

# The Place of Evaluations in Work Safety: What Can We Really Achieve?

Harry Shannon

Professor, McMaster University

Adjunct Scientist, Institute for Work & Health, Toronto

WoS October 2014

# Outline of Talk

---

- What is evaluation?
- Why evaluate?
- Types of evaluation study
- Published example
- What evaluation questions can we answer
- Some challenges

# An informal definition of evaluation ... with some add-ons

---

- To learn what does and doesn't 'work', and how well or badly ... (*Measure of effect*)
- And how and why? (*Mechanism*)
- For whom? Under what circumstances? (*Context*)

# Why evaluate?

---

- More good than harm?
- Where to spend resources?
- Side effects?
- How and why the intervention works (or doesn't work) ...
- ... and for whom?

# Why evaluate?

---

- More good than harm?
- Where to spend resources?
- Side effects
- How and why the intervention works (or doesn't work) ...
- ... and for whom?

In short, to be EVIDENCE-BASED

# Alternatives to 'Evidence-Based'

---

1. Eminence-based
2. Vehemence-based
3. Eloquence-based
4. Providence-based
5. Diffidence-based
6. Nervousness-based
7. Confidence-based

Isaacs and Fitzgerald

Seven alternatives to evidence-based medicine, BMJ 1999; 319:1618

# Types of Quantitative Evaluation Designs: Summary

---

**WEAKER**

- Non-experimental, e.g. before-after, interrupted time series
- Quasi-experimental, e.g. before-after with control group



**STRONGER**

- Experimental: randomized controlled trial (RCT)

“Stronger” means less susceptible to bias

# Can RCTs be done in practice?

---

- ... remembering they are the scientific ideal

Fairly straightforward for studies of individuals, or even work groups ...

... but more difficult for higher levels, e.g., companies, jurisdictions

‘But this measure must work,  
it’s obvious’

---

Implication – why waste time, etc., on an evaluation

# U.S. Study: Postal workers

---

- Educational intervention to prevent low back injuries (LBI)
- Hard to persuade unions and management to allow “true experiments”
- No significant difference between education group and control group for several measures

Daltroy et al. N Engl J Med 1997; 337:322-328

# Effectiveness of targeted consultation or inspection

---

- Ontario High Risk Firm Initiative
- Companies with poor safety records
- Companies (n=2,153) randomized to
  - Consultation
  - Inspection
  - Service as usual
- Injury rates for 21 months post-intervention
- No significant differences between the three groups

# Land mine risk education in Afghanistan

---

- Evaluations of land mine safety programmes in late 1990s
- Most relevant for farmers (and children)
- No one simple message from evaluation but ...
- Direct training appeared to lead to *increased* risk-taking and injuries

Andersson, Swaminathan, Whitaker, Roche,  
Third World Quarterly 2003; 24: 873-887

# Some historical interventions

---

Apparently not formally evaluated

# Deuteronomy 22:8 – fall protection

---

When thou buildest a new house, then thou shalt make a parapet for thy roof, that thou bring not blood upon thy house, if any man fall from thence.



# Goya's safety poster?

Francisco Goya  
The Drunken Bricklayer  
(detail)  
1786



# Goya's safety poster?

Francisco Goya  
The Injured Mason  
(detail)  
1786-7

# LEY DE SALUD LABORAL ¡YA!



*no podemos esperar*

DEFIENDE TU SALUD



Spanish safety campaign poster, 1992

# “10 Great Public Health Achievements” US 1900-1999

---

- Included ‘safer workplaces’

# 2014 – Centenary of Ontario's Workmen's Compensation Act

---

- Meredith Report 1913 led to the Act
- Employers could defend against suits with any of three arguments:
  - Workers took on risk by seeking employment with employer
  - Contributory negligence (of fellow worker)
  - Worker in some way at fault

# Radio City, New York, 1932





"Not enough money is being spent on safety, so be careful."

New Yorker

# Summary so far

---

- Many interventions, few evaluated
- Still, overall levels of work safety have dramatically improved
- Some interventions, when evaluated, have been shown to be useless or even harmful
- So good evaluations might help spend resources better

# What questions might feasibly be evaluated?

---

# Evaluation of 'multiple interventions' to improve safety

---

- Subsidies (financial incentives) from Dutch Ministry to companies to change management and organizational factors ( $\approx$  'safety culture')
- Companies had to provide before-after data on injuries
- Further research commissioned

Hale, Guldenmund, van Loenhout, Oh,  
Safety Science 2010; 48: 1026-1035,  
and other reports

# Questions for commissioned research

---

- How can success be measured ('operationalized') in such an organization/culture change study?
- Can 'success' be related to particular types of interventions?

# Defining 'success'

---

- Specific for each 'project' (n=17)
- Not based solely on injury rates
- Also used:
  - Increase in reporting hazards
  - Difference between minor and major injuries
  - Other company data

# Identified (mini-) interventions

---

- Specific changes made
- Three broad categories:
  - Technical changes – PPE, machinery, ...
  - Organizational change – procedures, systems, ...
  - Individual and group behavior – training, incentives, ...
- Mean number of changes  $\approx 16$  (range 3 – 28)

# Comparisons made

---

- Successful vs. not successful projects for % using each intervention
- E.g., ‘added inspections and audits’: 75% and 33%, respectively

# Comments in reports

---

- *“Changes in workplaces to improve safety do not come in single, neat packages ...”*
- *“Evaluation stands or falls with the existence of quantitative indicators of safety in companies”*
- *“Strictly controlled scientific methods of epidemiology can almost never be used ...”*
- *“The methodology was based on the highest level of scientific measurement available, namely a before and after study of the trends in performance indicators”*
- Lessons for companies, Ministry, safety science

# What questions might be answered?

---

# 1. Are the subsidies effective?

---

- The ‘how much effect on average?’ question
- Randomize companies
- Dutch Ministry could have randomized companies to:
  - Being offered subsidies
  - Not offered subsidies
- Answers big picture question – average magnitude of effect
- Likely of primary interest to Minister of Labour, Minister of Social Affairs and Employment

