Research Designs For Evaluating Healthcare Program Outcomes

Ariel Linden, Dr.P.H., M.S.
Why Bother?

- A program cannot be deemed "successful" without an evaluation.
- An evaluation cannot be successful if measurement was not planned in advance.
- The measurement cannot be successful if variables are not accurate or available.
The Five Program Domains

- Needs Assessment
- Program Design (Suitability of Services)
- Implementation and delivery (process)
- Program impact or outcomes
- Program efficiency (cost-effectiveness)
Threats to Validity

- Member
- Health Plan
- Physician/Provider
- Data/Measurement
- General
Member-Based Threats

- Selection Bias - motivation, sickness
- Loss to attrition - disenrollment, death
- Maturation - Progressive disease
- Benefit Designs - Cost sharing
Health Plan-Based Threats

- Regression to the mean
- Case-mix - health plan turnover rates
- Treatment interference - other programs
- Access - Provider access & availability
- New Technology - impact on utilization
Provider-Based Threats

- Hawthorne Effect - big brother!
- Reimbursement - practice patterns
Data-Based Threats

- Accurate sources of data
- Accurate identification of suitables
- Seasonality
- Validity/Reliability of outcome measures
Other Threats

- Secular trends
Research Design Categories

- Experimental designs - ability to control the *when* and *to whom* of exposure and measurement
- Quasi-experimental designs - can only control the *when* and *to whom* of measurement
Pre-Experimental Designs

- One-shot case study: $X \ O$

- One Group Pretest-posttest: $O_1 \ X \ O_2$

- Static-group Comparison: $X \ O_1 \ O_2$
**True Experimental Designs**

- **Pretest-posttest Control Group:**
  
  
  - R
  - O₁
  - X
  - O₂
  
  - R
  - O₃
  - O₄

- **Solomon 4-Group Design:**
  
  - R
  - O
  - X
  - O
  
  - R
  - O
  - O
  
  - R
  - X
  - O
  
  - R
  - O
Experimental Designs (cont’)

- Posttest-Only Control Group:  \( R \times O \)
  \( R \times O \)
## Design Validity

<table>
<thead>
<tr>
<th></th>
<th>Selection Bias</th>
<th>Loss to Attrition</th>
<th>Maturation</th>
<th>Benefit Design</th>
<th>Regression</th>
<th>Case-mix</th>
<th>Treatment Interfere.</th>
<th>Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>One Shot Case Study</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>One Group Pretest- Posttest Design</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Static Group Comparison</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pretest-Posttest Control Group Design</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Solomon 4-group Design</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Posttest-Only Control Group Design</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>New Tech</th>
<th>Hawthorne</th>
<th>Payment Methods</th>
<th>Accurate Data</th>
<th>Accurate Id'ing</th>
<th>Seasonal Factors</th>
<th>Validity</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>One Shot Case Study</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>?</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>One Group Pretest- Posttest Design</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Static Group Comparison</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pretest-Posttest Control Group Design</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Solomon 4-group Design</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Posttest-Only Control Group Design</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
Quasi-Experimental Designs

- **Time Series**: $O_1 O_2 O_3 O_4 X O_5 O_6 O_7 O_8$

- **Equivalent Time-Samples**: $X_1 O X_0 O X_1 O X_0 O$

- **Nonequivalent Control group**: $O O O O X O O O O O O O O O O$
### Design Validity

<table>
<thead>
<tr>
<th></th>
<th>Selection Bias</th>
<th>Loss to Attrition</th>
<th>Maturation</th>
<th>Benefit Design</th>
<th>Regression</th>
<th>Case-mix</th>
<th>Treatment Interfer.</th>
<th>Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Series Design</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Equivalent Time Samples Design</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Nonequivalent Control Group</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>New Tech</th>
<th>Hawthorne</th>
<th>Payment Methods</th>
<th>Accurate Data</th>
<th>Accurate Id'ing</th>
<th>Seasonal Factors</th>
<th>Validity Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Series Design</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Equivalent Time Samples Design</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Nonequivalent Control Group</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
Let’s Evaluate a Study!

“Do Disease Management Programs for Patients With Coronary Heart Disease Make a Difference? Experience of Nine Practices”

Am J Manag Care 2002;8:937-946
9 practice sites were chosen based on their willingness to implement a provider-defined DM program.

6 sites had practice coordinators, 3 used office staff, as available, to identify and enroll patients and provide patient education.
Although practice sites were encouraged to use treatment guidelines, no protocol-defined care was dictated.

The implementation plan, including the specific materials to be used for each intervention was determined by each site.
The main outcome measure was the proportion of patients with an LDL-C value less than 100 mg/dL at 6 and 12 months compared to baseline.

Secondary measures included lipid profile parameters, CAD risk factors, and current medications.
Results?

Practice sites chose a variety of patient intervention resources, with the most common being educational booklets, scripted calls to patients, and diaries.

