

Yes, When Will We Ever Learn? Exemplars of Strategies for Causal Attribution in Evaluation

Evaluation Cafe
The Evaluation Center
University of Western Michigan
25 April 2007

Associate Professor Patricia J. Rogers
CIRCLE at RMIT University, Australia

Overview

1. The story so far – and why we need exemplars
2. Exemplars supporting the need for more rigorous impact evaluation
 - Carrots
 - Ambulances
3. Exemplars raising concerns about RCTs
 - Parachutes
 - Plants
 - Comprehensive Child Development Program
4. Exemplars outlining alternative strategies
 - John Snow
 - Lung cancer
 - Sudden Infant Death Syndrome
 - New York crime rate
5. Learning lessons from exemplars



One way of summarising the story so far...

Episode 1:
A New Hope






The end of the paradigm wars

One way of summarising the story so far...

<p>Episode 1: A New Hope</p>		<p>The end of the paradigm wars</p>
<p>Episode 2: The Empire Strikes Back</p>		<p>The international move to advocate for RCTs as the most rigorous method for impact evaluation</p>

One way of summarising the story so far...

<p>Episode 1: A New Hope</p>		<p>Claims of the end of the paradigm wars</p>
<p>Episode 2: The Empire Strikes Back</p>		<p>The international move to advocate for RCTs as the most rigorous method for impact evaluation</p>
<p>Episode 3: Return of the Jedi</p>		<p>Reconciliation and a happy end</p>

Exemplars supporting the need for more rigorous impact evaluation

- Carrots
- Ambulances



CARROTS

FACT: 95%* of people who die have recently eaten carrots

ERRONEOUS CONCLUSION: Carrots are fatally dangerous

CORRECT CONCLUSION: Most people have recently eaten carrots and they are rarely related to death

(* actually this is a made up statistic)



CARROTS

FACT: 95%* of people who die have recently eaten carrots

ERRONEOUS CONCLUSION: Carrots are fatally dangerous

CORRECT CONCLUSION: Most people have recently eaten carrots and they are rarely related to death

LESSON FOR EVALUATION: Need comparison with those not receiving intervention or not having that outcome

(actually this is a made up statistic)

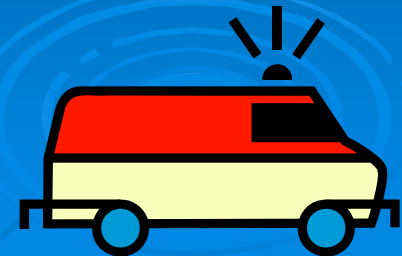


AMBULANCES

FACT: Patients who had travelled further in an ambulance were more likely to die

ERRONEOUS CONCLUSION: Lengthy transport is dangerous

CORRECT CONCLUSION: High-risk patients are more likely to be taken to centralised facilities not local ones



AMBULANCES

FACT: Patients who had travelled further in an ambulance were more likely to die

ERRONEOUS CONCLUSION: Lengthy transport is dangerous

CORRECT CONCLUSION: High-risk patients are more likely to be taken to centralised facilities not local ones

LESSON FOR EVALUATION: Need to consider systematic differences between the two groups



Exemplars raising issues about RCTs

- Parachutes
- Plants
- Comprehensive Child Development Program

PARACHUTES

FACT: A systematic review was unable to find any randomised controlled trials of parachute intervention.

ERRONEOUS CONCLUSION: Parachute use should be discontinued until evidence from RCTs supports their use

Smith and Pell (2003)



PARACHUTES

SUGGESTED CONCLUSION:

“Advocates of evidence based medicine have criticised the adoption of interventions evaluated by using only observational data. We think that everyone might benefit if the most radical protagonists of evidence based medicine organised and participated in a double blind, randomised, placebo controlled, crossover trial of the parachute.”

LESSON FOR EVALUATION: Lack of RCT evidence does not constitute lack of credible evidence

Smith and Pell (2003)



PLANTS

FACT: If plants are randomly assigned to a treatment providing daily water, or to a control that receives none, and both groups are placed in a dark cupboard, the treatment group does not have better outcomes than the control.

