



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 multifactorial (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...



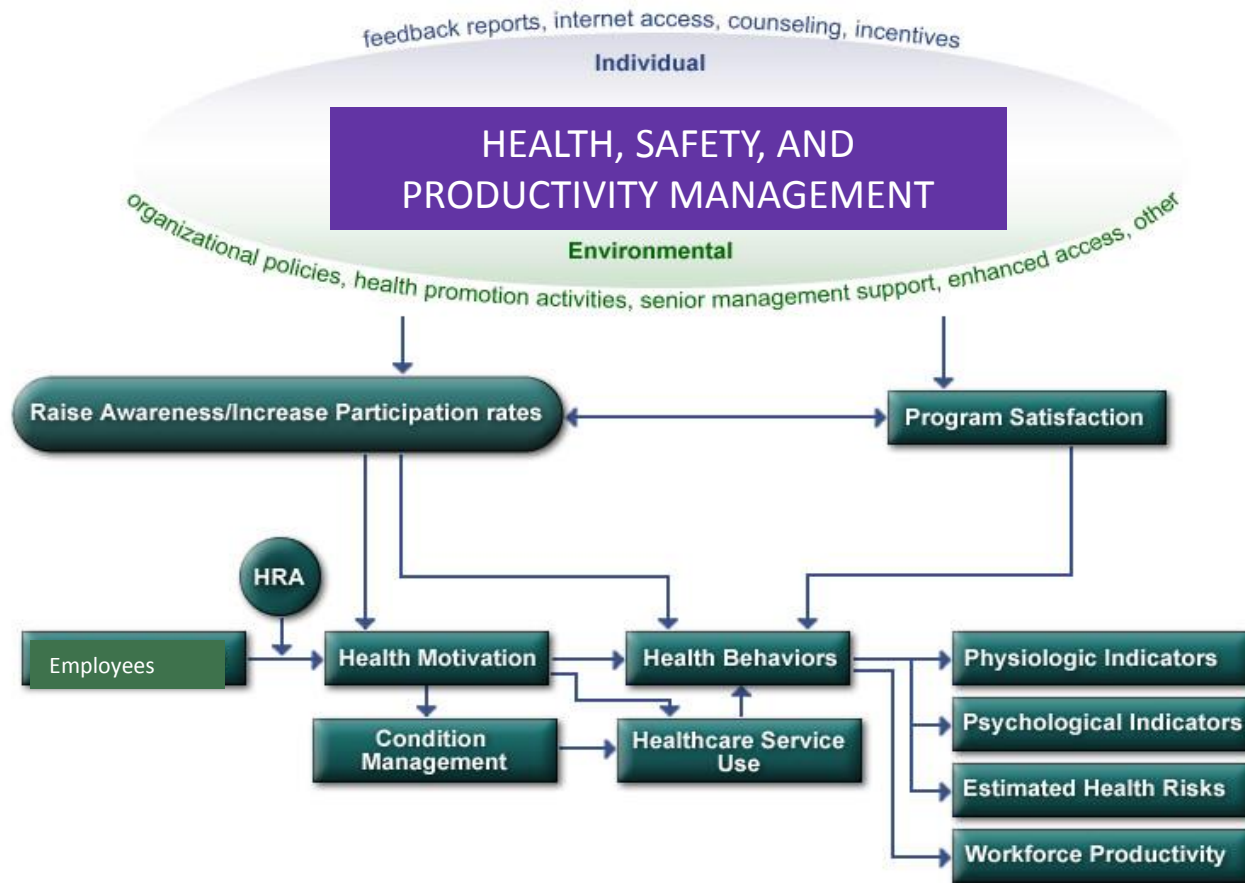
Evaluation “Buckets”

What should be evaluated?

