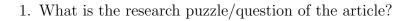# Journal Article Analysis worksheet

Intro to Political Science Research Methods

April 12, 2021

I Abstract/Introduction

1. What is the research puzzle/question of the article?

2. What is the thesis/answer to the research question?

   - hint: Unlike recommendations for your work, it might be multiple lines

3. Describe in your words what the article seems to be about, given the introduction.

  - What seems to be unclear to you as of this point?

- List words/terms in this section that are unclear at this point.

II Literature review

1. What is the dependent variable/outcome that the authors are interested in and describe?

2. What seem to be the major finding/paradigm to date?

    (a) i.e. Is there a specific theory/theories that frame the field? Seminal works the authors focus on?

3. What is the gap in the literature?

    - Are there any controversies/competing theories that the authors make note of?

4. Given the description, what seems to be the state of the field in your own words?

- List words/terms in this section that are unclear at this point.

III   Theory

1. What theory/framework do the authors propose to answer the research question and/or fill the gap in the state of the field so far?

   - Do they make use of an existing theory or propose a new one?

2. What hypotheses do the authors derive from the theory?

3. How decisive are the hypotheses in testing the theory?

   - Refer back to the s CDI Practice Paper by Melanie Punton and Katharina Welle (2015). Tests traditionally take the form of the following:
     (a) Straw in the wind
     (b) Hoop test
     (c) Smoking gun test
     (d) Doubly decisive test

4. What part of the theory do the authors focus on? What do they leave out?

   - Has the portion they left out been addressed by someone else? Or is it outstanding?

   - List words/terms in this section that are unclear at this point.

IV   Data and Methodology

Note: Methodology can be overwhelming at times, especially for models more complex than an OLS. Therefore, focus on the intuition and rationale of the methods. Keep careful note about what you don't understand, and go back to this and see if you understand it better at a later point.

1. What type of data do the authors make use of? These might include the following:

   (a) Survey data
   (b) Administrative data
   (c) Experimental
   (d) Textual (i.e. Twitter) data

2. What is the structure of the outcome variable (dependent variable) of interest?

3. What is the structure of the explanatory variable(s) of interest?

- How exogenous is the treatment?

4. What is the research design? Some designs might include the following:

   (a) Time series observational
   (b) Cross sectional observational
      - Multilevel?
   (c) Random assignment experimental
   (d) Weighted interview/archival research
      - Refer back to Andrew Bennett and process tracing

5. What potential confounding variables do the authors control for, why do they do so, and how?

6. What model do the authors use in their analysis?

   - What appears to be the rationale of the model?
   - What is confusing about it and why?

7. In your own words, what appear to be the limitations of the data and design, as you understand it?

   - List words/terms in this section that are unclear at this point.

V Results

1. Where and how does the study summarize the variables of interest?

2. Which hypotheses/expectations receive support?

3. What is the estimated impact of the explanatory variable on the dependent variable?

   - In which figures/tables are such findings presented?

4. How clear are the figures and tables to you?

- Are the axes labeled and described in text or caption?
- Are they overly cluttered?
- Did grayscale make the results undecipherable?
- Which is your favorite? Least favorite?

5. How clearly do the results relate back to the hypotheses and theory?

- List words/terms in this section that are unclear at this point.

VI Discussion and Conclusion

1. What do the authors conclude based on their results?

2. From their description, are the conclusions reasonable?

   - Do the authors overstep given their data and design?
   - Do they try to spin the results? If so, how?

3. How do these results offer to lay the foundation for future research?

4. What future research projects do the authors propose?

5. After completing the article, what is your takeaway? What do you remain confused about?

For additional materials, please see the following attached documents to supplement the journal article analysis worksheet:

- Huck, Schuyler W. 2004. *Reading Statistics and Research*. New York: Pearson.

- Punton, Melanie and Katharina Welle. 2015. "Straws-in-the-wind, Hoops and Smoking Guns: What can Process Tracing Offer to Impact Evaluation?" Centre for Development Impact, no. 10.

Reading Statistics and Research

FOURTH EDITION

Schuyler W. Huck

University of Tennessee

1

The Typical Format of a Journal Article

Almost all journal articles dealing with research studies are divided into different sections by means of headings and subheadings. Although there is variation among journals with respect to the terms used as the headings and the order in which different sections are arranged, there does appear to be a relatively standard format for published articles. Readers of the professional literature will find that they can get the most mileage out of the time they invest if they are familiar with the typical format of journal articles and the kind of information normally included in each section of the article.

We are now going to look at a particular journal article that does an excellent job of illustrating the basic format that many authors use as a guide when they are writing their articles. The different sections of our model article could be arranged in outline form as follows:

1. Abstract
2. Introduction
   a. Background
   b. Statement of purpose
   c. Hypotheses
3. Method
   a. Participants
   b. Materials
   c. Procedure
4. Results
5. Discussion
6. References

Let us now examine each of these items.

Abstract

An **abstract,** or *précis,* summarizes the entire research study and appears at the beginning of the article. Although it normally contains fewer than 150 words, the abstract usually provides the following information: (1) a statement of the purpose or objective of the investigation, (2) a description of the individuals who served as participants, (3) a brief explanation of what the participants did during the study, and (4) a summary of the important findings.

Excerpt 1.1 is the abstract from our model journal article. As in most articles, it was positioned immediately after the title and authors' names. This abstract was easy to distinguish from the rest of the article because it was indented and printed in a small font size. In some journals, the abstract will be italicized to make it stand out from the beginning paragraphs of the article.

EXCERPT 1.1 • *Abstract*

This study examined the anxiolytic (anxiety reducing) effects of exercise for elderly women engaging in a single bout of aqua aerobics. Volunteers ($N = 29$) completed questionnaires immediately before and after participating in an aqua aerobics class. The average age of participants was 66.4 yr. A brief form of Spielberger's State Anxiety Inventory and a question on demographic items were administered prior to engagement in exercise, and the brief form of the State Anxiety Inventory was administered again immediately after the exercise session. There was a significant difference on a *t* test between participants' ratings of anxiety before exercise ($M = 16.8$) compared to after exercise ($M = 13.9$); participants' ratings of state anxiety were somewhat lower after exercising. Weaknesses of the present study and suggestions for research are presented.

The sole purpose of the abstract is to provide readers with an overview of the material they will encounter in the remaining portions of the article. Even though the abstract from our model journal article contains two technical terms (*significant difference* and *t test*) and two abbreviations (*N* and *M*), you most likely can get a good sense for what issues were being studied, who supplied the data, and how the investigation turned out. On the basis of abstracts like the one shown in Excerpt 1.1,

you can decide that the article in front of you is a veritable gold mine, that it *may* be what you have been looking for, or that it is not at all related to your interests. Regardless of how you react to this brief synopsis of the full article, the abstract serves a useful purpose.

Introduction

The **introduction** of an article usually contains two items: a description of the study's **background** and a **statement of purpose.** Sometimes, as in our model journal article, a third portion of the introduction will contain a presentation of the researcher's **hypotheses.** These components of a journal article are critically important. Take the time to read them slowly and carefully.

Background

Most authors begin their articles by explaining what caused them to conduct their empirical investigations. Perhaps the author developed a researchable idea from discussions with colleagues or students. Maybe a previous study yielded unexpected results, thus prompting the current researcher to conduct a new study to see if those earlier results could be replicated. Or, maybe the author wanted to see which of two competing theories would be supported more by having the collected data conform to its hypotheses. By reading the introductory paragraph(s) of the article, you will learn why the author conducted the study.

In describing the background of their studies, authors typically highlight the connection between their studies and others' previously published work. Whether this review of literature is short or long, its purpose is to show that the current author's work has been informed by, or can be thought of as an extension of, previous knowledge. Such discussions are a hallmark of scholarly work. Occasionally, a researcher will conduct a study that is based on an idea that is not connected to anything anyone has investigated or written about; such studies, however, are rare.

