1

Introduction

1.1 Why study interventions?

Because solving problems requires interventions but not all interventions are effective!

Solving problems requires interventions.

Let's say you are confronting the problem of obesity in your clinic or community. A necessary first step is to understand the prevalence (frequency) of obesity, the characteristics of affected individuals (e.g., age, gender, geographic location), and how severely they are affected (e.g., presence of comorbid conditions such as diabetes). A second necessary step is to identify the risk factors for obesity, especially those that are amenable to change.

But sadly too many research agendas never move beyond this second step. Investigators conclude their manuscripts with the ubiquitous and meaningless phrase: "Interventions are needed to …. ," yet the intervention is never performed. A review of bibliography sources found that only 0.4 percent of academic research focused on public health interventions.[1] Although intervention research is more common with pharmaceuticals, this research is often limited to efficacy trials conducted under conditions that cannot easily be replicated in the real world.

Develop an intervention and change the world!

This does not have to be the case. Developing interventions can be more fulfilling than descriptive or risk-factor studies because they can directly change the world! Interventions can be drugs or medical devices, counseling or skill-building programs for individuals or groups, laws or changes in institutional practices.

Let's return to the problem of obesity. A large body of evidence has documented a major increase in obesity rates in developed countries, with serious sequelae including type 2 diabetes. Decreased activity, larger food portions, and high caloric foods have all been shown to be risk factors for obesity.

[1] Millward, L., Kelly, M., and Nutbeam, D. *Public Health Intervention Research: The Evidence*. London: Health Development Agency, 2003. www.nice.org.uk/niceMedia/documents/pubhealth.interventon.pdf. Accessed 3 March, 2008.

Farley and colleagues developed a common sense intervention to increase the activity of children and thereby diminish a known risk factor for obesity: they opened a schoolyard for play during non-school hours in a low-income neighborhood in New Orleans and provided attendants to ensure children's safety.[2] The schoolyard was immediately popular: 710 children were observed in the schoolyard at least once during a 12-month period; 66% of the children were physically active when observed in the schoolyard.

To evaluate the impact of the schoolyard on the physical activity of children in the community, the investigators compared children in this community to children from a neighboring community. Prior to the opening of the schoolyard the number of children observed to be active in the intervention community was lower than in the comparison community. After the intervention, the number of children observed to be active was greater in the intervention community than in the comparison community, **not counting the children in the schoolyard.**

Besides being easy to replicate, the intervention had the advantage that it does not single out children who are obese, which may harm self-image. Rather it takes advantage of the idea that all children should be active.

Consider evaluating an intervention developed by the government or other organization.

To ameliorate a health problem you don't have to be the one who develops the intervention. Many useful studies have evaluated interventions that were developed by the government (e.g., laws banning smoking at workplaces) or another organization (e.g., school-based physical education). For example, Hu and colleagues assessed the impact of taxation on cigarette sales.[3] They estimated that the 25-cent tax that California added in 1989 to each pack of cigarettes resulted in a reduction of 514 million packs sold over an 18-month period. Evaluations such as this have been successful in motivating other states to add tobacco excise tax.

Keep your eyes open for a natural experiment.

You may also find an unplanned opportunity, a "natural experiment," to evaluate whether a change in circumstances improves a health problem. For example, Costello and colleagues studied the impact of the opening of a casino on an Indian reservation on the mental health of Native American children.[4] The point of the casinos (the intervention) was not to improve children's mental health. However, it is known that psychopathology is higher in children from low-income families and that casinos increase the income of families living on

[2] Farley, T. A., Meriwether, R. A., Baker, E. T., Watkins, L. T., Johnson, C. C., and Webber, L. S. "Safe play spaces to promote physical activity in inner-city children: results from a pilot study of an environmental intervention." *Am. J. Public Health* **97** (2007): 1625–31.

[3] Hu, T., Sung, H. Y., and Keeler, T. E. "Reducing cigarette consumption in California: tobacco taxes vs an anti-smoking media campaign." *Am. J. Public Health* **85** (1995): 1218–22.

[4] Costello, E. J., Compton, S. N., Keeler, G., and Angold, A. "Relationships between poverty and psychopathology: a natural experiment." *JAMA* **290** (2003): 2023–9.

a reservation. Would the presence of the casinos improve the mental health of the children?