- Structure
- Process
- Outcomes



Workplace Health, Well-Being, and Safety Programs -- Logic Model



- STRUCTURE

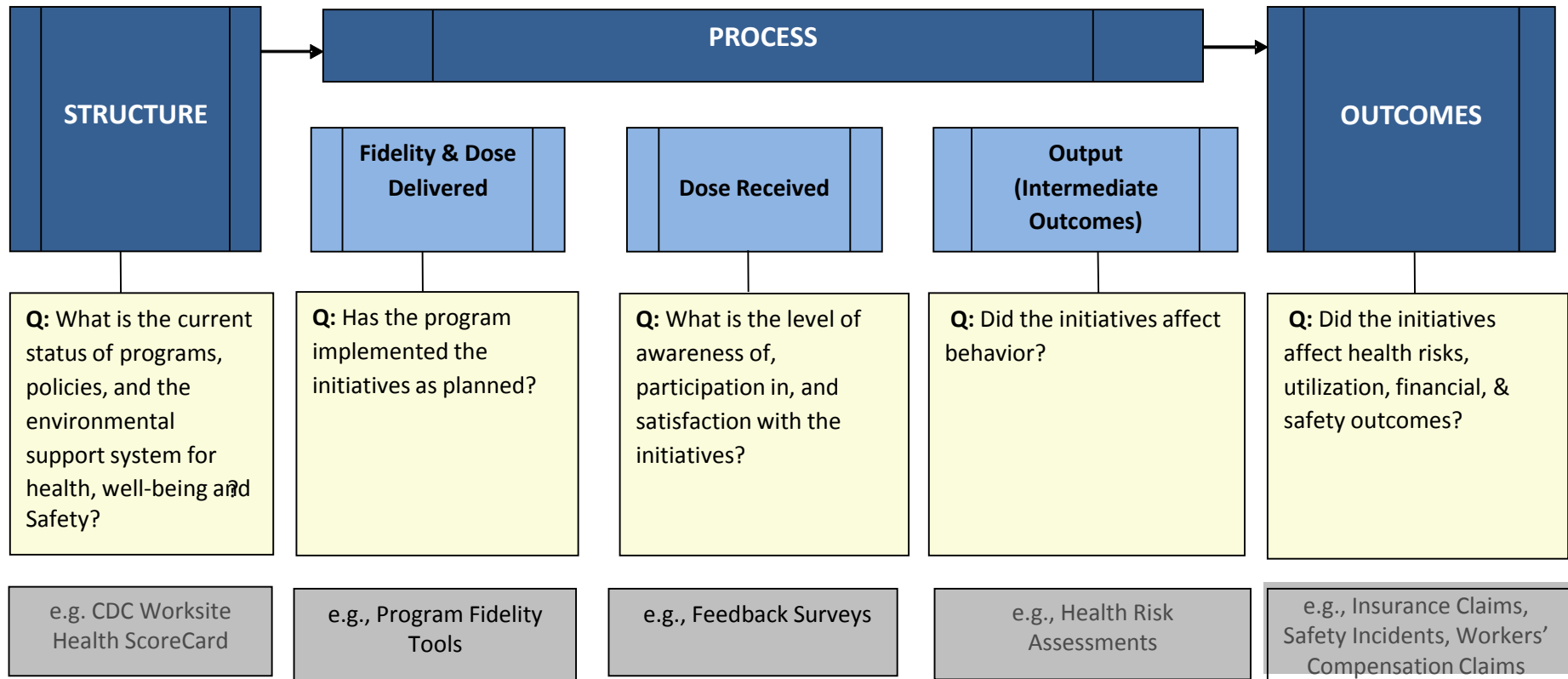
- PROCESS

- OUTCOMES

Modified Worksite Health Promotion (Assessment of Health Risk with Follow-Up) Logic Model

adopted by the CDC Community Guide Task Force

Evaluation Framework



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)