# Effectiveness of targeted consultation or inspection

---

- Ontario High Risk Firm Initiative
- Companies with poor safety records
- *Companies* (n=2,153) randomized to
  - Consultation
  - Inspection
  - Service as usual
- Compared reported injury rates for 21 months post-intervention
- No significant differences between the three groups

# RCT on safety violations when working at heights - protocol

---

- Study of construction companies in the Netherlands
- Randomizing 27 *cities* to one of three groups
  - Face-to-face guidance on hazards
  - Direct mailing of hazard information
  - No guidance
- Require 64 companies per group
- Inspections will produce scores out of 30 for safety violation

van der Molen and Frings-Dresen, BMC Public Health, 2014; 14: 541

## 2. What are characteristics of successful companies?

---

- The ‘for whom?’ question
- Identify ‘relevant’ characteristics
- Determine which characteristics distinguish successful and unsuccessful companies
- Done (qualitatively) in subsidies study – “6 prerequisites”
- Of particular interest to Ministry staff, inspectorate

# 3. What program components lead to success?

---

- The ‘what?’ question
- Question answered in subsidies study
- Determine ‘mini’-interventions that work
- Compare successful and unsuccessful companies on these variables
- Of particular interest to workplace parties – ‘what do we need to do to improve safety?’

# 3. What program components lead to success?

---

- Ontario study:

‘Limitations in implementation of the method of targeting workplaces, in intensity and duration of programme, and in outcomes available for evaluation may account, in part, for the absence of observed differences among the study groups.’

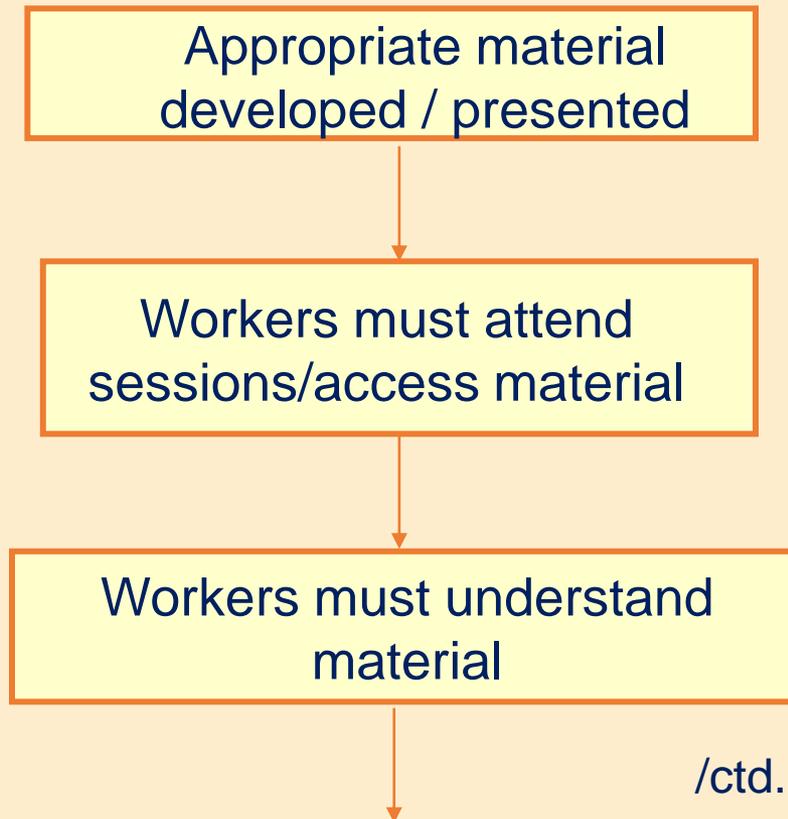
# Beyond the numbers (quantitative): Qualitative approaches

Inside the  
black box

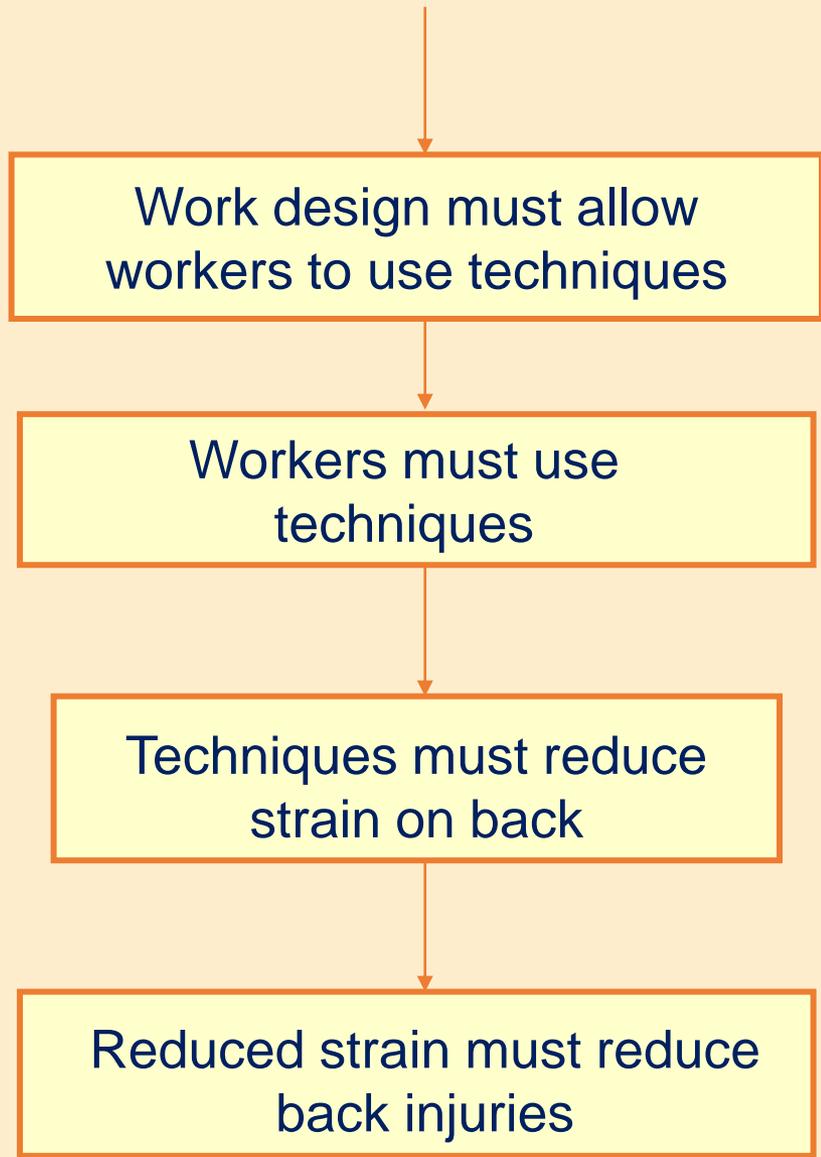
# Educational Program on Lifting Techniques

---

## What is Needed to Prevent Acute Back Injuries?



...