ERRONEOUS CONCLUSION: Watering plants is ineffective in making them grow.

CORRECT CONCLUSION: Plants need water AND light.

LESSON FOR EVALUATION: Need to pay attention to the whole causal package needed to produce the outcomes.

COMPREHENSIVE CHILD DEVELOPMENT PROGRAM

FACT: Families in the program improved their level of functioning in terms of intermediate and final outcomes (following program theory)

SPURIOUS CONCLUSION: Program was effective

St Pierre et al 1996, St Pierre and Rossi 2006



COMPREHENSIVE CHILD DEVELOPMENT PROGRAM

FACT: Families in the program improved their level of functioning in terms of intermediate and final outcomes (following program theory)

SPURIOUS CONCLUSION: Program was effective

ANOTHER FACT: Families which were randomly assigned to the control group improved by an equivalent amount

CONCLUSION: No evidence that the program was effective

POSSIBLE LESSON: Only an RCT can provide evidence of causal attribution

St Pierre et al 1996, St Pierre and Rossi 2006



COMPREHENSIVE CHILD DEVELOPMENT PROGRAM

FACT: Families in the program improved their level of functioning in terms of intermediate and final outcomes (following program theory)

SPURIOUS CONCLUSION: Program was effective

ANOTHER FACT: Families which were randomly assigned to the control group improved by an equivalent amount

CONCLUSION: No evidence that the program was effective

POSSIBLE LESSON: Only an RCT can provide evidence of causal attribution

ANOTHER FACT: Many control group families were able to obtain services on their own ('contamination')

CONCLUSION: The evaluation was not effective

POSSIBLE LESSON : Need to randomly assign groups
individuals

St Pierre et al 1996, St Pierre and Rossi 2006



COMPREHENSIVE CHILD DEVELOPMENT PROGRAM

FACT: “One of the 21 sites in the study had statistically significant and moderately large positive effects in several different outcome domains: children’s cognitive development; families’ employment , income, and use of federal benefits; and parenting attitude. No single factor can be pointed to as ‘the reason’ why CCDP was more effective on Site #2 than on other sites. The circumstances and context of Site #2 were probably unique, and certainly acted in concert to produce positive effects”.

CONCLUSION: No conclusion was possible

POSSIBLE LESSON: Success may not be due to a single factor. When something is working very well, we should try harder to learn from it

St Pierre et al 1996,

Limitations of RCTs

- False negatives - when an RCT evaluation wrongly concludes that an intervention doesn't work.
- False positives - when an RCT evaluation wrongly concludes that an intervention does work.
- Negative impact of the evaluation process itself

False negatives in RCTs

- Concluding there is no credible evidence of effectiveness when RCTs are not possible.
- The intervention by itself may not be sufficient to cause the outcomes, but also need the contribution of other interventions and/or favourable circumstances.
- The intervention may not be the only way to achieve the outcome – meaning that when the outcomes for those receiving the intervention are compared to those who did not receive it there is little evident difference
- Failing to detect inadequacies in implementation which are by themselves enough to prevent the outcomes from being achieved
- Failing to identify sub-groups or particular cases which have succeeded

False positives in RCTs

- Corruption of evaluation findings because of self-interest
- Mis-interpretation of statistically significant findings
- Knowing that they are receiving the intervention changes the way participants, staff and others behave, leading to positive outcomes that will not be sustained when the research component is removed
- Lack of external validity- inability to reproduce the favorable implementation environment of the experimental trial

An alternative approach

- List of Possible Causes (LOPC)
- General Elimination Methodology

Scriven (2007)



Mayne's Contribution Analysis: Addressing Attribution with Performance Measures

- Acknowledge the problem
- Present the logic of the program.
- Identify and document behavioural changes.
- Use discriminating indicators.
- Track performance over time.
- Discuss, and test alternative explanations.
- Gather additional relevant evidence.
- Gather multiple lines of evidence.
- When required, defer to the need for an evaluation.

