Study Design Considerations for Measuring the Effectiveness of TWH Programs

NIOSH 2nd International Symposium to Advance Total Worker Health
NIH Campus, Bethesda MD -- May 8-11, 2018

Ron Z. Goetzel, Ph.D.
Johns Hopkins Bloomberg School of Public Health and IBM Watson Health
Total Worker Health Evaluation Challenges

- TWH interventions occur in the “real-world” and are multi-factorial (e.g., change in work policies, health and safety programs, leadership, culture, reporting requirements, incentive structure, laws/regulations, etc.)
- Senior leadership buy-in is necessary
- Randomized controlled trials (RCTs) are complicated by legal, ethical, and practical concerns
- Workers are reluctant to submit personal data because of worries about privacy and possible impact on employment
- Fidelity of interventions across sites is difficult to maintain
- It takes time for interventions to take root and achieve a sufficient dose
It’s Messy…
Evaluation “Buckets”

What should be evaluated?

- Structure
- Process
- Outcomes
Workplace Health, Well-Being, and Safety Programs -- Logic Model

Modified Worksit Health Promotion (Assessment of Health Risk with Follow-Up) Logic Model
adopted by the CDC Community Guide Task Force
Evaluation Framework

**STRUCTURE**
- Fidelity & Dose Delivered
- Dose Received
- Output (Intermediate Outcomes)

**PROCESS**
- Q: What is the current status of programs, policies, and the environmental support system for health, well-being and Safety?
- e.g., CDC Worksite Health ScoreCard
- Q: Has the program implemented the initiatives as planned?
- e.g., Program Fidelity Tools
- Q: What is the level of awareness of, participation in, and satisfaction with the initiatives?
- e.g., Feedback Surveys
- Q: Did the initiatives affect behavior?
- e.g., Health Risk Assessments
- Q: Did the initiatives affect health risks, utilization, financial, & safety outcomes?
- e.g., Insurance Claims, Safety Incidents, Workers’ Compensation Claims
Basic Study Designs

- Pre-experimental
- Quasi-experimental
- True experimental

Validity of results increases as you move down this list
Notation In Study Design

- $X =$ Intervention or program
- $O =$ observation (data collection point)
Research Design: Non-Experimental (Pre-Experimental)

One group posttest only

X 0₂

One group before and after (pre-test/post-test)

0₁ X 0₂
Main Problem: Selection Bias and Regression to the Mean
Research Design: Experimental

TRUE EXPERIMENTAL – RANDOMIZED CLINICAL TRIAL (RCT)

\[ 0_1 \times 0_2 \quad \text{EXPERIMENTAL GROUP} \]

(R) \[ \quad \text{-------------} \]

\[ 0_1 \quad 0_2 \quad \text{CONTROL GROUP} \]
Challenges to Conducting RCTs

- Lack of leadership support for “research”
- Short intervention periods – potential attrition
- Spillover effects – fidelity, dose
- Deciding the unit of analysis for randomization – individual, business unit, organization, region?
- Stringent inclusion and exclusion criteria
- Other threats to validity – outside factors such as changes in laws, political/business climate
- Overall generalizability
Research Design: Quasi-Experimental

Quasi-experimental – Pre-test post-test with comparison group:

\[ O_1 \times O_2 \quad \text{EXPERIMENTAL GROUP} \]

\[ O_1 \quad O_2 \quad \text{COMPARISON GROUP} \]
Major Concern: Selection Bias

Pre-existing differences between groups

- Example: healthier, more motivated workers may enroll in health and safety programs initially – but then sicker and higher risk workers may join later

You can partially control for selection bias:

- Match subjects (treatment vs. comparison) – using readily available variables (e.g., age, gender, job level, tenure, location, plan design, baseline health status, medical costs) and potentially other variables (e.g., stress, attitude toward management, motivation to improve health)
Intent-to-Treat Study Overview

Matching Criteria
- Age
- Gender
- Total Costs
- ER, Inpatient, and Office Visit Use
- Drug Days Supply
- Risk Score
- Charlson Comorbidity Index
- Psychiatric Diagnostic Groups
- Non-Medicare
- Risk Score
- Readiness to Change