# 4. What are the mechanisms for success?

---

- The ‘how?’ question
- Use qualitative methods
- Interviews with workplace parties in successful AND unsuccessful companies asking ‘what happened and how?’
- Of particular interest to developers of interventions – ‘what’s right and wrong with the program we produced?’ and ‘do we need to modify the underlying theory behind the intervention?’

# 5. What is the cost-benefit ratio?

---

- The ‘is it worth the effort and expense?’ question
- Whose perspective?
- Identify costs such as
  - Time of all staff to implement intervention
  - Spending on safety
- Determine reductions in all injuries
- Reductions in costs of insurance premiums, etc
- Compare net costs with benefits

# What questions might be answered?

## Summary

---

1. Is the broad approach effective?
2. What are the characteristics of 'successful' companies?
3. What are the components that lead to success?
4. What are the mechanisms for success?
5. What is the cost-benefit of the policy?

# Further points

---

- Different stakeholders have different questions
- Different methods are needed for the various questions
- Still: Feasible to answer all questions with high (or at least good) degree of rigour

# Evaluation challenges - 1

---

- Creating a culture of evaluation – at every level
- Questions to ask:
  - Do our safety measures work?
  - Can we prove it?
  - If not, can we do an evaluation to find out?

# Evaluation challenges - 2

---

- Deriving valid measures of safety
  - Injury rates
  - Surrogate or intermediate measures
  - Leading or lagging indicators
  - Combinations of measures

# Evaluation challenges - 2

---

- Deriving valid measures of safety
  - Injury rates
  - Surrogate or intermediate measures
  - Leading or lagging indicators
  - Combinations of measures

*“The most satisfactory procedure would seem to be the use of as many criteria [measures] as possible in evaluating the success of a safety program”*

Hale and Hale, 1972.

A Review of the Industrial Accident Research Literature

# Evaluation challenges - 2

---

- IWH Organizational Performance Metric (OPM)
- 8–item questionnaire
- Scores have predicted injury rates

Amick et al., to be published;

<http://www.iwh.on.ca/the-iwh-opm-questionnaire>

# Evaluation challenges - 3

---

- Applying what has been ‘proven’ and evaluating the new implementation
  - Does the intervention work in the new context?
  - If not, why not?
  - (How) Can we modify the intervention?



PERGAMON

Safety Science 31 (1999) 161–179

---

---

SAFETY SCIENCE

---

---

## Methodological criteria for evaluating occupational safety intervention research<sup>1</sup>

Harry S. Shannon<sup>a,b,\*</sup>, Lynda S. Robson<sup>a</sup>, Stephen J. Guastello<sup>c</sup>

<sup>a</sup>*Institute for Work and Health, 250 Bloor St. E., 7th floor, Toronto, Ontario, Canada M4W 1E6*

<sup>b</sup>*McMaster University, Hamilton, Ontario, Canada*

<sup>c</sup>*Marquette University, Milwaukee, WI, USA*

Received 4 March 1998

---

### Abstract

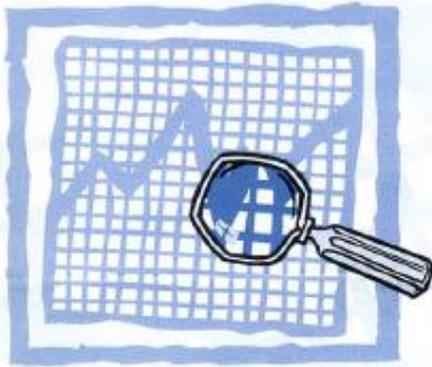
We describe the importance of evaluating workplace safety interventions. Based on the literature and other sources, we list eight areas for which readers can assess the quality of reports evaluating these interventions. The areas are: intervention objectives and their conceptual basis; study design; external validity; outcome measurement; use of qualitative data; threats to internal validity; statistical analysis; and study conclusions. Good quality evaluations can help avoid wasting limited time, money and effort on ineffective or even harmful interventions. © 1999 Elsevier Science Ltd. All rights reserved.

For guidance on  
*reading an*  
evaluation:

Safety Science 1999;  
34:161-179



CENTERS FOR DISEASE  
CONTROL AND PREVENTION



## **Guide to Evaluating the Effectiveness of Strategies for Preventing Work Injuries**

How to show whether a safety  
intervention really works



INSTITUTE FOR WORK & HEALTH  
INSTITUT DE RECHERCHE SUR  
LE TRAVAIL ET LA SANTÉ

For detailed guidance on  
*conducting* evaluations:

