

Method Topic 2: Discussion of appropriate control or comparison groups in studies assessing the efficacy and effectiveness of TWH interventions.

**Total Worker Health® Research Methodology Workshop
March 7-8, 2017 – University of Iowa**

*Ron Z. Goetzel, Ph.D.
Johns Hopkins University and Truven Health Analytics, an IBM Company*

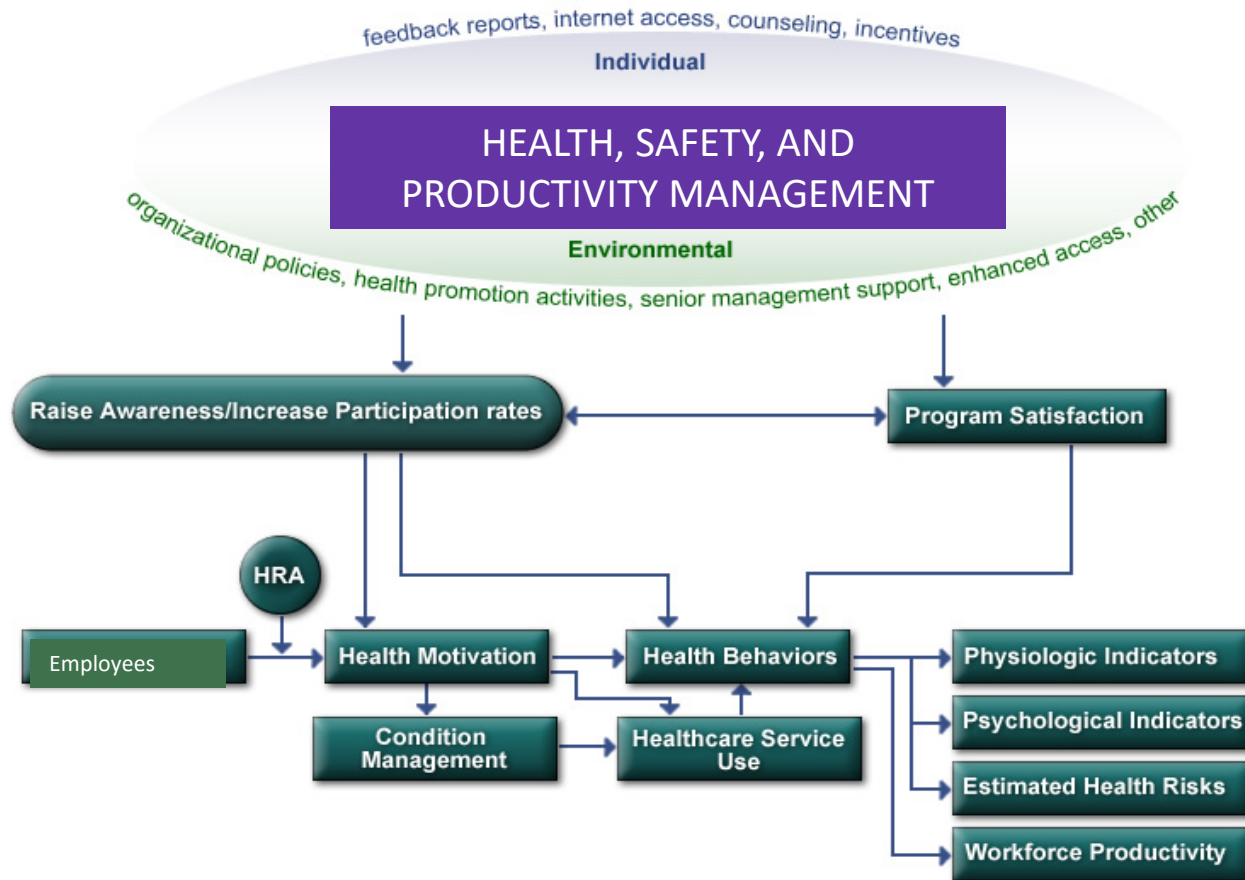
Methodology and Approach

What should be evaluated?

