International Journal of Social Research Methodology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tsrm20

From aggregations to multimethod case configurations. Case diversity in quantitative analysis when explaining COVID-19 fatalities

Philip Haynes & David Alemna

To cite this article: Philip Haynes & David Alemna (2022): From aggregations to multimethod case configurations. Case diversity in quantitative analysis when explaining COVID-19 fatalities, International Journal of Social Research Methodology, DOI: 10.1080/13645579.2022.2122224

To link to this article: https://doi.org/10.1080/13645579.2022.2122224

Published online: 13 Sep 2022.

INTERNATIONAL JOURNAL OF SOCIAL RESEARCH METHODOLOGY
https://doi.org/10.1080/13645579.2022.2122224

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

From aggregations to multimethod case configurations. Case diversity in quantitative analysis when explaining COVID-19 fatalities

Philip Haynes [a] and David Alemna [b]

[a]School of Humanities and Social Science, University of Brighton, Falmer, Brighton, UK; [b]School of Area Studies, History, Politics and Literature, University of Portsmouth, Portsmouth, UK

ABSTRACT
Three quantitative methods are compared for their ability to understand different COVID-19 fatality ratios in 33 OECD countries. Linear regression provides a limited overview without sensitivity to the diversity of cases. Cluster Analysis and Dynamic Patterns Synthesis (DPS) gives scrutiny to the granularity of case similarities and differences, and reveals case exceptions. Qualitative Comparative Analysis (QCA) develops causal theory about what conditions are sufficient for explaining outcomes by using robust and transparent conventions. Configurational case-based methods offer important advantages over inferential statistics when there is a need to focus on diversity in small n. These techniques can be combined as multi-methods. DPS and QCA can be used concurrently to aid research insights. These methods are also strengthened by additional qualitative evidence about the cases.

ARTICLE HISTORY
Received 19 July 2021
Accepted 18 August 2022

KEYWORDS
Regression; cluster analysis; Dynamic Pattern Synthesis (DPS); Qualitative Comparative Analysis (QCA); causal complexity; configurative comparative methods; COVID-19

Introduction

This paper examines the turn towards complex configurative methods. Aggregation methods model the characteristics of an archetypal case where all cases will perform the same way in given circumstances. Alternatively, configurative methods give credence to the diversity of cases. For example, outcomes can demonstrate equifinality, where different causal paths produce the same outcome (Livne-Tarandach et al., 2016).

This paper examines the claims of configurative quantitative approaches that they deal better with causal complexity (Bicket et al., 2020). For the purposes of this article, 'multimethod', is defined as the combination of two or more quantitative methods and 'mixed method' is defined as the combination of quantitative and qualitative approaches (Creswell & Plano Clark, 2011). In order to undertake this methodological evaluation, we consider the performance of three different quantitative methods when undertaking country comparisons: linear regression, Dynamic Patterns Synthesis (DPS – an extension of cluster analysis), and Qualitative Comparative Analysis (QCA).

Aggregation is the process of summarising a dataset with statistical procedures. Statistical aggregation is founded on the tendency towards material similarity. Quetelet's (1842) historical research into empirical observations of physical characteristics underpinned the conceptual basis of averages and standard deviations as reliable measures of demographics. When there are multiple variables for consideration, the focus moves to the overall experience of a case, in relation to

CONTACT Philip Haynes p.haynes@brighton.ac.uk School of Humanities and Social Science, University of Brighton. Falmer, Brighton BN1 9PH, United Kingdom

a combination of scores. A major challenge is developing a model that has predictive validity. This is difficult because of case diversity.

Linear models like regression are based on a number of assumptions. This includes that the dependent variable has a normal distribution, and that the predictor variables are not strongly correlated (multicollinearity). Also, the unexplained aspect of the linear model (the error, or residuals) should not conceal any pattern in the differences of variances (heteroscedasticity), or a correlation of the residuals for a subset of cases (autocorrelation). All this implies a risk that a linear model may be concealing unknown patterns.

Aggregation like this has two major statistical considerations. The first is, does the computation produce a chance result, or something more significant? Inferential statistics make such decisions. Probability theory eliminates the chance that the result is a random occurrence. The second consideration is the size of the effect. If the result is not a chance, is there a weak, moderate, or strong association?

Aggregation and considerations of chance and effect were transformational in the history of mathematics (Stewart, 2013). Their extension from demography to metaphysical social science is problematic as a form of social representation and led to 'the turn to the qualitative', seeing the importance of the interpretivist perspective (Yanow & Schwartz-Shea, 2015). Quantitative methods are needed with a greater sensitivity to case complexity (Byrne, 1998; Ragin & Becker, 1992).

