Measures of Fit

Patrick A. Mello
Willy Brandt School of Public Policy
University of Erfurt

Version: July 2019

Citation: Mello, Patrick A. (2019) "Measures of Fit." In: *Qualitative Comparative Analysis: Research Design and Application.* Washington, D.C.: Georgetown University Press (under contract).

set relations are important […] in the same way that assessments of significance and strength are important in the analysis of correlational connections.

Charles Ragin [1]

In the course of its development over the past 30 years, QCA has undergone substantial methodological sophistication. As a "Boolean approach" the method initially only worked with binary values and did not allow for "contradictory" truth table rows where some cases show the outcome and others do not – such truth table configurations needed to be resolved through measures of research design, before one could proceed with the analysis (Ragin et al. 1984; Rihoux and De Meur 2009; Rihoux and Ragin 2009).

However, perfect set relations can rarely be found in the social sciences. More often one may identify a relationship that approximates necessity or sufficiency, but there may always be cases in the empirical data that do not fit such a pattern. How to

[1] Ragin (2008, 45).

proceed under such circumstances? What proportion of cases will be "enough" to merit a set-theoretic relationship?

Hence, following the introduction of fuzzy sets (Ragin 2000), which allowed for differentiated degrees of set membership, Ragin (2006a) put forth the measures of fit "consistency" and "coverage" to assess imperfect set-theoretic relationships in empirical data. These were a major step forward in the development of QCA, fostering the evolution of transparent standards for replicable research.[2]

Consistency and coverage can be compared to the well-known statistical indicators of significance and strength, as highlighted in the introductory quote. Similar to statistical significance, *consistency* measures the degree to which an empirical relationship between a condition or combination of conditions and the outcome approximates set-theoretic necessity and/or sufficiency. Similar to statistical strength, *coverage* measures the empirical importance or relevance of a condition or combination of conditions (Ragin 2008, 45).

While consistency and coverage remain the principal indicators to formally assess set-theoretic relationships, over the years it became clear that these measures could not detect certain issues that might occasionally arise in empirical data. Hence researchers developed additional measures like the "proportional reduction in inconsistency" (PRI) to address simultaneous subset relations (Ragin 2006b; Schneider and Wagemann 2012), "relevance of necessity" (RoN) to further distinguish trivial from relevant necessary conditions (Schneider and Wagemann 2012),[3] and an alternative consistency measure to remedy an imbalance in the standard formula for the consistency of sufficient conditions (Haesebrouck 2015).

This chapter introduces these measures of fit, explains how they are calculated, and provides illustrative examples to demonstrate the issues at stake. Of course, for a standard application of QCA, users do not have to *manually* calculate these measures – that is what the software does – but users will benefit from going through

---

2 Following the increased popularity and usage of the method, Schneider and Wagemann (2010) were the first to formulate "standards of good practice" for QCA. Mello (2013) examines the extent to which empirical applications adhere to suggested guidelines.

3 On the methodology of necessary conditions, see also Braumoeller and Goertz (2000); Goertz (2006); Goertz and Starr (2003).

the calculations to understand how different scores are derived and how these should be interpreted.

Before proceeding, one caveat is in order. The measures of fit discussed in this chapter can help to identify set-theoretic relationships in empirical data, but any interpretation, particularly a *causal interpretation* of the identified patterns must always be grounded in theory and substantive knowledge (Ragin 2008, 54; Schneider and Wagemann 2012). Like the old adage, "correlation is not causation", we should acknowledge that "set relation is not causation".[4] Whether identified patterns are meaningful always depends on the design of a study, the strength of the empirical evidence, and its theoretical interpretation.