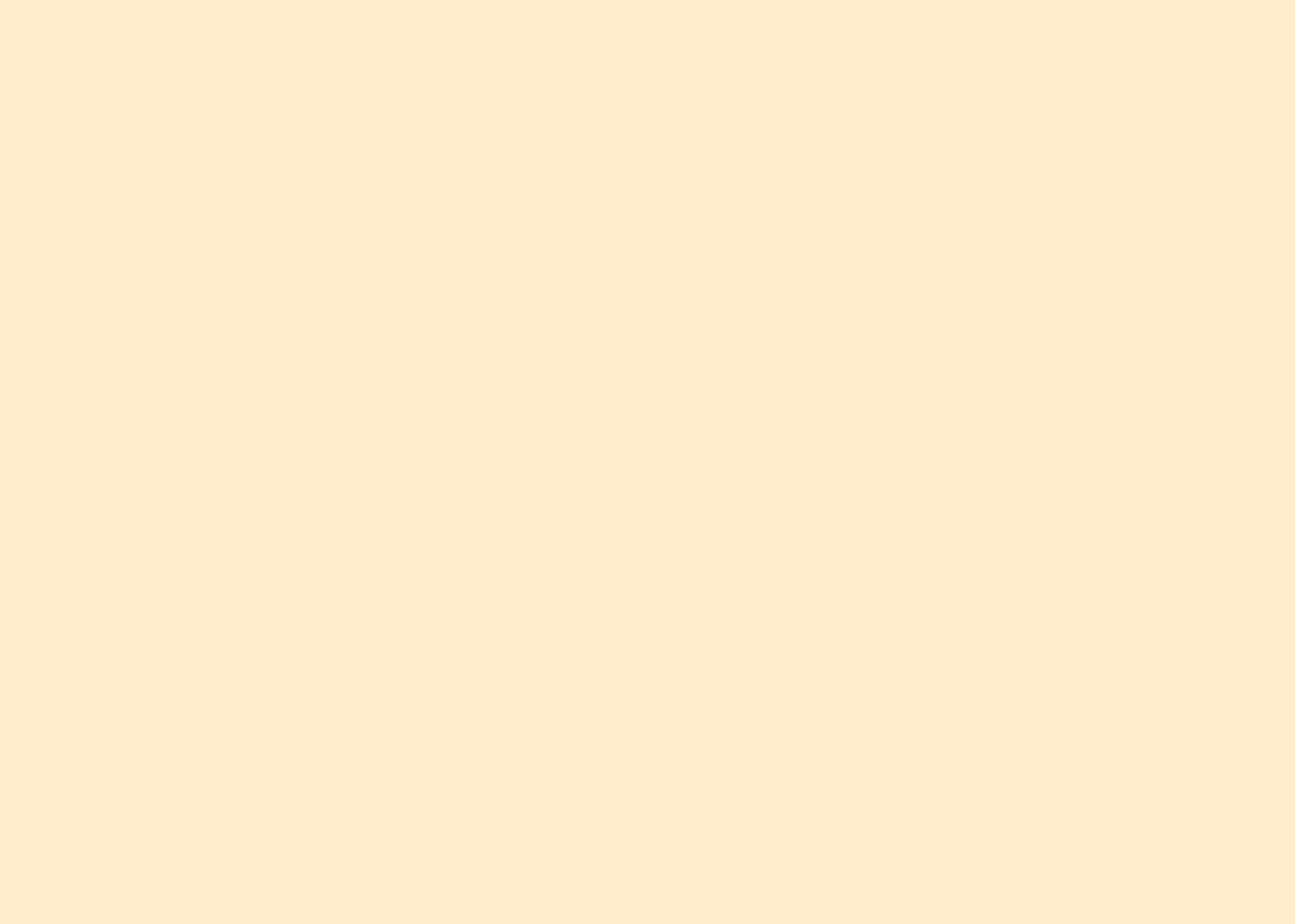
DHHS (NIOSH) Publication  
No. 2001-119

# Contact

---

[harry.shannon@mcmaster.ca](mailto:harry.shannon@mcmaster.ca)







# Complexity

---

- Feature of workplace studies
- For evaluations, subsidies policy can be seen as
  - Simple – offer of subsidies is ‘standardized’, how companies make changes is up to them
  - Complex – many different ways companies can implement changes, contexts differ, etc.
- Participating companies can be seen as
  - Simple – a company is a company
  - Complex – every company is unique, same interventions may have different effects

# Complexity

---

- For evaluations, subsidies policy can be seen as
  - Simple – offer of subsidies is ‘standardized’, how companies make changes is up to them
  - Complex – many different ways companies can implement changes, contexts differ, etc.
- If ‘simple’, can do RCT to estimate ‘average’ effect of subsidies
- But for policy adjustment, want to know *why* some companies are successful and others are not
- Implementers will want to know *what* to change
- Complexity leads to many evaluation questions

# Different sources of complexity – 1

## Characteristics of the intervention itself

---

- Multiple components – made up of various interconnecting parts
- Number of groups or organizational levels targeted by the intervention
- Degree of flexibility / tailoring of the intervention allowed
- Adaptivity, evolution over time

# Different sources of complexity – 2

## Characteristics of causal pathway

---

- Non-linear relationships
- Multiple mediators and pathways
- Feedback loops
- Synergy between components
- Number and variability of outcomes
- Connectivity – individual components of intervention are linked, so influence each other
- Interaction with context

# A Definition of Evaluation

---

A process that attempts to determine as *systematically and objectively* as possible the *relevance, effectiveness, and impact* of activities in the light of their *objectives*.

(emphasis added)

John Last

A Dictionary of Epidemiology, 1983

# Finnish study in shipyard

---

- Poster campaign on safety
- Widespread acceptance; viewed very positively by workforce
- But: no safety effect compared with control group
- “Although the campaign message was well received, it did not bring about changes in behavior.”

# Deaths of railway workers

---

- New helmet provided hearing protection
- Workers complained they couldn't hear anything
- Runaway train on line
- Workers apparently could not hear train approaching
- Mobile phone calls also not answered
- Four workers killed

theguardian.com 17 feb 2004  
(accessed 7 August 2014)

# Other high-level ‘interventions’

---

- Public pressure, e.g., ‘The Jungle’
- Health and safety legislation
- Government agency inspections
- No-fault workers’ compensation
- US Steel: ‘Safety pays’
- Change in attitude, from worker carelessness to managerial responsibility (late 19<sup>th</sup>, early 20<sup>th</sup> centuries) – ‘stunning reversal of earlier belief’ (Aldrich, 1997)
- UK HSC Vision ‘to see health and safety as a cornerstone of a civilized society’

# Low level interventions

---

- ‘Host of minor modifications’ in railways (Aldrich 2006, cited by Hemenway, 2009)
- US Navy ‘1001 Success Stories’

Little point in evaluating any one of them, but they can add up ...