Cluster analysis has had a longstanding influence on social science as a case-based classification (Blashfield & Aldenderfer, 1988). It examines the configurations of cases in regard to how they compare, rather than predicting from an aggregation of variable scores. Cluster analysis is a generic term for a multiplicity of techniques that classify cases (Everitt, 1993; Pastor, 2010). Some techniques require the researcher to specify the number of clusters in advance (Figueiredo Filho et al., 2014). Others produce hierarchical groupings, either agglomerative or divisive (Aldenderfer & Blashfield, 1984). Hierarchical output requires the researcher to decide which layer is of most theoretical validity. This determines the number of clusters of interest. This can be argued to be a subjective choice (Pastor et al., 2007), but such a decision can be viewed as the mixing of mathematical results with qualitative interpretations, as in 'mixed methods' (Tashakkori & Creswell, 2007).

Cluster analysis first became prominent in the physical sciences, for biological classifications (Sneath & Sokal, 1973). Different algorithms for measuring cluster similarity may provide different interpretations of where cases are situated (Aldenderfer & Blashfield, 1984). This has been suggested to be a weakness (Blashfield, 1976), unless researchers have a good prior knowledge about the cases. The overall importance of cluster analysis is exploration of empirical case comparisons for the subsequent development of theory. Dynamic Pattern Synthesis (DPS) is an extension of cluster analysis that includes more detailed analysis of each individual cluster and how they compare and contrast (Bicket et al., 2020; Haynes, 2017). It can also be used to compare longitudinal pattern changes across clusters.

A more recent advance in the relationship between theory and cases is Qualitative Comparative Analysis (QCA). Ragin (1987), its originator, was seeking empirical explanations of historical political comparisons. For example, he demonstrated how different patterns of conditions in a sample of countries (cases) could result in the same political outcome. Ragin (2014: xix) describes QCA as: 'techniques that both bridge and transcend the qualitative-quantitative divide'. Its contribution is the mathematical patterning of multiple causal paths. Instead of aggregating a typical case experience, QCA theorises configurational complexity and explains the diversity of case experiences.

The earliest QCA approach used the definition of 'crisp sets' to map configurations. Independent variables are conceptualised as 'conditions' for a binary outcome (Schneider & Wageman, 2010), and are defined by 'crisp set' binary categories, either extant (1) or negated (0; Rihoux & De Meur, 2009). Different combinations of the conditions become sets of possible explanations of the outcome.

The development of fuzzy set QCA (QCA_{fs}) allow for granularity in the definition of conditions and outcomes (Rihoux & Ragin, 2009). This is a popular method that calibrates conditions using fuzzy set scores between 0 and 1 (for example: 0.3, 0.6, 0.9). In all forms of QCA, Boolean algebra is used to summarise and simplify theoretical statements about the conditions for an outcome. These conditional patterns for a given outcome are known as 'sets' (Thomann, 2020). Minimisation is used to formally identify the most important patterns that are shared by cases. Short variable names are used with Boolean minimisation to construct 'solution formulas' (Schneider & Wageman, 2010: 18). The following symbols are used: ~ = negate, * = and, + = or, → = outcome. Boolean algebra is used for 'logical minimisation' of a set of conditions to find a concise theoretical summary of how conditions map with outcomes (Rihoux & De Meur, 2009). 'Prime Implicants' are the most parsimonious form of these summaries (Schneider & Wagemann, 2012: 109–110).

Qualitative Comparative Analysis has a growing influence on social science research (Bicket et al., 2021; Rihoux, et al, 2011). A recent example is the seminal work of the Centre for the Evaluation of Complexity Across the Nexus (CECAN) for the UK Civil Service (Bicket et al., 2020).

Example data analysis

The three methods are compared when analysing a dataset with R packages: gmodels (Warnes et al., 2018), cluster (Maechler et al., 2021), and QCA (Dusa, 2019). The example dataset investigates the difference in COVID-19 fatalities ratios across OECD countries during the first wave of the 2020 Covid-19 pandemic. Thirty-three OECD countries are included. The variables used are [*short name*]:

- Government COVID-19 response index at 20 days after the identification of the first case (source: Hale et al., 2020) [*GovResponse*]
- Millions, of international arrivals in 2018 (source: World Bank, 2020) [*IntArrivals*]
- Population ratio of known COVID-19 infections at 50 days after the first known case (source: Hale et al., 2020) [*Infections*]

The dependent/outcome variable is:

- Population ratio of COVID-19 fatalities (July 1st, 2020). (source: Hale et al., 2020) [*Fatalities*]

For comparative purposes, the dataset is standardised to values between 0 and 1. The Government COVID-19 response index is calculated by the Blavatnik School of Government at the University of Oxford. This index compares COVID-19 related government activity and was used in the development of the *Covidtracker* database from 2020. Examples of government activities include: school and workplace closing, restrictions on public events, the economic mitigation of pandemic impacts, and public health responses etc. This data is used to create a composite government response index (Hale et al., 2020). Scores are calculated on a daily basis allowing for longitudinal comparisons. Some countries have high scores because they act across the range of indices, while other countries choose to focus on a few indicators they see as particularly important for control of COVID-19. This raises some questions about the reliability and validity of the dataset. Nevertheless, other researchers have made use of this data (for example, Chisadza et al., 2021).