Mayne (2001)

Exemplars demonstrating an alternative to RCTs

- John Snow
- Lung cancer
- SIDS (Sudden Infant Death Syndrome)
- New York crime

Exemplar 1: John Snow

➤ The problem

Four cholera pandemics in London in the 1800s
14,600 deaths in London (6.2 per 1,000)

➤ Alternative theories as to cause

- Airborne spread of disease
- Waterborne spread of disease

Koch (2004)



Analysis

Cholera deaths in Soho mapped

“It might be noticed that the deaths are most numerous near to the pump in Broad Street”.



From Koch (2004)



Looking for exceptions

- Deaths in streets nearer to the Rupert Street pump
 - “some streets which are nearer to it on the map are in fact a good way removed on account of the circuitous road to it”.

Looking for exceptions

- Few deaths in the workhouse near the Broad Street pump:
 - Surrounded by houses with cholera deaths but only 5/535 inmates died from cholera
 - Had its own pump well on the premises and “the inmates never sent to Broad Street for water”

Looking for exceptions

- Few deaths among workers at the brewery near the epicentre of cholera deaths
 - No cholera among 70 workers
 - Had its own private deep well on-site
 - Workers never drank from the Broad Street pump

Looking for exceptions

- Deaths of 4 school children who did not live near the Broad Street pump
 - Drank from the pump on their way to school
- Deaths of 2 adults living further away
 - One temporarily living in Broad Street
 - One nursing a dying friend from Broad Street

Learning from the exemplar

- Search for exceptions and seek to explain them



Exemplar 2: Lung cancer

- The problem

 - From 1900 increase in numbers dying of lung cancer

- Methods

 - Retrospective studies

 - Prospective studies

 - Animal studies

Looking for exceptions

- Not everyone who gets lung cancer has smoked

Some of these have been affected by second-hand smoke – a different causal path caused by smoking

Looking for exceptions

- Not everyone who smokes gets lung cancer

Theory of cancer initiator and promoter – many things can initiate cancer but smoke is a major promoter

If lung cancer has begun, smoking will promote it

If lung cancer has not begun, smoking may not have an effect

LESSON FOR EVALUATION: Pay attention to the causal package

Information from documents from tobacco companies:	What the tobacco industry was saying publicly:	What the health authorities were saying:
<p>'At the best, the probabilities are that some combination of constituents of smoke will be found conducive to the onset of cancer or to create an environment in which cancer is more likely to occur.' 1963.</p>	<p>'The smoking of tobacco continues to be one of the subjects requiring study in the lung cancer problem, as do many other agents and influences in modern living. Science does not yet know enough about any suspected factors to judge whether they may operate alone, whether they may operate in conjunction with others, or whether they may affect or be affected by factors of whose existence science is not yet aware.' 1967.</p>	<p>Surgeon General: 'Cigarette smoking is causally related to lung cancer in men: the magnitude of the effect of cigarette smoking far outweighs all other factors. The data for women, though less extensive point in the same direction.' 'Cigarette smoking is a significant factor in the causation of cancer of the larynx (and) an association exists between cigarette smoking and cancer of the urinary bladder in men.' 1964.</p>

Information from documents from tobacco companies:	What the tobacco industry was saying publicly:	What the health authorities were saying:
<p>'Carbon monoxide will become increasingly regarded as a serious health hazard for smokers . . . Additional evidence of smoke-dose related incidence of some diseases associated with smoking has been published. But generally this has long ceased to be an area for scientific controversy.' 1978.</p>	<p>' . . . we believe there is sound evidence to conclude that the statement 'cigarettes cause cancer' is not a statement of fact but merely an hypotheses.' 1971.</p>	<p>Surgeon General: 'Tobacco use is associated with increased risk of coronary heart disease; stroke; aortic aneurism; peripheral vascular disease; chronic obstructive broncho-pulmonary disease; cancers of the lung, lip, larynx, oral cavity, oesophagus, urinary bladder, and pancreas; and gastrointestinal disorders such as peptic ulcer disease . . . In addition, maternal smoking during pregnancy retards foetal growth.' 1972.</p>

Information from documents from tobacco companies	What the tobacco industry was saying publicly:	What the health authorities were saying:
	'Cigarette smoking has not been scientifically established to be a cause of chronic diseases, such as cancer, cardiovascular disease, or emphysema. Nor has it been shown to affect pregnancy outcome adversely.' 1983.	Surgeon General 'Smoking is responsible for more than one of every six deaths in the United States. Smoking remains the single most important preventable cause of death in our society.' 1989.