- Structure
- Process
- Outcomes



Logic Model: Worksite Programs



- STRUCTURE

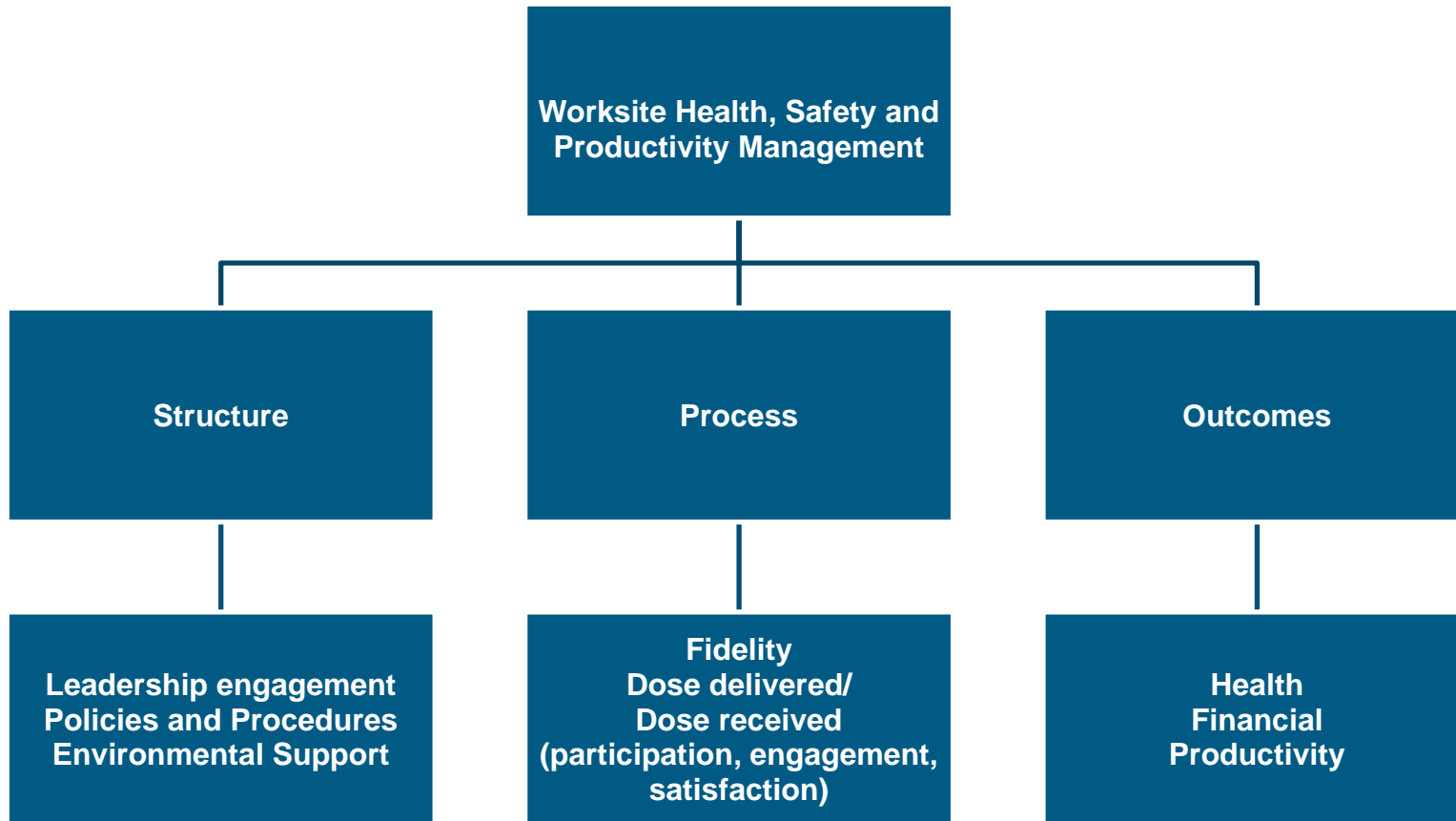
- PROCESS

- OUTCOMES

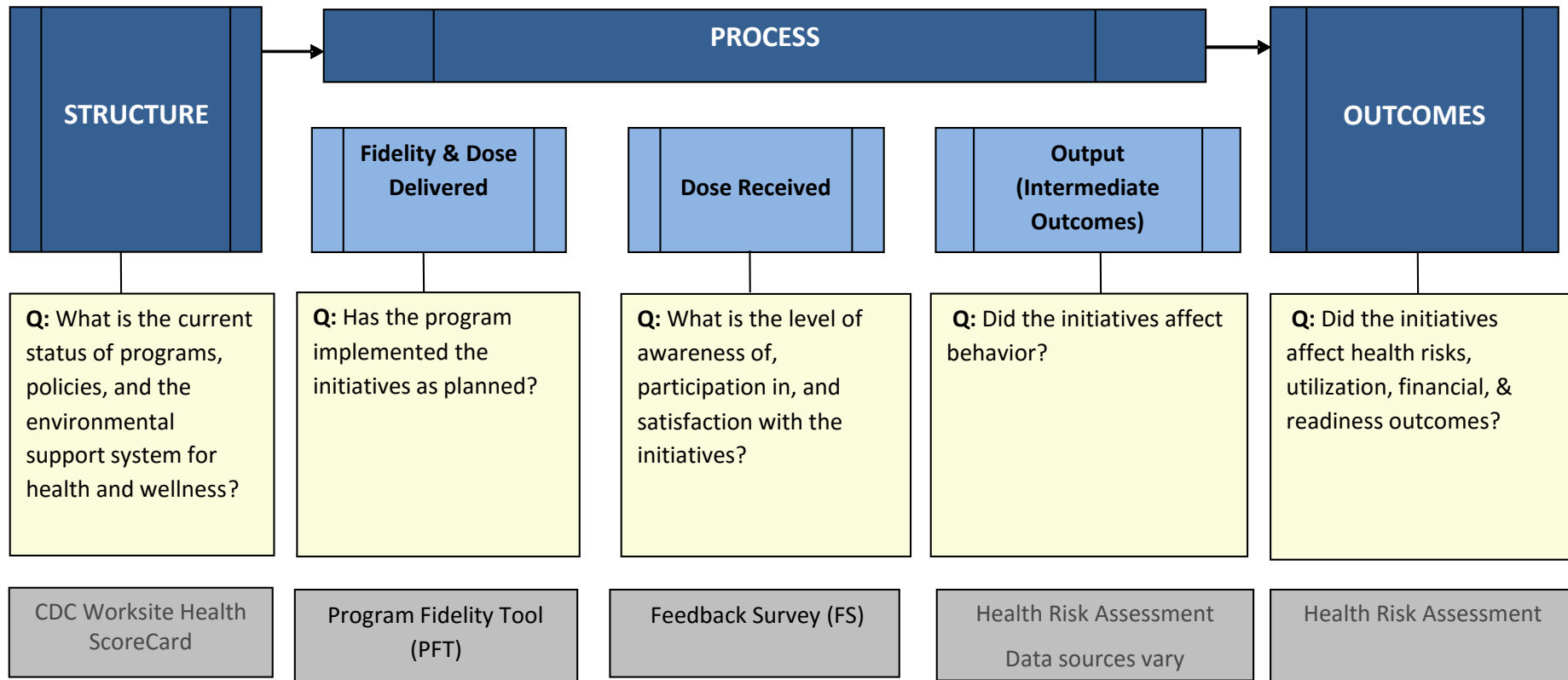
Modified Worksite Health Promotion (Assessment of Health Risk with Follow-Up) Logic Model

adopted by the CDC Community Guide Task Force

Measurement "Buckets"



Evaluation Framework



Research Methods -- Study Design 101