Consistency

The measure of consistency is used to determine the "fit of the empirical evidence" with an expected set-theoretic relationship (Mello 2017, 128). As such, the measure is applied both to assess the consistency of necessary conditions and of sufficient conditions. In formal terms, consistency is calculated to reflect the extent to which there is a subset relation between instances of a condition X and the outcome Y. When all values for the outcome Y are equal to or less than the respective values for X, then Y is a *subset* of X (vice versa, X is a superset of Y) and hence X is a *necessary condition* for Y. Likewise, if all values for Y are greater than or equal to the respective values for X, then Y is a *superset* of X (which also means that X is a subset of Y) and X is thus a *sufficient condition* for Y.[5]

Set-theoretic consistency for necessary and sufficient conditions is calculated with the following two formulae, which differ in the denominator (Ragin 2006a, 297):

---

[4] On different approaches to causation, see Goertz and Mahoney (2012), Pearl (2009), and Pearl and Mackenzie (2018).

[5] This means that when all values for the outcome Y and the condition X are exactly equal (in an XY plot with fuzzy data all points would be on the diagonal line), then X is both a necessary and sufficient condition for Y.

$$\text{Consistency}_{\text{Necessity}}(Y_i \leq X_i) = \frac{\sum \min(X_i, Y_i)}{\sum Y_i}$$

$$\text{Consistency}_{\text{Sufficiency}}(X_i \leq Y_i) = \frac{\sum \min(X_i, Y_i)}{\sum X_i}$$

Aimed to detect formal subset relationships, each formula divides the sum of the minimum values across the respective condition and the outcome by the sum of the values for the outcome (for necessary conditions), or by the sum of the values for the condition (for sufficient conditions). For perfect subset relations, this calculation yields consistency scores of 1. For imperfect set relations, where one or more cases violate a statement of necessity or sufficiency, the resulting scores will be less than 1. The similarity in the formulae reflects the inverse relationship between necessity and sufficiency.

How high does consistency have to be for a necessary or a sufficient condition? According to Ragin, "consistency scores should be as close to 1.0 (perfect consistency) as possible. When observed consistency scores are below 0.75, maintaining on substantive grounds that a set relation exists, even a very rough one, becomes increasingly difficult" (Ragin 2008, 46). While this statement underlines that there is no firm rule on consistency thresholds, it also shows that there is a corridor between 0.75 and 1.0 that researchers should aim for. For the truth table procedure, which is essentially an analysis of sufficient conditions, 0.75 is commonly used as a minimum threshold for rows to be included in the Boolean minimization (see Chapter 6). For necessary conditions, it is customary practice to use a *minimum threshold* of 0.90 consistency. Conditions that fall between 0.90 and 1.0 are considered "almost" necessary conditions.

You may wonder why there are different consistency thresholds for necessary and sufficient conditions. There are two reasons for this. First, the way necessary conditions are treated in social science theory, these usually focus on *single conditions* and there should be a strict standard to distinguish necessary or almost necessary conditions from other conditions. Moreover, the analysis of sufficiency is part of the *truth table minimization* and here a more relaxed standard can be applied for the first stage of the analysis, to keep coverage high by including as many positive cases as feasible. The eventual solution term should always strive for a consistency of *at least* 0.75, and ideally well above 0.80 consistency.

Of course, these thresholds are rules of thumb and one should always use individual judgment when interpreting empirical data.[6] Besides, one also has to take into account the number of cases involved: the smaller the number of cases, the higher the consistency thresholds should be set. For example, in a QCA study with only 12 cases one would expect intimate knowledge of these cases and thus there should be very high or even perfect consistency. On the contrary, for a QCA study with 50 or even 80 cases, a lower level of consistency can be expected and would certainly be acceptable (but consistency must still be above 0.75).

An example should help to illustrate the calculation of consistency values: Table 1 shows fuzzy-set values across eight hypothetical cases and the two conditions X_1 and X_2 and the outcome Y. The right-hand columns further list the minimum values across each condition and the outcome, information that is used for the numerator in both of the above formulae.