LESSON FOR EVALUATION: Pay attention to the incentives for discrediting evidence of impact

Exemplar 3: Sudden Infant Death Syndrome

➤ The problem

Unexplained death of babies

In the US and in Australia SIDS was the main cause of death among infants for most of the 1990s.

➤ Methods

Review of epidemiological data

Review of individual cases

Development of alternative hypotheses

Investigation of evidence

Development of guidelines for practice in the light of incomplete evidence

- At least 19 retrospective case-control studies demonstrated a higher risk of SIDS when infants slept prone, with odds ratios ranging from 1.2 to 14.1. Overall the studies showed a threefold or greater increased risk of SIDS when babies slept prone.

Public Health Association of Australia (1999/2005)

- An increased risk of SIDS when babies are exposed to cigarette smoke has been found in over 30 case-control and cohort studies. This finding is consistent over time and place. Many studies have reported a dose-response relationship.

Public Health Association of Australia (1999/2005)

Guidelines for practice

- Avoid maternal smoking
- Lie the baby to sleep on its back
- Avoid over-heating

FACT: In Australia, rates were higher in winter.

POSSIBLE EXPLANATION: Linked to winter respiratory diseases.

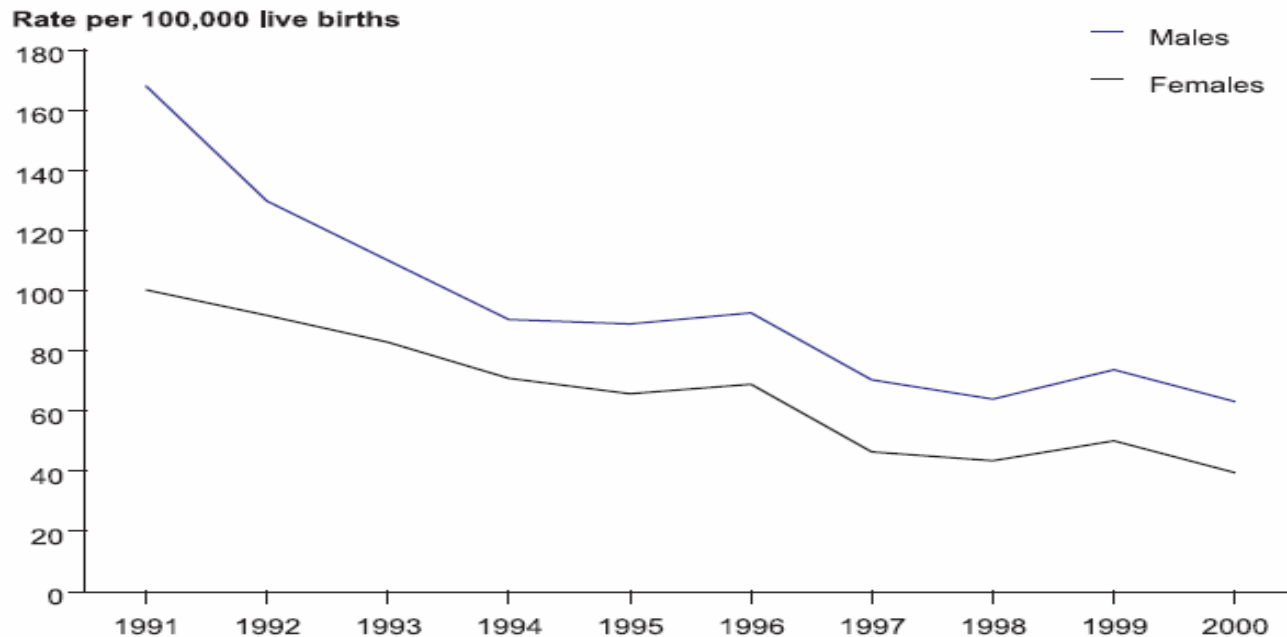
ANOTHER FACT: Rates were not higher in winter in Sweden – which has colder winters

ANOTHER FACT: A review of cases found a large number who had been put to bed fully dressed or with additional bedding

POSSIBLE EXPLANATION: In under-heated houses, parents over-dressed their children. Over-heating seemed to be a risk factor.