Excerpt 1.2 comes from our model article. Though only a single paragraph in length, this portion of the introduction sets the stage for a discussion of the authors' investigation.

Statement of Purpose

After discussing the study's background, an author usually states the specific purpose or goal of the investigation. This statement of purpose is one of the most important parts of a journal article since, in a sense, it explains what the author's "destination" was. It would be impossible for us to evaluate whether the trip was

EXCERPT 1.2 • *Background*

Many elderly (55 years of age and older) persons report anxiety. Based on survey research the prevalence has been estimated to be as high as 24% (Forsell & Winblad, 1988; Himmelfarb & Murrell, 1984). Previous studies have also indicated that ratings of state and trait anxiety are more prevalent among elderly women than men (Forsell & Winblad, 1988; Himmelfarb & Murrell, 1984; Schuab & Linden, 2000). One means of reducing state anxiety and improving one's mood is to engage in physical activity. Research with younger subjects has shown single bouts of exercise improve mood (Petruzzello, Landers, Hatfield, Kubits, & Salazar, 1991), yet little research has examined the effects of such exercise on mood among elderly people.

One drawback of exercise for elderly people is the potential for injury. However, the increasingly popular use of exercising in the water reduces the risk of weight-bearing injuries. Two studies with younger subjects have specifically examined the effects of acute bouts of water exercise and swimming on mood. Both reports indicated subjects had improved mood states after exercising (Berger & Owen, 1983; Crissman, 1999). Even though water-based exercise is a popular form of exercise for elderly persons and is often recommended by physicians, the author was unable to locate any studies investigating the effects of a single bout of water-based exercise on mood with elderly participants.

Source: S. R. Wininger. (2002). The anxiolytic effect of aqua aerobics in elderly women. *Perceptual and Motor Skills, 94*(1), p. 338.

successful—in terms of research findings and conclusions—unless we know where the author was headed.

The statement of purpose can be as short as a single sentence or as long as a full paragraph. It is often positioned just before the first main heading of the article, but it can appear anywhere in the introduction. Regardless of its length or where it is located, you will have no trouble finding the statement of purpose if the researcher begins a sentence with the words, "The purpose of this study was to . . ." or "This investigation was conducted in order to compare. . . ." In Excerpt 1.3, we see the statement of purpose from our model journal article.

Hypotheses

After articulating the study's intended purpose, some authors disclose the hypotheses they had at the beginning of the investigation. Other authors do not do this, either because they didn't have any firm expectations or because they consider it

EXCERPT 1.3 • *Statement of Purpose*

The purpose of this study was to examine the anxiolytic effects of exercise for elderly women engaging in a single bout of aqua aerobics.

Source: S. R. Wininger. (2002). The anxiolytic effect of aqua aerobics in elderly women. *Perceptual and Motor Skills, 94*(1), p. 338.

unscientific for the researcher to hold hunches that might bias the collection or interpretation of the data. Although there are cases (as you will see in Chapter 7) where a researcher can conduct a good study without having any hypotheses as to how things will turn out, and although it is important for researchers to be unbiased, there is a clear benefit in knowing what the researcher's hypotheses were. Simply stated, outcomes compared against hypotheses usually are more informative than are results that stand in a vacuum. Accordingly, I applaud those researchers who disclose in the introduction any *a priori* hypotheses they had.

Excerpt 1.4 comes from our model journal article, and it contains the single hypothesis that the researcher had in the study concerned with anxiety and exercise among elderly women. In light of the research findings cited in the article's first two paragraphs, it is not surprising that the hypothesis in this study was that anxiety scores would decrease following a session of aqua aerobics.

In most articles, the background, the statement of purpose, and hypotheses are not identified by separate headings, nor are they found under a common heading. If a common heading were to be used, though, the word *introduction* would probably be most appropriate because these three items set the stage for the substance of the article—an explanation of what was done and what the results were.

EXCERPT 1.4 • *Hypotheses*

The central hypothesis was that the average scores for state anxiety of the elderly women engaging in water aerobics would significantly decrease after a single exercise session.

Source: S. R. Wininger. (2002). The anxiolytic effect of aqua aerobics in elderly women. *Perceptual and Motor Skills, 94*(1), p. 339.

Method

In the **method** section of a journal article, an author will explain in detail how the study was conducted. Ideally, such an explanation should contain enough information to enable a reader to replicate (i.e., duplicate) the study. To accomplish this goal, the author will address three questions: (1) Who participated in the study? (2) What kinds of measuring instruments were used to collect the data? and (3) What were the participants required to do? The answer to each of these questions is generally found under an appropriately titled subheading of the method section.

Participants

Each of the individuals (or animals) who supplies data in a research study is considered to be a **participant** or a **subject.** (In some journals, the abbreviations *S* and *S*s are used, respectively, to designate one subject or a group of subjects.) Within this section of a journal article, an author usually indicates how many participants or subjects were used, who the participants were, and how they were selected.

A full description of the participants is needed because the results of a study will often vary according to the nature of the participants who are used. This means that the conclusions of a study, in most cases, are valid only for individuals (or animals) who are similar to the ones used by the researcher. For example, if two different types of counseling techniques are compared and found to differ in terms of how effective they are in helping clients clarify their goals, it is imperative that the investigator indicate whether the participants were high school students, adults, patients in a mental hospital, or whatever. What works for a counselor in a mental hospital may not work at all for a counselor in a high school (and vice versa).

It is also important for the author to indicate how the participants were obtained. Were they volunteers? Were they randomly selected from a larger pool of potential participants? Were any particular standards of selection used? Did the researcher simply use all members of a certain high school or college class? As you shall see in Chapter 5, certain procedures for selecting samples allow results to be generalized far beyond the specific individuals (or animals) included in the study, while other procedures for selecting samples limit the valid range of generalization.

Excerpt 1.5 comes from our model journal article. Labeled participants, it was the first portion of the article's method section.

Materials

This section of a journal article is normally labeled in one of four ways: **materials, equipment, apparatus,** or **instruments.** Regardless of its label, this part of the article contains a description of the things (other than the participants) used in the study. The goal here, as in other sections that fall under the method heading, is to

EXCERPT 1.5 • *Participants*

Twenty-nine elderly females voluntarily completed questionnaires immediately before and after participating in aqua aerobics classes. Their average age was 66.4 yr. (SD = 9.2). Participants were surveyed in aqua aerobics classes at two separate facilities, a university pool and a private club pool.

Source: S. R. Wininger. (2002). The anxiolytic effect of aqua aerobics in elderly women. *Perceptual and Motor Skills, 94*(1), p. 339.

describe what was done with sufficient clarity that others could replicate the investigation to see if the results remain the same.

Suppose, for example, that a researcher conducts a study to see if males differ from females in the way they evaluate various styles of clothing. To make it possible for others to replicate this study, the researcher would need to indicate whether the subjects saw actual articles of clothing or pictures of clothing (and if pictures, whether they were prints or slides, what size they were, and whether they were in color), whether the clothing articles were being worn when observed by participants (and if so, who modeled the clothes), what specific clothing styles were involved, how many articles of clothing were evaluated, who manufactured the clothes, and all other relevant details. If the researcher does not provide this information, it would be impossible for anyone to replicate the study.

Often, the only material involved is the measuring device used to collect data. Such measuring devices—whether of a mechanical variety (e.g., a stopwatch) or of a paper-and-pencil variety (e.g., a questionnaire)—ought to be described very carefully. If the measuring device is a new instrument designed specifically for the study described in the article, the researcher will typically report evidence concerning the instrument's technical psychometric properties. Generally, the author accomplishes this task by discussing the reliability and validity of the scores generated by using the new instrument.[1] Even if an existing and reputable measuring instrument has been used, the researcher ought to tell us specifically what instrument was used (by indicating form, model number, publication date, etc.). One would need to know such information, of course, before a full replication of the study could be attempted. In addition, the researcher ought to pass along reliability and validity evidence cited by those who developed the instrument. Ideally, the authors ought to provide their *own* evidence as to the reliability and validity of scores used in their study, even if an existing instrument is used.

