5

ANALYZING THE DATA—INITIAL ANALYSES

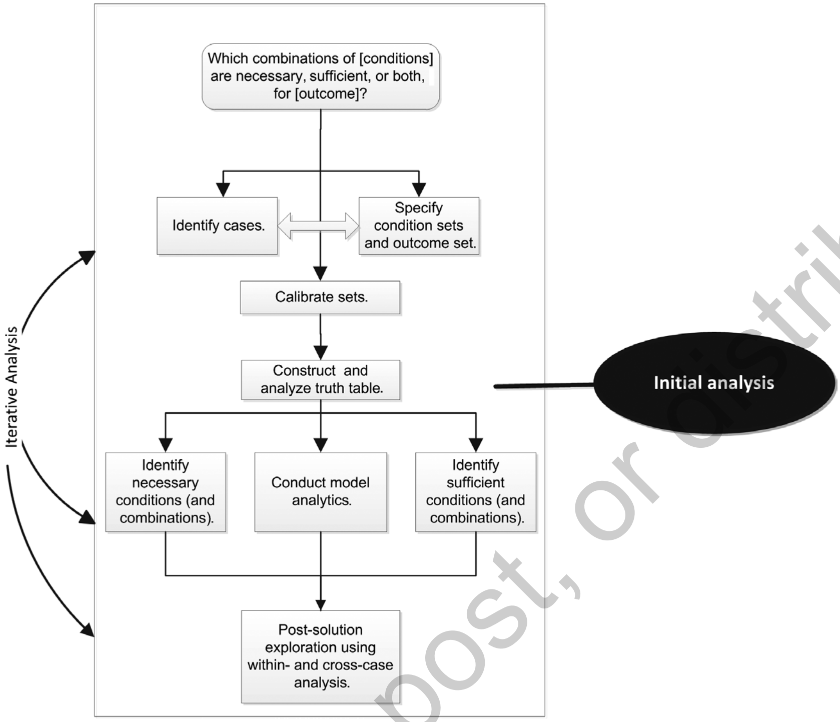
LEARNING OBJECTIVES

1. Transform a data matrix of set membership values into a truth table.
2. Employ strategies for managing contradictory truth table rows.
3. Revisit the data to manage contradictory truth table rows.
4. Inspect a truth table for potential issues.
5. Conduct an analysis of necessary conditions and combinations of conditions.
6. Conduct a *preliminary* sufficiency analysis of a truth table.

In this chapter, we initiate the analytic process. The chapter starts with an explanation of how to take the data matrix of set membership values (SMV) that the researcher has assembled to create a truth table. We carry over the school health features and academic performance example at the end of Chapter 4 to guide the reader through this process. In practice, one uses software to transform a data matrix into a truth table, but knowing the underlying process for creating a truth table will enable the reader to understand a core aspect of qualitative comparative analysis (QCA).

Next, we describe several analytic steps in detail. Conducting an analysis entails identifying the necessary conditions and combinations, reviewing the truth table, making preliminary analytic decisions, and conducting an initial analysis for sufficient conditions and combinations of conditions. We summarize key analytic decisions and explain practices for making them. We continue with the school health features and academic performance example to illustrate these steps.

FIGURE 5-1 ■ Guiding Heuristic: Initial Analysis



Adapted from: Kane, H., Lewis, M. A., Williams, P. A., & Kahwati, L. C. (2014). Using qualitative comparative analysis to understand and quantify translation and implementation. *Transl Behav Med,* 4(2), 201–208. doi: 10.1007/s13142-014-0251-6

Throughout this chapter, we describe the underlying mathematics that the analysis draws upon. However, we do not elaborate or demonstrate how to calculate solutions manually because software should be used to avoid errors. We refer curious readers to other textbooks that explain the underlying mathematics in detail. We recommend that readers be thoroughly familiar with concepts introduced in earlier chapters, such as sufficient and necessary relationships and consistency (Chapter 2), limited diversity (Chapter 3), and calibration (Chapter 4), before proceeding with this chapter.

OVERVIEW OF ANALYSIS

Prior chapters have focused on making design decisions, such as selecting cases and conditions, as well as organizing and transforming data for conducting a QCA. This chapter begins by explaining how to construct a truth table for analysis and how to conduct the initial analysis. As shown in *Figure 5-1,* the first step in preparing for an analysis entails transforming the matrix of set membership values (SMV) generated from raw data using a calibration rubric into a truth table—the central analytic device for QCA. Truth tables and the process for creating them are identical whether the researcher uses crisp or fuzzy sets. After describing how to construct a truth table, this chapter explains the initial steps in conducting a QCA. These steps involve assessing the truth table before conducting the analysis and analyzing the data for necessary conditions, followed by an analysis of sufficient conditions. In the analysis of necessary conditions, one evaluates whether the condition or combination of conditions is always present when the outcome is present. In the analysis of sufficient conditions, one conducts a truth table analysis to identify sufficient conditions and combinations of conditions; this also entails examining different solution types: the conservative, parsimonious, and intermediate solutions. As shown in the guiding heuristic (*Figure 5-1*), the analysis process is iterative, and what we describe in this chapter are only the initial steps.

TRANSFORM A DATA MATRIX INTO A TRUTH TABLE

In brief, the **truth table** displays all the possible combinations of conditions (i.e., configurations), shows the cases that belong to each of those possible configurations, and identifies the set relationship between each configuration and the outcome. Transforming a data matrix (i.e., the SMVs generated through the process of calibration) into a truth table involves three steps: (1) creating a truth table shell; (2) assigning cases from the data matrix to truth table rows, and (3) assigning an outcome value to each truth table row. Again, this is all done using software, but we show it in detail for deeper understanding and ultimately to help researchers avoid using software mechanistically.

Step 1: Creating a Truth Table Shell

Creating a truth table shell first involves constructing a table of all possible combinations of conditions (i.e., configurations) in an analysis. For ease of understanding, we return to the school health features and academic performance example we introduced in Chapter 4. In this crisp set example, the configural question was

> "What school-level health and wellness policies are found among demographically comparable schools that achieve adequate academic performance?"

The cases were demographically comparable schools (the first column of the data matrix, *Table 5-1*). The conditions were (1) serving healthy meals (MEAL),

(2) adequate levels of physical education instruction (PE), (3) adequate amounts of recess or unstructured physical activity time (RECESS), and (4) having a comprehensive school wellness policy (WELL). The outcome set was achieving adequate academic performance (OUT). *Table 5-1* is the data matrix of SMVs that we assigned for each condition and the outcome based on the calibration rubric we presented in Chapter 4. With this data matrix in hand, the first step is to build the truth table shell. Because this analysis uses four conditions, the truth table will have 16 rows, representing 16 possible configurations of the four conditions. Recall, that the number of logically possible configurations for the truth table is equal to 2^k, where k is the number of conditions (see Chapter 3.)

TABLE 5-1 ■ Data Matrix of Schools, Health, and Wellness Policies (Conditions), and Academic Performance (Outcome): Crisp Set Example

Case	MEAL	PE	RECESS	WELL	OUT
Springfield	1	0	0	0	0
Creekside	1	1	1	0	1
Westside	1	1	1	1	1
Ellis	1	0	0	1	0
Smithton	1	0	1	1	1
Harrison	1	0	0	1	0
Tubman	1	0	0	0	1
Curie	1	1	1	0	1
Tahoe	0	0	0	0	0
Obama	1	0	1	1	1
Cardinal	1	0	0	1	1
Watt	1	1	1	1	1
Parks	1	1	1	0	1
Fletcher	0	0	0	0	0
Victory	1	1	1	0	1
Goodall	1	0	0	0	0
Kruse	0	1	1	0	1

Pine	1	1	0	0	1
Flora	1	1	0	0	1
Lincoln	1	0	1	1	1
DuBois	1	0	1	1	1
Lovelace	1	1	0	1	0
Parkwood	1	1	0	1	1
Grove	1	1	0	1	1

Creating a truth table shell for this analysis involves listing each condition name at the top of a column. Then, one writes out the logically possible configurations of condition SMVs with 0s or 1s for each row. A truth table always uses 0 and 1 to indicate set membership in the condition, regardless of whether it is a crisp or fuzzy set analysis. *Table 5-2* shows the truth table shell with 16 logically possible configurations for our example. Although the condition names may be different, any crisp or fuzzy set analysis with 4 conditions will have the same combinations of 0s and 1s in the truth table shell.