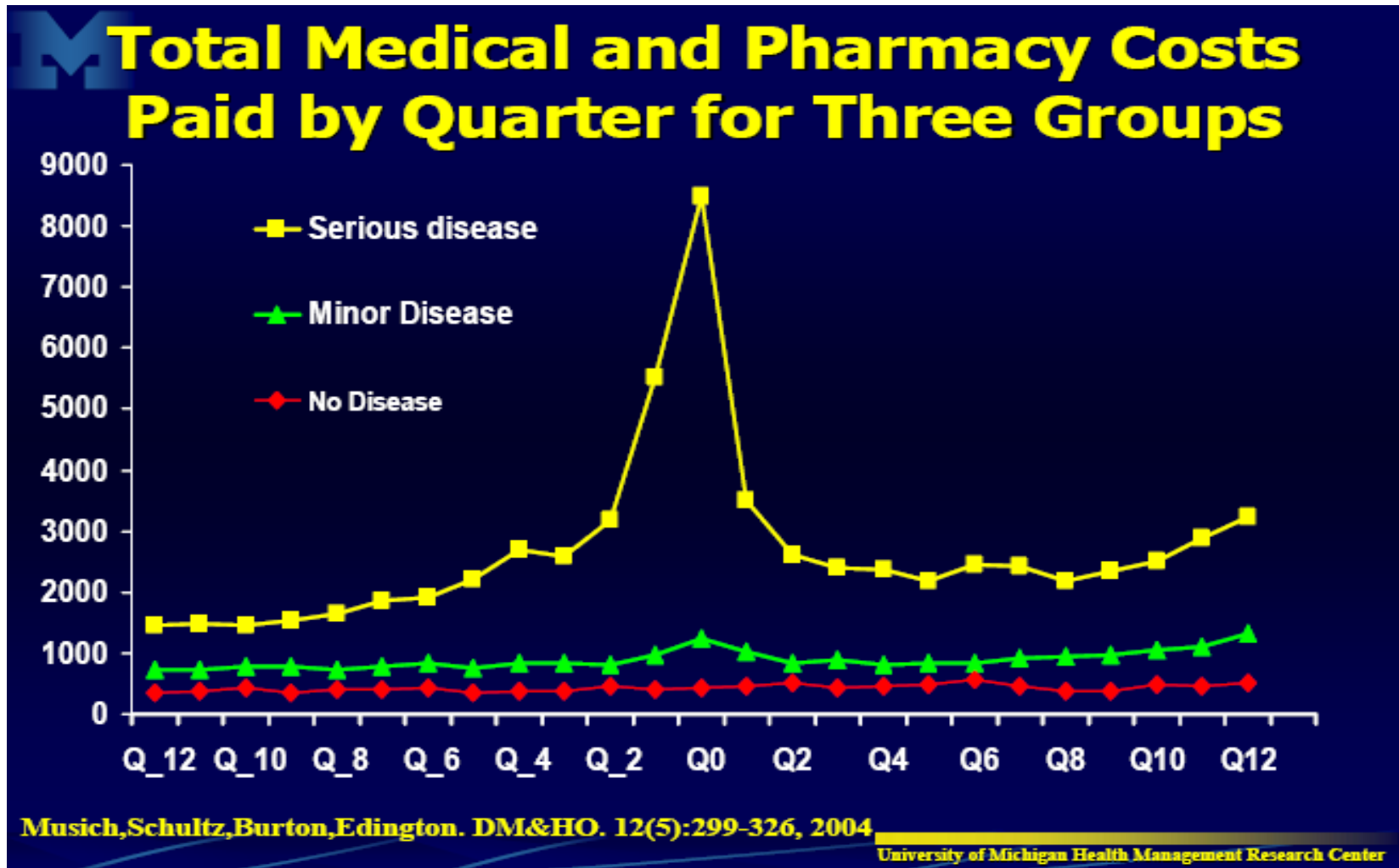
One group posttest only

X O_2

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

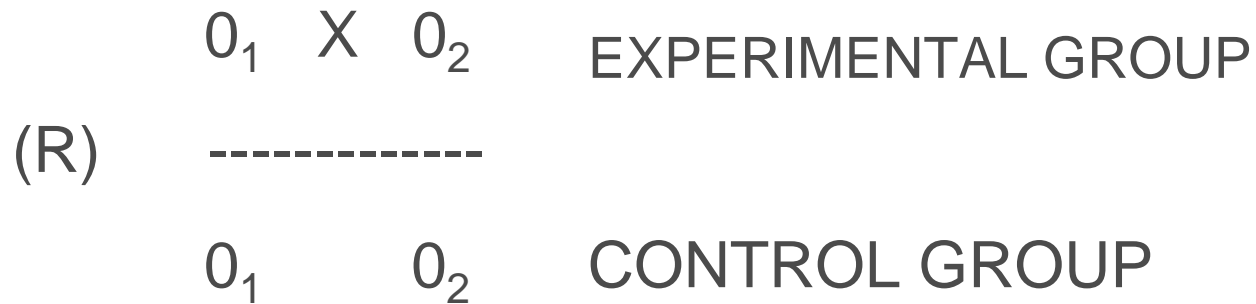
O_1 X O_2

Main Problem: Selection Bias and Regression to the Mean



Research Design: Experimental

TRUE EXPERIMENTAL – RANDOMIZED CLINICAL TRIAL (RCT)



