% of data randomly missing from 3 covariates
  > Complete data in \(100 \times 0.95^3 = 86\%\) of patients
  > Corrected sample size is \(300 / 0.86 = 350\)
- This approach is valid for complete-case analysis as well as multiple imputation.
Propensity Scores

• Used to adjust for non-randomized treatment assignments
• Estimated probability of receiving new treatment
  > Calculated for all treatment groups
  > Calculated using logistic regression
• Scores calculated for each patient after data collected
• Comparisons between patients with comparable propensity scores
• Incorporate scores in analysis:
  > Matching (quintiles, distance between scores)
  > Weighting
  > Scores as covariates

Propensity Score Ranges
Propensity Score Example

• Observational study comparing two treatments (new vs. standard) for rheumatoid arthritis.

• Primary endpoint is change in the DAS28-ESR score at six months.

• If this were a randomized trial, the sample size required per treatment group would be 448 (alpha = 0.05, 2-sided test, power = 80%).

• Sample size increased to 560, assuming 80% propensity score overlap.

Sample Size Re-Estimation Example

• Long-term longitudinal oncology study of treatment response and its association with genomic variants

• Real-world personalized medicine study with ‘standard-of-care’ treatments

• Initially, treatment patterns and distribution of genomic variants largely unknown

• Also, uncertain estimates for treatment response and dropout rates in initial sample size calculations

• Plan to re-estimate sample size after 10% and 20% of patients enrolled.

• No adjustment in P-value is planned
Time to Event

- Time to first (AE, MACE, Death, Progression)
  - Binary
  - 1 event same as 3 events

- Hazards Ratio – Ratio of the Hazard (event) rates
  - Patient-Years
  - 100 pts for 5 years = 500 pts over 1 year.

- Power
  - 90% or 80%
An Example – Oral Contraceptive

- New versus Established

- Primary endpoint:
  - Venous Thromboembolism (VTE)
  - Deep Vein Thrombosis

- Literature VTE:
  - No hormones = 1-5 in 10,000 women-years
  - Established OC = 2-9 in 10,000 women-years
  - Pregnancy = 5-20 in 10,000 women-years
  - Postpartum = 40-65 in 10,000 women-years

- Endpoint: Exclude two-fold risk of VTE
  - $H_0$: HR = 2
  - $H_1$: HR > 2
  - 1-sided
  - 90% Power – reject $H_0$, $H_1$ is true

Design and Analytic Approaches

- Recruitment = 3 years
- Follow-up = 5 years
- Drop-out rate = 5%
- Planned Analyses
  - Test based on exponential survival, accrual period and dropouts

Method

\[
event = \frac{z(1-\alpha) + z(1-\beta)}{\left[\pi_1 \pi_2 \ln(HR)\right]^2}
\]

\[
event \approx \frac{34}{[\ln(HR)]^2}
\]

- 90% Power, 1-sided
  - \( z(1-\alpha) = 1.64 \) (5%, 1-sided)
  - \( Z(1-\beta) = 1.28 \) (90% Power)

- \( HR = \) Hazard Ratio = 2 (note: \( HR \neq 1 \))
- Total Events = 71
  - \( \pi = 0.5 \) (1:1, ratio of sample in each group)

Method

\[
E(P) = \frac{\lambda_1}{\lambda_1 + d} \left[ 1 - \frac{e^{-(\lambda_1 + d)(T-T_0)} - e^{-(\lambda_1 + d)T}}{(\lambda_1 + d)T_o} \right]
\]

- \( T_0 = \) Accrual rate = 3 years
- \( T = \) Follow-up rate = 5 years
- \( D = \) drop out rate = 5%

- \( \lambda = \) event rate (events per women-year)
- \( E(P) = \) expected event proportion (events per women)
- \( E(P1, \text{new}) = 12.7 \) (events per 10,000W) in study
- \( E(P2, \text{est}) = 6.39 \) (events per 10,000W) in study
Method

\[
\text{Total} = \text{Event} \left( \frac{\pi_1}{E(P_1)} + \frac{\pi_2}{E(P_2)} \right)
\]

Total = Event \times \pi \left( \frac{10,000}{12.7} + \frac{10,000}{6.39} \right)

Total = 71(394 + 782) = 83,354W

- Total = number of events * number of women for 1 event
- \( E(P_{1,\text{new}}) = 12.7 \) events per 10,000 Women
- \( E(P_{2,\text{est}}) = 6.39 \) events per 10,000 Women
- \( n \) per group = 41,677 Women
- Total Sample = 83,354 Women

- Note: \( \pi_1 = \pi_2 = 0.5 \) (1:1, ratio of sample in each group)