Step 2: Assign Cases From the Data Matrix to a Truth Table Row

For crisply calibrated sets, assigning cases to truth table rows is straightforward. All cases in the data matrix will be represented by one and only one of the rows in the truth table shell. This process entails matching the configuration of each case's SMVs (i.e., its 0s and 1s) to the appropriate truth table row. For example, to place the first case (Springfield Elementary) from *Table 5-1* into a truth table row, one would look for the truth table row with the combination of a SMVs equal to 1 for MEAL, 0 for PE, 0 for RECESS, and 0 for WELL. This combination of SMVs is represented by Row 9 of the truth table shell (*Table 5-2*). Creekside Elementary (Case 2) can be placed into Row 15 of the truth table shell, based on its combination of SMVs for MEAL (1), PE (1), RECESS (1), and WELL (0). This process is repeated for all the cases until each one is assigned to the truth table row that represents its configuration of SMVs. *Table 5-3* shows the nearly final truth table, where all cases are assigned to a row, and we have added columns to tally the number of cases in each row and the names of the cases that belong to each row.

At the end of this step, it is possible that one truth table row may have multiple cases assigned to it (e.g., most of the rows in *Table 5-3*); this simply means those cases share the same configuration for all specified conditions. It is also possible that some of the rows in the truth table may not have any cases assigned

TABLE 5-2 ■ Truth Table Shell for the 4 Conditions of the School Health Features and Academic Performance Example

Row #	MEAL	PE	RECESS	WELL
1	0	0	0	0
2	0	0	0	1
3	0	0	1	0
4	0	0	1	1
5	0	1	0	0
6	0	1	0	0
7	0	1	1	0
8	0	1	1	1
9	1	0	0	0
10	1	0	0	1
11	1	0	1	0
12	1	0	1	1
13	1	1	0	0
14	1	1	0	1
15	1	1	1	0
16	1	1	1	1

(e.g., Rows 2, 3, 4, 5, 6, 8, 11 of *Table 5-3*); this means there are no cases with the specific configuration represented by that row. These empty rows are called **logical remainders**, and when a truth table has logical remainders, this means the truth table has limited diversity (Chapter 3). We will return to logical remainders in a later part of the analysis. *Table 5-4* provides a fuzzy set data matrix for the school health features and academic performance example. The fuzzy set data matrix mirrors the crisp set data matrix (*Table 5-1*); for example, if a case had a crisp SMV of 1, it has a fuzzy SMV above the 0.5 crossover point. Also, to distinguish the fuzzy conditions from the crisp ones, we have added "F" in front of the condition and outcome set names. The process for transforming a data matrix derived from fuzzy sets into a truth table is similar in that each case is assigned to one row in the truth table. However, unlike cases with crisp sets that can be matched easily to a row based on the configuration of 0s and 1s, cases calibrated

TABLE 5-3 ■ Truth Table With Cases Assigned to Each Row for the School Health Features and Academic Performance Example (Crisp Sets)

Row #	MEAL	PE	RECESS	WELL	NUMBER OF CASES	CASES
1	0	0	0	0	2	Tahoe, Fletcher
2	0	0	0	1	0	---
3	0	0	1	0	0	---
4	0	0	1	1	0	---
5	0	1	0	0	0	---
6	0	1	0	0	0	---
7	0	1	1	0	1	Kruse
8	0	1	1	1	0	---
9	1	0	0	0	3	Springfield, Tubman, Goodall
10	1	0	0	1	3	Ellis, Harrison, Cardinal
11	1	0	1	0	0	---
12	1	0	1	1	4	Smithton, Obama, Lincoln, DuBois
13	1	1	0	0	2	Pine, Flora
14	1	1	0	1	3	Lovelace, Parkwood, Grove
15	1	1	1	0	4	Creekside, Curie, Parks, Victory
16	1	1	1	1	2	Westside, Watt

with fuzzy sets have partial set membership with SMVs between 0 and 1, which means they partially belong to multiple rows.

Recall from Chapter 2, if a case has a SMV of 0.2 in a condition called "A", it simultaneously has a SMV of 0.8 in the condition "NOT A." The QCA software uses Boolean algebra to determine the case's SMV in each row. A case will have a SMV of greater than 0.5 in only one row of the table. The case is then assigned to that row. We refer the reader to other QCA texts for a more detailed discussion of the underlying mathematics of creating the truth table with fuzzy sets (Ragin, 2000, 2008b; Schneider & Wagemann, 2012).

TABLE 5-4 ■ Fuzzy Set Data Matrix for the School Health Features and Academic Performance Example

Case	FMEAL	FPE	FRECESS	FWELL	FOUT
Springfield	0.6	0.2	0.2	0.4	0.2
Creekside	0.8	0.6	0.8	0.2	0.8
Westside	1	1	1	0.8	1
Ellis	0.6	0.4	0.2	0.6	0
Smithton	0.8	0.4	0.6	1	0.8
Harrison	0.6	0	0.4	0.8	0
Tubman	0.6	0.4	0.4	0.4	0.6
Curie	0.8	0.8	0.6	0.4	0.8
Tahoe	0.2	0.2	0	0	0
Obama	0.8	0.4	0.8	0.6	1
Cardinal	0.6	0.4	0.2	0.6	0.6
Watt	1	0.8	1	0.6	0.8
Parks	0.6	0.8	0.8	0.4	0.8
Fletcher	0.2	0.2	0.2	0.2	0
Victory	1	0.6	0.8	0.4	0.8
Goodall	0.8	0	0	0	0
Kruse	0.4	0.6	0.8	0.2	0.6
Pine	1	1	0.2	0.2	1
Flora	0.8	0.8	0.2	0.2	1
Lincoln	0.8	0	0.8	1	0.8
DuBois	1	0.2	1	0.6	1
Lovelace	0.6	0.6	0	0.6	0
Parkwood	0.6	0.6	0	0.6	0.6
Grove	0.6	0.6	0	0.6	0.6

Step 3: Assign an Outcome Value to Each Truth Table Row

The last step in creating a truth table entails using the outcome SMV from the cases within each row to assign an outcome value for the row. Each truth table row represents a set of cases with membership in a particular configuration of conditions. To determine the outcome value of the row, one examines the parameter of fit called row consistency (note that some software packages refer to row consistency as inclusion or "incl"). We introduced the concept of consistency in Chapter 2 in the context of the relationship between a single condition and an outcome; here we consider the consistency of the sufficiency relationship between the complex set (i.e., configuration) represented by the truth table row and the outcome set. In simple terms the row consistency is the portion of the cases in the configuration that are also in the outcome set. Consistency can range between 0 and 1, with 0 signifying no subset relationship and thus no relationship of sufficiency, and 1 indicating a perfect subset relationship and thus a strong relationship of sufficiency. In general, a consistency of 0.8 to 1 demonstrates a strong sufficiency relationship, meaning nearly all cases with the configuration of conditions are in the outcome set. Consistency between 0.6 to 0.8 indicates a modest sufficiency relationship. Values below 0.5 to 0.6 represent weak sufficiency relationships.

When assigning an outcome value to a row, the researcher compares the row consistency against a prespecified **row consistency** threshold. The row consistency threshold refers to the acceptable strength of the sufficiency relationship the researcher is going to requires for the analysis; an acceptable **consistency threshold** depends on the study. Selecting a consistency threshold will be discussed in more detail below. One uses the row consistency threshold to assign either a 1 or a 0 to the row; if the row consistency is below the threshold, a value of 0 is assigned as the outcome value. If the row consistency is at or above the threshold, a value of 1 is assigned as the outcome value. This step results in a complete truth table, where rows meeting the acceptable threshold for sufficiency are identified with an outcome value of 1. These are the rows that one uses in subsequent parts of the sufficiency analysis.

Returning to our school health features and academic performance example, the truth table (*Table 5-5*) now includes a column for the row consistency and for the outcome value. First, we will explain how to calculate the row consistency; then, we elaborate how the researcher uses the row consistency to determine the outcome value to assign to the row.