The investigators found that after the opening of the casino the psychopathology level of the previously poor children improved to the level of the children who were never poor. The study was feasible only because a cohort study was in progress prior to the opening of the casino. The beauty of the study is that it overcomes the limitations of other possible designs: it is impossible to randomize families to higher income; a longitudinal observational cohort study looking at the connection between income and children's mental health would have the challenge of separating the impact of income gains from the impact of the factors that led to the gain in income (e.g., new job, second parent working).

1.2 How can you tell whether an intervention is effective?

It is not always easy! Of course, if you develop a new treatment for rabies (a disease that is almost uniformly fatal without prompt treatment) and your first ten patients all survive, you may have enough evidence to prove your case. But, most of the health problems that plague the twenty-first century world do not have a single cause; most of the outcomes don't occur quickly or predictably; and no intervention is expected to be near 100% effective. Rather, problems like obesity, violence or substance abuse have multiple interrelated causes; outcomes occur over a period of years; and an intervention that reduced the prevalence of any of these conditions by 15% would be heralded as a major breakthrough.

There are a wide variety of study designs available for evaluating the effectiveness of interventions, each with their own advantages and disadvantages.[5] Studies may be randomized or nonrandomized, prospective or retrospective, clustered or nonclustered. Investigators may use complicated analytic tools such as time-series analysis or multivariable modeling, or simply compare percentages. Regardless of what study designs and analytic tools are employed, determining whether the intervention works will be based on answering one (or more) of the following three questions:

1 Is the post-intervention assessment significantly different from the pre-intervention assessment?

[5] The term evaluation encompasses a broad set of activities including assessing whether a problem exists, how well a program is functioning (e.g., number of clients being served, length of time it takes to serve a client, client satisfaction with the service), what the program costs, etc. The focus of this book is the efficacy or effectiveness of the intervention. For an excellent book on the full spectrum of evaluation activities see: Berk, R. A. and Rossi, P. H. *Thinking about Program Evaluation 2*. Thousand Oaks: Sage Publications, 1999; also Shadish, W. R. "The common threads in program evaluation." *Prev Chronic Dis* **391** (2006): A03 at http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=16356356.

2 Is the change between the pre-intervention and the post-intervention assessment of the intervention group significantly different than that of the comparison group?
3 Is the outcome for the intervention group significantly different than for the comparison group?

In the next section I will review the data elements you will need to answer each of these questions. Following that I will discuss development of interventions (Chapter 2), evaluation of interventions (Chapter 3), and then compare randomized designs (Chapter 4) to nonrandomized designs (Chapter 5). In Chapter 6 I will return to how to answer these three questions from a statistical point of view.

1.2.A Is the post-intervention assessment significantly different from the pre-intervention assessment?

An intuitively simple way of determining whether an intervention works is to assess a group of people prior to it (pre-intervention) and again after it (post-intervention) (Figure 1.1). This is referred to as a one-group pre-intervention versus post-intervention design.

Longitudinal cohort studies repeatedly assess the same individuals over time.

If you are assessing the same individuals on more than one occasion, your study is a longitudinal cohort study. If you are sampling from the same population but not necessarily the same individuals on more than one occasion, you are performing a serial cross-sectional study. By definition one-group pre-intervention versus post-intervention designs are nonrandomized designs because you need at least two groups to randomize assignment.

Serial cross-sectional studies repeatedly sample from the same population over time.

Regardless of whether you are assessing the same individuals over time, or samples of the same population over time, you will be testing the null hypothesis; in this case, that there is no difference between the two assessments. If the difference between the pre-intervention and post-intervention assessments is sufficiently great that it is unlikely that the difference could have occurred by chance alone, you will consider the alternative hypothesis: that the intervention worked. If the pre-intervention and post-intervention assessments are similar, you will conclude that the null hypothesis was correct: the intervention did not work.

Let's look at a very important intervention evaluated with a one-group pre-intervention versus post-intervention design. Pronovost and colleagues designed an intervention to decrease catheter-related blood infections among patients cared for in the intensive care unit.[6] The intervention included

[6] Pronovost, P., Needham, D., Berenholtz, S., *et al.* "An intervention to decrease catheter-related bloodstream infections in the ICU." *N. Engl. J. Med.* **355** (2006): 2725–32.