---

[1]Later, in Chapter 4, we will talk more about the kinds of evidence researchers normally offer to document their instruments' technical merit.

Excerpt 1.6 contains the materials section from our model article. By reading this excerpt, we learn that the elderly women who participated in this study completed a demographic form and the brief form of Spielberger's State Anxiety Inventory. If more space had been allocated to this article in the journal, the author undoubtedly would have described more fully these two measuring instruments. In that more extended description, the author would have (1) pointed out what information (in addition to age) was gathered on the demographic form, (2) presented a sample item from the State Anxiety Inventory, and (3) indicated that Spielberger's anxiety instruments are considered to be extremely good for measuring people's anxiety levels.

EXCERPT 1.6 • *Materials*

The questionnaires assessed state anxiety and demographic characteristics. State anxiety was measured using a 10-item brief form of the State Anxiety Inventory (Spielberger, 1979). Scores on this version range from 10 to 40, with higher scores indicating greater state anxiety. Evidence of the validity and reliability are discussed by Spielberger (1979). Estimates of internal consistency for this inventory in this study before and after exercise were acceptable as coefficients alpha were .80 and .71.

Source: S. R. Wininger. (2002). The anxiolytic effect of aqua aerobics in elderly women. *Perceptual and Motor Skills, 94*(1), p. 339.

In most empirical studies, the **dependent variable** is closely connected to the measuring instrument used to collect data. In fact, many researchers operationally define the dependent variable as being equivalent to the scores earned by people when they are measured with the study's instrument. Though this practice is widespread (especially among statistical consultants), it is *not* wise to think that dependent variables and data are one and the same.

Although there are different ways to conceptualize what a dependent variable is, a simple definition is useful in most situations. According to this definition, a dependent variable is simply a characteristic of the participants that (1) is of interest to the researcher; (2) is not possessed to an equal degree, or in the same way, by all participants; and (3) serves as the target of the researcher's data-collection efforts. Thus, in a study conducted to compare the intelligence of males and females, the dependent variable would be intelligence.

In the study associated with our model article, there was one dependent variable: state anxiety. (As opposed to *trait anxiety* which remains high, medium, or low over time, *state anxiety* fluctuates depending on the circumstances of the moment.) That was the target of the primary measuring instrument used in this study. By citing Spielberger's validity and reliability findings and by presenting his own evidence on internal consistency (by citing two alpha coefficients), the author's final two sentences in Excerpt 1.6 represent his claim that the collected scores did, in fact, reveal the participants' levels of state anxiety.

Procedure

How the study was conducted is explained in the **procedure** section of the journal article. Here, the researcher explains what the participants did—or what was done to them—during the investigation. Sometimes an author will even include a verbatim account of instructions given to the participants.

Remember that the method section is included so as to permit a reader to replicate a study. To accomplish this desirable goal, the author must outline clearly the procedures that were followed, providing answers to questions such as these: Where was the study conducted? Who conducted the study? In what sequence did events take place? Did any of the subjects drop out prior to the study's completion? (In Chapter 5, we will see that subject dropout can cause the results to be distorted.)

Excerpt 1.7 is the procedure section from our model article. Even though this section is extremely brief, it provides information regarding who collected the data, when this took place, and where this was done. In addition, the researcher points out that permission was granted to collect the data.

EXCERPT 1.7 • *Procedure*

Permission to administer the surveys was granted by the supervisors and instructors at both facilities. One instructor who taught the aqua aerobics classes at both facilities introduced the researcher who briefly described the survey and asked for volunteers. Prior to exercising the participants completed a questionnaire on demographic characteristics and the State Anxiety Inventory. Immediately following the approximately 60-min. session of water aerobic exercise, participants took the State Anxiety Inventory again. There was no control group.

Source: S. R. Wininger. (2002). The anxiolytic effect of aqua aerobics in elderly women. *Perceptual and Motor Skills, 94*(1), p. 339.

Results

There are three ways in which the results of an empirical investigation will be reported. First, the results can be presented within the text of the article—that is, with only words. Second, they can be summarized in one or more tables. Third, the findings can be displayed by means of a graph (which is technically called a **figure**). Not infrequently, a combination of these mechanisms for reporting results is used to help readers gain a more complete understanding of how the study turned out. In Excerpt 1.8, we see that the author of our model article presented his results by means of a single paragraph of text.

EXCERPT 1.8 • *Results*

There was a significant difference between participants' state anxiety before ($M = 16.8$, $SD = 5.9$) and after exercise ($M = 13.9$, $SD = 4.1$, $t_{28} = 3.76$, $p = .001$). On the average, the effect size for the decrease in participants' scores on state anxiety was .71 standard deviation units (pooled $SD = 4.27$, M difference $= 3.0$). It is also interesting to note that the distribution of the anxiety scores changed before and after exercise. Scores before exercise had a skewness value of .37, while scores after exercise were more skewed at 1.07, that is, the majority of the subjects rated their anxiety low after exercising.

Source: S. R. Wininger. (2002). The anxiolytic effect of aqua aerobics in elderly women. *Perceptual and Motor Skills, 94*(1), p. 339.

Although the **results** section of a journal article contains some of the most crucial information about the study (if not *the* most crucial information), readers of the professional literature often disregard it. They do this because the typical results section is loaded with statistical terms and notation not used in everyday communication. Accordingly, many readers of technical research reports simply skip the results section because it seems as if it came from another planet.

If you are to function as a discerning "consumer" of journal articles, you must develop the ability to read, understand, and evaluate the results provided by authors. Those who choose not to do this are forced into the unfortunate position of uncritical acceptance of the printed word. Researchers are human, however, and they make mistakes. Unfortunately, the reviewers who serve on editorial boards do not catch all of these errors. As a consequence, there sometimes will be an inconsistency between the results discussed in the text of the article and the results presented in the tables. At times, a researcher will use an inappropriate statistical test. More often than you would suspect, the conclusions drawn from the statistical results will extend far beyond the realistic limits of the actual data that were collected.

You do not have to be a sophisticated mathematician in order to understand and evaluate the results sections of most journal articles. However, you must be-

come familiar with the terminology, symbols, and logic used by researchers. This text was written to help you do just that.

Look at Excerpt 1.8 once again. The text material included in this excerpt is literally packed with information intended to help you. Unfortunately, many readers miss out on the opportunity to receive this information because they lack the skills needed to decode what is being communicated or are intimidated by statistical presentations. One of my goals in this book is to help readers acquire (or refine) their decoding skills. In doing this, I hope to show that there is no reason for anyone to be intimidated by what is included in technical research reports.

After reading subsequent chapters in this book, you will be able to decipher easily all of the information presented in Excerpt 1.8. Once you get these chapters under your belt, you also will be in a position to critically evaluate this investigation and judge for yourself whether the researcher's claims ought to be believed. More important, you will be able to decipher and critique the majority of the research reports you encounter in printed and electronic journals.

Discussion

The results section of a journal article contains a technical report of how the statistical analyses turned out, while the **discussion** section is usually devoted to a non-technical interpretation of the results. In other words, the author will normally use the discussion section to explain what the results mean in regard to the central purpose of the study. The statement of purpose, which appears near the beginning of the article, usually contains an underlying or obvious research question; the discussion section ought to provide a direct answer to that question.

In addition to telling us what the results mean, many authors use this section of the article to explain *why* they think the results turned out the way they did. Although such a discussion will occasionally be found in articles where the data support the researchers' hunches, authors are much more inclined to point out possible reasons for the obtained results when those results are inconsistent with their expectations. If one or more of the scores turn out to be highly different from the rest, the researcher may talk about such serendipitous findings in the discussion section.

Sometimes an author will use the discussion section to suggest ideas for further research studies. Even if the results do not turn out the way the researcher had hoped they would, the study may be quite worthwhile in that it might stimulate the researcher (and others) to identify new types of studies that need to be conducted. Although this form of discussion is more typically associated with unpublished master's theses and doctoral dissertations, you will occasionally encounter it in a journal article.