For crisp sets, calculating row consistency is a simple proportion; the row consistency is the proportion of cases within the row that also have membership in the outcome set. Row 9 in *Table 5-5* has a 0.33 row consistency. One school (Tubman) has a SMV of 1 for the outcome, but Springfield and Goodall have a SMV of 0 for the outcome. The consistency value of 0.33 represents the portion of cases that have membership in the outcome set (1/3 = 0.33). Based on this low row consistency, we would conclude that the configuration of conditions in Row 9 has a very weak sufficiency relationship with the outcome, and

TABLE 5-5 ■ Complete Truth Table for the School Health Features and Academic Performance Example (Crisp Sets)

Row #	MEAL	PE	RECESS	WELL	Number of Cases	Row Consistency	Outcome Value	Cases
1	0	0	0	0	2	0.00	0	Tahoe, Fletcher
2	0	0	0	1	0	---	?	---
3	0	0	1	0	0	---	?	---
4	0	0	1	1	0	---	?	---
5	0	1	0	0	0	---	?	---
6	0	1	0	0	0	---	?	---
7	0	1	1	0	1	1.00	1	Kruse
8	0	1	1	1	0	---	?	---
9	1	0	0	0	3	0.33	0	Springfield, Tubman, Goodall
10	1	0	0	1	3	0.33	0	Ellis, Harrison, Cardinal
11	1	0	1	0	0	---	?	---
12	1	0	1	1	4	1.00	1	Smithton, Obama, Lincoln, DuBois
13	1	1	0	0	2	1.00	1	Pine, Flora
14	1	1	0	1	3	0.67	0	Lovelace, Parkwood, Grove
15	1	1	1	0	4	1.00	1	Creekside, Curie, Parks, Victory
16	1	1	1	1	2	1.00	1	Westside, Watt

PRACTICE TIP 5-1

ASSIGNING OUTCOME VALUES TO ROWS IN INTERMEDIATE- AND LARGER-N STUDIES

In intermediate (50–100 cases) and larger-N (over 100 cases) studies, the researcher may wish to exclude rows (i.e., assign an outcome value of 0) with a very small number of cases (<5), even if the consistency value is high because the cases may be outliers. However, if the mixed method study design (e.g., explanatory sequential) accommodates additional data collection, a researcher could purposively sample those cases for further exploration. In Chapter 6, we discuss dropping or adding cases to assess robustness of results, and in Chapter 7, we discuss further exploration of cases after conducting a QCA using within- or cross-case analysis.

we would assign a value of 0 to the outcome column of the truth table row to indicate that this combination of conditions is not sufficient for membership in the outcome set.

Row 9 is also called **contradictory truth table row** because the cases in the row contradict each other (some have membership in the outcome, some do not). In contrast, Rows 7 and 13 have perfect consistency. In row 7, Kruse exhibits the outcome and has a SMV of 1 for the outcome (1/1 = 1.0). Likewise, in Row 13, two out of the two schools (Pine and Flora) have a SMV of 1 for the outcome; this results in a consistency of 1.0 for the row (2/2 = 1.0). We would conclude that the configuration of conditions in Rows 9 and 13 has a strong (in fact perfect) sufficiency relationship with the outcome and would assign a value of 1 to the outcome column of the truth table.

Logical remainders (truth table rows without any cases) cannot be assigned an outcome value at the outset of creating the truth table. Thus, the outcome value for these rows is represented by a "?" and signifies that the outcome value is unknown for these rows because no empiric data are available to assess.

For fuzzy sets, one uses a more complex formula to calculate consistency of a truth table row (though this formula also works for crisp sets). This formula takes the sum of the minimum values between each case's membership score for the row and membership in the outcome, and then divides that sum by the sum of each case's membership score for the row. In *Box 5-1,* we provide the equation, but in practice, the QCA software computes the row consistency for truth table rows and will assign an SMV of 1 or 0 to each truth table row based on the row consistency threshold the researcher specifies for the analysis. We refer readers to other textbooks for more detail on calculating row consistency with fuzzy sets (Ragin, 2008b; Schneider & Wagemann, 2012).

BOX 5-1 Equation for Calculating Consistency of Sufficiency for Truth Table Rows

$$\frac{\sum_{i=1}^{I} \min(Xi, Yi)}{\sum_{i=1}^{I} Yi}$$

Xi indicates the condition SMV for each case

Yi indicates the outcome SMV for each case

Table 5-6 shows the truth table generated from the fuzzy set data matrix, and in comparing it to the crisp set truth table (*Table 5-5*), one will note that the combinations with 0s and 1s are the same, the same cases appear in the same rows, and the number of cases in each row is the same. However, the row consistency values change because cases calibrated with fuzzy sets hold membership in multiple rows simultaneously. Membership in multiple rows contributes to increasing (or decreasing) the row consistency. For example, in the crisp truth table (*Table 5-5*), Row 1 has a consistency value of 0; in the fuzzy truth table (*Table 5-6*), the consistency value for Row 1 has increased to 0.5 because other cases with a high SMV in the outcome now have partial membership in Row 1. Alternately, in the crisp set table (*Table 5-5*), Row 16 has a consistency value of 1.0, but the value drops to 0.93 in the fuzzy set truth table (*Table 5-6*), because other cases with lower SMV in the outcome set now have partial membership in Row 16.

STRATEGIES FOR MANAGING CONTRADICTORY TRUTH TABLE ROWS

Although selecting an acceptable row consistency threshold to assign the outcome value to rows depends on the study and its purpose, researchers have several choices determining what to do with contradictory truth table rows (i.e., rows with less than perfect consistency of sufficiency) before proceeding with the next part of the sufficiency analysis. One can (1) include rows with high but perhaps not perfect consistency and exclude rows with low consistency, (2) include all contradictory rows by ignoring consistency entirely, or (3) exclude all contradictory rows by requiring perfect (1.0) consistency (Ragin, 2000; Schneider & Wagemann, 2012).

Including Only Rows With High Consistency

The most commonly used approach for both crisp and fuzzy set analyses and the one we illustrated in the previous section of this chapter, is to include

TABLE 5-6 ■ Truth Table for Fuzzy Set Data for the School Health Features and Academic Performance Example

Row #	MEAL	PE	RECESS	WELL	Number of Cases	Consistency	Outcome Value	Cases
1	0	0	0	0	2	0.5	0	Tahoe, Fletcher
2	0	0	0	1	0	---	?	---
3	0	0	1	0	0	---	?	---
4	0	0	1	1	0	---	?	---
5	0	1	0	0	0	---	?	---
6	0	1	0	0	0	---	?	---
7	0	1	1	0	1	0.87	1	Kruse
8	0	1	1	1	0	---	?	---
9	1	0	0	0	3	0.57	0	Springfield, Tubman, Goodall
10	1	0	0	1	3	0.67	0	Ellis, Harrison, Cardinal
11	1	0	1	0	0	---	?	---
12	1	0	1	1	4	0.86	1	Smithton, Obama, Lincoln, DuBois
13	1	1	0	0	2	0.80	1	Pine, Flora
14	1	1	0	1	3	0.79	0	Lovelace, Parkwood, Grove
15	1	1	1	0	4	0.93	1	Creekside, Curie, Parks, Victory
16	1	1	1	1	2	0.93	1	Westside, Watt

contradictory rows with high consistency. In doing so, the researcher selects a row consistency threshold to determine which row consistency values will be assigned an outcome value of 1 versus 0. A common row consistency threshold for including the row is 0.8. At this threshold, rows with consistency of 0.8 or higher are considered sufficient, and an outcome value of 1 is assigned, whereas rows with a row consistency less than 0.8 are assigned an outcome value of 0. Although 0.8 is a typical row consistency threshold, the researcher can use a lower or higher threshold depending on the context of the configural question and study. One should almost never use a consistency threshold below 0.75 (Ragin, 2008b); if a researcher has a compelling reason for doing so, then she will need to provide a rationale. For example, if someone wanted to examine programs that encourage people to walk more frequently, then perhaps 0.70 would be adequate. In contrast, some configural questions or areas of research may demand near perfect sufficiency relationships for policy action and or decision making, which suggests the use of a 0.9 or 0.95 consistency threshold. For example, if a researcher wanted to assess interventions for keeping people from getting injured, then she might establish a higher consistency threshold for accepting a truth table row as sufficient. In all QCA software packages, the researcher provides the specific row consistency threshold that the software should use, and the software assigns the outcome value to each row of the truth table based on that threshold.

Including All Contradictory Rows

Ignoring consistency entirely and including all contradictory rows is generally not recommended. Nevertheless, the researcher may have an interest in identifying all the potential ways of achieving an outcome, even if some of those ways do not produce the outcome with a high degree of consistency. For example, a researcher may wish to identify all possible combinations of conditions that lead to people contracting a disease, even if the method of contracting the disease does not always or even frequently result in a person getting the disease.

Excluding All Contradictory Rows

Alternately, one could decide to exclude all contradictory rows (i.e., exclude all rows with a consistency value less than 1.0). This is usually an unnecessary precaution; yet, in some instances, a researcher may have an interest in identifying only the combinations that produce the outcome 100% of the time. For example, when assessing airplane flight safety, a researcher might want to accept only combinations that have a perfect (1.0) consistency value.

Reflection

- Think of work in your substantive area. When might a higher row consistency threshold be needed? Why? When might a lower row consistency threshold be considered acceptable? Why?

REVISITING THE DATA TO MANAGE CONTRADICTORY TRUTH TABLE ROWS

In the previous section, we discussed using row consistency thresholds to manage contradictory truth table rows. When possible, Schneider and Wagemann (2010) recommend that researchers think through a few additional strategies to manage contradictory rows before assigning the outcome value to the row. These strategies entail revisiting the data and re-considering some previous analytic choices.