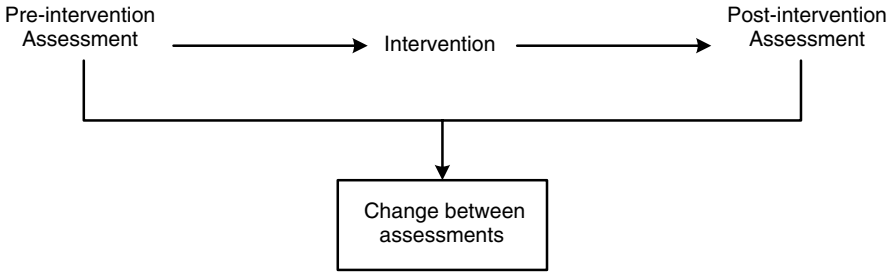

Figure 1.1 Schematic of one-group pre-intervention versus post-intervention design.

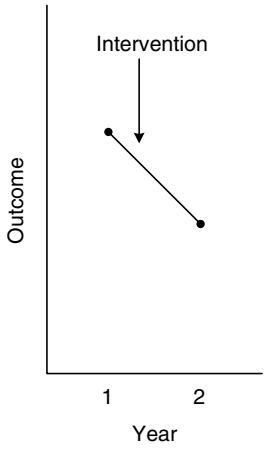

Figure 1.2 Hypothetical data with a pre-intervention and a post-intervention assessment.

strategies to increase hand-washing by medical personnel, use of barrier precautions, antibacterial cleaning of the catheter site, avoiding use of the femoral site, and removing unnecessary catheters.

Three months after the intervention there were significantly fewer infections (0 infections per 1000 catheter days) than before the intervention (2.7 infections per 1000 catheter days). The probability was small ($P \leq 0.002$) that the change in the infection rate between the pre-intervention and the post-intervention assessments occurred by chance.

To appreciate one of the weaknesses of a pre-intervention versus post-intervention evaluation such as this one, look at the hypothetical data in Figure 1.2. It appears that there has been a major drop in outcomes (e.g., infection, cancer, heart disease) following institution of an intervention.

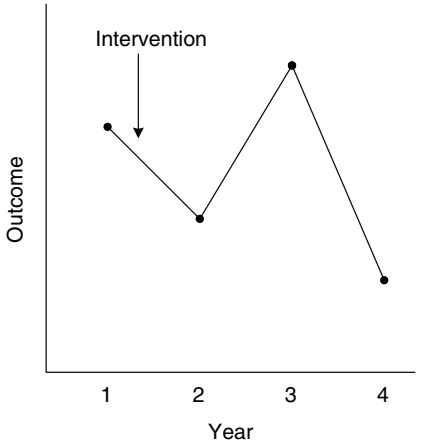

Figure 1.3 Hypothetical data with a pre-intervention assessment and three post-intervention assessments.

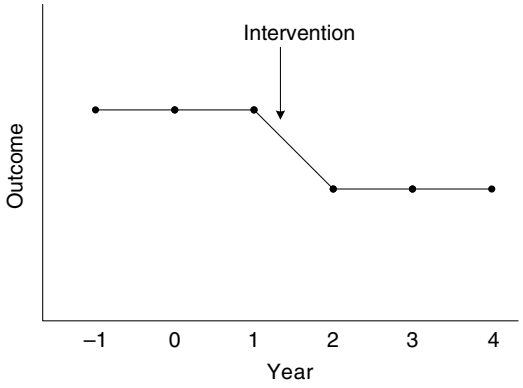

Figure 1.4 Hypothetical data with three pre-intervention assessments and three post-intervention assessments.

Before concluding that the intervention worked, look at Figure 1.3. It is the continuation of the same study. You can see that outcomes were back up by year 3 and back down by year 4. Remember: two points do not make a trend.

Two points do not make a trend.

One way to improve the strength of a single group pre-intervention versus post-intervention design is to get additional data points. For example, if the frequency of outcome had been measured several times prior to the intervention and was stable, and then changed precipitously following the intervention, and the change was sustained at several points after the intervention, as shown in Figure 1.4, you would have much greater confidence that the decrease in outcome was due to the intervention.

Table 1.1. Rate of catheter-related infection at baseline, during the intervention, and after the intervention.