In addition, daily fatalities rates from each country are inputted using international public health sources like the World Health Organisation (WHO) and are incorporated in the University of Oxford's COVID-19 data directory known as the 'Oxford Supertracker' (Daly et al., 2020). International travel movements are an important consideration. A reputable international indicator of the movement of people into a country is the World Bank count of the annual volume of international arrivals. This counts individual arrival events. Contextually, the aim of this paper is to compare methodological approaches to small-N quantitative datasets, and not to draw conclusions on the effectiveness of government responses. We present our results below.

Comparisons of results

Linear regression

First, multiple regression is used to explain the relationship between three predictors and the ratio of COVID-19 fatalities. The model reveals a low and significant effect size (R = 0.67. R Square = 0.45 and Adjusted R Square = 0.39: ANOVA, F (29,3) = 7.90, p = 0.001). Standardised regression coefficients show the substantive contribution of GovResponse (ß − 0.62, t = −3.54, F = 8.94, p = 0.01, CI = −0.97 to −0.26) and Infections, (ß 0.56 t = 3.28, F = 10.77, p = 0.003, CI = 0.24 to 1.02). The coefficient contribution of IntArrivals is ß 0.34 (t = 2.30, F = 4.00, p = 0.05, CI = 0.04 to 0.68). A higher fatalities rate is associated with a lower government response, higher known infections, and higher international arrivals. There is no autocorrelation (Durbin Watson = 2.30), and no multicollinearity (tolerance: GovResponse = 0.62, Infections = 0.65, IntArrivals = 0.87). There is no evidence of heteroscedasticity.

Cluster analysis and DPS

The cluster method used is agglomerative Hierarchical Cluster Analysis (HCA), applying Ward's method of finding the smallest Error Sum of Squares (Ward, 1963). Ward's method is preferred because it produces more homogenous clustering (Aldenderfer & Blashfield, 1984: 43). It is common practice to standardise the variables, so that the construction of clusters is not impacted by elevations of scale (Romesburg, 2004: 78). The standardisation applied is between 0 and 1. This mirrors the calibration used in the second configuration method, QCA_{fs}.

The HCA computation starts by finding the two most similar cases. It then compares all cases and allocates them into hierarchical groups until all are agglomerated. The first stage provides the maximum potential clusters where the clusters are at their most homogeneous. Cluster analysis is often the first stage in a multimethod. Other techniques validate the clusters and demonstrate how the cases in clusters are similar. Haynes (2017) has proposed a combination of cluster analysis with other configurational methods to understand complex case patterns, for example, DPS (Alemna et al., 2021; Bicket et al., 2020; Taylor et al., 2021). Table 1 shows the resulting cluster patterns using the four variables and is sorted by the cluster memberships in the final column. There are eight clusters, but Luxembourg is a singularity.

Table 1 shows the variable scores for each case. The shading shows threshold similarities (above or below the mean) for scores contributing to a cluster membership. White text on a black background shows where all cases in the cluster share above average scores. Black text on a grey background shows clusters where all cases share below average scores. The final column of Table 1 gives a text-based summary for a cluster. For example, in cluster one, all countries have below mean average IntArrivals and an above average ratio of Infections. The GovResponses are close to the mean, but Switzerland is marginally below. This demonstrates that in addition to the observation of above and below mean scores, aspects like the range of scores can also be important to consider. The population mean, median and standard deviation are shown at the bottom of Table 1. It is notable when case exceptions are important, for example, all but one member of a larger cluster may have a similar variable pattern, while one is an exception (but the case is similar to its cluster on other variables scores). In this way, DPS investigates the complex configurations of clusters.

This adds value to the overall general pattern suggested by the regression. In Table 1, there are subgroups of cases where the association between Fatalities and GovResponse is evidenced better than in the regression analysis (i.e. clusters 7 and 8). Similarly, differences in Infections and

Table 1. Dynamic Pattern Synthesis: cluster configurations and summaries.