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 X O_2 EXPERIMENTAL GROUP
—————

O_1 O_2 COMPARISON GROUP



No (R)

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

Recruited
Group



Closely
Matched
"Twins"



Comparison
Group



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

Propensity Score Matching Results – Example Johnson & Johnson

EXHIBIT 1

Johnson & Johnson And Comparison-Group Medical Care Sample Before And After Matching

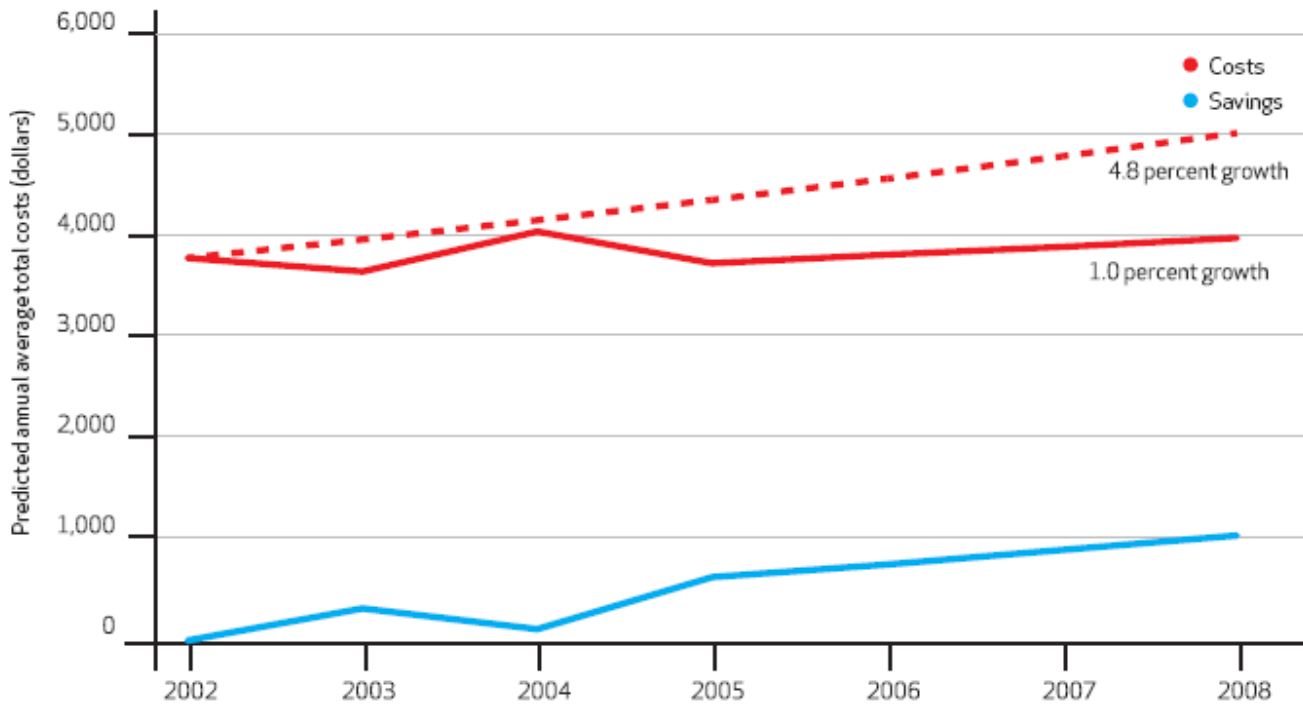
Characteristic/sample	Before match	Standardized difference ^a	After match	Standardized difference ^a
NUMBER				
Johnson & Johnson	32,478	— ^b	31,823	— ^b
Comparison	473,213	— ^b	31,823	— ^b
AGE (YEARS)				
Johnson & Johnson	39.7	14.1	39.6	1.4
Comparison	41.1		39.4	
PERCENT FEMALE				
Johnson & Johnson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

Source: Henke RM, Goetzel RZ, McHugh J, Isaac F. Recent Experience in Health Promotion at Johnson & Johnson: Lower Health Spending, Strong Return on Investment. *Health Aff (Millwood)*. 2011 Mar;30(3):490-9.