It should be noted that some authors use the term **conclusion** rather than discussion to label this part of the journal article. These two terms are used interchangeably. It is unusual, therefore, to find an article that contains both a discussion section and a conclusion section.

Excerpt 1.9 contains the discussion section that appeared in our model journal article. Notice how the author used the first paragraph to argue that (1) his hypothesis was supported by the empirical evidence of the study and (2) there was a meaningful *effect size* (i.e., impact) associated with the exercise session. In the second paragraph, the author addresses an apparent weakness in the study's design—the lack of a control group—by presenting evidence suggesting that the study's main finding is still valid even though no control group was used. In the final paragraph, an alternative explanation for the obtained results is discussed, with suggestions offered for further research. This researcher deserves high praise for discussing these concepts of effect size, control groups, and alternative explanations.

EXCERPT 1.9 • *Discussion*

The hypothesis was supported as participants' scores on state anxiety were significantly lower after exercising than before exercising. The effect size for the lower scores was .71 standard deviation units and indicates the magnitude of the decrease is large compared to the average effect size of .23 reported in previous studies on the anxiolytic effect of exercise when the State Anxiety Inventory was given to young and middle-age adults (Petruzzello, *et al.*, 1991). An effect size of this magnitude indicates a medium to large difference (Cohen, 1988). One may infer that a single bout of water aerobic exercise reduced scores on anxiety more for these elderly women than occurred for younger adults.

The main weakness of this study was the absence of a control group. The major drawback of a one-group pre-posttest design is that it does not control for regression to the mean, which might account for the present result. However, the large effect size found suggests the validity of the current finding. It is important to note that some of the studies reviewed in the same meta-analysis by Petruzzello, *et al.* compared anxiety scores before and after exercise while other studies included control groups and exercise groups. Comparisons of the effect sizes for these two types of studies indicates that the pre-posttest studies had an average effect size of .47 while the treatment and control group designs had an average effect size of .22 to .26. The authors of the meta-analysis suggest that the decrease may be related to controlling for the threats to internal validity. However, they point out the exercise groups did report lower scores on anxiety than the control groups.

Another potential explanation for the observed result could be that the reduction in anxiety was associated with social interaction rather than exercising. Researchers could improve designs by creating several different control groups to control for social interaction. For example, one could design a study with a group of elderly people in a social setting who are not engaging in physical activity, elderly people exercising alone, and elderly people who are not exercising or engaging in social interactions with others. Researchers should also include elderly men and ascertain the effects on the anxiolytic effect of exercise of potential covariates such as personality types.

Source: S. R. Wininger. (2002). The anxiolytic effect of aqua aerobics in elderly women. *Perceptual and Motor Skills, 94*(1), p. 339–340.

References

A journal article normally concludes with a list of the books, journal articles, and other source material referred to by the author. Most of these items were probably mentioned by the author in the review of the literature positioned near the beginning of the article. Excerpt 1.10 is the **references** section of our model article.

EXCERPT 1.10 • *References*

Berger, B. G., & Owen, D. R. (1983). Mood alteration with swimming—swimmers really do "feel better." *Psychosomatic Medicine, 45,* 425–433.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, NJ: Erlbaum.

Crissman, D. E. (1999). Collegiate females' state anxiety levels following water exercise intervals. *Dissertation Abstracts International, 59*(10-A), 3873.

Forsell, Y., & Winblad, B. (1998). Feelings of anxiety and associated variables in a very elderly population. *International Journal of Geriatric Psychiatry, 13,* 454–458.

Himmelfarb, S., & Murrell, S. A. (1984). The prevalence and correlates of anxiety symptoms in older adults. *Journal of Psychology, 116,* 159–167.

Petruzzello, S. J., Landers, D. M., Hatfield, B. D., Kubitz, K. A., & Salazar, W. (1991). A meta-analysis on the anxiety-reducing effects of acute and chronic exercise: outcomes and mechanisms. *Sports Medicine, 11,* 143–182.

Schuab, R. T., & Linden, M. (2000). Anxiety and anxiety disorders in the old and very old—results from the Berlin Aging Study (BASE). *Comprehensive Psychiatry, 41,* 48–54.

Spielberger, C. D. (1979). Preliminary manual for the State-Trait Personality Inventory (STPI). (Unpublished manuscript, Univer. of South Florida, Tampa).

Source: S. R. Wininger. (2002). The anxiolytic effect of aqua aerobics in elderly women. *Perceptual and Motor Skills, 94*(1), p. 339–340.

The references can be very helpful to you if you want to know more about the particular study you are reading. Journal articles and convention presentations are usually designed to cover one particular study or a narrowly defined area of a subject. Unlike more extended writings (e.g., monographs and books), they include only a portion of the background information and only partial descriptions of related studies that would aid the reader's comprehension of the study. Reading books and articles listed in the references section will provide you with some of this information and probably give you a clearer understanding as to why and how the author conducted the particular study you have just read. Before hunting down any particular reference item, it is a good idea to look back into the article to reread the sentence or paragraph containing the original citation. This will give you an idea of what is in each reference item.

Notes

Near the beginning or end of their research reports, authors sometimes present one or more notes. In general, such notes are used by authors for three reasons: (1) to thank others who helped them with their study or with the preparation of the technical report, (2) to clarify something that was discussed earlier in the journal article, and (3) to indicate how an interested reader can contact them to discuss this particular study or other research that might be conducted in the future. In our model journal article, there were three notes. The researcher used these notes to thank two people who helped with the study, to indicate that Charles Spielberger had given permission to use the State Anxiety Inventory, and to provide his postal address for anyone who wished to contact him.

Two Final Comments

As we come to the end of this chapter, I would like to make two final points. One concerns the interconnectedness among the different components of the research summary. The other concerns the very first of those components: the abstract.

In this chapter, we have dissected a journal article that summarizes a research study that focused on exercise and anxiety among elderly women. In looking at this particular article section by section, you may have gotten the impression that each of the various parts of a research article can be interpreted and evaluated separately from the other sections that go together to form the full article. You should not leave this chapter with that thought, because the various parts of a well-prepared research report are tied together so as to create an integrated whole.

In our model journal article, the researcher had one principal hypothesis. That hypothesis appears in Excerpt 1.4. That same hypothesis is the focus of the first two sentences of the results section of the journal article (see Excerpt 1.8), the first paragraph of the discussion section (see Excerpt 1.9), and the fifth sentence of the abstract (see Excerpt 1.1). The author who prepared this journal article deserves high marks for keeping focused on the study's central hypothesis and for showing a clear connection between that hypothesis and his findings. Unfortunately, many journal articles display very loose (and sometimes undetectable) connections between the component parts of their articles.

My final comment takes the form of a warning. Simply stated, do not read an abstract and then think that you understand the study well enough to forgo reading the entire article. As was stated earlier, an abstract gives you a thumbnail sketch of a study, thus allowing you to decide whether the article fits into your area of interest. If it does not, then you rightfully can move on. On the other hand, if an abstract makes it appear that the study is, in fact, consistent with your interests, you need to then read the entire article for two reasons. First, the results summarized in the abstract may not coincide with the information that appears in the results section of

the full article. Second, you cannot properly evaluate the quality of the results—even if they are consistently presented in the abstract, results, and discussion sections of the article—without coming to understand who or what was measured, how measurements were taken, and what kinds of statistical procedures were applied.

If you read an abstract (but nothing else in the article) and then utilize the abstract's information to bolster your existing knowledge or guide your own research projects, you potentially harm rather than help yourself. That is the case because the findings reported in many abstracts are simply not true. To be able to tell whether or not an abstract can be trusted, you will need to read the full article. The rest of this book has been written to help make that important task easier for you.