The most used physician intervention resources were the CV risk factor flow sheet, cholesterol-lowering algorithm, and guideline pocket guides.
## Clinical Results?

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>6 mos.</th>
<th>12 mos.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDL &lt;100</td>
<td>29.9</td>
<td>53.9</td>
<td>59.0</td>
</tr>
<tr>
<td>Total Chol &lt; 200</td>
<td>56.1</td>
<td>76.0</td>
<td>80.3</td>
</tr>
<tr>
<td>Triglycerides &lt; 200</td>
<td>67.6</td>
<td>71.9</td>
<td>74.9</td>
</tr>
<tr>
<td>On lipid therapy</td>
<td>73.7</td>
<td>79.5</td>
<td>81.4</td>
</tr>
<tr>
<td>On therapy LDL &gt;130</td>
<td>84.4</td>
<td>86.7</td>
<td>86.2</td>
</tr>
<tr>
<td>On beta-blocker</td>
<td>54.4</td>
<td>53.2</td>
<td>51.5</td>
</tr>
<tr>
<td>DM meds if HbA1c &gt; 7%</td>
<td>88.0</td>
<td>89.5</td>
<td>77.8</td>
</tr>
</tbody>
</table>
Conclusions

“These provider-defined disease management programs generally displayed a positive impact on reducing CV risk-factors in a group of patients at high risk for cardiac events.”
What Design Was Used?

Simple Pretest-posttest:

\[ O_1 \times O_2 \]
\[ O_1 \times O_2 \]
\[ O_1 \times O_2 \]
\[ O_1 \times O_2 \]
Threats to Validity?

- Selection Bias?
- Loss to Attrition?
- Maturation?
- Benefit Design?
- Regression to the mean?
- Case-mix?
- Treatment Interference?
- Access to care?
Threats to Validity? (cont’)

- New Technology?
- Hawthorne Effect?
- Reimbursement?
- Data Accuracy?
- Accurate Id of eligibles?
- Seasonality?
- Validity/reliability of measures?
- Secular trends?
Let's Do Another Study!

“Outcomes of an Educational Component of a Disease Management Program for Hypertension”

*Manag Care Interface 1999:70-73*
BlueCare HMO members with high-risk Hypertension (ages 21-64 yrs)

619 members completed a two-day educational program, given over the course of a two-week period.

All participants received quarterly newsletters.
Outcome variables were measured, via survey, at baseline and 6 months after completing the program.

Measures included exercise level, controlled BP, mental health summary score, and physical health summary score (SF-36).
Results

Of the 619 participants only 355 completed both the baseline and post-program survey.
## Clinical Results?

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>6 mos.</th>
<th>Change%</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Exercise &gt; 3days/wk</td>
<td>61</td>
<td>68</td>
<td>11</td>
</tr>
<tr>
<td>% BP in Control</td>
<td>40</td>
<td>59</td>
<td>48</td>
</tr>
<tr>
<td>Mental Health Score*</td>
<td>50.6</td>
<td>52.6</td>
<td>4</td>
</tr>
<tr>
<td>Physical Health Score*</td>
<td>49.5</td>
<td>50.5</td>
<td>2</td>
</tr>
</tbody>
</table>

* U.S. population mean = 50
Conclusions

“Based on the positive results of the participants who completed the educational program and the high percentage of members who believe that they are not having difficulty coping with their hypertension, the need for a population-based education program is evident.”
What Design Was Used?

Simple Pretest-posttest: O X O O
Threats to Validity?

Selection Bias? ✓
Loss to Attrition? ✓
Maturation? ?
Benefit Design? ?
Regression to the mean? ✓
Case-mix? ?
Treatment Interference? ?
Access to care? ✓
Threats to Validity? (cont’)

New Technology? -
Hawthorne Effect? ✓
Reimbursement? ✓
Data Accuracy? ✓
Accurate Id of eligibles? -
Seasonality? ?
Validity/reliability of measures? ✓
Secular trends? ✓
Final Study Design!

HEDIS
Who/What?

- HEDIS measures require a random sample N of 411.
- Measures are usually stated as “the proportion of members receiving a given service/all eligible members”
- Measures are compared year-over-year for changes.
What Design Was Used?

Pretest-posttest design with randomization of representative sample:

\[ R - O \times O \]
Threats to Validity?

Selection Bias?
- (randomization)

Loss to Attrition?
-

Maturation?
?

Benefit Design?
?

Regression to the mean?
?

Case-mix?
?

Treatment Interference?
?

Access to care?
✓
Threats to Validity? (cont’)

- New Technology?
- Hawthorne Effect?
- Reimbursement?
- Data Accuracy?
- Accurate Id of eligibles?
- Seasonality?
- Validity/reliability of measures?
- Secular trends?
In an ideal world, all studies will have randomly chosen subjects placed in either an experimental or control group.

Always look for the most feasible design to implement that will have the fewest threats to validity.

Control for as many factors as possible, but at the very least, recognize them!