- Pre-experimental
- Quasi-experimental
- True experimental



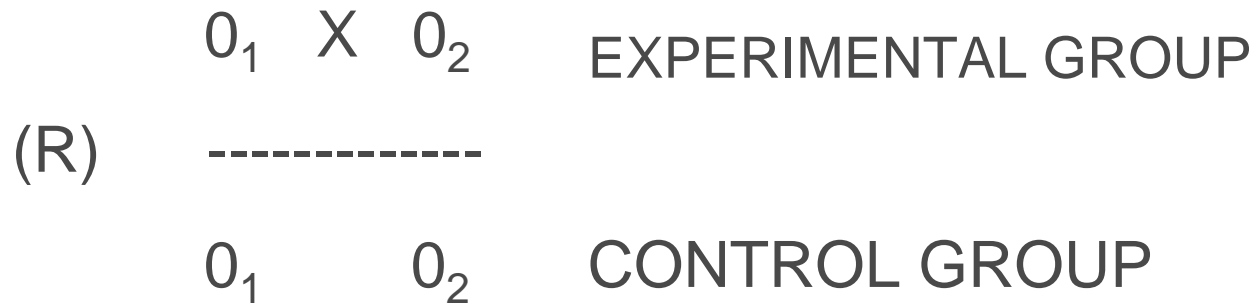
Validity of results
increases as you move
down this list

NOTATION IN STUDY DESIGN

- X=Intervention or program
- O=observation (data collection point)

RESEARCH DESIGN: EXPERIMENTAL

TRUE EXPERIMENTAL – RANDOMIZED CLINICAL TRIAL (RCT)



RESEARCH DESIGN: NON-EXPERIMENTAL (PRE-EXPERIMENTAL)