RECOMMENDATION FOR PRACTICE: Avoid overheating: The infant should be lightly clothed for sleep, and the bedroom temperature should be kept comfortable for a lightly clothed adult. Overbundling should be avoided, and the infant should not feel hot to the touch.

SIDS rates in Australia

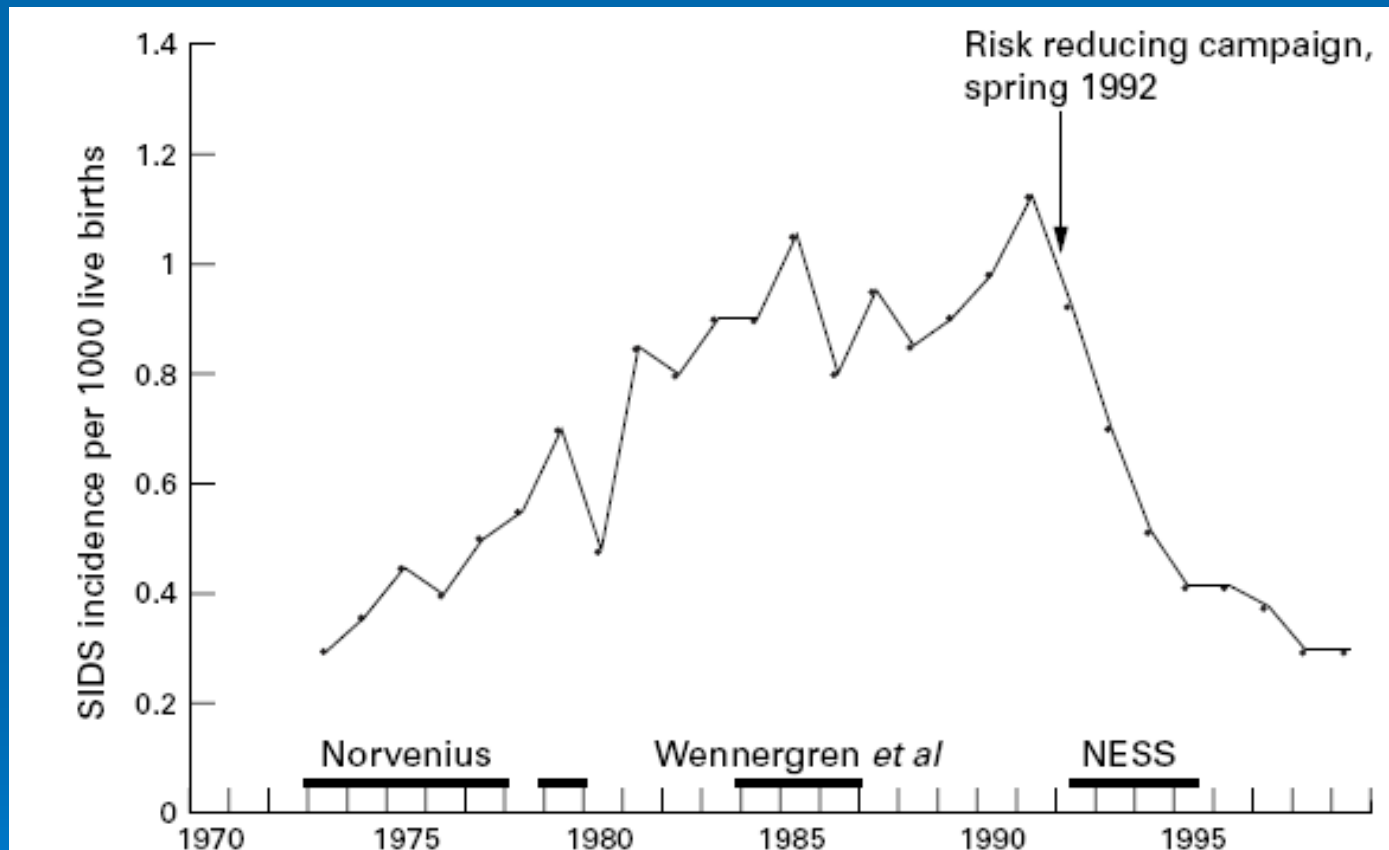


Note: ICD-9 code 798.0 (1991 to 1996) and ICD-10 code R95 (1997 to 2000).

Source: AIHW Mortality Database [Table A8.1].