Review Terms

Abstract	Notes
Dependent variable	Participants
Discussion	Procedure
Figure	References
Hypotheses	Results
Materials	Subject
Method	

The Best Items in the Companion Website

1. An important email message sent by the author at the beginning of the semester to students enrolled in his statistics and research course.
2. An interactive online quiz (with immediate feedback provided) covering Chapter 1.
3. Gary Gildner's wonderful poem entitled "Statistics."
4. Five misconceptions about the content of Chapter 1.

Fun Exercises inside Research Navigator

1. What kind of "no-suicide" contract do college students prefer?

In this study, each of 112 college students evaluated three different "no-suicide" contracts on how well each one might help prevent suicides on college campuses. The three contracts differed in length and complexity. One contained a single sentence; one was made up of two sentences; one had nine sentences covering six points. Each research participant rated each contract on several criteria (e.g., effectiveness in stopping suicidal thoughts, potential for lessening depression). Which of the three contracts do you think was evaluated

as being the best? Do you think the subgroups of participants who had or had not contemplated suicide felt the same way about which contract was best? To find out the research-based answers to these questions, locate the PDF version of the research report in the Helping Professions database of ContentSelect and read (on page 588) the first paragraph of the section entitled "Discussion." In addition, take a look at *all* of the article's sections and headings. This will give you a feel for the way most research articles are organized.

G. Buelow & L. M. Range. No-suicide contracts among college students. *Death Studies*. Located in the HELPING PROFESSIONS database of ContentSelect.

2. **Do college students get anxious about aging when taking a course on aging?**

In this investigation, 256 college students were measured as to their knowledge about aging, their attitudes toward elderly people, and their anxiety about becoming old. Half were seniors who had just completed a course on aging; the other half were freshmen who had never taken such a course. The researchers compared the seniors with the freshmen on each of the study's three dependent variables: knowledge, attitude, and anxiety. As expected, the seniors did better on the knowledge-of-aging measure. But what about the other two measures? Do you think the seniors differed from the freshmen on either the attitude measure or the anxiety measure? To find out, locate the PDF version of the research report in the Nursing, Health, and Medicine database of ContentSelect and read (on page 664) the first two paragraphs of the "Discussion." Also, skim through the entire article, noting the different sections and headings. Most articles are set up like this one.

L. A. Harris & S. Dollinger. Participation in a course on aging: Knowledge, attitudes, and anxiety about aging in oneself and others. *Educational Gerontology*. Located in the NURSING, HEALTH, AND MEDICINE database of ContentSelect.

Review Questions and Answers begin on page 516.

CDI Centre for Development Impact
PRACTICE PAPER
Innovation and learning in impact evaluation

Straws-in-the-wind, Hoops and Smoking Guns: What can Process Tracing Offer to Impact Evaluation?

Abstract This CDI Practice Paper by Melanie Punton and Katharina Welle explains the methodological and theoretical foundations of process tracing, and discusses its potential application in international development impact evaluations. It draws on two early applications of process tracing for assessing impact in international development interventions: Oxfam Great Britain (GB)'s contribution to advancing universal health care in Ghana, and the impact of the Hunger and Nutrition Commitment Index (HANCI) on policy change in Tanzania. In a companion to this paper, Practice Paper 10 Annex describes the main steps in applying process tracing and provides some examples of how these steps might be applied in practice.

The move to Sustainable Development Goals in 2015 reflects a wider shift towards more multifaceted and complex ambitions in international development. This trend poses new challenges to measuring impact. For example, how do we measure outcomes such as empowerment, or attribute policy changes to specific advocacy initiatives? The evaluation community is increasingly recognising the limits of classic impact evaluation methodologies based on counterfactual perspectives of causality (for example, randomised controlled trials), implying the need for methodological innovation in the field.

Process tracing is a qualitative method that uses probability tests to assess the strength of evidence for specified causal relationships, within a single-case design and without a control group. It offers the potential to evaluate impact[1] (including in *ex post* designs) through establishing *confidence* in *how* and *why* an effect occurred. This CDI Practice Paper explains the methodological and theoretical foundations of process tracing, and discusses its potential application in international development impact evaluations. The paper draws on two early applications of process tracing for assessing impact in international development interventions: Oxfam Great Britain (GB)'s contribution to advancing universal health care in Ghana, and the impact of the Hunger and Nutrition Commitment Index (HANCI) on policy change in Tanzania.

1 What is process tracing?

Process tracing is part of a wider effort in the social sciences to systematise qualitative methods (Collier 2011), by adopting a *generative* perspective of causality. The strengths of qualitative methods are that they can assist in explaining *how* a given input (resource, activity) led to an observed effect (*ibid.*), an aspect that is often lacking in quantitative methods. Box 1 compares different perspectives on causality.

As a social science research method, process tracing is relatively recent, and its application still requires further development and refinement. It was originally used to provide theoretical explanations of historical events (Falleti 2006). In the social sciences, process tracing has been used by scholars who want to go beyond identifying statistical correlations – for example to better understand the relationship between democracy and peace (Beach and Pedersen 2013). Early contributions to the articulation of process tracing in political science stem from Alexander George and Andrew Bennett (Bennett and George 1997, 2005; Bennett 2008, 2010). Process tracing was further elaborated by David Collier (2011) and a recent book by Derek Beach and Rasmus Brun Pedersen (2013) provides a detailed articulation of the theoretical groundings of process tracing as well as step-by-step guidance to its application.[2] Beach and Pedersen emphasise the potential

Box 1 **Different perspectives on causality**

Counterfactual Based on Mill's method of difference, the counterfactual perspective of causal inference uses a control group to isolate the effect of an intervention. A comparison is made between two otherwise identical cases in which one received an intervention or treatment and the other one did not. This framework is frequently used in clinical trials for medical research. However, counterfactual causal inference does not explain *how* a specific effect came about.

Regularity Originating from Mill's method of agreement, regularity frameworks use the frequency of association between two observations to assess an effect. Regularity is the basis for making causal claims in many statistical approaches to evaluation. However, regularity frameworks do not identify the direction of change (which observation is the cause and which the effect), and cannot answer questions about *how* and *why* change happens. Their application is also problematic in complex situations where it is difficult to single out specific cause and effect factors.

Configurational Drawing on the concepts of *necessity* and *sufficiency*, configurational frameworks describe a number of causes that lead to a specific effect, and identify specific configurations of causal factors that are associated with it. The configurational view of causation recognises that more than one constellation of causes can lead to the same effect, and that similar constellations can lead to different, even opposite effects. Configurational frameworks are used in Qualitative Comparative Analysis. The sets of conditions identified in these frameworks go some way in answering 'how' a specific effect occurred.

Generative The distinctive feature of generative frameworks is that they provide a detailed description of a causal mechanism that led to a specific effect, and by doing so demonstrate the causal relation. Through a fine-grained explanation of what happens between a cause and an effect, generative mechanisms help to explain 'why' a certain effect occurred.

Source: Based on Befani (2012) and Collier (2011).

of process tracing as a qualitative method for assessing causal inference through the analysis of *causal mechanisms* in a *single-case design*. In our summary of the process tracing methodology we mainly draw on Beach and Pedersen's approach, and discuss its practical application in impact evaluation practice.

Process tracing involves articulating the steps between a hypothesised cause (for example, a development intervention) and an outcome. This involves unpacking the causal mechanism that explains what it is about a cause that leads to an outcome: the causal force or power that links cause A with outcome B (Beach and Pedersen 2013). The concept of 'causal mechanism' is central to the generative framework underpinning process tracing (see Box 1), but can cause confusion as there is no clear consensus in the literature on what exactly a mechanism is (Shaffer 2014). In Beach and Pedersen's description of process tracing, a mechanism is the causal chain or story linking event A with outcome B. A mechanism is made up of a number of 'parts' composed of *entities* (for example, people, organisations, systems) that engage in *activities* (for example, protesting, researching, campaigning); and each part of the mechanism is *necessary* to give rise to the subsequent part. This differs from the definition of mechanism used elsewhere, for example in realist evaluation approaches where a researcher may examine multiple different mechanisms within a single case (Westhorp 2014).