[6] The set-theoretic requirement of 0.90 consistency for necessary conditions is fairly demanding. This means that it is rarely reached by empirical cases, even when there is a pattern in the data. For example, in a study of military coalition defection (Mello 2019), the condition "upcoming elections" yielded a consistency of 0.83 and was thus formally speaking not necessary. However, a standard chi-square test of association showed that there was a statistically significant difference between groups that faced elections and those who did not (16 out of 18 cases showed the pattern).

Table 1: Calculation of Set-Theoretic Consistency

Case	X_1	X_2	Y	$min(X_1, Y)$	$min(X_2, Y)$
Case 1	1.0	0.7	0.9	0.9	0.7
Case 2	1.0	0.9	1.0	1.0	0.9
Case 3	0.8	0.6	0.7	0.7	0.6
Case 4	0.7	0.4	0.6	0.6	0.4
Case 5	0.3	0.0	0.0	0.0	0.0
Case 6	0.3	0.1	0.2	0.2	0.1
Case 7	0.6	0.3	0.4	0.4	0.3
Case 8	0.3	0.2	0.3	0.3	0.2
Sum	5.0	3.2	4.1	4.1	3.2

Given this data, we can calculate the set-theoretic consistency of X_1 and X_2 as necessary and/or sufficient conditions. For set-theoretic *necessity*, the outcome should be a perfect subset of the condition. For X_1 this is calculated by taking the sum of the minimum values across X_1 and Y and dividing this score by the sum of the values for the outcome. Taking the values found in the bottom line of Table 1, we divide 4.1 by 4.1, which equals 1. This means that X_1 is a perfect necessary condition for Y. For X_2 we divide 3.2 by 4.1, which equals 0.78. This is well below the threshold of 0.9 and hence X_2 should not be treated as a necessary condition for Y.

$$\text{Consistency}_{\text{Necessity}}(X_1) = \frac{4.1}{4.1} = 1.0 \qquad \text{Consistency}_{\text{Necessity}}(X_2) = \frac{3.2}{4.1} = 0.78$$

For set-theoretic *sufficiency*, the relationship should be inverted, expecting the respective condition to be a subset of the outcome. We calculate this taking the sum of the minimum values across the condition and the outcome and dividing this score by the sum of the values for the condition. For X_1, we thus divide 4.1 by 5.0, which is 0.82. This is above the commonly-used threshold of 0.75 for set-theoretic consistency, which means that X_1 is an almost sufficient condition for Y. For X_2, we divide 3.2 by 3.2, which equals 1. Hence X_2 is a perfect sufficient condition for Y.

$$\text{Consistency}_{\text{Sufficiency}}(X_1) = \frac{4.1}{5.0} = 0.82 \qquad \text{Consistency}_{\text{Sufficiency}}(X_2) = \frac{3.2}{3.2} = 1.0$$

Coverage

The measure of coverage is used to determine "the *relevancy* of a condition in empirical terms" (Mello 2017, 128). For sufficient conditions, coverage indicates "how much" of the empirical evidence is explained by a given condition. For instance, if there are twenty cases that show the outcome and ten of these can be accounted for with a given solution, then the coverage of that solution will be 0.50. Coverage can also be calculated for individual solution paths, which will be discussed at the end of this section. For necessary conditions, coverage helps to distinguish *relevant* from *trivial* necessary conditions. For example, a condition may formally fulfill the criteria for set-theoretic necessity (X being a perfect *superset* of the outcome Y), but this might still be a trivial finding because the condition is almost always present and thus it is difficult to establish a causal link between the condition and the outcome.

Set-theoretic coverage for necessary and sufficient conditions are calculated with the following two formulae, which differ in their denominator (Ragin 2006a, 63):

$$\text{Coverage}_{\text{Necessity}}(Y_i \leq X_i) = \frac{\sum \min(X_i, Y_i)}{\sum X_i}$$

$$\text{Coverage}_{\text{Sufficiency}}(X_i \leq Y_i) = \frac{\sum \min(X_i, Y_i)}{\sum Y_i}$$

It is apparent from the formulae for consistency and coverage that these are *inversely related.* Yet, it must be noted that consistency is the *primary* measure of fit and it should always be calculated first. Coverage is only meaningful when a consistent set-theoretic relationship has been identified in the analysis. As Ragin (2008, 55) puts it, it is "pointless" to examine and interpret coverage for a condition which is not a consistent subset or superset of the outcome. Yet, once consistency is established, then the calculation of coverage can help to assess the empirical relevance of the identified set-theoretic relationship.