# Work accident fatalities in the UK

---

	Number	Rate per 100,000 employees
1974	651	2.9
2012/13	95	0.4

<http://www.hse.gov.uk/statistics/history/historical-picture.pdf>

# Work permits for teenagers

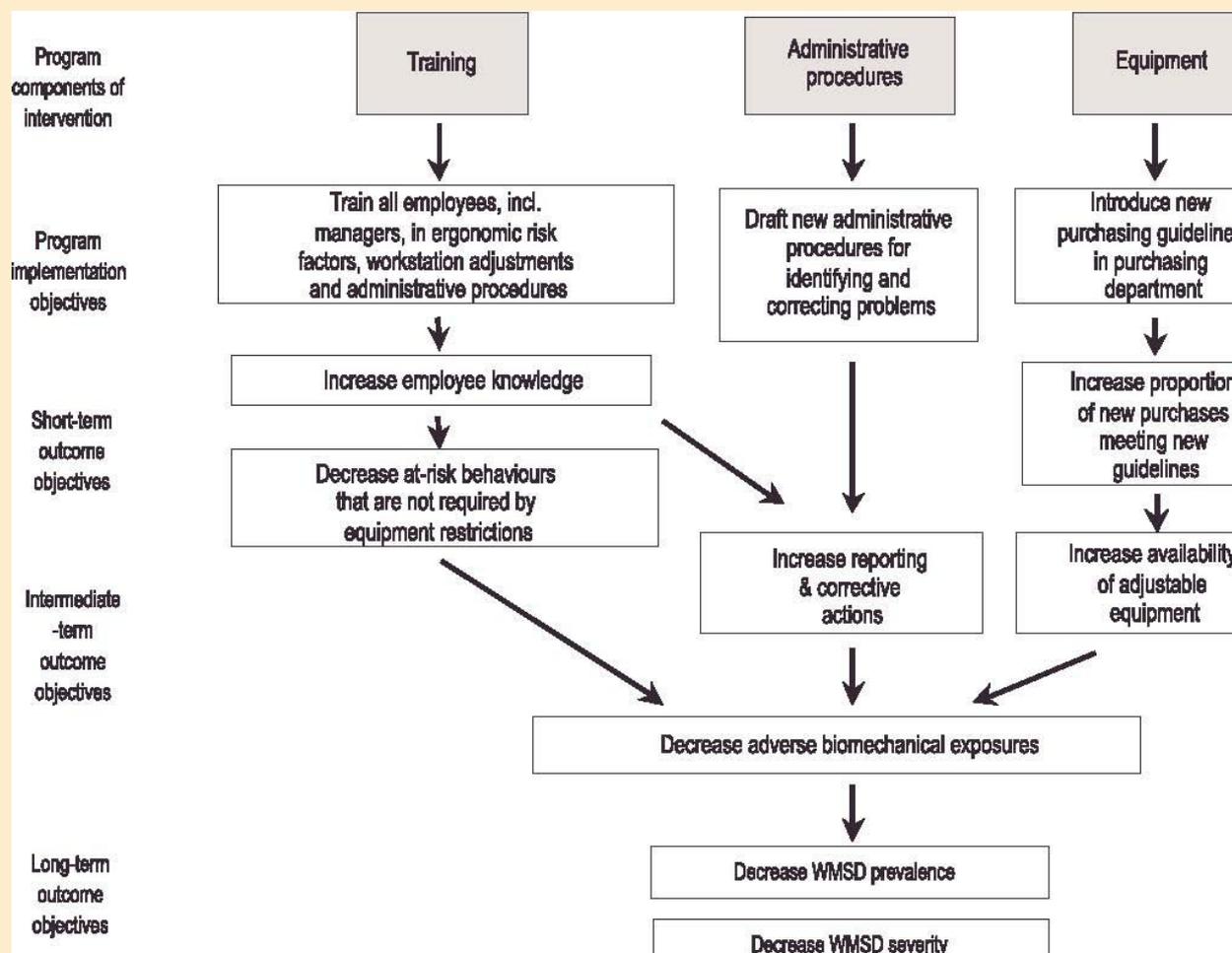
---

‘Although teens with work permits were more likely to report receiving safety training than teens without permits, the *occurrence of injury was no different among the groups.*’

(emphasis added)

Zierold and Anderson 2006; Am J Indust Med 49: 360-366

# Example of a program logic model for an ergonomic program



Source: Robson et al., 2001: 11

# Models in Planning Evaluations

---

- Conceptual Model

Shows presumed causal relationship based on theory, etc.

[Kurt Lewin: There's nothing as practical as a good theory]

- Program Logic Model

Shows how objectives (short-, medium-, long-term) may follow from the components of the safety measure.

# The Importance of Theory / Logic Model

---

- Helps establish *how* the outcome should change

Can protect against deluding yourself and being led astray by extraneous influences and chance occurrences

- E.g., suppose after the lifting program there was no change in behaviour, but fewer back injuries

Attribution to the program would be dubious

# Methodological criteria

---

## **Program objectives and conceptual basis**

- Were the program objectives stated?
- Was the conceptual basis of the program explained and sound?

## **Study design**

- Was an experimental or quasi-experimental design employed instead of a non-experimental design?

## **External validity**

- Were program participants/study population fully described?
- Was the intervention explicitly described?
- Were contextual factors described?

# Methodological criteria

---

## **Outcome measurement**

- Were all relevant outcomes measured?
- Was the outcome measurement standardized by exposure?
- Were the measurement methods shown to be valid and reliable?

## **Qualitative data**

- Were qualitative methods used to supplement quantitative data?

## **Threats to internal validity**

- Were the major threats to internal validity addressed in the study?

# Methodological criteria

---

## **Statistical analysis**

- Were the appropriate statistical analyses conducted?
- If study results were negative, were statistical power or confidence intervals calculated?

## **Conclusions**

- Did conclusions address program objectives?
- Were the limitations of the study addressed?
- Were the conclusions supported by the analysis?
- Was the practical significance of the result discussed?