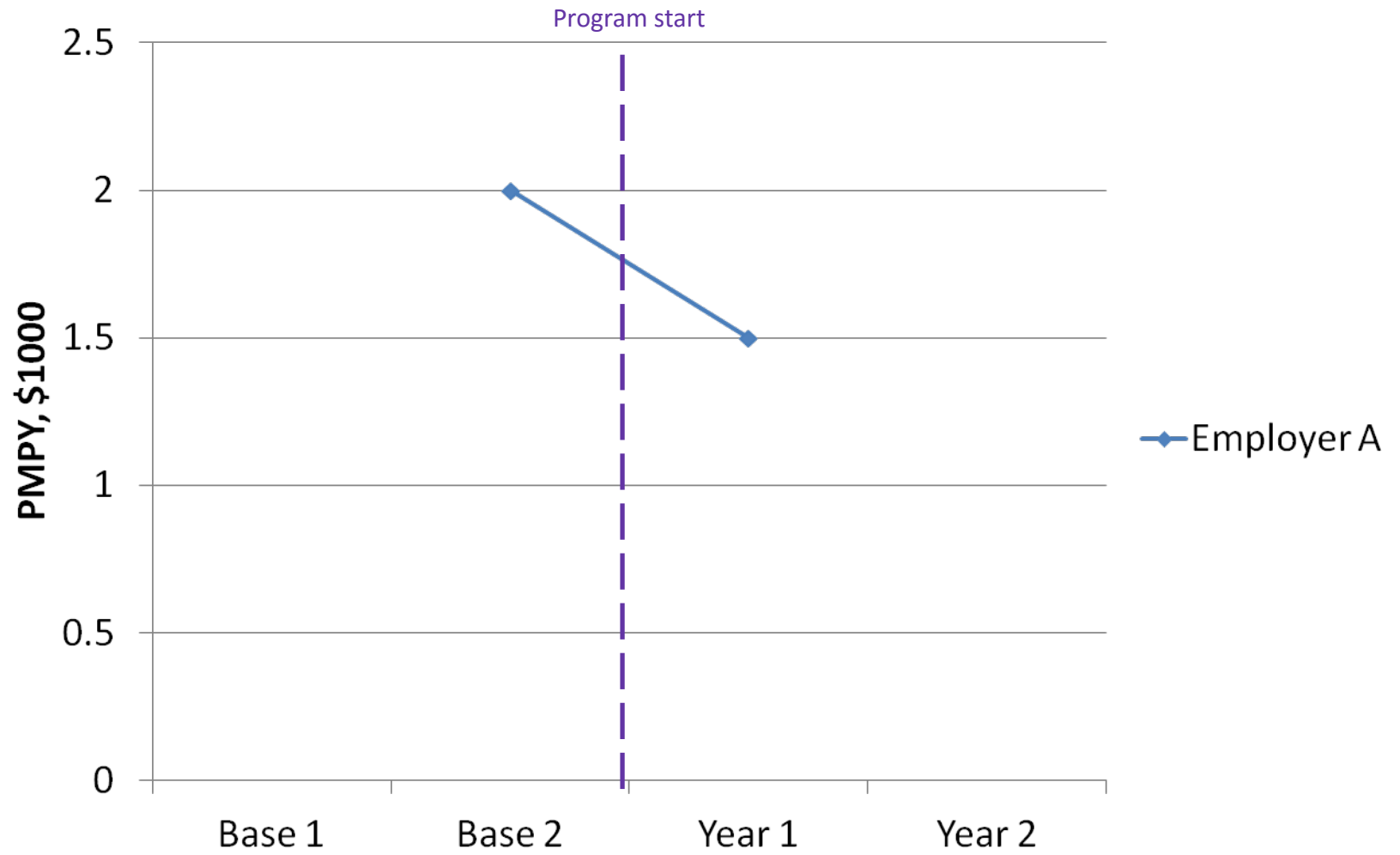
One group posttest only

X O₂

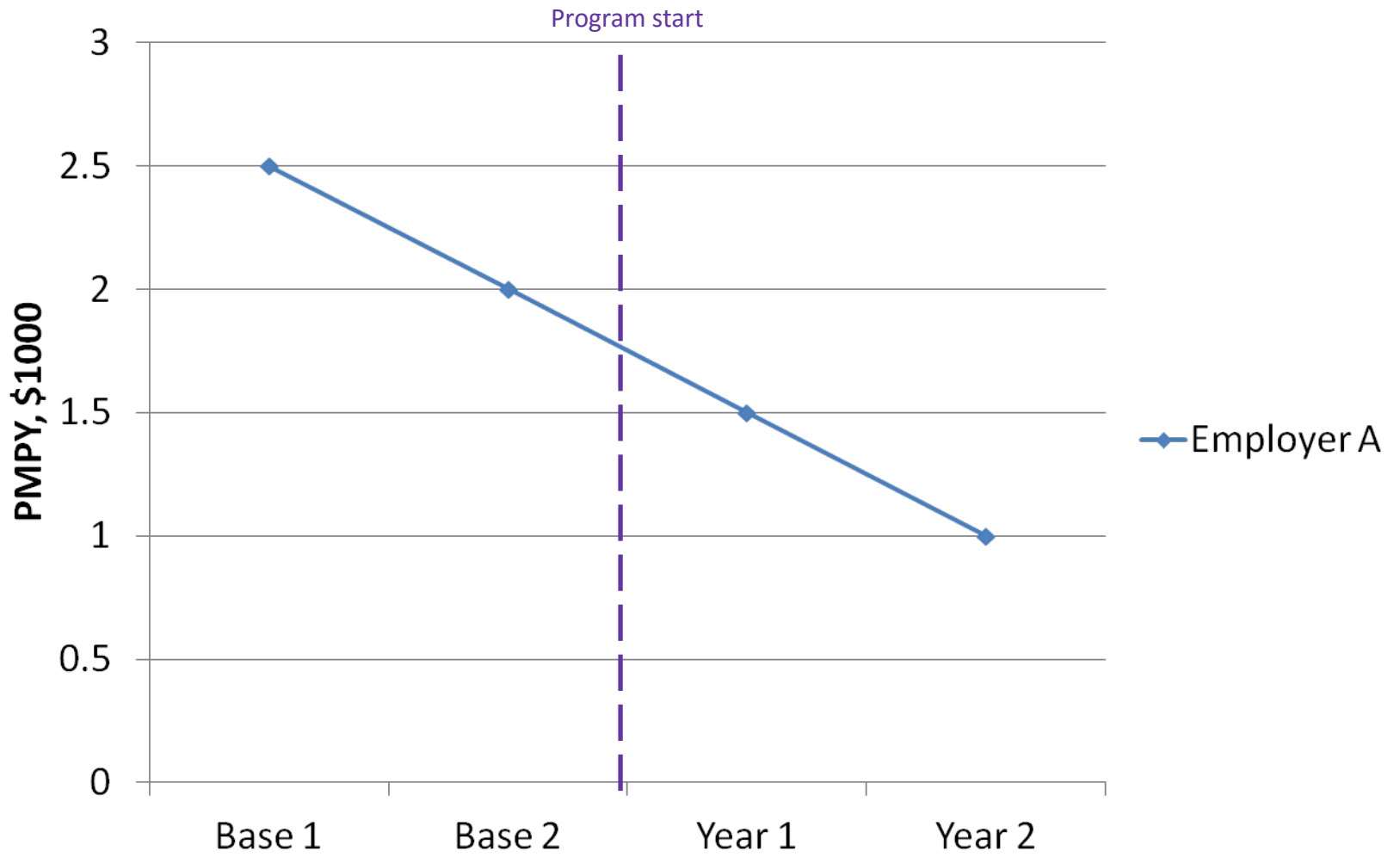
One group before and after (pre-test/posttest)