Again, let us illustrate the calculation of coverage using the data provided in Table 1. In the first stage, we identified X_1 as a necessary condition. Let us now calculate its coverage. For this we take the sum of the minimum values across the condition and the outcome and divide it by the sum of the minimum values for the condition. This means we divide 4.1 by 5.0, which yields 0.82. This high coverage value reflects a close fit between the condition and the outcome. We can also calculate coverage for the sufficient condition X_2. For this we divide the sum of the minimum values across the condition and the outcome (3.2) by the sum of the values for the outcome (4.1), which equals 0.78.

$$\text{Coverage}_{\text{Necessity}}(X_1) = \frac{4.1}{5.0} = 0.82 \qquad \text{Coverage}_{\text{Sufficiency}}(X_2) = \frac{3.2}{4.1} = 0.78$$

The latter calculation means that X_2 covers 78% of the *set-membership values* of the outcome Y. This is an important distinction: at first, one may think that coverage simply returns the *percentage* of the covered cases. With fuzzy sets, the calculation is different because most cases will hold partial set membership and cases with less than 0.5 membership are included in the calculation. We can confirm this with a look at the data in Table 1, where three out of the first four cases that show the outcome (Cases 1–4) are covered by X_2, suggesting a "coverage" of 75%. Yet, we see that the calculation takes all fuzzy values into account and thus yields an actual coverage of 0.78.

For QCA solution terms, we can calculate coverage for the entire solution, but also for individual paths. While *solution coverage* indicates how comprehensively the observed phenomenon is accounted for with a given solution, *raw coverage* designates how much any single path accounts for the outcome. Moreover, *unique coverage* measures how much is only covered by a single path and not by any overlapping paths.

This should become clearer with an example. Table 2 shows three solution paths from a study on military coalition defection (Mello 2019). Besides other measures of fit, the table also lists raw coverage and unique coverage for each path, and solution coverage for the entire parsimonious solution.[7] As such, Path 1 has a raw coverage of 0.31 and a unique coverage of 0.15. This means that it covers 31% of the case membership values but about half of these are also covered by the other two paths, hence its unique coverage is only 15%. In total, the solution has a coverage of 0.80, which means that 80% of the membership values of the outcome are accounted for.

---

7 The parsimonious solution term is taken from the study's supplementary document, accessible at: https://doi.org/10.7910/DVN/8UWS1R (last accessed 12 July 2019).

Why is there such a difference between the overall coverage and the coverage values for the individual paths? In short, this is because of *empirical overlap*. Cases can hold partial set membership in various solution paths. We can see from Table 2 that the three paths are not mutually exclusive and only those cases printed in bold are uniquely covered by a single path. We can assess the extent to which the paths overlap by adding up their unique coverage values: 0.15 + 0.09 + 0.16 = 0.40. This means that half of the total solution coverage of 0.80 is attained by overlapping paths, whereas the other half is accounted for by the individual paths.[8]

Table 2: Solution Coverage, Raw Coverage, and Unique Coverage (Mello 2019)

	Path 1	Path 2	Path 3
Leadership Change			⊗
Upcoming Elections	⊗	⊗	⊗
Leftist Partisanship	●		
Low Commitment			
Fatalities		●	
Consistency	0.94	0.99	0.98
PRI	0.94	0.99	0.98
Raw Coverage	0.31	0.41	0.47
Unique Coverage	0.15	0.09	0.16
Covered Cases /	**AU2**	BG1	BG1
Uniquely Covered	ES2	DK1	DK1
Cases (Bold)	**HU2**	ES2	**HN1**
	IT2	JP1	JP1
	NZ1	**LV2**	**NI1**
	SK2	NL1	NL1
		PH1	**NO1**
		RO2	NZ1
		SK2	PH1
Solution Consistency		0.96	
Solution PRI		0.96	
Solution Coverage		0.80	

[8] See Duşa (2019) for a technical discussion of how coverage types are calculated. Ragin (2008, 63-68) provides crisp-set and fuzzy-set examples.