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

EXHIBIT 2

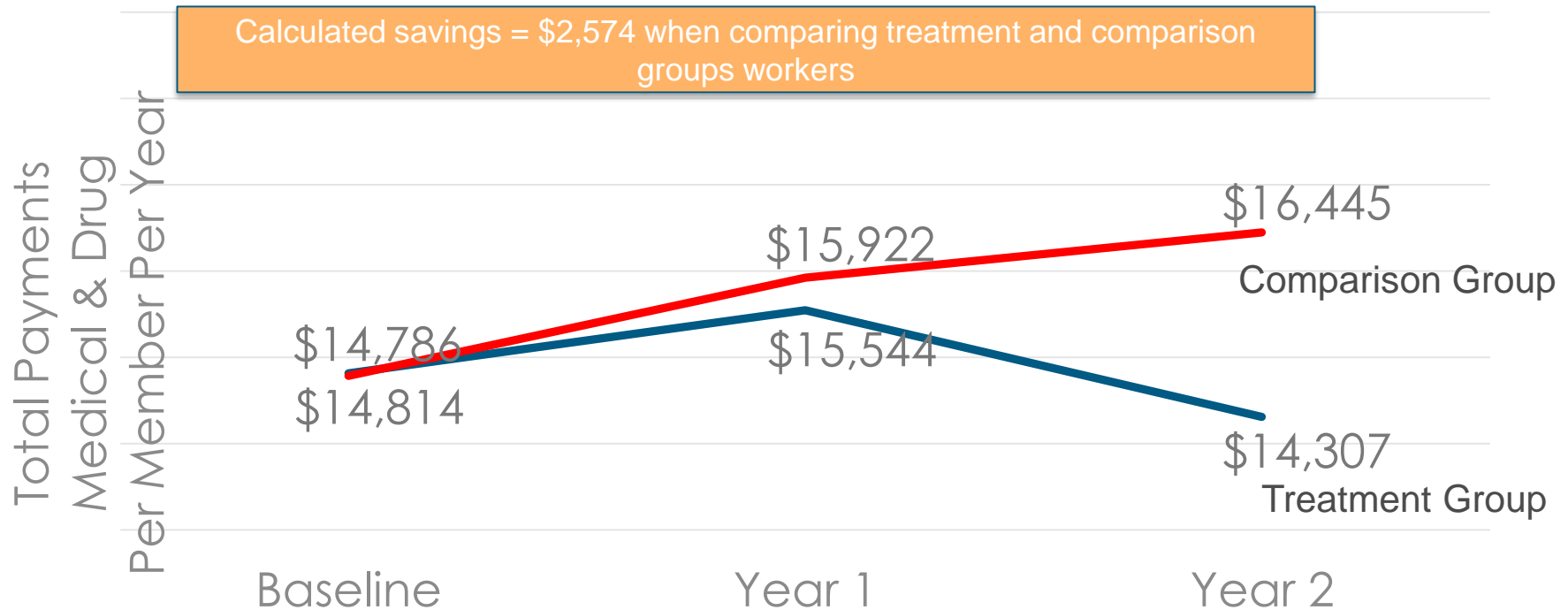
Johnson & Johnson Adjusted Medical And Drug Costs Versus Johnson & Johnson Expected Medical And Drug Costs With Comparison-Group Trend



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



Note: Unadjusted annual results.

Source: Henke R, Lenhart G, Berko J, Cutler E, Goetzel RZ. Return on Investment for a Payer-Provider Partnership to Improve Care Management of Employees and Early Retirees. Journal of Occupational and Environmental Medicine. January 2018. doi: 10.1097/JOM.0000000000001279.

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!