Adding a Condition

First, a researcher can consider adding a condition to the analysis to distinguish the cases that produce the outcome in a row from the cases that do not. For Row 9 of *Table 5-5,* the researcher may try to identify what factor differentiates Tubman from Springfield and Goodall. Tubman has membership in the outcome (SMV = 1); Springfield and Goodall do not (SMV = 0). Say, for instance, an outside organization offers an afterschool program at Tubman which nearly all students participate in and that provides many opportunities for physical activity, but the program is not an official part of the school curriculum (and is, therefore, outside the calibrated definition of offering PE or recess). One could add a condition “having an afterschool play program” as a condition. Adding a condition would move Tubman out of the same row as Springfield and Goodall, and this row would no longer be contradictory.

However, adding conditions to an analysis comes with a trade-off, as doing so increases the number of possible combinations and likely the number of logical remainder rows, which results in greater limited diversity. In this hypothetical example, adding a fifth condition increases the number of possible combinations from 16 to 32. Because the number of cases in this analysis is only 24 schools, the analysis will have a minimum of eight logical remainder rows. Recall from Chapter 3 that generally one should have three to four cases for each condition in the analysis; this, of course, is not a fixed rule, but it can minimize potential issues that arise from having too much limited diversity.

Revisiting Case Selection

A second strategy is to revisit case selection. To do this, a researcher assesses whether all the cases are appropriate to include in the analysis or whether a case is missing from the analysis. If cases are not entirely appropriate, excluding less relevant cases (and providing an appropriate rationale for doing so) strengthens the analysis, even if it means reducing the number of cases. Using the school health features and academic performance example, the researcher may determine that Tubman does not belong in the same analysis as the other schools, if, for instance, its special afterschool program for physical activity makes it too different from the other schools to include in the analysis. Dropping cases also creates a tradeoff as one loses cases for the analysis, but thoughtfully reconsidering the relevance of a case and dropping irrelevant ones may be warranted.

If more cases (and resources for data collection) are available, another strategy is to add cases. This strategy could clarify the relationship of rows with marginal consistency (i.e., close to the row consistency threshold) because an additional case can increase or decrease row consistency and enable the researcher to make decisions based on more information. However, this strategy does not eliminate the contradictory rows. For instance, if the researcher had another school that shared the same truth table row as Tubman, Springfield, and Goodall and that school had a SMV of 0 for the outcome, the consistency would drop to 0.25 (1/4 = 0.25). This would provide more evidence that this row should be assigned an outcome value of 0. When adding a case to manage contradictory rows, one would prioritize adding cases to contradictory rows with a row consistency value that is close to the selected row consistency threshold; this strategy would reduce contradiction in the row by increasing (or decreasing) the row consistency and its relationship to the selected row consistency threshold (Rihoux & Ragin, 2009). Prioritizing cases for rows with a consistency value close to the row consistency threshold aids in corroborating the extant data in the analysis and in establishing reliability.

Revisiting the Definition and Calibration of the Conditions and Outcomes

Finally, a researcher can review how she defined and calibrated the condition and outcome sets and consider re-defining one or more of them. Deciding to recalibrate or redefine a condition or outcome should be grounded in substantive knowledge or a deep understanding of the cases. For example, say the researcher knows from interviews that Tubman has an extensive afterschool playtime program that nearly all students participate in. Because of that afterschool program, the school board decided that less recess and PE time were needed. Thus, most of the Tubman students have a lot of physical activity and unstructured play time, but not within the context of the formal school curriculum. With this knowledge, the researcher may decide to expand her definition of the "RECESS" condition to include afterschool programs. Re-defining and recalibrating RECESS would give Tubman the combination of having healthy meals and not having physical education and having recess and not having a comprehensive wellness program (MEAL*~PE*RECESS*~WELL). Recalibrating RECESS would move Tubman out of the row it shares with Springfield and Goodall (Row 9) and move Tubman into Row 11. As discussed in Chapter 4, one should document such decisions in the calibration rubric.

After inspecting the truth table to establish a row consistency threshold and determine whether one should make different decisions about conditions, outcomes, and case selection, one may need to update the data matrix, create a new truth table, and inspect the truth table once more. Before moving to the next part of the analysis, one examines the truth table for a few additional potential issues:

clustering of cases in a few rows (i.e., limited diversity) and unlikely combinations of conditions. We discuss these issues in the next section.

Reflection

- What may foster or limit a researcher from adding cases to manage contradictory rows?
- What situations might foster or limit a research from redefining the condition and outcome definitions and calibrations?

INSPECT THE TRUTH TABLE FOR POTENTIAL ISSUES

As the previous sections suggest, a QCA should never be mechanistic. Good practice involves carefully considering the study purpose when selecting the row consistency threshold and reviewing one's analytic choices to manage contradictory truth table rows before beginning an analysis. Researchers should also implement two additional quality checks on the truth table before proceeding in the analysis. First, one should examine whether most of the cases appear in only a few rows (i.e., clustering), and second, one should assess whether some rows represent combinations of conditions that appear to be unlikely combinations (and whether cases appear in those rows).

In the first instance, if cases cluster into only a few rows, this signals a high amount of limited diversity. This clustering indicates that all cases appear to be nearly identical with regards to the conditions selected for the analysis. A high amount of limited diversity may point to a problem with case or condition selection or calibration. Managing this challenge depends upon the underlying problem. If a researcher selected homogeneous cases, then she will want to add other cases to the analysis. One would recognize that the cases are homogeneous if they have the same combination and consistently produce the outcome (i.e., row consistency is 1.0 or very close to it) or fail to produce the outcome (i.e., row consistency is 0.0 or very close to it).

Alternately, if one determines that the apparently identical cases produce a contradictory truth table row, then a researcher should rethink the choice of conditions or the calibration of the conditions. A contradictory truth table row may suggest that something differentiates the cases, but the researcher has not yet captured it in the selection of conditions or condition calibration. The earlier example using Row 9 of the truth table and differentiating Tubman from Goodall and Springfield illustrates this insight. Those three schools appear identical on MEAL, PE, RECESS, and WELL, whereas Tubman demonstrates the outcome, the other two do not. Re-defining or adding a condition differentiated cases that

BOX 5-3 Example from the Field: Using Mixed Methods to Learn from Unlikely Combinations

Anater, A., Chambard, M., Council, M., Emery, K., Kandefer, S., Kane, H., . . . Zaccaro, D. (2014). *Understanding the interdependencies among three types of coping strategies used by very low food security households with children*. Report for University of Kentucky, Center for Poverty Research. Research Program on Childhood Hunger—Large Grants Program.

Dr. Anater and her colleagues used an explanatory sequential mixed method study design to explore how households with very low food security (VLFS) and with children used multiple coping strategies to manage not having enough food. These coping strategies included participating in federal feeding and financial assistance programs (i.e., safety net programs), obtaining food from charitable or other non-governmental food providers (e.g., food pantries), and using individually developed food acquisition strategies. The study also examined how individuals in VLFS households make decisions around how and when to use particular strategies.

The research team administered 320 surveys with clients of food providers who offer assistance to people with limited resources within eight North Carolina counties with the highest food insecurity rates. The survey captured sociodemographic information, the number of adults and children in the household, household food security level, and coping strategies used in the past 30 days. Using a longitudinal design, more than a year later, the team completed in-depth interviews with a cohort (n = 28) of survey participants. The team purposively sampled individuals from households with VLFS to participate in interviews to discuss how they made decisions to use coping strategies.

At the completion of the data collection, the team developed a truth table to examine the household characteristics of the 28 interview and survey respondents using variables that are commonly associated with food security. The conditions included (1) household has any adults employed full-time, (2) household has single head of household, (3) household receives Women, Infants, and Children (WIC) benefits, and (4) household includes a young child (age 0–5 years). The truth table included cases in truth table rows that represented unlikely combinations of conditions in the opinion of the researchers. In three rows, several households without children under 5 also received WIC benefits. Generally, WIC benefits are for low-income households with children age 5 and under. Although some of the survey respondents could have been pregnant women who would be WIC eligible, the study team found those rows unusual and worth exploring.

The team ruled out coding errors, as the data had already been assessed for quality. The study team then examined the interview data to understand why cases like this could exist. From the interviews, the team learned that some individuals in households with VLFS manage not having enough food by having children stay part-time with grandparents who can provide meals. Thus, some survey respondents, who received WIC and did not report having children under 5 in the household, may not have counted their child as a full-time member of the household because the child stayed more than 50% of the time (e.g., during the work day) with the grandparents. By examining the unlikely combinations of conditions, the team learned how using kinship networks enabled people to cope with food insecurity.

were apparently identical in the truth table but were not so similar when one revisited the data from the cases.