Proportional Reduction in Inconsistency

Proportional reduction in inconsistency (PRI) is a measure to identify *simultaneous subset relations* in the analysis of sufficient conditions. What does this mean? Simultaneous subset relations may be the case when the consistency and coverage measures indicate that a condition X is sufficient for the outcome Y and at the same time also sufficient for the non-outcome ~Y. This would be a *logical contradiction* but based solely on the measures of consistency and coverage it would be difficult to determine whether X should be treated as sufficient for the outcome or the non-outcome. To identify such situations and provide guidance for the correct interpretation of the set-theoretic relationship, the PRI measure was introduced into the fs/QCA software by Ragin (2006b) and first described in the textbook by Schneider and Wagemann (2012, 237-44).[9] The PRI measure is now a standard feature in most QCA software.

To illustrate simultaneous subset relations, let us take a simple example, involving the condition A, the outcome Y, and the non-outcome ~Y.

Table 3: Simultaneous Subset Relations

Case	*A*	*Y*	*~Y*	*min(A, Y)*	*min(A, ~Y)*	*min(A, Y, ~Y)*
France	0.2	0.3	0.7	0.2	0.2	0.2
Greece	0.5	0.6	0.4	0.5	0.4	0.4
Italy	0.4	0.5	0.5	0.4	0.4	0.4
Portugal	0.6	0.7	0.3	0.6	0.3	0.3
Spain	0.3	0.4	0.6	0.3	0.3	0.3
Sum	2.0	2.5	2.5	2.0	1.6	1.6

[9] On the PRI measure, see also Duşa (2019, 134-36).

Let us now calculate the consistency for A as a *sufficient condition* for the *outcome* Y and the non-outcome ~Y. We use the standard consistency formula introduced earlier in this chapter.

$$\text{Consistency}_{\text{Sufficiency}}(A \leq Y) = \frac{2.0}{2.0} = 1.0 \qquad \text{Consistency}_{\text{Sufficiency}}(A \leq \sim Y) = \frac{1.6}{2.0} = 0.80$$

Taking the sum values from the bottom line of Table 3, we divide the sum of the minimum values across A and Y (2.0) by the sum of the values for A (2.0), which equals 1.0. This means that A is a perfect sufficient condition for Y. What about the *non-outcome*? Here, we divide the sum of the minimal values across the condition and the non-outcome (1.6) by the sum of the values for A (2.0), which yields 0.80. Because the consistency value for A and ~Y is above the 0.75 threshold, we might treat A as an "almost" sufficient condition for the non-outcome, especially if this score referred to a truth table row. But this would mean that the condition equally leads to the outcome and its negation, which would be a logical contradiction.

How to resolve this paradox? The measures of fit consistency and coverage do not help us in this scenario, but PRI detects the problem. The formula for the proportional reduction in inconsistency is (Schneider and Wagemann 2012, 242):

$$\text{PRI} = \frac{\sum \min(X_i, Y_i) - \sum \min(X_i, Y_i, \sim Y_i)}{\sum X_i - \sum \min(X_i, Y_i, \sim Y_i)}$$

We apply this formula to assess the relationship between A and Y as well as A and ~Y:

$$\text{PRI}_{(A,\ Y)} = \frac{(2.0 - 1.6)}{(2.0 - 1.6)} = \frac{0.4}{0.4} = 1.0 \qquad \text{PRI}_{(A,\ \sim Y)} = \frac{(1.6 - 1.6)}{(2.0 - 1.6)} = \frac{0}{0.4} = 0$$

The PRI value for A as a sufficient condition for Y is 1.0, whereas the PRI for A and ~Y is 0. While consistency and coverage are relatively similar for both outcome and non-outcome (see the summary in Table 4), the PRI scores give a clear indication that we should treat A only as a valid sufficient condition for Y and not as a sufficient condition for ~Y.