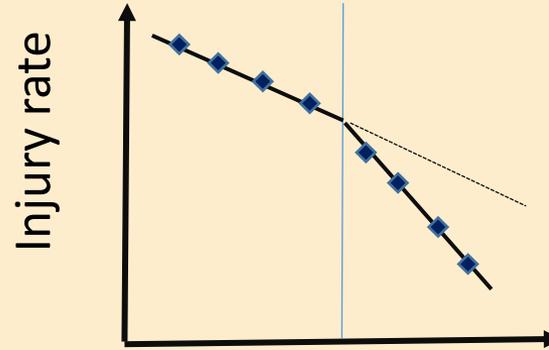
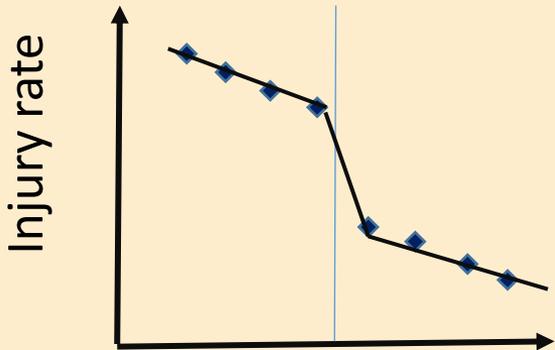
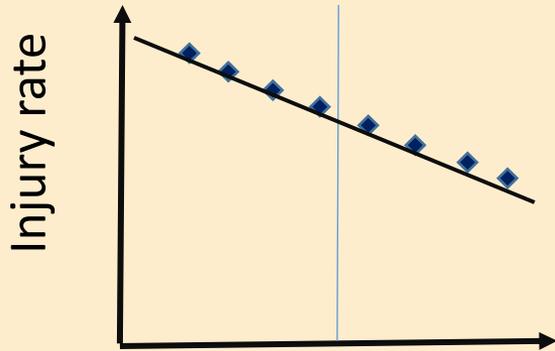
# Improving the Design: Turning a non-experimental study into a quasi-experiment

---

1. Add a control group
2. Take more measurements
3. Stagger the intervention timing in different groups (multiple baseline)
4. Add a “reversal” of the intervention
5. Use additional outcome measures
6. Combinations of 1-5

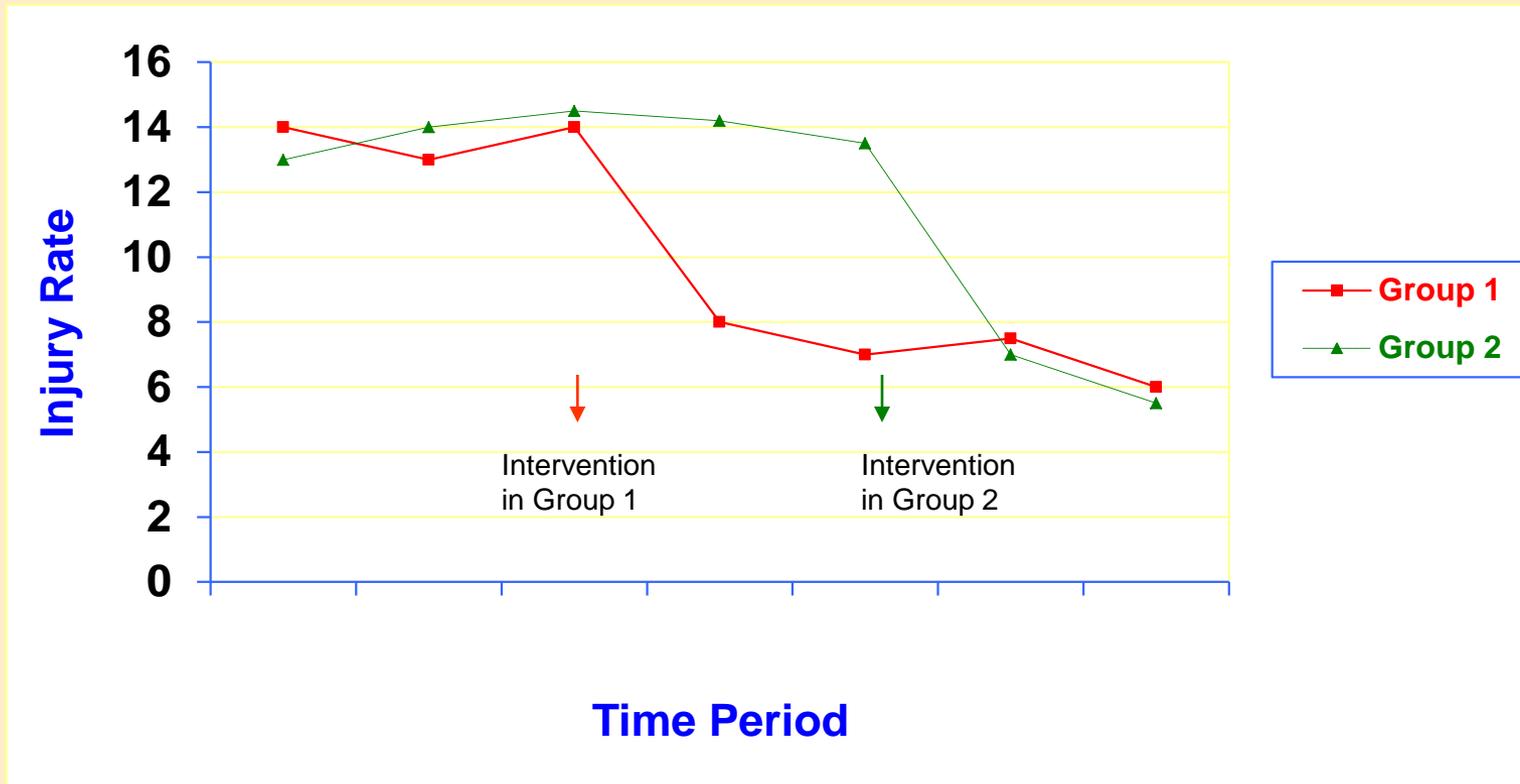
# Strategy 2: more measurements

---



‘Interrupted time series’

# Strategy 3: How Staggering the Intervention Can Help



# Validity Threats in Studies with Control Groups

---

- Arise when control and intervention groups differ by more than the intervention
- Types of threat:
  - Selection
  - Selection Interaction
  - Diffusion/Contamination
  - Rivalry/Resentment