	Rate of infection (95% confidence intervals)	*P* value for comparison with baseline rate
Baseline	2.7 (0.6–4.8)	–
During intervention	1.6 (0–4.4)	≤0.05
After intervention		
0–3 mo	0 (0–3.0)	≤0.002
4–6 mo	0 (0–2.7)	≤0.002
7–9 mo	0 (0–2.1)	≤0.002
10–12 mo	0 (0–1.9)	≤0.002
13–15 mo	0 (0–1.6)	≤0.002
16–18 mo	0 (0–2.4)	≤0.002

Data from Pronovost, P., *et al.* "An intervention to decrease catheter-related bloodstream infections in the ICU." *N. Engl. J. Med.* **355** (2006): 2725–32.

In the case of the study to decrease infections in the ICU, the investigators had only one measurement of infection prior to the intervention, but they had additional data points: one during the implementation period and five additional points in the post-intervention period (Table 1.1). These data points increase confidence that the intervention worked. When you have lots of consecutive data points over a period of time you can analyze your data using time series analysis (Chapter 8).

However, no matter how many data points you have prior to, during, and after an intervention, with only one group your study has a serious limitation: there is always the possibility that any observed change occurred for a reason other than the intervention. Returning to the example of the study of catheter-related infections in the ICU, perhaps infections decreased due to media attention on hospital infection rates or changes in physician prescribing practices with respect to antibiotics. To overcome this limitation we need a design that includes a comparison group. This design is discussed in the next subsection.

1.2.B Is the change between the pre-intervention and the post-intervention assessment of the intervention group significantly greater (lesser) than that of the comparison group?

Adding one or more comparison groups to a pre-intervention versus post-intervention assessment results in a stronger evaluation design than having a single group. If the subjects are assigned by random to the groups and the subjects are followed prospectively, you have a randomized controlled trial

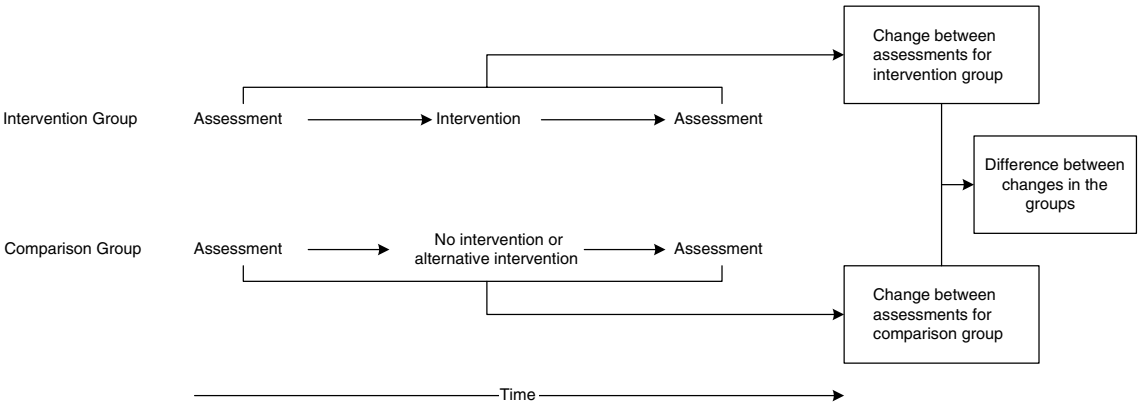

Figure 1.5 Schematic diagram of a pre-intervention versus post-intervention design with comparison group.

(Chapter 4). There are also a number of ways of assembling a comparable control group without randomization (Chapter 5).

Whether the comparison group is assigned randomly or not, the major question is whether the change that occurs between the pre-intervention and the post-intervention assessment is greater (lesser) than the change over the same period of time in a comparison population (Figure 1.5).

To illustrate the benefits of adding a comparison group to a pre-intervention versus post-intervention assessment, let's look at a study evaluating whether providing hospitals with confidential information on their performance improves the care of patients having coronary artery bypass grafting (CABG).[7] The intervention was performed at 20 hospitals in Alabama.

In Table 1.2 you can see the impact of the intervention on four important process measures. Following the intervention the rate of internal mammary artery use, the percentage of patients discharged on aspirin, and the percentage of patients who were intubated for less than 6 hours increased, and the median time patients were intubated decreased. But as with the study of ICU catheter infections, is it possible that these changes occurred for some reason other than the intervention? Indeed, three and a half years passed between the start of the pre-intervention assessment and the end of the post-intervention assessment. Perhaps improvements in surgical technique or medical practice during these years are the actual cause of the observed improvement in practice.