Country	CountryCode	GovResponse	IntArrivals	Infections	clustermembership	Fatalities	Summary notes
Switzerland	CHE	37	10.36	3003	1	194.46	Lower international arrivals, higher known infections. Gov. response close to average
Ireland	IRL	46	10.93	2987	1	351.57	
Luxembourg	LUX	81	1.02	5706	2	175.73	Outlier, highest infections, similar to C3
Turkey	TUR	62	45.77	1409	3	60.84	Lower fatalities, higher gov. responses
Austria	AUT	73	30.82	1604	3	78.28	
Portugal	PRT	79	16.19	2032	3	154.56	
Slovenia	SVN	75	4.43	648	3	53.39	
Czech Republic	CZE	69	10.61	636	3	32.59	
Slovak Republic	SVK	70	2.26	249	3	5.13	
Hungary	HUN	65	17.55	234	3	60.56	
Netherlands	NLD	55	18.78	1686	4	356.76	Lower international arrivals. Above average gov. responses, (exception is CHL, marginally below average). Tendency to below average fatalities (exceptions are NLD and CHL – above average, but not as high scores as countries in C7 and C8)
Chile	CHL	36	5.72	572	4	297.55	
Norway	NOR	57	5.69	1249	4	46.12	
Denmark	DNK	58	12.75	1148	4	104.45	
Israel	ISR	39	4.12	1150	4	36.97	
Estonia	EST	41	3.23	1081	4	52.02	
Finland	FIN	10	3.22	67	5	59.20	Lower fatalities, lower known infections. The countries with a higher range of known infections in the cluster have relatively higher gov. response scores in cluster.
Australia	AUS	24	9.25	8	5	4.08	
South Korea	KOR	33	1.53	143	5	5.50	
New Zealand	NZL	41	3.69	221	5	4.56	
Poland	POL	44	19.62	260	5	38.66	
Greece	GRC	38	30.12	205	5	18.42	
Mexico	MEX	6	41.31	54	6	215.38	Lower government responses, lower known infections. Fatalities are lower, or close to average.
Canada	CAN	8	21.13	7	6	227.62	
Japan	JPN	12	31.19	2	6	7.70	
Germany	DEU	18	38.88	72	6	107.24	
Spain	ESP	18	82.77	674	7	606.46	Higher fatalities, lower government responses, higher international arrivals, lower known infections
Italy	ITA	24	61.57	681	7	575.02	
United States	USA	8	79.75	2	7	384.92	
France	FRA	10	89.32	55	7	457.20	
United Kingdom	GBR	18	36.32	49	8	644.17	Higher fatalities, lower government response, lower known infections
Sweden	SWE	10	7.44	158	8	528.06	
Belgium	BEL	8	9.12	589	8	841.62	
Mean		***39***	***23.23***	***868***		***205.66***	**Key**
Median		*38*	*12.75*	*572*		*104.45*	Cluster scores are all above mean average
St dev		*24*	*24.16*	*1172*		*223.18*	Cluster scores are all below mean average

IntArrivals relative to Fatalities, are specific to clusters. In Clusters 7 and 8, Infections are lower, suggesting limited tracing activity. GovResponse is lower and Fatalities are higher. In Cluster 7, IntArrivals are also higher. Conversely, in Cluster 3, IntArrivals are lower, while GovResponses are higher, and Fatalities lower. The configuration in cluster 3 has an association of lower Fatalities and higher GovResponses. Cluster 5 has lower Fatalities and Infections.

The cluster configurative approach enables specific conceptualisations of the relationship of cases with variable configurations. Patterns of association seen in some clusters are not uniformly evidenced for all. This adds a case-by-case understanding.

Qualitative Comparative Analysis (QCA)

An alternative configurational research approach is QCA. This method is explanatory rather than exploratory in that it seeks a set-theoretical causal conclusion, in relation to a specified outcome. A prerequisite to such explanation is good knowledge of the cases and their context (Ragin, 2014; Rihoux & Ragin, 2009). The data is prepared using the fuzzy set calibration outline by Ragin (2017) and this requires a standardised dataset in the range 0–1. This is the standardisation used for all the methods compared in this paper.

Fuzzy set QCA uses explanatory formulas to discover the conditions for an outcome. The principle of these formulaic calculations is that they relate to the selection of minimum scores from combinations of conditions, rather than aggregating from scale variable measurements. This is an important difference to the methods used in linear regression, and cluster analysis. Ragin (2009: 96) gives an example:

> 'If a country's membership in the set of poor countries is 0.7 and its membership of a set of democratic countries is 0.9, its membership in the set for countries that are both poor and democratic is the smaller of these two scores, 0.7.'

The resulting computations are: the *consistency* of conditions for a specific outcome, and *coverage* (how many cases from the dataset share one set of conditions). These computations return a value between 0 and 1, where higher values indicate substantive scores. Researchers usually aim to find consistency scores of > 0.8 to conclude that conditions for an outcome are substantive and conceptually important (Schneider & Wageman, 2010: 10).

Consistency scores measure the number of cases in a row that share the same outcome. There are several measures of consistency and Proportional Reduction Interpretation (PRI) eliminates unreliable aspects of the default consistency algorithm (Schneider & Wagemann, 2012). It produces slightly lower scores than the default and is considered more reliable. Identifying necessary conditions is the first step in a QCA_{fs} analysis (Schneider & Wageman, 2010). A necessary condition is one that is present with an outcome but does not assure it will happen. The criteria for necessary conditions is set at >0.9 for consistency and >0.6 for coverage. Necessary conditions are not present in the Covid-19 dataset, and the analysis progresses to sufficiency (conditions that are usually present with an outcome).