# Even better than a quasi-experimental design

---

- A **true experiment** – i.e., a randomized controlled trial (RCT)

# Effectiveness of government safety inspections

---

- California OSHA *randomly* selected workplaces in high-injury industries for inspections
- Avoided bias of selecting only companies with high injury rates
- Matched each company with one not inspected
- Matching by industry, region of state, size, ...
- Compared injuries and costs up to 4 years post-inspection
- Also examined employment, sales, credit ratings, ‘survival of firms’
- Called for qualitative research to understand ‘process’

Levine, Toffel, Johnson. *Science* 2012; 336: 907-911

# Cluster RCT of job rotation to prevent musculoskeletal disorders - protocol

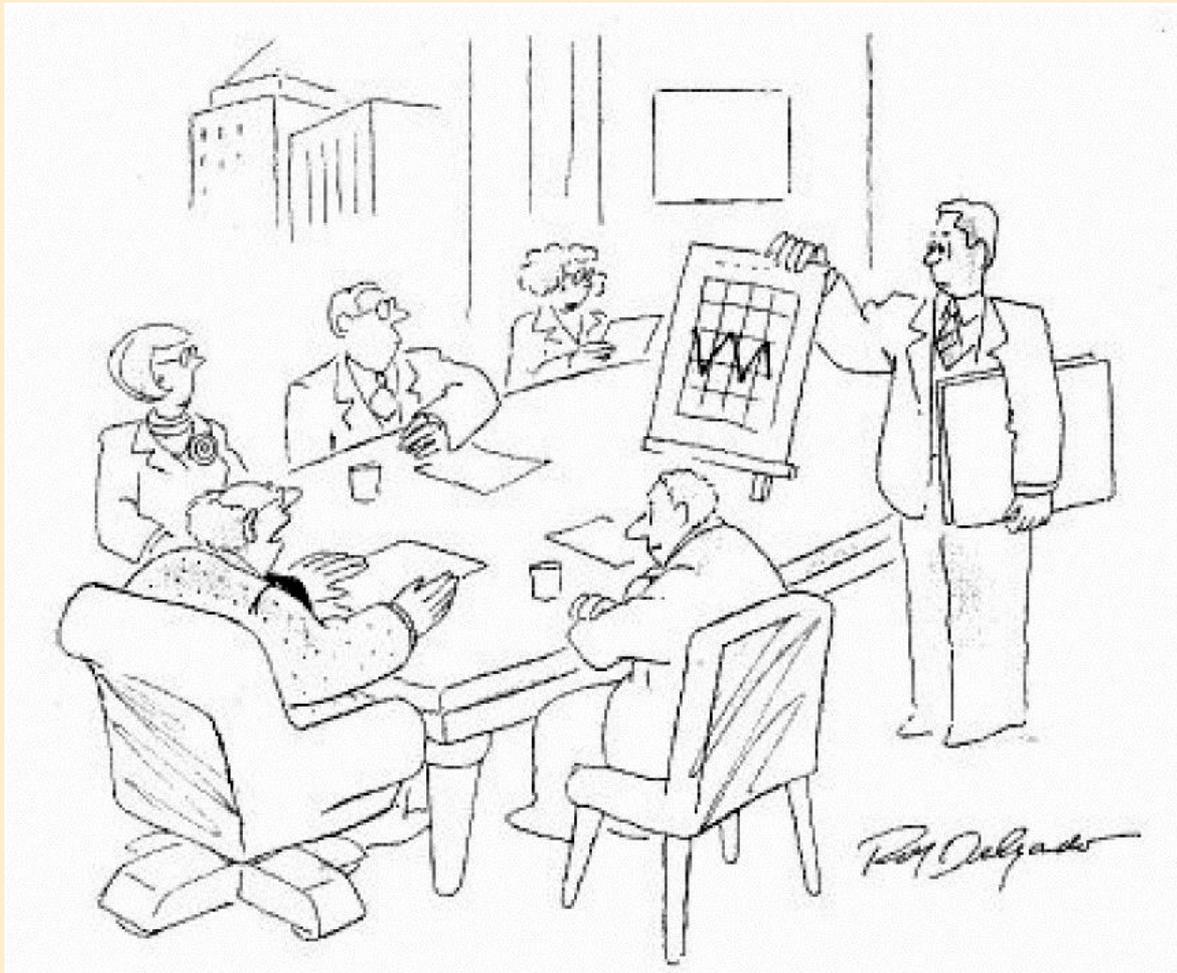
---

- Production line workers
- Randomized *departments* to one of two groups
  - Training on ergonomic guidelines
  - Training on ergonomic guidelines + job rotation
- Job rotation based on ergonomic assessments using standardized measures
- Outcome : working hours lost to MSKs over next 12 months
- Protocol recently published

# When might there be problems doing an RCT?

---

- Intervention is 'complex'
- Intervention is legislation
- Ethical concerns about randomization
- When the units are sectors of the economy



"Your facts are impressive, but we need instinct to back them up."

New Yorker

# “No control group”?

---

No comparison group for the intervention companies

But for the specific interventions ...

# “No control group”?

As shown in paper for  
“Competence / motivation of safety staff”

	Success		
Intervention	Yes	No	Total
Yes	2	3	5
No	6	6	12
Total	8	9	17



Of successful companies, intervention used in  $2/8 = 25\%$

Of not successful companies, intervention used in  $3/9 = 33\%$

# “No control group”?

My interpretation

	Success		
Intervention	Yes	No	Total
Yes	2	3	5
No	6	6	12
Total	8	9	17



In intervention companies, success in  $2/5 = 40\%$

In control companies, success in  $6/12 = 50\%$

# Additional topics

---

- How to improve evaluations
  - Methodological
  - Practical
  - Questions to ask
- Importance of theory
- Role of complexity ('realist evaluations')

# Improving evaluations - 1

## Methodological

---

- Sufficient power to deal with clustering
- Proper analysis: unit targeted, clustering, etc.
- Replication in new settings
- Base evaluations on underlying theory
- Consider implementation climate (readiness-for-change) and quality
- Better reporting of intervention itself

# IWH Organization Performance Metric

---

1. Formal safety audits at regular intervals are a normal part of our business
2. Everyone at this organization values ongoing safety improvement in this organization
3. This organization considers safety at least as important as production and quality in the way work is done
4. Workers and supervisors have the information they need to work safely
5. Employees are always involved in decisions affecting their health and safety
6. Those in charge of safety have the authority to make the changes they have identified as necessary.
7. Those who act safely receive positive recognition.
8. Everyone has the tools and/or equipment they need to complete their work safely.