The investigators addressed this concern by including CABG patients from a comparison state. As you can see in Table 1.3 the changes over time were more

[7] Holman, W. L., Allman, R. M., Sansom, M., *et al.* "Alabama coronary artery bypass grafting project: results of a statewide quality improvement initiative." *JAMA* **285** (2001): 3003–10.

Table 1.2. Impact of a quality improvement program in Alabama hospitals.

Measure	Pre-intervention	Post-intervention	*P* value
Internal mammary artery use, %	73	84	≤0.001
Aspirin therapy, %	88	92	≤0.001
Intubation time, median, h	12	7	≤0.001
Intubation <6 h, %	9	41	≤0.001

Data from: Holman, W. L., *et al.* "Alabama coronary artery bypass grafting project: results of a statewide quality improvement initiative." *JAMA* **285** (2001): 3003–10.

Table 1.3. Quality indicators for CABG surgery in Alabama compared to another state.

	Alabama		Comparison state		*P* value for difference between states
Measure	Pre-intervention	Post-intervention	Baseline	Follow-up	
Internal mammary artery use, %	73	84	48	55	= 0.001
Aspirin therapy, %	88	92	86	82	<0.001
Intubation time, median, h	12	7	7	8	<0.001
Intubation <6 h, %	9	41	40	39	<0.001

Data from: Holman, W. L., *et al.* "Alabama coronary artery bypass grafting project: results of a statewide quality improvement initiative." *JAMA* **285** (2001): 3003–10.

favorable in the intervention state (Alabama) than in the comparison state. Temporal improvements in medical practice would have likely influenced the outcomes in the comparison state as well.

Note also from Table 1.3 the importance of having two measurements from both of the states. If you had only the second measurements for the median intubation time and for the percentage of patients who were intubated for less than six hours, you might conclude that there was no difference between the two states. If you had only the second measurements for internal mammary artery use you would think that the intervention was a much greater success that it was – even before the intervention, the use of the internal mammary artery as part of the CABG procedure was substantially higher in Alabama hospitals than in the comparison state.

Although the inclusion of a comparison group (in this case concurrent controls) greatly strengthens the model, there remains an important potential weakness. Is the comparison group comparable to the intervention group? As you can see from Table 1.4, patients from the intervention group and the comparison group are similar on most characteristics at both baseline and follow-up assessments. Nonetheless there are differences (for example, the percentage with left main disease is lower in Alabama than in the comparison state).

Table 1.4. Characteristics of patients receiving a CABG in Alabama compared to patients receiving a CABG in another state

	Alabama		Comparison state	
Characteristics	Pre-intervention	Post-intervention	Baseline	Follow-up
Patients	4090	1694	2288	926
Mean age, y	69.9	70.7	70.6	71.4
Male, %	65	55	66	66
White, %	91	91	94	93
CAD				
Left main, %	16	19	23	27
MI within 3 days, %	9	14	8	9
MI within 6 months, %	22	26	21	24
CHF	16	22	12	19
Poor LVF, %	26	29	20	19
Cardiogenic shock, %	3	3	3	2
COPD, %	25	30	23	31
Diabetes mellitus, %	29	32	27	32
Dialysis, %	2	1	1	2

Data from: Holman, W. L., *et al*. "Alabama coronary artery bypass grafting project: results of a statewide quality improvement initiative." *JAMA* **285** (2001): 3003–10.

Could some other difference between Alabama hospitals and those in the comparison state explain the differences that we are attributing to Alabama's intervention? Yes. The only way to be certain that there are no important differences between the intervention and the comparison group is to randomize assignment to the groups (Chapter 4).

Still, it is important to note that this intervention, which had an important impact, would have been very difficult to perform using a randomized design. Randomization of individual patients would have been impossible because the intervention was being performed on the level of physicians and hospitals.

A clustered randomization design (Section 4.5), where the hospitals were randomized to be in the intervention, would have been superior from a design point of view. However, this study was possible because of a statewide initiative in Alabama to improve the care of patients receiving CABG in all hospitals and had the cooperation of the state peer review organization. The comparison state only had to agree to release medical record information.

Randomizing hospitals may have created another problem: it may have increased the diffusion of the intervention to non-intervention hospitals because physicians often work at more than one hospital and share information through local and state professional associations. Diffusion of the intervention would have been good for care of patients receiving CABG but would