As a general rule, one should always observe whether there is a substantial difference between PRI and consistency. When that is the case, one should also examine the non-outcome for simultaneous subset relations.

Table 4: Simultaneous Subset Relations and PRI

	$A \rightarrow Y$	$A \rightarrow \sim Y$
Consistency	1.00	0.80
Coverage	0.80	0.64
PRI	1.00	0.00

Relevance of Necessity

As introduced above, the standard coverage measure helps to distinguish trivial from relevant necessary conditions. However, there can be circumstances where potentially trivial necessary conditions in the empirical data would not be detected with the existing formula. Therefore, several alternative measures of necessary conditions' relevance and trivialness have been suggested (Braumoeller and Goertz 2000; Goertz 2006; Schneider and Wagemann 2012).

Against the backdrop of the "trivialness of necessity" measure suggested by Goertz (2006), Schneider and Wagemann (2012) put forth the "relevance of necessity" (RoN) measure, which has become a standard indicator in testing for necessary conditions. The formula for the calculation of RoN is (Schneider and Wagemann 2012, 236):

$$\text{Relevance of Necessity} = \frac{\sum(1 - X_i)}{\sum(1 - \min(X_i, Y_i))}$$

By conception, the relevance of necessity measure can yield values between 0 and 1, where lower scores indicate *trivialness* and higher scores denote *relevance*.

Let us look at an example, involving the condition X, the outcome Y, and hypothetical data on four cases. Table 4 shows that, formally speaking, the values for X are almost always larger than or equal to the values for Y, which indicates that X is an almost perfect superset of Y. Hence it should be an (almost) necessary condition.

Table 5: Relevance of Necessity

	X	Y	*Test for Necessity: X ⟵ Y*	
Case 1	1.0	0.9	Consistency	0.95
Case 2	0.9	0.4	Coverage	0.59
Case 3	1.0	0.3	Relevance of Necessity	0.38
Case 4	0.3	0.4		

The right-hand side of Table 4 shows the results for the calculation of the consistency and coverage for necessary conditions and the relevance of necessity. As expected, we can see that with 0.95 the set-theoretic consistency is very high, satisfying the formal threshold for necessary conditions (equal to or above 0.90). The coverage is lower, but at 0.59 it would not immediately prompt concern. However, we can see that the relevance of necessity indicator is closer to 0 than to 1, suggesting a trivial necessary condition. Why is that?

This being a simple example, we have no substantive knowledge of the underlying data. But what we can see is that X comes close to a *constant*, where three out of four cases show values equal to or close to 1. With data patterns like this, the consistency measure would always satisfy the criterion for a necessary condition. However, the RoN measure suggests that we should be cautious before treating it as a relevant necessary condition for the outcome.

Ultimately, dealing with data patterns like this is a matter of interpretation. There can be situations where a condition is almost a constant (for empirical reasons: the time period chosen, social or political circumstances, and so forth), but still the condition may have substantive importance and relevance as a necessary condition. However, such an interpretation would have to be justified explicitly.