Figure 8.1: SIDS rates for infants, 1991-00

SIDS rates in Sweden



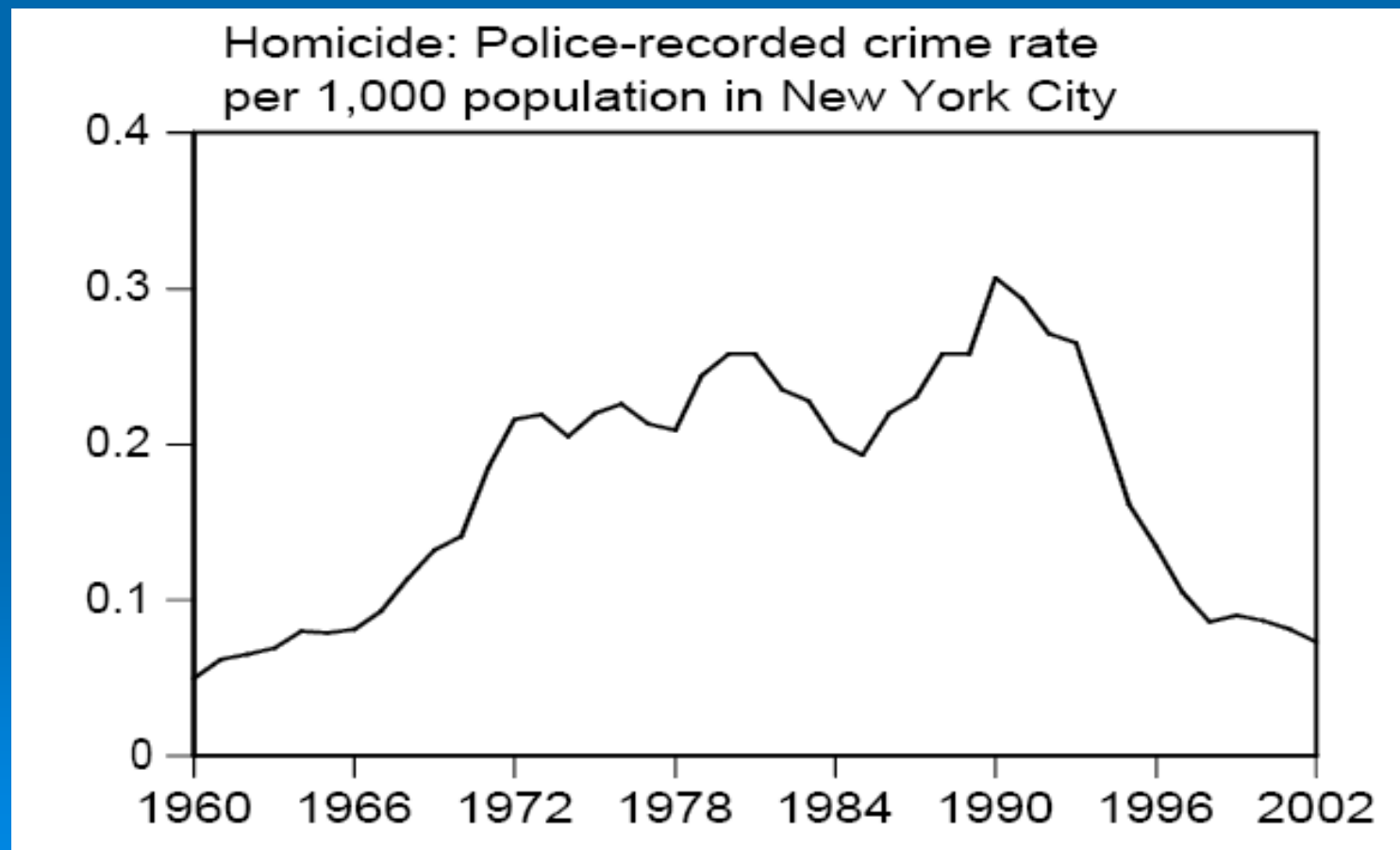
Alm et al 2001

- **LESSONS FOR EVALUATION:** Less than perfect information can be useful



Exemplar 4: New York crime rate

FACT: Crime rates fell



POSSIBLE EXPLANATIONS:

New approach to policing – zero tolerance and CompStat

Would expect decline to follow implementation in 2004, and not be present in other cities

Increased number of police

Would expect decline to follow increased police from 1991

Ageing of the population – fewer young men

Would expect a gradual decline as population aged

Improved access to community colleges

Would expect a decline linked to the time of introduction

Wider access to abortions from 1973

Would expect decline to have started earlier in states which introduced this earlier, and to be greater in states with high abortion rates

“Over a five year period, New York experienced a precipitous drop in the burglary rate (53%) a 54% drop in reported robberies, and an incredible 67% drop in the murder rate. ... These extraordinary achievements were realized in large part due to the department’s innovative model of police management, known as CompStat.”

(O’Connell, 2001, p.8) (Quoted in Kusek and Rist, 2004, p. 141)

What we can learn from these exemplars



Strategies for impact evaluation of programs to address:

- Obesity and related diseases?
- Climate change?
- World poverty?

Patricia.Rogers@rmit.edu.au

CIRCLE at RMIT University

Collaborative Institute for Research, Consulting and Learning in Evaluation,

Royal Melbourne Institute of Technology

Melbourne, AUSTRALIA



References and Links