Recruited Group

Comparison Group

Closely Matched “Twins”
## Propensity Score Matching Results – Example Johnson & Johnson

### EXHIBIT 1

<table>
<thead>
<tr>
<th>Characteristic/samples</th>
<th>Before match</th>
<th>Standardized difference$^a$</th>
<th>After match</th>
<th>Standardized difference$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NUMBER</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnson &amp; Johnson</td>
<td>32,478</td>
<td>—$^b$</td>
<td>31,823</td>
<td>—$^b$</td>
</tr>
<tr>
<td>Comparison</td>
<td>473,213</td>
<td>—$^b$</td>
<td>31,823</td>
<td>—$^b$</td>
</tr>
<tr>
<td><strong>AGE (YEARS)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnson &amp; Johnson</td>
<td>39.7</td>
<td>14.1</td>
<td>39.6</td>
<td>1.4</td>
</tr>
<tr>
<td>Comparison</td>
<td>41.1</td>
<td></td>
<td>39.4</td>
<td></td>
</tr>
<tr>
<td><strong>PERCENT FEMALE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnson &amp; Johnson</td>
<td>45.5</td>
<td>38.2</td>
<td>45.2</td>
<td>4.2</td>
</tr>
<tr>
<td>Comparison</td>
<td>27.4</td>
<td></td>
<td>43.1</td>
<td></td>
</tr>
<tr>
<td><strong>PERCENT IN EACH REGION</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North Central</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnson &amp; Johnson</td>
<td>4.0</td>
<td>59.8</td>
<td>14.7</td>
<td>0.4</td>
</tr>
<tr>
<td>Comparison</td>
<td>23.8</td>
<td></td>
<td>14.6</td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnson &amp; Johnson</td>
<td>55.3</td>
<td>96.6</td>
<td>45.0</td>
<td>1.7</td>
</tr>
<tr>
<td>Comparison</td>
<td>13.9</td>
<td></td>
<td>45.9</td>
<td></td>
</tr>
<tr>
<td>South</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnson &amp; Johnson</td>
<td>1.8</td>
<td>123.5</td>
<td>11.5</td>
<td>2.2</td>
</tr>
<tr>
<td>Comparison</td>
<td>46.8</td>
<td></td>
<td>15.1</td>
<td></td>
</tr>
<tr>
<td>West</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnson &amp; Johnson</td>
<td>14.8</td>
<td>9.3</td>
<td>14.9</td>
<td>0.7</td>
</tr>
<tr>
<td>Comparison</td>
<td>11.6</td>
<td></td>
<td>15.1</td>
<td></td>
</tr>
<tr>
<td><strong>PERCENT ENROLLED IN POINT-OF-SERVICE WITHOUT CAPITATION OR PREFERRED PROVIDER ORGANIZATION</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnson &amp; Johnson</td>
<td>85.0</td>
<td>21.9</td>
<td>86.7</td>
<td>1.2</td>
</tr>
<tr>
<td>Comparison</td>
<td>91.9</td>
<td></td>
<td>87.1</td>
<td></td>
</tr>
<tr>
<td><strong>YEARS OF DATA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnson &amp; Johnson</td>
<td>3.7</td>
<td>12.1</td>
<td>3.7</td>
<td>1.0</td>
</tr>
<tr>
<td>Comparison</td>
<td>3.9</td>
<td></td>
<td>3.7</td>
<td></td>
</tr>
</tbody>
</table>

Adjusted Medical and Drug Costs vs. Expected Costs from Comparison Group

Average Savings 2002-2008 = $565/employee/year
Estimated ROI: $1.88 - $3.92 to $1.00
Matching On Baseline Variables – Including Costs

**Annual Per Member Medical & Prescription Drug Claims**

Calculated savings = $2,574 when comparing treatment and comparison groups workers

Note: Unadjusted annual results.

Summary

• Real world research is messy

• Aim for a triangulation approach to measurement and evaluation – by collecting data on:
  – Structure
  – Process
  – Outcomes

• Apply the Goldie Locks Model to study design
  – Pre-experimental (too cold)
  – True experimental (too hot)
  – Quasi experimental (just right)

• Good luck!