An Alternative Consistency Measure

The consistency measure for sufficient conditions, presented above, is the standard formula presented in the prevailing textbooks (Ragin 2008; Rihoux and Ragin 2009; Schneider and Wagemann 2012) and part of the default settings of R packages (Duşa 2019; Oana and Schneider 2018). However, as Haesebrouck (2015) noted, the formula yields higher consistency values when inconsistent cases have large set membership scores, as opposed to cases with lower set membership scores. Let us illustrate this with the following example, which draws upon data similar to Haesebrouck (2015, 68):

Table 6: Limits of the Standard Consistency Measure

Case	X_1	Y_1	$min(X_1, Y_1)$	X_2	Y_2	$min(X_2, Y_2)$
1	1.0	1.0	1.0	1.0	1.0	1.0
2	0.3	0.3	0.3	0.3	0.3	0.3
3	0.3	0.3	0.3	0.3	0.3	0.3
4	1.0	0.7	0.7	0.3	0.0	0.0
Sum	2.6		2.3	1.9		1.6
	Consistency$_{Suff.}$ X_1: 2.3 / 2.6 = 0.88			Consistency$_{Suff.}$ X_2: 1.6 / 1.9 = 0.84		

The data in the two pairs of condition and outcome is identical except for case 4. However, the inconsistency between the condition and the outcome amounts to 0.3 in both scenarios (1.0 to 0.7 and 0.3 to 0.0). Yet, we can see that the standard formula yields a higher consistency for X_1 than X_2 (calculation in the bottom row). This is because X_1 has a higher set membership value with a larger consistent part and thus it counts more heavily in the calculation.

What to do about this? Haesebrouck (2015, 69) suggests a refined calculation for the consistency of sufficient conditions, which "attributes greater impact to large inconsistent cases":

$$\text{Consistency}_{\text{Sufficiency}}\,(X_i \leq Y_i) = \frac{\sum \min(X_i, Y_i)}{\sum(X_i, Y_i) \; + \sqrt{(\max(X_i - Y_i, 0) * X_i)}}$$

Haesebrouck (2015) applies his alternative consistency formula to two published studies (Mello 2014; Schneider and Makszin 2014) to show how their results would change with the alternative calculation. We can see the difference also by taking the data from the previous example:

Table 7: Alternative Consistency Calculation

Case	X_1	Y_1	$min(X_1,Y_1)$	$\sqrt{(max(X_1-Y_1, 0) * X_1)}$	X_2	Y_2	$min(X_2,Y_2)$	$\sqrt{(max(X_2-Y_2, 0) * X_2)}$
1	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
2	0.3	0.3	0.3	0.0	0.3	0.3	0.3	0.0
3	0.3	0.3	0.3	0.0	0.3	0.3	0.3	0.0
4	1.0	0.7	0.7	0.5	0.3	0.0	0.0	0.3
Sum	2.6		2.3	0.5	1.9		1.6	0.3

$Consistency_{Suff.}$ X_1: 2.3 / (2.3 + 0.5) = 0.81

$Consistency_{Suff.}$ X_2: 1.6 / (1.6 + 0.3) = 0.84

With the alternative measure of consistency, larger inconsistencies weigh more heavily in the calculation, as shown in the lower consistency for X_1 (0.81) as opposed to the standard formula, which yielded 0.88 for this condition.

So far, Haesebrouck's measure has not yet become standard in QCA applications but it is included as a function in the R package "SetMethods" (Oana and Schneider 2018) and can also be implemented in the "QCA" package (Duşa 2019, 253), if users wish to use an alternative measure or want to explore the robustness of their results.

Summary

With *consistency* and *coverage*, this chapter has introduced the two principal set-theoretic measures of fit. These remain the primary indicators to assess the strength and relevance of QCA results. The chapter further discussed the important complementary measures of *proportional reduction in inconsistency* (PRI, Ragin 2006b) and *relevance of necessity* (RoN, Schneider and Wagemann 2012), which help to, respectively, detect simultaneous subset relations in sufficient conditions and trivialness of necessary conditions. Finally, an alternative measure of consistency (Haesebrouck 2015) was presented, which overcomes certain limitations of the standard formula for the consistency of sufficient conditions.

As a cautionary note, it is important to highlight that QCA remains a *qualitative* approach. Hence, besides the importance of numerical measures, emphasis should always be placed on the qualitative interpretation of analytical results, rather than merely striving for high measures of fit. These measures should serve as the foundation for a rigorous empirical analysis, but they cannot assess the substantive relevance of QCA findings or replace a sound interpretation of one's empirical results.