Next, one assesses whether the truth table contains unlikely combinations. Unlikely combinations are combinations of conditions that are nearly impossible to occur, such as having a school without teachers or an illiterate college professor. When those rows have no cases, then those rows should be dealt with as untenable assumptions, which is discussed at length in Chapter 6. Finding cases in such rows may point to a coding error in the data; in that circumstance, the researcher should return to the data for a quality check. However, if the data have been entered correctly, then such cases may yield special insights that can be explored in a mixed methods study. An explanatory sequential design can readily support assessment of unlikely combinations of conditions. If a researcher examines combinations of characteristics among cases after quantitative data collection and identifies cases that are unlikely to exist, she can use qualitative data collection, such as interviews, to explore the conditions under which such unlikely combinations may exist. *Box 5-3* provides an example from the field of how unlikely combinations in one explanatory sequential, mixed methods study, led to new insights about the data.

Once the researcher decides that she has done everything feasible for resolving difficulties in the truth table given the constraints of data and resources, then she can move to the analysis. In the analysis, the first steps involve conducting separate analyses for necessary conditions followed by the analysis of sufficient conditions. Schneider and Wagemann (2010) recommend conducting the analysis of necessary conditions before the analysis of sufficient conditions, because inferring the presence or lack of necessary conditions from the sufficient solution terms can lead to incorrect conclusions. The next section describes the analysis of necessary conditions.

CONDUCT AN ANALYSIS OF NECESSARY CONDITIONS AND COMBINATIONS OF CONDITIONS

Chapter 2 explained that a necessary condition or combination of conditions must be present for the outcome to occur; its absence guarantees the outcome will not occur. Thus, a necessary condition implies that if the outcome is present (or present at a high level), then the condition is also present (or present at a high level). Thus, we can infer a necessary condition from an analysis when the necessary condition is in a superset relationship to the outcome set.

When analyzing data to identify necessary conditions, one looks for set relationships between condition sets (and condition complement sets) and the outcome set (and outcome complement set). Recall that the rules of logic dictate that cases that are nonmembers of a condition set are members of the condition's complement set. Similarly, if a case is a member of the outcome set, then it is a nonmember of the outcome's complement set. As a result, four possible relationships of necessity could exist, where X can be an individual condition or combination of conditions. These are listed in *Box 5-4*.

In addition to looking for necessary set relationships between individual conditions and the outcome set, one should also assess for set relationships between any necessary combinations of conditions that one deems theoretically relevant to the outcome. However, the rules of logic dictate that the only way a combination of conditions can be necessary is if both conditions are individually necessary. Thus, in practice, if at least two conditions are not individually necessary, it is not possible to have any necessary combination of conditions.

Although one can use a visual inspection of the truth table by looking at all cases where the outcome occurs for determining individual necessary conditions, this step typically uses software. Recall from Chapter 2 that for crisp sets, the consistency of a necessary condition takes on a value from 0 to 1 and can be determined by simple analysis of 2X2 tables. To calculate consistency, we divided

BOX 5-4 Assessing Necessary Conditions and Combinations of Conditions

1. Membership in a condition set may be necessary for membership in the outcome set (X ←Y).
2. Nonmembership in a condition set (i.e., membership in the condition set complement) may be necessary for membership in the outcome set (~X←Y).
3. Membership in a condition set may be necessary for membership in the complement of the outcome set (X←~Y).
4. Nonmembership in a condition (i.e., membership in the condition set complement) may be necessary for membership in the complement of the outcome set (~X←~Y).

the number of cases with membership in both the condition and outcome set by the number of cases with membership in the outcome set. The upper panel of *Box 5-5* provides the equations for calculating consistency of a necessary condition from the condition set SMV (X in the formula) and outcome set SMV (Y in the formula)—this formula applies to both crisp and fuzzy sets. Experts recommend using a consistency threshold of at least 0.9 for establishing relationships of necessity (Ragin, 2008b; Schneider, 2018).

If a condition has high enough consistency to be considered necessary, the next step is to evaluate its coverage and assess for trivialness. In some circumstances, a condition that is necessary based on a high consistency value may, in fact, be empirically irrelevant or a **trivial necessary condition**. Coverage of a necessary condition captures the degree to which a necessary condition is empirically relevant (i.e., observed in the data) (Ragin, 2006). The middle panel of *Box 5-5* provides the formula for coverage developed by Ragin, which is used by most software programs. Values closer to 1 indicate that a necessary condition is highly empirically relevant. Values closer to 0 indicate that a necessary condition is not relevant (i.e., observed infrequently in the data).

BOX 5-5 Equations for Calculating Consistency and Coverage of Necessary Conditions

Consistency of a necessary condition

$$\frac{\sum_{i=1}^{I} \mathrm{min}(Xi,Yi)}{\sum_{i=1}^{I} Yi}$$

Coverage of a necessary condition (Ragin, 2006)

$$\frac{\sum_{i=1}^{\mathrm{I}} \mathrm{min}(Xi,Yi)}{\sum_{i=1}^{\mathrm{I}} Xi}$$

Relevance of necessity (Schneider & Wagemann, 2012)

$$\frac{\sum_{i=1}^{\mathrm{I}} (1-Xi)}{\sum_{i=1}^{\mathrm{I}} (1-\mathrm{min}(Xi,=Yi))}$$

***Xi* indicates the condition SMV for each case**

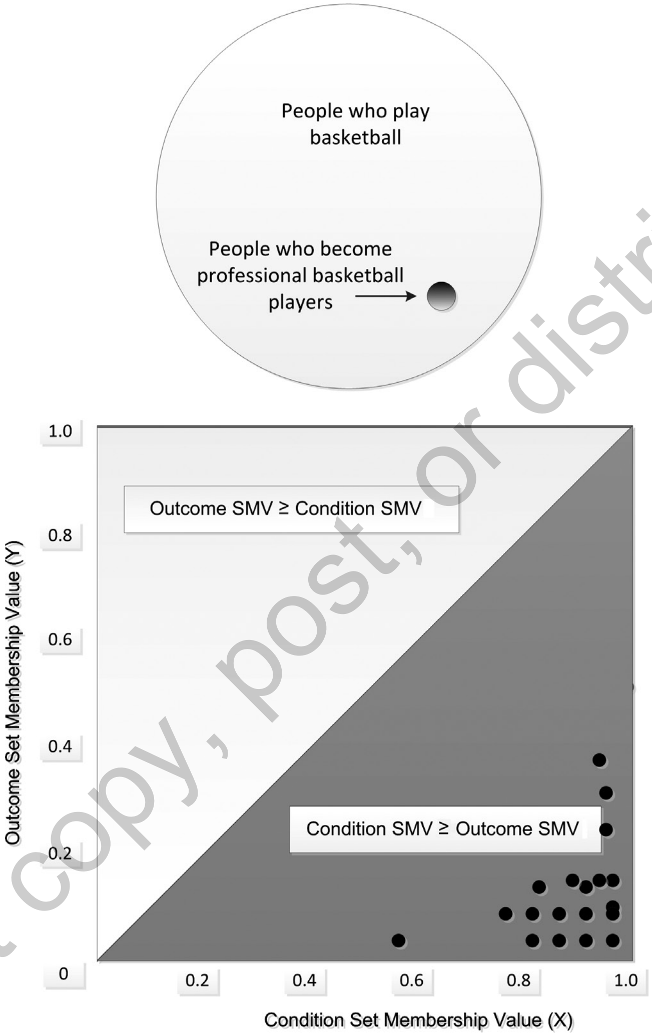
***Yi* indicates the outcome SMV for each case**

A necessary condition is trivial when either the outcome represents a very small subset of the condition as illustrated in *Figure 5-2,* or the outcome and conditions represent very large sets and are nearly "constants" as illustrated in *Figure 5-3*. In the first circumstance, the outcome is a rare event (X>>>Y) that renders the necessary relationship virtually unimportant as a finding or for policy relevance; the SMV of the cases with the condition is almost always 1 or very near 1, regardless of the SMV of the outcome. For example, playing basketball is a necessary condition for becoming a professional basketball player; although one must play basketball to become a professional player, the size of the set of professional athletes relative to the size of the set of people who play basketball is very small. The upper panel of *Figure 5-2* provides a Venn diagram of this example, which is based on crisp sets. The lower panel of *Figure 5-2* depicts how this trivial necessary condition would appear on an X-Y plot using fuzzy sets. The Ragin formula for coverage in this instance will be low, correctly identifying this condition as a trivial (i.e., irrelevant) necessary condition.

The second type of trivial necessary condition results when the condition is a common contextual factor that is nearly always present in the background such that there is almost no instance of its absence (X and Y >>>~X). For example, identifying "laws requiring school attendance" as necessary for membership in the outcome set of students with high school attendance would be silly because nearly all children are subject to school attendance laws; although a superset relationship exists (i.e., high consistency), this condition is a trivial necessary condition for high student attendance. *Figure 5-3* provides a Venn diagram of this example for crisp sets, as well as shows how a trivial necessary condition would appear on an X–Y plot with fuzzy SMVs in the lower panel. For this second type of trivialness, the Ragin formula for coverage is not ideal as it overestimates the empirical relevance of the necessity relationship. Thus, Schneider and Wagemann (2012) proposed a formula for assessing coverage and the second type of trivialness, referred to as the "relevance of necessity" (lower panel of *Box 5-5*). Like Ragin's coverage parameter, the relevance of necessity values also range from 0 to 1, with 0 indicating irrelevance (i.e., trivialness) and 1 indicating high relevance. The relevance of necessity parameter can better diagnose the circumstance where the condition is close to a constant; it will not generate an artificially high coverage value. Currently, the relevance of necessity calculation is only available in the QCA and SetMethods R packages.