Results

| Two group test of equal exponential survival (a large), exponential dropout |
|-------------------|---|---|---|
| Test significance level, \( \alpha \) | 0.050 | 0.050 | 0.050 |
| 1 or 2 sided test? | 1 | 1 | 1 |
| Length of accrual period | 3.00 | 3.00 | 3.00 |
| Maximum length of followup | 5.00 | 5.00 | 5.00 |
| Common exponential dropout rate, \( d \) | 0.0500 | 0.0500 | 0.0500 |
| Group 1 exponential parameter, \( \lambda_1 \) | 0.0004 | 0.0006 | 0.0008 |
| Group 2 exponential parameter, \( \lambda_2 \) | 0.0002 | 0.0003 | 0.0004 |
| Hazard ratio, \( h_{\lambda_1/\lambda_2} \) | 2.000 | 2.000 | 2.000 |
| Power (\%) | 80 | 80 | 80 |
| \( n \) per group | 41,861 | 27,914 | 20,941 |
| Total number of events required, \( E \) | 21 | 21 | 21 |

- Drop-out = 5%
- Patient Accrual = 3 years
- Follow-up = 5 years
- Excel
  - \( n \) per group = 41,677W
  - Total Sample = 83,354W
- nQuery
  - \( n \) per group = 41,861W
  - Total Sample = 83,734W
Summary for Time–to-Event

• Case Study
  > VTE incidence rates per Women years
  > Hazard Ratio

• Method
  > Events
  > Expected proportion of events
  > Total number

• Results
  > nQuery

Sample Size Estimation for a Single Cohort

Dr. Pablo Mallaina
Andrew Burgess, BSc
CV Risk assessment among smokers in Primary care in Europe

Study design:
> Observational, multi-centre, European, cross-sectional study (single cohort)

Objectives:
> To evaluate the CVD risk among smokers at PC in Europe using standard risk assessment tools
  - Multi-factorial risk models
    » Framingham Risk Score
    » Systemic Coronary Risk Evaluation (SCORE)
    » Progetto CUORE
> To estimate the CVD risk attributable to smoking

Sample size calculation:
> Deterministic sample size using Framingham 10-year CVD risk

Sample Size Based on Precision of Estimate

• Simple case: No comparisons, binary outcome

• No power calculation – no hypothesis

• Considerations
  > Desired level of precision
  > Expected endpoint (percentage/rate)
  > Cost and logistics of recruiting subjects

• Range of sample sizes: balance costs versus precision.
Assumptions

• 95% confidence interval
• 10-year CVD risk for smokers range from 5% to 20%
• Large n = normal approximation
• 2-sided interval
• nQuery 7.0

Method

\[
CI = p \pm z \cdot se(p)
\]

Precision

• \( p \) = measured event (CVD risk)
• \( z \approx 2 \)
• \( se(p) = \) standard error
Example: CV Aspire

nQuery screenshot of n = 1000

<table>
<thead>
<tr>
<th>Sample</th>
<th>Size</th>
<th>Expected proportion (%)</th>
<th>9%</th>
<th>10%</th>
<th>15%</th>
<th>20%</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>Precision (%)</td>
<td>1.91</td>
<td>2.63</td>
<td>3.13</td>
<td>3.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CI (% )</td>
<td>(3.09, 6.91)</td>
<td>(7.37, 12.63)</td>
<td>(11.87, 18.13)</td>
<td>(16.49, 23.51)</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>Precision (%)</td>
<td>1.35</td>
<td>1.86</td>
<td>2.21</td>
<td>2.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CI (% )</td>
<td>(3.65, 6.35)</td>
<td>(8.14, 11.86)</td>
<td>(12.79, 17.21)</td>
<td>(17.52, 22.48)</td>
<td></td>
</tr>
<tr>
<td>1500</td>
<td>Precision (%)</td>
<td>1.10</td>
<td>1.52</td>
<td>1.81</td>
<td>2.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CI (% )</td>
<td>(3.90, 6.10)</td>
<td>(8.48, 11.52)</td>
<td>(13.19, 16.81)</td>
<td>(17.89, 22.02)</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>Precision (%)</td>
<td>0.96</td>
<td>1.31</td>
<td>1.56</td>
<td>1.75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CI (% )</td>
<td>(4.04, 5.96)</td>
<td>(8.69, 11.31)</td>
<td>(13.44, 16.56)</td>
<td>(18.25, 21.75)</td>
<td></td>
</tr>
</tbody>
</table>

Results
CV Aspire : Summary

• Number of patients in the study = 1,439
• The CVD risk among smokers is = 21.2%
  > (Framingham 10-year CVD score)
• The final precision = 2.1%
• 95% Confident that actual risk = 19.1% to 23.3%

• Objective:
  > The relative increase in CVD risk attributable to smoking = 31.9%
    - The CVD risk in simulated non-smoker = 16.0%

Summary for Precision

• Case Study
  > CV Aspire
• Method
  > Number of patients - precision
• Results
  > nQuery
What does it mean for the sample to be ‘representative’?