O₁ X O₂

NON-EXPERIMENTAL DESIGN -- (PRE-EXPERIMENTAL)

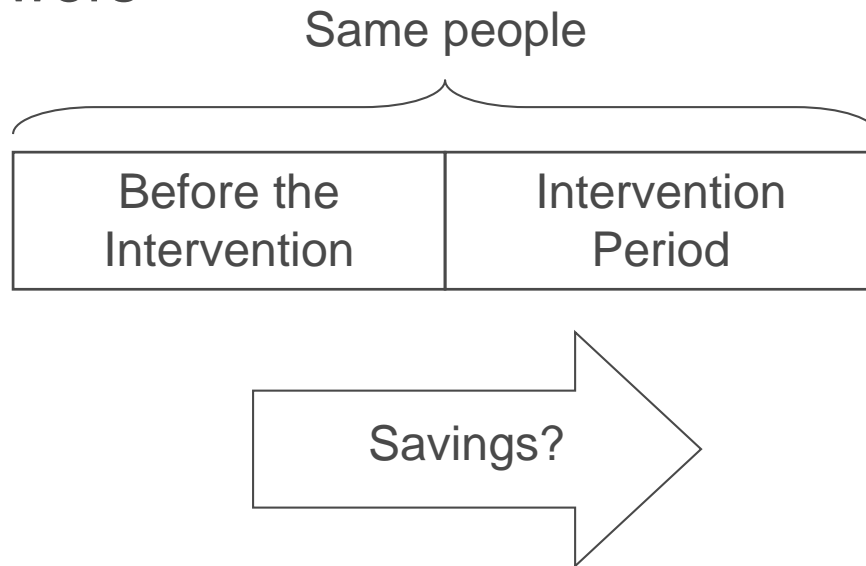


GENERAL TREND OR PROGRAM EFFECT?