Byard et al. SIDS rates and risk campaigns [electronic response to Goldwater PN; Sudden infant death syndrome: a critical review of approaches to research] archdischild 2004

<http://adc.bmjournals.com/cgi/eletters/archdischild;88/12/1095#593>

Goldwater, P. N. (2003) Sudden infant death syndrome: a critical review of approaches to research. Arch. Dis. Child. 2003;88;1095-1100

Koch, T. (2004) The Map as Intent: Variations on the Theme of John Snow. CARTOGRAPHICA, 39, # 4, WINTER 2004: 1-14.

Kusek and Rist (2004) Ten Steps to a Results-Based Monitoring and Evaluation System.

Levitt and Dubner (2005) Freakonomics.

Mayne, John (2001) Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly. Canadian Journal of Program Evaluation, 19 (1): 31-60.

Quit SA (2006) The tobacco industry. Available at http://www.quitsa.org.au/asp/the_tobacco_industry.aspx

References and Links

Rogers, P. J. (2000). *Causal models in program theory evaluation*. In P. J. Rogers, T. A. Hacsí, A. Petrosino, & T. A. Huebner (Eds.), *Program theory in evaluation: Challenges and opportunities*. *New Directions for Evaluation*, 87, 47-55.

Scriven, M. (2007) *The logic of causal investigations*.

Smith, G. and J. Pell (2003). "Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials." *BMJ* 327(7429): 1459-61.

St Pierre et al, 1996 *Report on the National Evaluation of the Comprehensive Child Development Program*. Summary and links to reports available at http://www.researchforum.org/project_abstract_166.html

St Pierre and Rossi 2006. *Early Childhood Programs Randomize Groups, Not Individuals: A Strategy for Improving Early Childhood Programs Evaluation Review*; 30; 656