When interpreting whether a condition is necessary, one should assess the consistency, coverage, and relevance of necessity values. In general, one should use a high consistency threshold for identifying a condition (or combination) as necessary, such as 0.9 consistency (Ragin, 2006; Schneider & Wagemann, 2012). The coverage and relevance of necessity values should also be high, but the exact value is study dependent. Although one can use thresholds below 0.9 consistency and lower coverage and relevance of necessity values, doing so will depend on providing a good theoretical rationale. In addition to assessing whether a condition

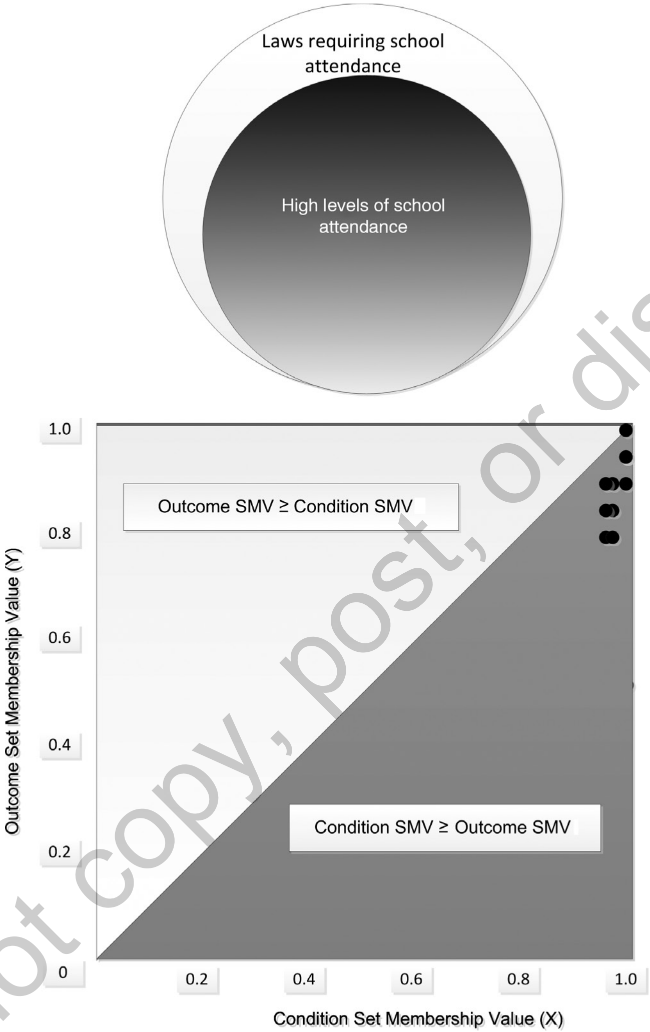
FIGURE 5-2 ■ Trivial Necessary Condition–Outcome Is a Small Subset of the Condition



The top panel depicts a trivial necessary condition with crisp sets, and the bottom panel depicts a trivial necessary condition with fuzzy sets.

is necessary, one should also confirm that the identified necessary conditions are not logically inconsistent. First, the same condition cannot be necessary for both the outcome and the outcome complement, and second, the condition and its complement cannot both be a necessary condition for the outcome. For the

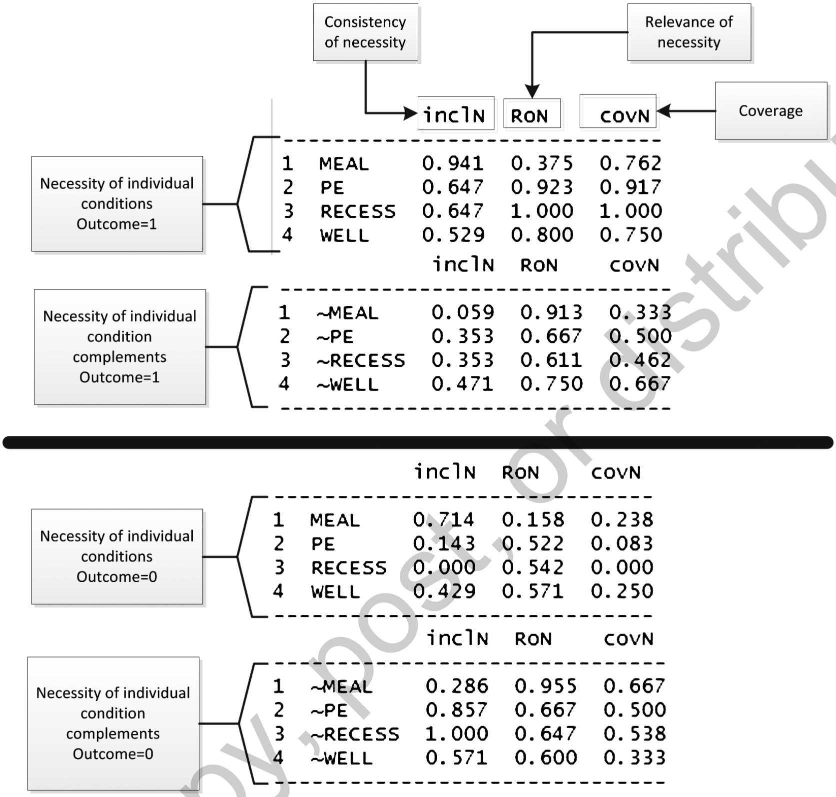
FIGURE 5-3 ■ Trivial Necessary Condition–Outcome and Condition Set Are Much Bigger Than the Condition Complement



The top panel depicts a trivial necessary condition with crisp sets, and the bottom panel depicts a trivial necessary condition with fuzzy sets.

first case, the one cannot logically say that serving healthy meals is necessary for adequate academic performance and also say serving meals is necessary for *not* having adequate academic performance (i.e, MEAL←OUTCOME and MEAL← ~ OUTCOME). For the second case, one cannot say that having recess

FIGURE 5-4 ■ Output for Necessity Tests for the School Health Features and Academic Performance Example (Crisp Set)



```
              inclN  RoN    covN
-------------------------------
1  MEAL       0.941  0.375  0.762
2  PE         0.647  0.923  0.917
3  RECESS     0.647  1.000  1.000
4  WELL       0.529  0.800  0.750
               inclN  RoN    covN
-------------------------------
1  ~MEAL       0.059  0.913  0.333
2  ~PE         0.353  0.667  0.500
3  ~RECESS     0.353  0.611  0.462
4  ~WELL       0.471  0.750  0.667
-------------------------------
```

```
              inclN  RoN    covN
-------------------------------
1  MEAL       0.714  0.158  0.238
2  PE         0.143  0.522  0.083
3  RECESS     0.000  0.542  0.000
4  WELL       0.429  0.571  0.250
-------------------------------
               inclN  RoN    covN
-------------------------------
1  ~MEAL       0.286  0.955  0.667
2  ~PE         0.857  0.667  0.500
3  ~RECESS     1.000  0.647  0.538
4  ~WELL       0.571  0.600  0.333
-------------------------------
```

Abbreviations: "InclN" refers to consistency of necessity; "RoN" refers to relevance of necessity, which is based on the Schneider and Wagemann equation in *Box 5-5*. The "covN" refers to coverage of the necessary condition, which is based on the Ragin equation in *Box 5-5*.

is necessary for adequate academic performance, and *not* having recess is necessary for having adequate academic performance (i.e., RECESS←OUTCOME and ~RECESS←OUTCOME).

Returning to the school health features and academic performance example, we conducted all four necessity tests using the QCA package in R (Dusa, 2017b) and provide the output in *Figure 5-4*. In reviewing the output, one might consider MEAL as a necessary condition for the outcome or ~RECESS as a necessary condition for the complement of the outcome. Both conditions have high consistency of necessity (labeled as "incl" on the output) values, but when one reviews the coverage and relevance of necessity (RoN), those values are low. Thus, neither should be identified as a necessary condition.

One might be tempted to infer a necessary condition if one saw a condition as a part of every sufficient solution term (e.g., ABC + AD + AE→OUTCOME, Condition A would appear necessary). However, making that inference can be faulty when one has less than perfect consistency of sufficiency. When one has contradictory truth table rows (i.e., less than perfect consistency of sufficiency), "false" necessary conditions can appear, or true necessary conditions disappear in the solution. For this reason, the analysis of necessary conditions must be conducted first and separately from the analysis of sufficient conditions. Once the analysis of necessary conditions and combinations is complete, one can conduct the analysis of sufficient conditions and combinations, described in the next section.