The truth table (Table 2) is constructed using the method outlined by Ragin (2009: 109–111). The table simplifies the conditions and outcomes from fuzzy scores to crisp set scores of 0 (negated) and 1 (extant). In truth tables, the consistency scores represent the reliability and validity of the theoretical sets of conditions displayed in rows. While these rows summarise the overall pattern of conditions in a set for a given outcome, it is possible in truth tables for cases with a negated outcome to be located in a set row with cases argued to have an extant outcome (and visa versa; Ragin, 2009: 109; Schneider & Wagemann, 2012: 103). These are contradictory outcomes. There are several examples in Table 2. For illustration, in the set presented in the second row: InArrivals * ~GovResponse * ~Infections, where the outcome is extant COVID-19 fatalities, Greece and Japan have the contradictory outcome <0.5 (~Fatalities). This 'contradictory configuration' (Berg-Schlosser, De Meur, Rihoux & Ragin, 2009: 15) is reflected in the low PRI consistency score of 0.502 for that row (Table 2). Therefore, the set is not valid for a concluding theoretical model. Truth

Table 2. Truth Table: Covid-19 fatality outcomes.

IntArrivals	GovResponse	Infections	N	Fatality	Cases	Raw consist.	PRI consist.
1	0	1	2	1	ITA, ESP	**0.898**	**0.674**
1	0	0	7	1	FRA, DEU, GRC, JPN, MEX, GBR, USA	0.732	0.502
1	1	1	2	0	AUT, TUK	0.724	0.032
0	0	1	6	0	BEL, CHL, EST, IRL, ISR, CHE	0.694	0.321
0	1	1	7	0	CZE, DNK, LUX, NLD, NOR, PRT, SVN	0.567	0.113
0	0	0	7	0	AUS, CAN, FIN, NZK, POL, KOR, SWE	0.525	0.238
0	1	0	2	0	HUN, SVK	0.456	0.015
IntArrivals	GovResponse	Infections	N	~Deaths	Cases	raw consist.	PRI consist.
0	1	0	2	1	HUN, SVK	**0.992**	**0.985**
1	1	1	2	1	AUT, TUK	**0.991**	**0.968**
0	1	1	7	1	CZE,DNK,LUX,NLD,NOR,PRT,SVN	**0.873**	**0.740**
0	0	0	7	1	AUS,CAN,FIN,NZK,POL,KOR,SWE	0.838	0.741
0	0	1	6	1	BEL,CHL,EST,IRL,ISR,CHE	0.829	0.622
1	0	1	2	0	ITA, ESP	0.789	0.326
1	0	0	7	0	FRA,DEU,GRC,JPN,MEX, GBR,USA	0.679	0.404

(Logical remainder 1 1 0)

tables also demonstrate 'logical reminders' where a set of possible conditions has no cases. In the example (Table 2) the logical remainder is: IntArrivals * GovResponse * ~Infections. Logical remainders reveal the limited diversity of cases and support the extension of theory beyond observed cases (De Meur et al., 2009).

There is only one small set defined as: IntArrivals * ~GovResponse * Infections (Italy and Spain) in Table 2 that is of sufficient consistency, and without contradictions, to be of theoretical interest for explaining extant Fatalities outcomes (consistency = 0.898, PRI consistency 0.674). The other row sets in Table 2 do not consistently calculate which cases have higher Fatalities. There are contradictions. For example, in the set in the second row of Table 2 for extant outcomes, Greece and Japan, two countries with low Fatalities, appear with countries with high Fatalities like Britain, France and the USA (IntArrivals * ~GovResponse * ~Infections).

Table 2 also computes negated Fatalities outcomes and this shows more validity, and has less contradictions. The first two rows of negated outcomes show strong consistency scores for two pairs. Row one is Hungary and Slovakia (~IntArrivals * GovResponse * ~Infections: consistency = 0.992, PRI = 0.985) Row two is Austria and Turkey (IntArrivals * GovResponse * Infections: consistency = 0.991, PRI = 0.968). The third row has seven cases and explains negated outcomes for Czech Republic, Denmark, Luxembourg, the Netherlands, Norway, Portugal and Slovenia (~IntArrivals * GovResponse * Infections: consistency = 0.873, PRI consistency = 0.740). The Netherlands is a contradictory case with Fatalities outcome >0.5. If one examines The Netherlands in Table 1, it has above average Fatalities, but it is not one of the highest scoring countries. For this reason, the set is included in the final theoretical model.

Row 4 has marginally lower consistency and two contradictory configurations, Canada, and Sweden. The latter is a country with substantially higher comparative Fatalities (see, Table 1). For this reason, row 4 is not included in the final theoretical model. The first three negated rows of Table 2 are reduced to a parsimonious solution term (Schneider & Wagemann, 2012: 18) using Boolean logical minimisation to create the expression:

GovResponse * (~IntArrivals * ~Infections) + (IntArrivals * Infections) + (~IntArrivals * Infections) → ~Fatalities

This is confirmed by a sufficiency calculation with R QCA (Consistency = 0.992, PRI = 0.986, Coverage = 0.434).

Minimisation in R QCA reduces this to: ~IntArrivals * GovResponse + Infections * GovResponse → ~Fatalities (Consistency = 0.906, PRI = 0.850, Coverage = 0.550).

For the logical remainder with no cases, it is difficult to conclude on the substantive significance, given the limitations of the data discussed earlier in the paper.