There is also a strong diagnostic element to process tracing. A causal chain linking cause A and outcome B is developed, and Bayesian probability logic is followed in order to assess the strength of the evidence of each part of the chain. Contrary to statistical methods, the quality of the evidence is not judged by sample size (the number of observations) but rather the *probability* of observing certain pieces of evidence. Assessments of probability in process tracing are not necessarily quantitative. Rather, the nature and assessment of evidence has parallels to a law court: evidence consists of empirical observations combined with knowledge of contextual factors (such as prior knowledge, timing, and the ways in which facts emerge) (Befani and Mayne 2014). The investigator works in a similar way to a detective (literature on process tracing often references Sherlock Holmes), looking for evidence to increase confidence that an outcome was caused in a particular way. Using probability logic, the investigator then systematically assesses the evidence in order to test hypotheses at each stage of the theory, including hypotheses representing alternative causal explanations.

One difference to the workings of a detective is that more than one causal chain may contribute to the effect under investigation (Bennett and George 2005). For example, evidence might give us high confidence that our advocacy intervention caused a policy change, but this does not rule out the possibility that other factors

external to the intervention also contributed to the outcome. This has an important repercussion for the use of process tracing in impact evaluation: it allows for judgements on contribution rather than attribution.

2 Applying process tracing tests

In assessing the probability that the hypothesised causal chain led to an isolated effect, the investigator compares alternative causal sequences, through:

A Reviewing the evidence under the assumption that the hypothesised causal sequence holds: cause A led to outcome B in the theorised way.

B Reviewing the evidence under the assumption that the hypothesised causal sequence does not hold: an alternative causal sequence explains the outcome.

The investigator examines the available evidence to test the inferential weight of evidence for each of these

---

Box 2 **Illustrations of the four process tracing tests**

Straw-in-the-wind test (low uniqueness, low certainty). This is the weakest of the four tests, neither necessary nor sufficient to confirm a hypothesis.

> **Example hypothesis** John shot Mary because he discovered her having an affair.
> **Evidence constituting this type of test** Evidence that affair was taking place – for example, a hotel receipt, suggestive text messages.
> **What happens if the hypothesis passes the test (i.e. reliable evidence of this type exists)?** The investigator can be slightly more confident in the hypothesis, but this is not enough to conclusively prove it or to disprove alternative hypotheses. However, straw-in-the-wind tests can provide a valuable benchmark, and if a hypothesis passes multiple tests this can add up to important evidence.
> **What happens if the hypothesis fails the test (i.e. reliable evidence of this type does not exist)?** This slightly raises doubts about the truth of the hypothesis, but is not enough to rule it out.

Hoop test (high certainty: necessary to confirm hypothesis).

> **Example hypothesis** John shot Mary.
> **Evidence constituting this type of test** John lacks a good alibi for the night of the murder – for example, he claims he was alone.
> **What happens if the hypothesis passes the test?** It does not significantly raise the investigator's confidence that the hypothesis is true. John lacking a good alibi is not enough on its own to prove the hypothesis.
> **What happens if the hypothesis fails the test?** It disconfirms the hypothesis. If John has a watertight alibi, we can be confident that he did not shoot Mary. Because of this, hoop tests are often used to exclude alternative hypotheses.

Smoking gun test (high uniqueness: sufficient to confirm hypothesis).

> **Example hypothesis** John shot Mary.
> **Evidence constituting this type of test** John was found holding a smoking gun over Mary's body.
> **What happens if the hypothesis passes the test?** The investigator can be confident that the hypothesis is true – John did indeed shoot Mary.
> **What happens if the hypothesis fails the test?** It does not significantly decrease confidence in the hypothesis. John may have shot Mary and escaped undetected.

Doubly decisive test (high certainty, high uniqueness). This is the most demanding test, both necessary and sufficient to confirm a hypothesis.

> **Example hypothesis** John shot Mary.
> **Evidence constituting this type of test** John was caught on a high-resolution, tamper-proof CCTV camera committing the crime.
> **What happens if the hypothesis passes the test?** We can be confident that the hypothesis is true, and that all alternative hypotheses are false. John did indeed shoot Mary.
> **What happens if the hypothesis fails the test?** It depends on the nature of the test. If someone else was caught on CCTV committing the crime, it would disconfirm the hypothesis. But if there simply was not a camera, it does nothing to increase or decrease our confidence in the hypothesis.

Source: Beach and Pedersen (2013) and Collier (2011).

causal sequences. Four 'tests' have been developed to assist with this process: 'straw-in-the-wind' tests, 'hoop' tests, 'smoking gun' tests and 'doubly decisive' tests (Bennett 2010; Collier 2011; Van Evera 1997). These tests are based on the principles of certainty and uniqueness; in other words, whether the tests are *necessary* and/or *sufficient* for inferring the evidence. Tests with high uniqueness help to strengthen the confirmatory evidence for a particular hypothesis, by showing that a given piece of evidence was sufficient to confirm it. Tests with high certainty help to rule out alternative explanations by demonstrating that a piece of evidence is necessary for the hypothesis to hold (Beach and Pedersen 2013; Befani and Mayne 2014). The four tests are illustrated in Box 2. In a companion to this paper, CDI Practice Paper 10 Annex describes the main steps in applying process tracing and provides some examples of how these steps might be applied in practice.

3 Applications of process tracing within international development impact evaluation

Impact evaluation designs based on counterfactual and regularity frameworks frequently encounter limitations when applied in the field of international development. For example, they may not be appropriate to measure initiatives that aim to achieve change through advocacy and policy influence, because the pathways of change are usually unpredictable, highly dependent on changing circumstances and often need to respond to changing goalposts (Tsui, Hearn and Young 2014).

While process tracing has predominantly been applied as a social science research method, the approach is currently being explored in several international development impact evaluations. A recent workshop on process tracing in impact evaluation organised by the Centre for Development Impact (CDI) brought together a number of evaluators who currently apply or intend to apply this method (Barnett and Munslow 2014). Two applications of process tracing are discussed below: a completed evaluation of the Universal Health Care Campaign in Ghana funded by Oxfam GB (Stedman-Bryce 2013); and an ongoing evaluation of the Hunger and Nutrition Commitment Index (HANCI).[3]

Case study 1: Assessing the Universal Health Care Campaign in Ghana

The Universal Health Care Campaign in Ghana is a collaborative advocacy effort by civil society organisations and networks to promote universal free access to health care, which received core funding from Oxfam GB until 2013. The campaign aimed to take advantage of the window of opportunity created by national elections in 2012 in order to advocate for universal free health care by 2015, through a combination of lobbying, popular mobilising, media and research.

An effectiveness review of the campaign was commissioned by Oxfam GB and conducted in 2012–13 (Stedman-Bryce 2013). The review is based on a process tracing protocol developed by Oxfam GB (2011), which incorporates elements of both process tracing and contribution analysis.[4] The protocol focuses on elaborating and testing a small number of (not necessarily directly connected) outcomes within a larger project theory of change. It involves three elements:

■ Shortlisting one or more evidenced explanations for the outcome in question (which may or may not include the intervention).

■ Ruling out alternative competing explanations incompatible with the evidence.

■ If more than one explanation is supported by the evidence, estimating the level of influence each had on bringing about the change in question.

In line with this protocol, the effectiveness review compares alternative causal sequences and attempts to weight evidence, but it does not explicitly apply process tracing tests. It also does not specify or test a full causal chain. This means that, although the evaluation establishes a degree of confidence in the contribution of the campaign to a number of distinct outcomes, it does not demonstrate whether the programme *as a whole* contributed to the final outcome.

The evaluation used an existing project theory of change, revised following conversations between the evaluator and project staff. The evaluator stressed that translating project staff's informal understanding about how change happened into an explicit, formal theory was a crucial and challenging aspect of the evaluation. The analysis drew on 21 key informant interviews, mainly with members of the campaign itself and with government representatives. It also drew on project documentation and data (for example, news articles) accessed online. In particular, the evaluator highlighted the value of a Facebook page created by the campaign in helping to reconstruct a timeline of events and in accessing online evidence.