CONDUCT ANALYSIS OF SUFFICIENT CONDITIONS AND COMBINATIONS OF CONDITIONS

This section explains how to conduct the analysis for sufficient conditions and combinations, sometimes referred to as the truth table analysis. This analytic step entails minimizing the truth table to identify sufficient conditions or combinations of conditions for the outcome referred to as a **solution**. Individual solution terms comprise the solution. (The distinction between solution and solution terms is explained and illustrated below.) When minimizing the truth table, a researcher will make additional decisions about the logical remainder rows, which will generate different—but logically consistent—solutions. This section will describe the process of minimizing the truth table and the relationships among the different solution terms that are produced in the process.

Each truth table row with a high-level of consistency is a sufficient combination, and combining all sufficient combinations with an "OR" represents the most complex solution possible. However, including each truth table row as a term in the solution would create a very complex solution and primarily would merely describe the cases, rather than identify the most salient combinations of conditions for the outcome. Thus, truth table analysis involves simplifying those complex solution terms (i.e., sufficient rows) into fewer solution terms with a smaller number of conditions (if possible). The process for generating fewer, less complex solutions terms is known as **logical minimization,** sometimes referred to as Boolean reduction or minimization of the truth table.

In the minimization process, the software uses an algorithm designed to systematically pair rows with an outcome value of 1 and determine whether a condition can be eliminated or "reduced" out of the combination. Most software designed to conduct QCA uses the Quine–McCluskey algorithm (or an enhanced version of this algorithm) to perform the logical minimization process. We illustrate this process with an abridged example from the full truth table of our school health features and academic performance example (*Table 5-5*); in the

TABLE 5-7 ■ Abridged Truth Table From School Health Features and Academic Performance Example (Crisp Sets)

Row #	MEAL	PE	RECESS	WELL	Number of Cases	Consistency	Outcome Value	Cases
15	1	1	1	0	4	1.00	1	Creekside, Curie, Parks, Victory
16	1	1	1	1	2	1.00	1	Westside, Watt

abridged version (*Table 5-7*), we have extracted Rows 15 and 16 from the full truth table. Both rows have an outcome SMV of 1.

Comparing Rows 15 and 16, the software would recognize that the only difference between Row 15 and Row 16 is whether the school has a comprehensive wellness policy. Cases (i.e., schools) in both rows have healthy meals (MEAL) and offer physical education classes (PE) and have recess (RECESS). Cases in row 15 have a comprehensive wellness policy (WELL); cases in row 16 do not (~WELL). In those two rows and cases, having a wellness policy did not matter for whether the case has the outcome. The software would then minimize or reduce by "cancelling out" WELL and ~WELL. The minimized combination would be "having healthy meals AND offering physical education classes AND having recess" (MEAL*PE*RECESS). The reduced combination of Rows 15 and 16 is sufficient for the outcome. The computer would continue this process with the other rows and may also pair rows 15 and 16 with other rows to reduce the rows even further. Doing this process by hand without errors is difficult, and software is essential for this step. We refer the reader to the guide by Thomann, Oana, & Wittwer (2018) to learn how to conduct this analysis in R (see end of this book for additional software resources).

When conducting the truth table minimization, one generates three solutions: (1) the conservative solution, (2) the parsimonious solution, and (3) the intermediate solution. These different solutions pertain to how the software handles the logical remainder rows (i.e., the rows without cases that have "?"s for outcome value). Of course, if a truth table has no logical remainders, then one does not need to generate all three solutions because they will be the same.

Conservative Solution

When generating the **conservative solution**, the researcher configures the software to ignore all logical remainder rows. The truth table is minimized using only those rows with cases that have outcome values equal to 1, which are the rows that the researcher has deemed sufficient based on a consistency value at or above the selected row consistency threshold. Some QCA textbooks may refer to this solution as the *complex* solution. Others prefer conservative, as any of the three solutions can be complex in terms of number of conditions and operators included.

Parsimonious Solution

When generating the **parsimonious solution**, the researcher configures the software to use the logical remainder rows in whatever way it can to achieve the fewest terms (i.e., conditions and operators) in the solution. A logical remainder used by the software to minimize the truth table is called a **simplifying assumption**; it is simplifying because it usually helps to create a less complex solution (i.e., fewer conditions or operators). It is an assumption because using a row without any cases requires an assumption about whether hypothetical cases that would belong to the row would have membership in the outcome set. If assuming membership in the outcome set produces a simpler solution, then the software will make that assumption. Alternatively, if assuming nonmembership in the outcome set yields a simpler solution, then the software will make that assumption. Thus, parsimony drives the assumptions made during minimization to generate the most parsimonious solution. This has advantages and consequences which we will explore in further detail in Chapter 6.

Intermediate Solution

Finally, for the **intermediate solution**, the researcher uses theory to guide the software about handling the logical remainders during minimization. The researcher provides **directional expectations** for each condition in the analysis; the precise way in which this is done varies by software package. A directional expectation indicates whether the condition theoretically should or should not contribute to a case having membership in the outcome set. One typically defines expectations in the calibration rubric (Chapter 4). Alternatively, a researcher can leave a directional expectation unspecified for one or more conditions in the analysis.

For instance, in the school health features and academic performance example, the calibration rubric indicated that one might expect that serving healthy meals, providing children with regular physical education, offering regular recess, and having a comprehensive wellness policy would enable students to perform better academically. Once directional expectations are set by the researcher within the software, minimization proceeds including only those logical remainder rows that result in a simpler solution and that are consistent with directional expectations. Thus, the intermediate solution prioritizes consistency with directional expectations over parsimony to arrive at a simpler solution. In the intermediate solution, the minimization process may not use all the simplifying assumptions used to generate the parsimonious solution because some of the simplifying assumptions may not be consistent with the directional expectations set by the

PRACTICE TIP 5-2

DOCUMENTING SIMPLIFYING ASSUMPTIONS JOURNAL MANUSCRIPTS

QCA experts recommend documenting the simplifying assumptions used in all analyses for transparency. However, the ability to easily do this varies by software packages (Ragin & Rihoux, 2004; Schneider & Wagemann, 2010).

researcher. It is possible for the parsimonious solution and intermediate solution to be identical, which would mean the same simplifying assumptions were used to generate both solutions.

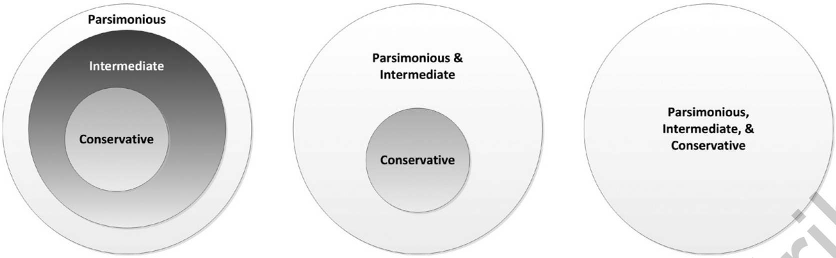
Although the three solutions that are generated will result in somewhat different solution terms, none will contradict the empirical case information because they are generated from the same truth table. Thus, the three solutions share a logically consistent relationship. The conservative solution is a subset of the intermediate solution, and the intermediate solution is a subset of the parsimonious solution as illustrated in *Figure 5-5*. The left image in the figure shows a typical relationship among the solutions. In this image, the parsimonious and intermediate solutions did not use the same logical remainder rows because some of the logical remainder rows were not in line with directional expectations. As before, the conservative solution used none of the logical remainder rows. In the center image, the parsimonious and intermediate solutions used the same logical remainder rows (i.e., the simplifying assumptions were all in line with directional expectations). The image on the right shows that the conservative, intermediate, and parsimonious solutions are identical because the truth table was fully specified, meaning it had cases in each row and thus had no logical remainder rows.

The researcher is free to interpret and report any one of the three solutions generated although many reports we have seen rely on the intermediate solution. However, when disseminating findings from an analysis, all three solutions should be made available to readers for transparency (Schneider & Wagemann, 2010). We will discuss this aspect of interpretation further in Chapter 7.

Now, we return to the school health features and academic performance example and examine the solutions that we generated. In the figures that follow, we provide the output from the R QCA package that we used to conduct the truth table analysis (Dusa, 2017b). We provide the conservative, parsimonious, and intermediate solutions in *Figure 5-6*. In these figures, we will focus on only the solution terms, which are annotated; we will cover the other output values in Chapters 6 and 7.

For clarity of terminology, we refer to the individual complex sets within a solution as a solution term and refer to all of the solution terms together as the solution. The solution terms represent two or more truth table rows that

FIGURE 5-5 ■ Set Relationships Among the Conservative, Intermediate, and Parsimonious Solutions



Left figure: Conservative solution is a subset of the intermediate solution, which is a subset of the parsimonious solution.