The conclusion for the sufficiency of the QCA_{fs} explanation is that the outcome of lower Fatalities is more conclusive, in terms of case coverage, than understanding the outcome of higher Fatalities. QCA's theoretical conclusions are not always symmetrical (Schneider & Wagemann, 2012: 113). There are numerous countries where lower Fatalities are associated with a relatively strong GovResponse. A lower level of IntArrivals also evidences lower Fatalities in some countries. Having a higher knowledge of Infections (by implication through a tracing and testing system) also evidences low Fatalities in some countries. Qualitative Comparative Analysis adds value to DPS by formulising a theoretical approach that commits to simplified statements about necessary and sufficient conditions for an outcome. It goes beyond exploration to explanation.

Discussion

Linear regression seeks a consistent relationship between the IVs with the DV so that any relationship is applied to all cases. A more detailed examination of the inferential probabilistic aspects of the outcome variable reveals limitations. The small sample, and inconsistent linearity between the IV and DV across all cases, means that causal inference is partial. The confidence intervals of the regression coefficients show that the range of likely errors is considerable. This illustrates the error range, if the model's equation is used to predict the Fatalities of any other countries not included in the OECD dataset.

Larger samples reduce issues with statistical significance, but rarely solve challenges about understanding the substantive effects. Low effect results are an indication of the likelihood of additional complex case configurations. These require different methods for an explanation. Low effect sizes in regression may indicate asymmetrical results for the outcome. For example, the independent variables are better at predicting below average rather than above average outcomes. Substantial variations in the confidence intervals for each country, due to the small sample size, suggests that it is problematic to use the regression analysis to examine each case independently with regards to their fit on the regression line. Configurative methods do not assume symmetrical relationships between conditions and an outcome as demonstrated in the more valid negation formula for Fatalities in the QCA example, when compared to extant Fatalities (Table 2). Configurative methods explain more than asymmetry and reveal aspects of equifinality, where different independent variable patterns (QCA conditions) have the same or similar outcomes, and multifinality, where the same or similar independent variable patterns (QCA conditions) have different outcomes.

Cluster analysis does not preference an outcome, unless used with an additional method. It requires the addition of a multimethod (Haynes, 2017), to explore cluster detail and outcomes, but does not have formal techniques for comparing outcomes. Qualitative Comparative Analysis provides a robust procedure for explaining outcomes. Equifinality is revealed in different sets that have the same outcome (rows 1, 2, and 3 for negated fatalities in Table 2) and multifinality is evidenced where a single set has contradictory outcomes (row 2 for extant fatalities in Table 2).

Clusters and their interpretation with DPS place emphasis on case exceptionalism, rather than the specifics of complexity of outcome. Different variable patterns are evidenced for each cluster and there are relative degrees of exceptionalism within these clusters. Japan (cluster 6) has a low GovResponse score, and low Infections and Fatalities. It has some key differences to other members of its cluster and this is better conceptualised as exceptionalism than multifinality. Dynamic Pattern Synthesis provides important forensic detail for understanding the diverse impact of variables in different clusters by disaggregating cases into groups and highlighting the related impact of variable score ranges within groups. An example is the relationship of IntArrivals with Fatalities. Dynamic Pattern Synthesis reveals some key detail in this respect through the analysis of the data ranges within and between the clusters. Cluster 5 (Table 1) shares below average Fatalities. Several cases in Cluster 5, have an interaction with low levels of IntArrivals. IntArrivals, however, are not below average for the whole cluster, because of Greece being an exception. Cluster 4 (Table 1) shares below average IntArrivals, but not below average Fatalities, because of two exceptions with higher scores: Netherlands and Chile. Nevertheless, these two scores are not as high as the scale range of Fatalities evidenced in clusters 7 and 8. These two clusters tend to share high IntArrivals (especially cluster 7), but there are two exceptions in cluster 8 with below average IntArrivals (Belgium and Sweden) and DPS allows for detailed case-by-case forensic analysis.

While QCA_{fs} provides a formulaic approach to case differences this creates a degree of aggregation of sets. This can result in contradictory outcomes, where individual cases must be explained further. Such explanation is more precise with the addition of the forensic analysis offered by DPS. This is not to argue that cluster analysis and DPS are uniformly better for understanding case configurations. A weakness of imposing cluster boundaries from the cluster algorithms is that cases may be more valid if in another cluster. If different algorithms are applied, different memberships might result and expose the more ambiguous of members (Aldenderfer & Blashfield, 1984).

In this sense, the clustering algorithm might be argued to be a form of aggregation, as it is a data summary. Any negative impact of this aggregation is lessened by the demonstration of the full case-by-case variable scores in Table 1. The table information is used to make qualitative judgements about the degree to which members of clusters are similar. Clusters are not mutually exclusive, and they have 'fuzzy' characteristics. The forensic detail of the DPS in Table 1 shows where the association between GovResponse and Fatalities has the most validity (in clusters 2, 3, 7 and 8, Table 1). The cases in clusters 2 and 3 are all covered in the scope of the concluding QCA negated Fatalities formula, as represented in rows 1–3 of the negated truth table (Table 2).