# Improving evaluations – 2

## Practical

---

- Involving stakeholders (be aware they may feel threatened)
- Create culture of evaluation – persuade policy makers, stakeholders to do evaluations
- Build evaluation into intervention
- Take advantage of opportunities (e.g., Levine et al.)
- Train people with expertise in evaluation

# Improving evaluations – 3

## Interventions to study

---

- Do more evaluations of ‘knowledge transfer’, implementation of ‘proven’ approaches
- More evaluations on broadly applied interventions (cf post-marketing surveillance)
- Ask the right questions
  - Is it an important question?
  - Has question been answered?

# Improving evaluations – challenges

## Methodological

---

- Inverse relationship between rigour and feasibility, between rigour and level of intervention
- Single companies can only do Before-After studies (but also incorporate qualitative data)
- Sample size issues
- Valid measures of safety (under-reporting of injuries?)
  - surrogate or intermediate?
  - leading or lagging indicators?

# Improving evaluations – challenges

## Practical

---

- Lack of expertise (core competency for OHS certification?)
- Mindset (culture)
- Unwillingness to learn program not effective
- Small businesses
- Ethical issues
- Getting companies to participate - incentives?
  - Donnerlein and Sparer – selling point
  - Reduced premiums

# Some extra points 1

---

- Complex interventions
- ‘Realist evaluation’
- Applying what we know, barriers to action
- Systematic reviews – many conclude: few hi qual studies, more needed
- Possibilities of RCTs?
- How to improve – TIDieR
- How good is good enough? Don’t let best be enemy of the good – balance rigour and feasibility

# Some extra points 2

---

- Few evaluations – they are difficult, can be ruined mid-stream
- Correlational studies
- Separating out ‘active ingredients’
- Sample size for outcomes or intermediate measures
- Proper statistical analyses (esp clustering)
- Generalizability – need evaluations across industries, cultures, jurisdictions

# Some extra points 3

---

- Contrast:
  - low level interventions, hardly worth evaluating
  - High level interventions, important to evaluate, but can be hard (or impossible?)
- Context

# How low can they go?

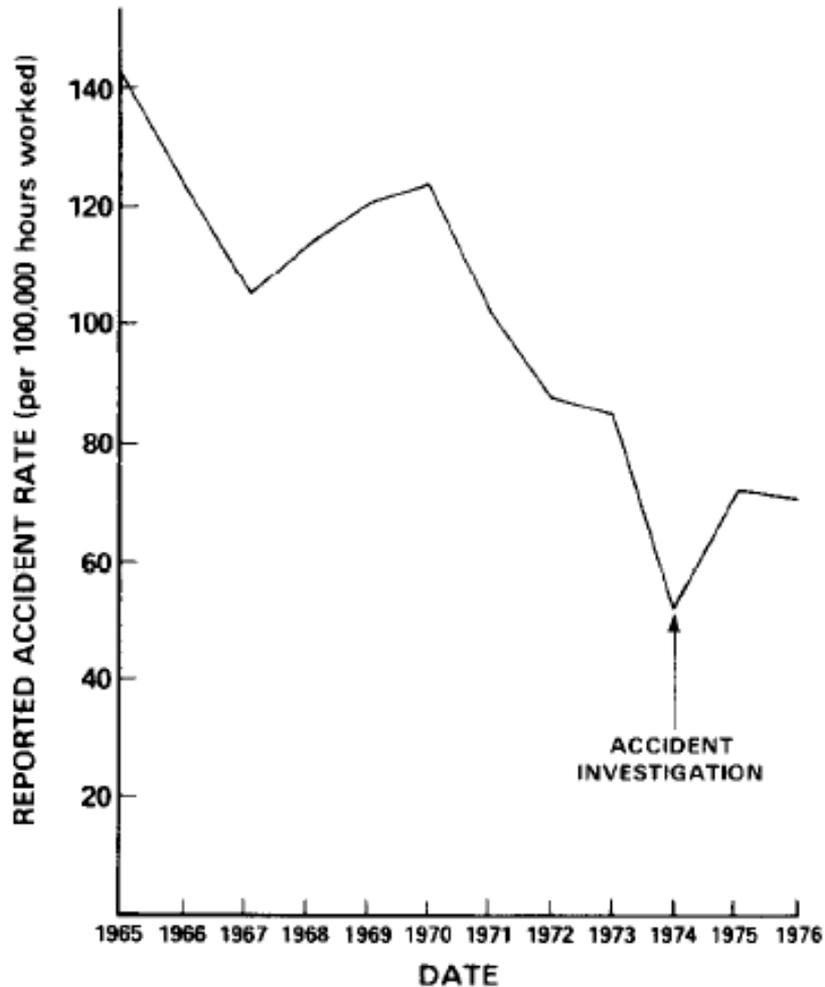
## Potential for reduction in injury rates

---

- Ontario workers' compensation data
- Grouped companies by type of work, size
- Identified injury rate at 25<sup>th</sup> percentile
- Estimated number of injuries if all companies experienced that rate
- Represented 42% reduction in injuries
- Eminently achievable using current approaches
- Need to understand barriers to doing this

Shannon and Vidmar, Injury Prev, 2004; 10: 290-295

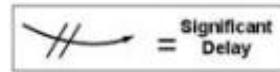
# Under-reporting of injuries



Drop in 1974 resulted from doing the study

Shannon and Manning, 1979

# Afghanistan Stability / COIN Dynamics



- Population/Popular Support
- Infrastructure, Economy, & Services
- Government
- Afghanistan Security Forces
- Insurgents
- Crime and Narcotics
- Coalition Forces & Actions
- Physical Environment