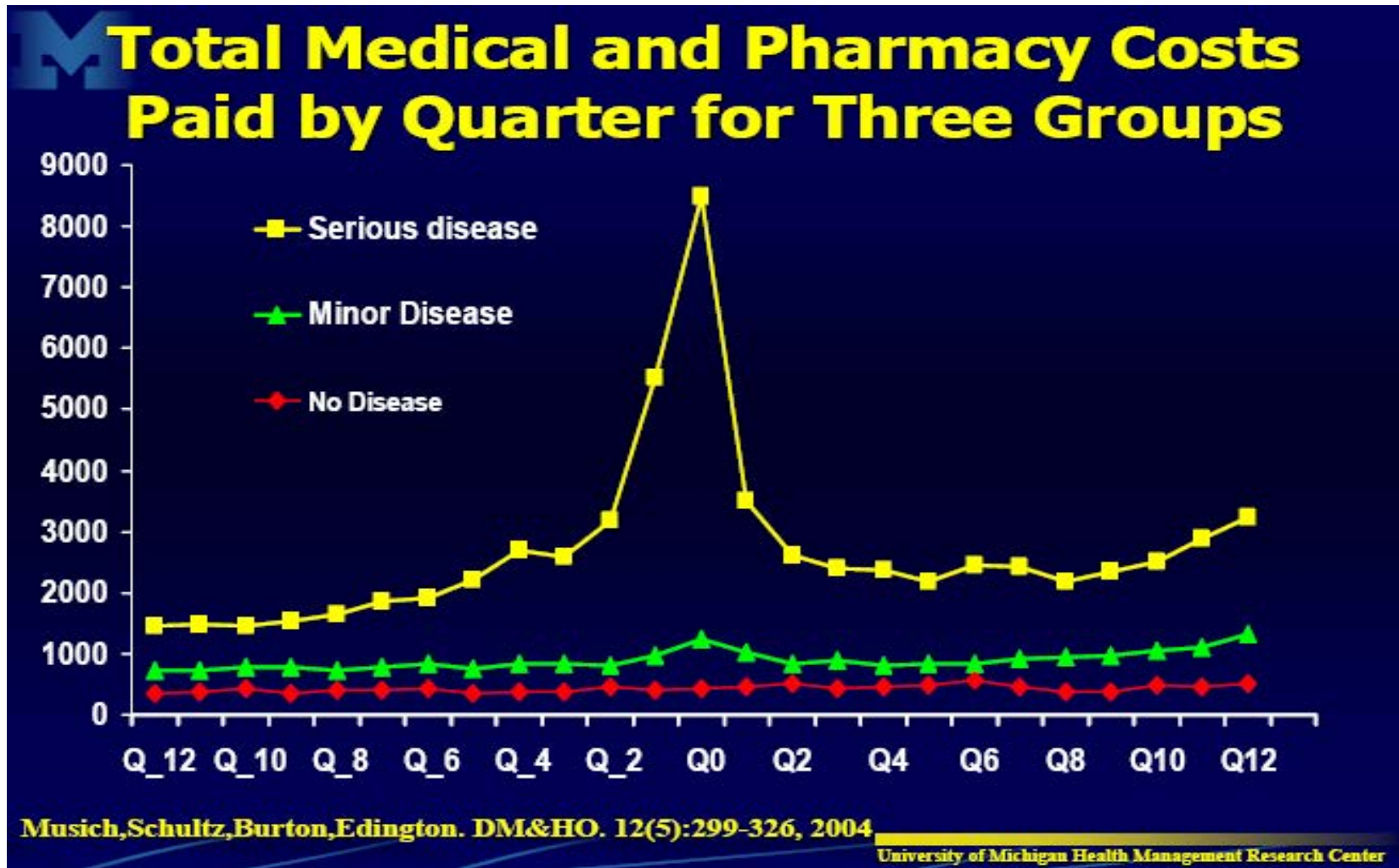


PROBLEMS WITH A PRE-EXPERIMENTAL DESIGN: REGRESSION TO THE MEAN/ SELECTION BIAS

- The most simple analysis may produce the wrong answers

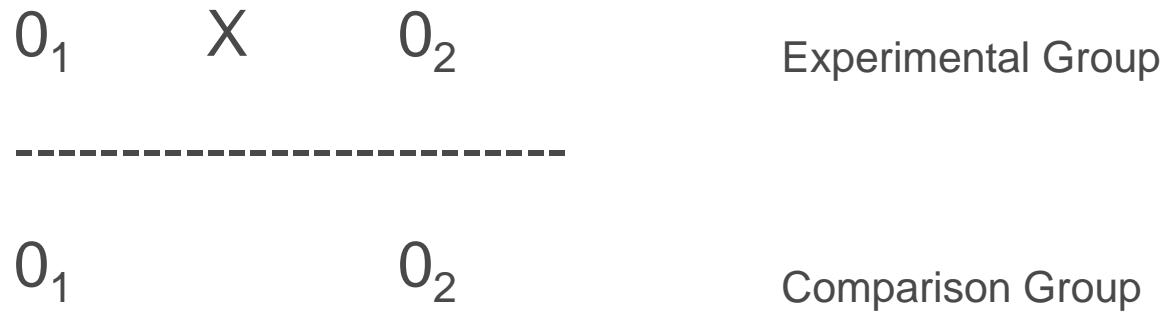


EXAMPLE OF REGRESSION TO THE MEAN



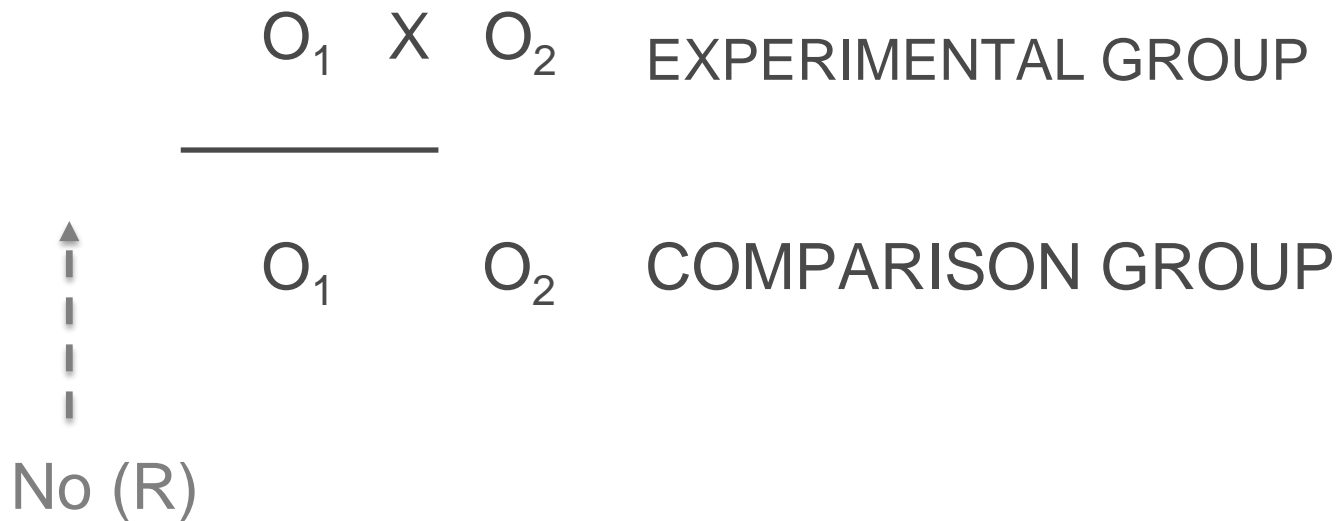
RESEARCH DESIGNS: QUASI-EXPERIMENTAL

Pretest posttest with comparison group



Research Design: Quasi-experimental

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



Definitions of Validity:

- The extent to which one is measuring what one thinks one is measuring (Kerlinger, 1973)
- The best available approximation to the truth of propositions, including propositions about cause (Conrad et al., 1991)
- Internal vs. external validity:
 - **Internal validity**: the truth about causal statements made about the *population being studied*
 - **External validity**: the truth about causal statements *beyond the study population*- i.e., how generalizable are the study's conclusions to other populations in other settings

Addressing threats to internal validity

- Selection bias
- Maturation
- History
- Instrumentation
- Regression to the mean
- Treatment fidelity
- Diffusion of treatments
- Testing

Selection bias

Preexisting differences between groups

- Example: healthier more motivated population volunteers for health promotion program

You can partially control for selection bias:

- Match subjects (treatment v. control) – but this can only be done on a limited number of variables (e.g., age, gender, job level, tenure) but not others (e.g., stress, drug use, attitude toward management, motivation, etc.).