**Study Objective**
- Treatment Patterns
- Health Economic
- Disease / Epidemiology

**Target Population Level**
- Site / Physician
- Patient

**Possible Strata**
- Physician Specialty
- Geography
- Practice Size
- Age, Gender
- Disease & Tx Duration
VALUE Case Study

**Background**

VALUE: An Observational Study to Assess the Cost of Venous Leg Ulcer.

Venous leg ulcers treated by a disparate group of physician types in the U.S.

Treatment patterns believed to vary by physician specialty and region

Study results used to inform a pricing and reimbursement strategy for a product at peri-approval stage.

**Study Description**

Disease and Cost of Treatment Registry

12-week observation of the resource utilization involved in the treatment of venous leg ulcers and the associated clinical and patient reported outcomes.

Variable number of visits

10 sites/50 patients in the U.K.

15 sites/100 patients in U.S. -nationally representative

Estimate the Distribution of Venous Leg Ulcer Patients across the U.S.

- From the National Ambulatory Medical Care Survey (NAMCS), records with one of the following ICD-9-CM diagnosis codes listed as their primary (i.e., first-listed) diagnosis were extracted for analysis.

<table>
<thead>
<tr>
<th>ICD-9-CM</th>
<th>Diagnosis Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>454.0, 454.2</td>
<td>Varicose veins of lower extremities, with ulcer</td>
</tr>
<tr>
<td>459.81</td>
<td>Venous (peripheral) insufficiency, unspecified</td>
</tr>
<tr>
<td>707.12</td>
<td>Ulcer of calf, except decubitus</td>
</tr>
<tr>
<td>707.13</td>
<td>Ulcer of ankle, except decubitus</td>
</tr>
</tbody>
</table>

## Target Patient Distribution Dictated by Geography and Physician Surgical Status

<table>
<thead>
<tr>
<th></th>
<th>Northeast</th>
<th>South</th>
<th>Midwest</th>
<th>West</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgical¹</td>
<td>306,944</td>
<td>177,583</td>
<td>197,561</td>
<td>62,757</td>
<td>744,845</td>
</tr>
<tr>
<td></td>
<td>(18.95%)</td>
<td>(10.96%)</td>
<td>(12.20%)</td>
<td>(3.87%)</td>
<td>(45.98%)</td>
</tr>
<tr>
<td>Non-Surgical²</td>
<td>87,354</td>
<td>304,167</td>
<td>153,043</td>
<td>330,321</td>
<td>874,885</td>
</tr>
<tr>
<td></td>
<td>(5.39%)</td>
<td>(18.78%)</td>
<td>(9.45%)</td>
<td>(20.39%)</td>
<td>(54.01%)</td>
</tr>
<tr>
<td>Total</td>
<td>394,298</td>
<td>481,750</td>
<td>350,604</td>
<td>393,078</td>
<td>1,619,730</td>
</tr>
<tr>
<td></td>
<td>(24.34%)</td>
<td>(29.74%)</td>
<td>(21.65%)</td>
<td>(24.26%)</td>
<td>(100.00%)</td>
</tr>
</tbody>
</table>

Source: National Ambulatory Medical Care Survey (NAMCS), 2003-2004
1 Vascular surgeon; general surgeon
2 Dermatologist; general practitioner; internal medicine; family practitioner; geriatrician

## Target U.S. Enrollment Counts by Physician Surgical Status and Geography

<table>
<thead>
<tr>
<th></th>
<th>Northeast</th>
<th>South</th>
<th>Midwest</th>
<th>West</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgical¹</td>
<td>19</td>
<td>11</td>
<td>12</td>
<td>4</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(36, 56)</td>
</tr>
<tr>
<td>Non-Surgical²</td>
<td>5</td>
<td>19</td>
<td>10</td>
<td>20</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(44, 64)</td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>30</td>
<td>22</td>
<td>24</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>(16, 34)</td>
<td>(21, 40)</td>
<td>(14, 31)</td>
<td>(16, 34)</td>
<td></td>
</tr>
</tbody>
</table>

1 Vascular surgeon; general surgeon
2 Dermatologist; general practitioner; internal medicine; family practitioner; geriatrician
Source: 2003-2004 NAMCS