References

Braumoeller, Bear F., and Gary Goertz. 2000. "The Methodology of Necessary Conditions." *American Journal of Political Science* 44 (4): 844–58.

Duşa, Adrian. 2019. *QCA with R. A Comprehensive Resource*. Cham: Springer.

Goertz, Gary. 2006. "Assessing the Trivialness, Relevance, and Relative Importance of Necessary or Sufficient Conditions in Social Science." *Studies in Comparative and International Development* 41 (2): 88–109.

Goertz, Gary, and James Mahoney. 2012. *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton, NJ: Princeton University Press.

Goertz, Gary, and Harvey Starr. 2003. "Introduction: Necessary Condition Logics, Research Design, and Theory." In *Necessary Conditions*, edited by Goertz, Gary and Harvey Starr. Lanham: Rowman & Littlefield, 1–23.

Haesebrouck, Tim. 2015. "Pitfalls in QCA's Consistency Measure." *Journal of Comparative Politics* (2): 65-80.

Mello, Patrick A. 2013. "From Prospect to Practice: A Critical Review of Applications in Fuzzy-Set Qualitative Comparative Analysis." 8th Pan-European Conference on International Relations, Warsaw, 18-21 September.

Mello, Patrick A. 2014. *Democratic Participation in Armed Conflict: Military Involvement in Kosovo, Afghanistan, and Iraq*. Basingstoke: Palgrave Macmillan.

Mello, Patrick A. 2017. "Qualitative Comparative Analysis and the Study of Non-State Actors." In *Researching Non-State Actors in International Security: Theory & Practice*, edited by Kruck, Andreas and Andrea Schneiker. Oxon: Routledge, 123-42.

Mello, Patrick A. 2019. "Paths towards Coalition Defection: Democracies and Withdrawal from the Iraq War." *European Journal of International Security* (14 June): https://doi.org/10.1017/eis.2019.10.

Oana, Ioana-Elena, and Carsten Q. Schneider. 2018. "SetMethods: An Add-on R Package for Advanced QCA." *The R Journal* 10/1 (July): 507-33.

Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.

Pearl, Judea, and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books.

Ragin, Charles C. 2000. *Fuzzy-Set Social Science*. Chicago, IL: University of Chicago Press.

Ragin, Charles C. 2006a. "Set Relations in Social Research: Evaluating Their Consistency and Coverage." *Political Analysis* 14 (3): 291–310.

Ragin, Charles C. 2006b. *User's Guide to Fuzzy-Set / Qualitative Comparative Analysis*. Irvine, CA: University of California.

Ragin, Charles C. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago, IL: University of Chicago Press.

Ragin, Charles C., Susan E. Mayer, and Kriss A. Drass. 1984. "Assessing Discrimination: A Boolean Approach." *American Sociological Review* 49 (2): 221-34.

Rihoux, Benoît, and Gisèle De Meur. 2009. "Crisp-Set Qualitative Comparative Analysis (csQCA)." In *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*, edited by Rihoux, Benoît and Charles C. Ragin. Thousand Oaks: Sage, 33-68.

Rihoux, Benoît, and Charles C. Ragin, ed. 2009. *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques.* Thousand Oaks, CA: Sage.

Schneider, Carsten Q., and Kristin Makszin. 2014. "Forms of Welfare Capitalism and Education-Based Participatory Inequality." *Socio-Economic Review* 12 (2): 437-62.

Schneider, Carsten Q., and Claudius Wagemann. 2010. "Standards of Good Practice in Qualitative Comparative Analysis (QCA) and Fuzzy-Sets." *Comparative Sociology* 9 (3): 397–418.

Schneider, Carsten Q., and Claudius Wagemann. 2012. *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*. New York, NY: Cambridge University Press.