TREATMENT VS. COMPARISON EMPLOYEES AT BASELINE

Variable	Unweighted			Propensity Score Weighted		
	Buy-Up	Standard	p> t	Buy-Up	Standard	p> t
Age	49.8	50.3	0.002	50.0	50.0	0.937
Gender						
Female	57.0%	58.7%	0.046	58.1%	58.1%	0.985
Male	43.0%	41.3%		41.9%	41.9%	
Relationship Code						
Employee/Self	54.2%	53.5%	0.434	54.4%	53.8%	0.583
Spouse/Partner	41.0%	42.1%	0.218	41.1%	41.5%	0.608
Child/Other Dependent	4.8%	4.4%	0.299	4.6%	4.6%	0.918
Clinical Flags - 12 month baseline						
Heart Disease	11.7%	12.5%	0.182	11.9%	12.4%	0.450
Diabetes	15.8%	15.7%	0.796	16.1%	15.3%	0.296
Pregnancy	3.2%	3.5%	0.474	3.2%	3.6%	0.284
Musculoskeletal	55.1%	55.7%	0.481	55.9%	54.6%	0.197
Mental Health	18.0%	15.3%	0.000	18.7%	15.3%	0.000
DCG Relative Risk Scores -12 month baseline						
Concurrent Nonrescaled	356	352	0.568	354	352	0.854
Prospective Explanatory Nonrescaled	298	300	0.587	298	300	0.727
Utilization - 12 month baseline PMPY						
Acute Admissions	0.20	0.20	0.918	0.20	0.19	0.685
Emergency Department Visits	0.39	0.45	0.001	0.42	0.43	0.494
Office Visits	15.01	14.83	0.475	15.00	14.88	0.671
Number of Prescriptions	25.42	25.05	0.388	25.96	24.55	0.003
Days Supplied	1122.30	1098.40	0.209	1140.10	1075.11	0.002
Expenditures - 12 month baseline PMPY						
Medical and Prescription Drug	\$ 16,342	\$ 16,935	0.327	\$ 16,298	\$ 16,922	0.307

CHARACTERISTICS USED IN MATCHING SUBJECTS – AIM IS TO SHOW PARTICIPANTS AND NON-PARTICIPANTS ARE NOT STATISTICALLY DIFFERENT

Overall Comparison

Calendar Year 2001	All Participants	Non-Participants	
	N = 1890	N = 1890	P-value
Male, n (%)	484 (25.6)	484 (25.6)	0.98
Age, 2001 mean years	41.7	41.6	0.94
Net payments for healthcare expenditures in 2001, mean	\$1,414	\$1,318	0.94
Comorbidity Prevalence, %			
Heart disease, n(%)	183 (9.7)	184 (9.7)	
Diabetes, n(%)	13 (0.7)	13 (0.7)	0.99
CCI Group 1 comorbidity, n(%)	849 (44.9)	849 (44.9)	0.98
CCI Group 2 comorbidity, n(%)	528 (27.9)	528 (27.9)	0.98
CCI, median (range)	1.75 (0-17)	1.75 (0-18)	0.97

CCI = Charlson comorbidity index; Group 1 comorbidity includes presence of any of these: chronic obstructive pulmonary disease, rheumatologic disease stomach ulcer or dementia, all as coded by using the Charlson index; Group 2 comorbidity includes presence of any of these: cancer, renal failure, liver disease or cirrhosis, autoimmune disease.

EXTERNAL COMPARISON GROUP

- MarketScan Normative Database
 - Mostly large, established, Fortune 500 employers.
 - Industries: telecommunications, finance, oil and gas, pharmaceuticals, retail, manufacturing, consumer goods.
 - Self-insured
 - Aggressive attention to health care utilization and costs
 - Average to sophisticated health & productivity management programs

ELIGIBILITY CRITERIA

- Study-specific eligibility criteria:
 - 18-64 years old
 - Live in U.S.
 - Not pregnant
 - Active employee (excluded dependents)
 - Continuously enrolled for each study year examined
 - Required 2 years continuous enrollment for adjusted analysis and 6 years of continuous enrollment in cohort analyses
 - Not enrolled in a health plan with capitated services
 - Have RX data

CONTROL VARIABLES

- To ensure comparability of treatment and comparison employees in all adjusted analyses, we matched subjects on the following variables at baseline:
 - Age, gender, region, employee pay (hourly/salary), union status
 - We also conducted analyses that matched employees on baseline medical and drug cost / utilization patterns and baseline clinical severity (CCI, CDS, PDG) to further equalize the groups

PROPENSITY SCORE MATCHING RESULTS

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	