Data were analysed by assessing their explanatory power in relation to two rival causal sequences for the identified outcomes, followed by an assessment of the contribution of each causal sequence to the observed change.

An example of the analytical process used to weigh alternative causal sequences

One of the outcomes examined was: 'the current National Health Insurance Scheme (NHIS) is shown to be an ineffective vehicle to deliver free universal health care in Ghana' (Stedman-Bryce 2013: 4). An important milestone related to this outcome was a highly controversial report published by the campaign, which contended that the

number of people enrolled under the NHIS was inaccurate and needed to be revised downwards. Several months after the report was published, the government department responsible for the NHIS (the National Health Insurance Authority, or NHIA), revised its approach to counting NHIS membership, resulting in a decrease in official statistics on membership from 67 per cent to 34 per cent.

The two rival causal sequences examined in the evaluation were:

- The methodology revision occurred as a result of pressure exerted by the campaign.

- The revisions occurred based on the NHIA's own plans and timetable.

The evidence used to evaluate these alternative causal sequences was as follows:

- Evidence regarding the level of attention the campaign's report received – for example, quotations from key informant interviews, media articles and blogs (including several published responses by the department refuting the report's claims). The evidence suggested that the report did indeed dominate the health sector debate in Ghana for some time.

- Testimonies from campaign members affirming that the NHIA revised its methodology based on the public uproar caused by the report. This evidence is enhanced by consideration of the context, particularly the additional pressure on the government – exerted by forthcoming elections – to respond to allegations of corruption and inefficiency.

- A statement by the Ghana delegation at an international meeting on Universal Health Care in Geneva, confirming that the campaign's report 'was very helpful and prompted us to revise our figures'. Although the report did not explicitly apply tests, this is a clear example of a 'smoking gun'. It is highly unlikely that the delegation would make this statement if the report had not influenced them, particularly since the NHIA had dismissed the report during the national health sector debate that ensued in Ghana after its publication. The evidence therefore has high uniqueness, and significantly increases confidence in causal sequence 1.

The evaluator then goes on to test the rival causal sequence – that the methodology was revised based on the NHIA's own plans and timetable. He finds that there is no convincing evidence to that end, and infers from the timing of the campaign, and the contestation of any flaws in methodology from the NHIA itself only weeks before its revision, that this rival sequence does not hold.

Case study 2: Framing hunger and nutrition as political commitment: an intervention assessment of HANCI in Tanzania

The Hunger and Nutrition Commitment Index (HANCI) ranks governments in 45 countries on their political commitment to tackling hunger and undernutrition.[5] One of HANCI's main aims is to reshape the debate around nutrition, in order to frame the solution to hunger as political rather than purely technical. At the time of writing this paper, researchers at the Institute of Development Studies (IDS) were conducting an intervention assessment of HANCI's policy impact, using process tracing methods. Programme staff selected process tracing for its potential to trace the ways in which HANCI *contributed* to policy change, given the recognition that change cannot be attributed to a single programme in this complex area.

Specifying the outcome(s)

The assessment aimed to examine the contribution of HANCI to the framing of nutrition policy in Tanzania. However, specifying what this outcome would look like in empirical terms was difficult, given the fact that the project was ongoing, and that the nature and framing of policy discussions are emergent and unpredictable factors. This means that at the beginning of the assessment, programme staff did not know what the outcome they wanted to test would look like – which complicated data collection. As a result, the theory was eventually split into two outcomes, which will be considered in two separate forthcoming papers. The intermediate outcome was that *partners find evidence generated by HANCI credible, and use it in their policy advocacy.* The final outcome was that *HANCI influenced the framing of nutrition problems and solutions during the drafting of political party manifestos in the run up to elections in September 2015.*

Developing the theory

HANCI programme staff developed a theory of change to posit how the HANCI intervention was likely to influence the outcomes. From this, a causal mechanism was developed to provide a plausible explanation describing the link between the intervention and the outcome. The mechanism was constructed using an iterative approach, through two parallel processes:

- A substantial literature review was conducted to identify theories exploring the role of advocacy in promoting policy change. These theories helped inform the creation of the causal mechanism. The literature review also collated empirical evidence from other interventions which supported or challenged the HANCI causal mechanisms. Distilling evidence from the wider literature in this way proved time consuming – given the large number of potentially relevant theories, the indistinct boundaries and overlaps between them, and the subsequent difficulty in classifying evidence in order to use it to support or challenge the HANCI mechanism.

■ Pre-existing evidence (collected throughout the programme) was examined – including workshop reports and surveys, interview transcripts and country learning reports produced by partners. This both helped develop the theory of how HANCI led to change and provided evidence to test this theory.

Applying the four process tracing tests to the causal mechanism proved challenging. The main constraint was that the evidence was vast and included findings at organisational and national levels – for example, strategy documents and national media releases. Obtaining evidence for alternative causal sequences (in which HANCI did *not* lead to changes in framing at the national level) was also hard to come by, and given the scope of the study (national-level policy change) it was difficult to test and eliminate all possible alternative explanations. The tests were therefore applied in a limited way, in that only empirical observations that passed the 'hoop' test were considered (observations that were *necessary* for the causal mechanism to hold).

4 Reflections on the application of process tracing in impact evaluation

This section discusses some of the emerging reflections and lessons on the practical application of process tracing in impact evaluation.

Advantages of process tracing
Process tracing offers a rigorous approach to assess causal change, including through an *ex post* design without a control group. Time and resources are major obstacles to many organisations wishing to measure the impact of their work. Methods based on a counterfactual causal framework (involving baseline and endline data collection, and comparing change across beneficiary and counterfactual groups) can be time consuming and expensive. Process tracing therefore offers great potential as a rigorous method appropriate for *ex post* evaluations, without the requirement for baseline or counterfactual data (although it is possible to use both, for example within process tracing tests).

Process tracing offers potential for examining causality in programmes where attribution is difficult. Process tracing focuses on breaking down an intervention into its constituent causal parts, investigating each part and exploring how and why they link together. This approach is particularly relevant to development interventions where the pathways of change are not always predictable or certain; and where multiple processes occur in parallel and so change cannot be easily attributed to a particular cause. As the case studies demonstrate, policy and advocacy interventions are particularly conducive to this approach. Another advantage of process tracing is that it provides evidence on *how* and *why* an intervention led to change.

This is particularly relevant in new or complex interventions in which the causal pathways are not well known.

There is potential to combine aspects of process tracing with other theory-based evaluation approaches. Process tracing can provide a degree of confidence in a particular causal sequence, but cannot demonstrate *how important* a particular cause was to the achievement of the outcome relative to other causes. In other words, process tracing alone does not provide evidence on the degree or weight of contribution. However, the potential of combining process tracing with contribution analysis is discussed by Befani and Mayne (2014), who argue that the Bayesian logic of inference within process tracing (i.e. the process tracing tests) can complement an assessment of the relative contribution of different causal factors within an overarching contribution analysis framework – resulting in stronger inferences than either process tracing or contribution analysis can provide alone. This is similar to (although more in-depth and systematic than) the method suggested in Oxfam GB's process tracing protocol.

There may also be potential to apply the process tracing tests to systematically and transparently weigh and test qualitative evidence within other theory-based qualitative evaluation approaches, such as realist evaluation. Although time consuming to apply, the tests have the advantage of being fairly intuitive, perhaps given the long exposure many of us have to Sherlock Holmes stories or courtroom dramas. This is demonstrated above in the Oxfam GB case study, in which certain evidence can be retrospectively linked to various tests.