The descriptive focus of the detail of DPS remains close to the case granularity. Qualitative Comparative Analysis seeks to be more explanatory by use of its conventions, and formulaic outcomes. Sets of cases are compared logically with an outcome in a systematic manner. The truth table approach of QCA_{fs} leaves the researcher with the task of deciding what degree of contradiction to tolerate. For example, consistency scores for a set can be relatively high, but a single contradictory case may still be present. During the interrogation of DPS, focus is placed on exceptional cases, and to consider exactly how the variable scores are different for that case when compared with others.

For both configurative approaches, it can be argued that the solution to contradictions and exceptions is to study the cases in more qualitative detail (Miller, 2018). This is important with COVID-19 comparative research. There is much interest in why Japan is different. Only rich qualitative data and analysis can provide in-depth answers. Tashiro and Shaw (2020) argue that Japan has longstanding cultural factors that make physical contact between its population less likely, and that the government's national response may have low comparative scores, but this may hide a focus on localised responses that balance with an apparent lack of coordinated national lockdown.

The contemporary approaches to evaluating quantitative configurations offer a greater ability to understand case diversity in smaller datasets when compared to traditional inferential statistical approaches. Much of the degree of success with new configurative approaches relies on their telescopic attention to small groups of cases. Aggregate statistical models are important for summarising at a larger scale. Cluster analysis and QCA can be used with larger samples (Cooper & Glaesser, 2011) but the calculation of aggregation at scale in configurative methods is challenging with regard to understanding the granularity of all cases (Cooper & Glaesser, 2016). An example is the use of consistency scores in QCA, given multiple contradictory case configurations in large n, and if inferential and effect statistics are used to explain cluster differences from large n. In these circumstances, important case contradictions and exceptions might be missed.

Conclusion

Configurative approaches offer advantages when used with small n that are historically a strength of qualitative research rather than quantitative research. Multi-methods (combinations of quantitative methods) with mixed methods (use of consecutive or concurrent quantitative and qualitative) add to the understanding of case diversity, exceptions, and contradictions. Combinations of methods can help to solve complex social research questions that exhibit case diversity, such as comparative government responses to the COVID-19 pandemic. Different case-based methods can be used concurrently to enhance research design.

Dataset

Country	Country Code	Fatalities	IntArrvials	GovResponse	Infections
Australia	AUS	0.05	0.12	0.15	0.05
Austria	AUT	0.18	0.57	0.9	0.65
Belgium	BEL	0.95	0.12	0.06	0.51
Canada	CAN	0.58	0.38	0.06	0.05
Chile	CHL	0.65	0.08	0.28	0.51
Czech Republic	CZE	0.08	0.14	0.86	0.52
Denmark	DNK	0.27	0.18	0.68	0.59
Estonia	EST	0.11	0.06	0.35	0.58
Finland	FIN	0.13	0.06	0.06	0.07
France	FRA	0.79	0.95	0.06	0.06
Germany	DEU	0.28	0.65	0.11	0.07
Greece	GRC	0.06	0.56	0.31	0.14
Hungary	HUN	0.13	0.28	0.8	0.17
Ireland	IRL	0.7	0.15	0.43	0.81
Israel	ISR	0.09	0.07	0.32	0.59
Italy	ITA	0.86	0.84	0.15	0.53
Japan	JPN	0.05	0.57	0.07	0.05
Luxembourg	LUX	0.48	0.05	0.95	0.95
Mexico	MEX	0.57	0.68	0.05	0.06
Netherlands	NLD	0.71	0.31	0.62	0.66
New Zealand	NZL	0.05	0.07	0.35	0.16
Norway	NOR	0.1	0.08	0.66	0.61
Poland	POL	0.09	0.34	0.4	0.19
Portugal	PRT	0.48	0.25	0.94	0.71
Slovak Republic	SVK	0.05	0.06	0.87	0.18
Slovenia	SVN	0.12	0.07	0.91	0.52
South Korea	KOR	0.05	0.05	0.24	0.1
Spain	ESP	0.88	0.94	0.11	0.53
Sweden	SWE	0.84	0.1	0.06	0.11
Switzerland	CHE	0.48	0.14	0.3	0.81
Turkey	TUR	0.13	0.72	0.75	0.63
United Kingdom	GBR	0.89	0.63	0.11	0.06
United States	USA	0.73	0.93	0.06	0.05

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

David Almena is an ESRC funded Postdoctoral Fellow with the South Coast Doctoral Training Partnership [Project Reference: ES/W005743/1].

Notes on contributors

Philip Haynes is Professor of Public Policy at the University of Brighton. His research includes the use and development of methods that better represent the complexity of social and economic policy systems.

David Alemna is ESRC South Coast Doctoral Training Partnership Post Doc Fellow at the University of Portsmouth. He completed is PhD research on comparative public policy and policy transfer, using configurative methods in 2020, and has also worked on policy research for an international consultancy.

ORCID

Philip Haynes http://orcid.org/0000-0002-9880-356X
David Alemna http://orcid.org/0000-0002-9319-3758

References

Aldenderfer, M., & Blashfield, R. (1984). *Cluster analysis*. Sage.