Center figure: Conservative solution is a subset of intermediate and parsimonious solutions; the simplifying assumptions were in line with the directional expectations.

Right figure: The solutions are identical because the truth table is fully specified (i.e., no logical remainder rows).

have been minimized to their simplest form. For example, in the conservative solution, MEAL*PE*~WELL is a solution term. MEAL*PE*~WELL + MEAL*RECESS*WELL + PE*RECESS*~WELL is the solution.

In comparing the solutions, one can see that the parsimonious solution terms are super sets of the intermediate solution terms. For example, RECESS is a super set of PE*RECESS and MEAL*RECESS*WELL, and PE*~WELL is a super set of MEAL*PE*~WELL. Two of the intermediate solution terms are identical to the conservative solution terms, and one of the intermediate terms (PE*RECESS) is a super set of the conservative solution terms (PE*RECESS*~WELL). These superset and subset relationships demonstrate that the solutions are logically consistent with each other.

Figure 5-7 provides a Venn diagram depicting the set relationships for this example. The 13 schools indicated with an **X** are schools that were in sufficient truth table rows that were used in the minimization process to generate the solution. The four schools indicated with a O are schools that had membership in the outcome set but were located in contradictory truth table rows not used in the minimization process because these rows had a row consistency value below our prespecified threshold (i.e., the truth table rows showed a weak sufficiency relationship with the outcome). Thus, these four schools are not "covered" by the solution terms. In Chapter 7, we will discuss ways of exploring these kinds of cases.

FIGURE 5-6 ■ Solutions for the School Health Features and Academic Performance Example (Crisp Sets)

Conservative Solution

Solution

Solution Terms and Cases Covered by Each Term

```
M1: MEAL*PE*~WELL + MEAL*RECESS*WELL + PE*RECESS*~WELL => OUT

                         inclS  PRI    covS   covU   cases
 ------------------------------------------------------------------------------------------
 1  MEAL*PE*~WELL        1.000  1.000  0.353  0.118  Pine,Flora; Creekside,Curie,Parks,Victory
 2  MEAL*RECESS*WELL     1.000  1.000  0.353  0.353  Smithton,Obama,Lincoln,DuBois; Westside,Watt
 3  PE*RECESS*~WELL      1.000  1.000  0.294  0.059  Kruse; Creekside,Curie,Parks,Victory
 ------------------------------------------------------------------------------------------
    M1                   1.000  1.000  0.765
```

Parsimonious Solution

```
M1: RECESS + PE*~WELL => OUT

             inclS  PRI    covS   covU   cases
----------------------------------------------------------------------------------------------------------------------
1  RECESS    1.000  1.000  0.647  0.353  Kruse; Smithton,Obama,Lincoln,DuBois; Creekside,Curie,Parks,Victory; Westside,Watt
2  PE*~WELL  1.000  1.000  0.412  0.118  Kruse; Pine,Flora; Creekside,Curie,Parks,Victory
----------------------------------------------------------------------------------------------------------------------
   M1        1.000  1.000  0.765
```

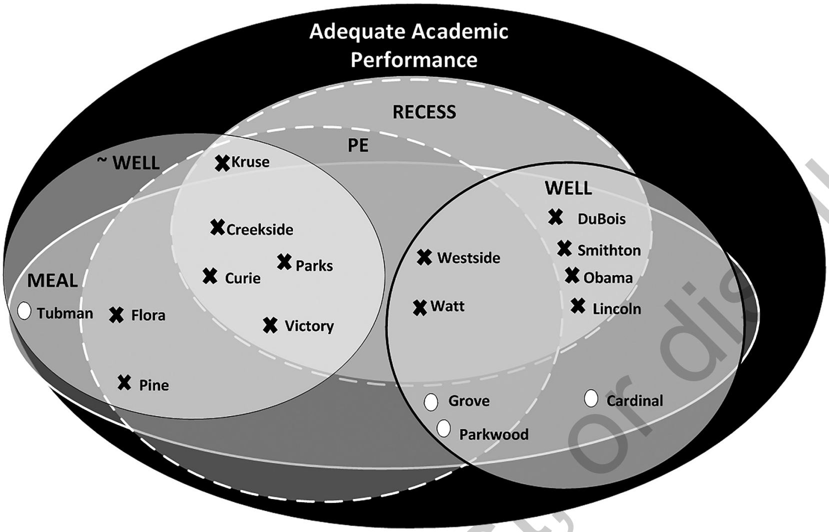
Intermediate Solution

```
M1:    PE*RECESS + MEAL*PE*~WELL + MEAL*RECESS*WELL => OUT

                      inclS  PRI    covS   covU   cases
-----------------------------------------------------------------------------------------------
1  PE*RECESS          1.000  1.000  0.412  0.059  Kruse; Creekside,Curie,Parks,Victory; Westside,Watt
2  MEAL*PE*~WELL      1.000  1.000  0.353  0.118  Pine,Flora; Creekside,Curie,Parks,Victory
3  MEAL*RECESS*WELL   1.000  1.000  0.353  0.235  Smithton,Obama,Lincoln,DuBois; Westside,Watt
-----------------------------------------------------------------------------------------------
   M1                 1.000  1.000  0.765
```

Abbreviations: M1 = solution; inclS = consistency of sufficiency; PRI = Proportional Reduction in Inconsistency; covS = coverage of sufficiency; covU = unique coverage

FIGURE 5-7 ■ Venn Diagram Depicting Set Relationships Among the Solution Terms for the School Health Features and Academic Performance Example



X Indicates a case (school) covered by one or more solution terms

O Indicates a case (school) that is *not* covered by any solution terms but is in the outcome set.

Reflection

- Transparency is key for reporting QCA results. What analytic decisions should the researcher report for transparency?
- What are the benefits of being transparent?

CONCLUSIONS

This chapter explained the initial analysis, including how to transform the raw data into a truth table, how to assess the truth table before conducting an analysis, how to assess necessity, and how to generate the conservative, parsimonious, and intermediate solutions. We provide a checklist in *Box 5-6* to summarize these initial analysis steps presented in this chapter. In the next chapter, we will explain how to conduct model analytics of these initial analyses to assess parameters of fit and test the logical consistency of one's analytic decisions and robustness of the solutions generated.

BOX 5-6 Checklist of Steps for Analysis—Initial Analysis

1. Create the truth table from the data matrix.
    a. Create a truth table shell.
    b. Assign cases from data matrix to truth table row.
    c. Assign an outcome value to each truth table row.
2. Inspect the truth table.
    a. Examine where the cases fall (e.g., do all or most fall into just a few rows?).
    b. Examine the consistency values of each row.
    c. Assign an outcome value to each row based on a row consistency threshold.
    d. Try to resolve contradictory truth table rows.
    e. Check for coding errors if cases appear in unlikely truth table rows.
3. Test for individual necessary conditions (and combinations of necessary conditions if applicable).
    a. Examine for necessity of both the condition and its complement with the outcome and its complement (i.e., four tests of necessity)
    b. Consider conditions necessary only when they surpass a consistency value of 0.9 and have high coverage and relevance of necessary parameters.
4. Minimize the truth table with software to generate the conservative, parsimonious, and intermediate solutions.

Summary and Key Points

This chapter explained how to transform the data matrix into a truth table, review the truth table, and conduct an analysis of necessary and sufficient conditions and combinations of conditions. The key points for this chapter are as follows:

- Transforming a data matrix involves three steps: (1) creating a truth table shell, (2) taking the data matrix (i.e., the raw data) and assigning cases to truth table rows, and (3) assigning an outcome value to each truth table row based on a selected row consistency threshold.

- Strategies for managing contradictory truth table rows include adding a condition, revisiting case selection, and revisiting the definition and calibration of the conditions and outcomes.
- One should inspect the truth table for limited diversity and unlikely combinations of conditions (i.e., configurations or truth table rows).
- One should conduct an analysis of necessary and sufficient conditions and combinations separately. The analysis of necessary conditions should precede sufficiency analyses.
- When a truth table has logical remainder rows, the logical minimization involves generating three solutions: (1) the conservative solution, (2) the most parsimonious solution, and (3) the intermediate solution. These solutions differ with respect to how the software handles the rows without empirical information (i.e., the logical remainder rows).
- A logical remainder used to reduce the truth table is called a simplifying assumption.
- One should use software to create the truth table and conduct all analyses.

Supplementary Digital Content

Using the datasets provided at the book's companion website, readers can use the checklist provided in *Box 5-6* to practice the process of creating truth tables, conducting an analysis of necessary and sufficient conditions, and generating the conservative, parsimonious, and intermediate solutions. Please visit **study.sagepub.com/researchmethods/mixed-methods/kahwati**