Challenges of process tracing
Process tracing can be time intensive. Developing a causal mechanism takes significant time and may require considerable stakeholder involvement and/or review of secondary literature, as emphasised in the two case studies above. Similarly, collecting the right amount and type of information to construct various tests requires considerable knowledge and understanding of the project, and sufficient capacity and time to analyse the data. For example, a key limitation identified in the Oxfam GB assessment was the small number of interviews with government representatives, given the difficulty of accessing these stakeholders in the time available. This highlights the risk that process tracing may provide inconclusive results if the evidence collected cannot fully support a causal sequence. To thoroughly test alternative hypotheses, the evaluator needs to have access to a range of stakeholders and to published and unpublished material.

There are challenges in applying process tracing where an outcome is not fully known. In Beach and Pedersen's description of process tracing, the outcome is known in advance. This poses challenges in the context of impact

evaluation, where the outcome is *not* known until the end of an (often multi-year) evaluation process and where there are multiple outcomes (which may remain somewhat uncertain), as was the case in the Oxfam GB evaluation. It also proved challenging in the HANCI evaluation, where it took time to specify evidence that might support the outcome of policy change while this change was still unfolding. Despite these challenges, it seems plausible that process tracing could be used as part of a mixed-methods evaluation design, applied at the end or at a mid-point of the evaluation to provide evidence of how and why the intervention led to a particular outcome (which might be established and verified through other methods).

It may also be possible to conduct process tracing alongside an intervention, by developing a causal mechanism *before or during* the project; and then collecting evidence to test parts in the mechanism as the intervention unfolds. However, this represents a potentially major risk. In complex interventions (such as policy and advocacy initiatives) objectives can be fluid, and the final outcome is frequently quite different to that initially envisaged. A clear causal mechanism may be difficult to develop in a situation where multiple factors combine and accumulate to lead to tipping points; or where feedback loops mean that later events reinforce earlier events and processes (Ramalingam 2013), as was the case in HANCI. This means that it is highly likely that the mechanism developed at the beginning of the project

would change over time – and this could mean that data collected during earlier stages of an evaluation is not the right evidence to test the revised mechanism at the end.

5 Conclusion

Process tracing has major potential as a rigorous method for qualitative impact evaluation, using probability tests based on Bayesian logic of inference to establish *confidence* in *how* and *why* an effect occurred. It may be of particular interest as a method appropriate for *ex post* evaluations which lack baseline data or a control group – although it certainly does not offer a quick or easy evaluation solution. The process tracing tests (straw-in-the-wind, hoop, smoking gun and doubly decisive) are a particularly intriguing aspect of the method, drawing on relatively intuitive concepts of uniqueness and certainty to systematically and transparently weigh and test qualitative evidence. So far the applications of process tracing within impact evaluation in international development are limited, although a number of ongoing evaluations are attempting to apply the method. The two case studies discussed in this paper illustrate some of the challenges faced and choices made along the way. There are still unanswered questions around the utility of process tracing within impact evaluation; for example, in relation to evaluating complex interventions, or applying it in circumstances where the outcome is not yet known. However, overall process tracing represents a valuable methodological approach to add to the evaluator's toolbox.

Notes

1 Impacts are defined by the Organisation for Economic Co-operation and Development (OECD 2010: 24) as 'positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended'. Impact evaluations attempt to identify a clear link between causes and effects and explain how the intervention worked and for whom.

2 A new book on process tracing was published in November 2014: *Process Tracing. From Metaphor to Analytic Tool* (Bennett and Checkel 2014). This was not available to the authors at the time of writing this paper and therefore is not taken into consideration here.

3 Another example of a discussion on the application of process tracing is a recent journal article by Befani and Mayne (2014). This, however, explores a hypothetical case study that combines process tracing and contribution analysis.

4 Contribution analysis assesses the contribution of an intervention to observed results; through verifying the theory of change behind the intervention (using logic, secondary evidence and empirical evidence) and considering the role of other influencing factors (see Befani and Mayne 2014; Mayne 2008).

5 See www.hancindex.org and te Lintelo *et al.* (2014).

References

Barnett, C. and Munslow, T. (2014) *Process Tracing: The Potential and Pitfalls for Impact Evaluation in International Development*, Evidence Report 102, Brighton: IDS

Beach, D. and Pedersen, R.B. (2013) *Process-Tracing Methods: Foundations and Guidelines*, Ann Arbor MI: University of Michigan Press

Befani, B. (2012) 'Models of Causality and Causal Inference', in E. Stern; N. Stame; J. Mayne; K. Forss; R. Davies and B. Befani (eds), *Broadening the Range of Designs and Methods for Impact Evaluations*, DFID Working Paper 38, London: Department for International Development

Befani, B. and Mayne, J. (2014) 'Process Tracing and Contribution Analysis: A Combined Approach to Generative Causal Inference for Impact Evaluation', *IDS Bulletin* 45.6: 17–36

Bennett, A. (2010) 'Process Tracing and Causal Inference', in H.E. Brady and D. Collier (eds), *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, 2nd ed., Lanham MD: Rowman and Littlefield: 207–19

Bennett, A. (2008) 'Process Tracing: A Bayesian Perspective', in J.M. Box-Steffensmeier, H. Brady and D. Collier (eds), *The Oxford Handbook of Political Methodology*, Oxford: Oxford University Press: 702–21

Bennett, A. and Checkel, J.T. (eds) (2014) *Process Tracing. From Metaphor to Analytic Tool*, Cambridge: Cambridge University Press

Bennett, A. and George, A.L. (2005) 'Process-Tracing and Historical Explanation', in A. Bennett and A.L. George (eds), *Case Studies and Theory Development in the Social Sciences*, Cambridge MA: MIT Press: 205–32

Bennett, A. and George, A.L. (1997) 'Process Tracing in Case Study Research', paper presented at the MacArthur Foundation Workshop on Case Study Methods , Belfer Center for Science and International Affairs (BCSIA), University of Harvard, Cambridge MA

Collier, D. (2011) 'Understanding Process Tracing', *Political Science & Politics* 44.4: 823–30

Falleti, T. (2006) 'Theory-guided Process Tracing in Comparative Politics: Something Old, Something New', *APSA-CP: Newsletter of the American Political Science Association* 17.1: 9–14

Mayne, J. (2008) 'Contribution Analysis: An Approach to Exploring Cause and Effect', *ILAC Brief* 16, International Learning and Change

OECD (2010) *Glossary of Key Terms in Evaluation and Results Based Management*, www.oecd.org/dac/evaluation/2754804.pdf (accessed 2 February 2015)

Oxfam GB (2011) 'Process Tracing: Draft Protocol', Oxfam GB

Ramalingam, B. (2013) *Aid on the Edge of Chaos: Rethinking International Cooperation in a Complex World*, Oxford: Oxford University Press: 480

Shaffer, P. (2014) 'Two Concepts of Causation: Implications for Poverty', *Development and Change* 46.1: 148–66

Stedman-Bryce, G. (2013) *Health for All: Towards Free Universal Health Care in Ghana, End of Campaign Evaluation Report*, Oxford: Oxfam GB

te Lintelo, D.; Haddad, L.; Lakshman, R. and Gatellier, K. (2014) *The Hunger And Nutrition Commitment Index (HANCI 2013): Measuring the Political Commitment to Reduce Hunger and Undernutrition in Developing Countries*, IDS Evidence Report 78, Brighton: IDS

Tsui, J.; Hearn, S. and Young, J. (2014) *Monitoring and Evaluation of Policy Influence and Advocacy*, Working Paper 395, London: Overseas Development Institute

Van Evera, S. (1997) *Guide to Methods for Students of Political Science*, Ithaca NY: Cornell University Press

Westhorp, G. (2014) *Realist Impact Evaluation: An Introduction*, London: Methods Lab/Overseas Development Institute

" Process tracing represents a valuable addition to the evaluator's toolbox, although there are still unanswered questions around how best to apply it to international development initiatives. "

Institute of Development Studies, Brighton BN1 9RE, UK
T +44 (0) 1273 915637 **F** +44 (0) 1273 621202 **E** ids@ids.ac.uk **W** www.ids.ac.uk

CDI PRACTICE PAPER 10 April 2015   **www.ids.ac.uk/cdi**