Alemna, D., Artaraz, K., Haynes, P., & Mwale, S. (2021). The complexity and instability of policy conditionality and transfer. *IMF Interventions in The Political Economy of South America. Complexity, Governance and Networks*, *6*(1), 14–31. https://doi.org/10.20377/cgn-103

Berg-Schlosser, D., De Meur, G., Rihoux, B., & Ragin, C. (2009). Qualitative Comparative Analysis (QCA) as an approach. In B. Rihoux & C. Ragin (Eds.), *Configurative comparative methods* (pp. 1–18). Sage.

Bicket, M., Christie, I., Gilbert, N., Hills, D., Penn, A., & Wilkinson, H. (2020). *Handling complexity in policy evaluation: Supplementary Guide to Magenta book*. HM Treasury.

Bicket, M., Hills, D., Wilkinson, H., & Penn, A. (2021). Don't panic: Bringing complexity thinking to UK Government evaluation guidance. *Evaluation*, *27* (1), 18–31. https://doi.org/10.1177/1356389020980479

Blashfield, R. (1976). Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychological Bulletin*, *83*(3), 377–388. https://doi.org/10.1037/0033-2909.83.3.377

Blashfield, R., & Aldenderfer, M. (1988). The methods and problems of cluster analysis. In J. Nesselroade & R. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 447–473). Springer.

Byrne, D. (1998). *Complexity theory and the social sciences*. Routledge.

Chisadza, C., Clance, M., & Gupta, R. (2021). Government Effectiveness and the COVID-19 Pandemic. *Sustainability*, *13*(6), 3042. https://doi.org/10.3390/su13063042

Cooper, B., & Glaesser, J. (2011). (2011) Using case-based approaches to analyse large datasets: A comparison of Ragin's fsQCA and fuzzy cluster analysis. *International Journal of Social Research Methodology*, *14*(1), 31–48. https://doi.org/10.1080/13645579.2010.483079

Cooper, B., & Glaesser, J. (2016). Exploring the robustness of set theoretical findings from a large n fsQCA: An illustration from the sociology of education. *International Journal of Social Research Methodology*, *19*(4), 445–459. https://doi.org/10.1080/13645579.2015.1033799

Creswell, J. W., & Plano Clark, V. (2011). *Designing and conducting mixed methods research*. Sage.

Daly, M., Ebbinghaus, B., Lehner, L., Naczyk, M., & Vlandas, T. (2020). *Tracking Policy Responses to COVID-19: Opportunities, Challenges and Solutions*. Department of Social Policy and Intervention, University of Oxford. https://supertracker.spi.ox.ac.uk/

De Meur, G., Rihoux, B., & Yamasaki, S. (2009). Addressing the Critiques of QCA. In B. Rihoux & C. Ragin (Eds.), *Configurational comparative methods. Qualitative comparative analysis (QCA) and related techniques* (pp. 147–166). Sage.

Dusa, A. (2019). *QCA with R. A Comprehensive Resource*. Springer.

Everitt, B. S. (1993). *Cluster Analysis*. Arnold.

Figueiredo Filho, D. B., da Rocha, E. C., da Silva Júnior, J. A., Paranhos, R., da Silvada Silva, M. B., & Duarte, B. S. (2014). Cluster analysis for political scientists. *Applied Mathematics*, *5*(15), 2408–2415. https://doi.org/10.4236/am.2014.515232

Hale, T., Angrist, N., Kira, B., Petherick, A., Phillips, T., & S, W. (2020). *Variation in government response to COVID-19. Version 6*. Oxford: Blavatnik School of Government Working Paper May 25. www.bsg.ox.ac.uk/covidtracker

Haynes, P. (2017). *Social Synthesis: Finding Dynamic Patterns in Complex Social Systems*. Routledge.

Livne-Tarandach, R., Hawbaker, B., Lahnneman, B., & Jones, C. (2016). Qualitative comparative analysis: Opportunities for case-based research. In K. Elsbach & R. M. Kramer (Eds.), *Handbook of qualitative organizational research: Innovative pathways and methods* (pp. 156–167). Routledge.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2021). *cluster: cluster analysis basics and extensions. R package version 2.1.2*. https://CRAN.R-project.org/package=cluster

Miller, D. (2018). Challenging trends in configuration research: Where are the configurations? *Strategic Organization*, *16*(4), 453–469. https://doi.org/10.1177/1476127017729315

Pastor, D. (2010). Cluster Analysis. In G. Hancock & R. Mueller (Eds.), *The reviewer's guide to quantitative methods* (pp. 41–54). Routledge.

Pastor, D., Barron, K., Miller, B., & Davis, S. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemporary Educational Psychology*, *32* (1), 8–47. https://doi.org/10.1016/j.cedpsych.2006.10.003

Quetelet, L. A. (1842). *History of psychology series.A treatise on man and the development of his faculties (A facsimile reproduction of the English translation of 1842 with an introduction by Solomon Diamond)*. Facsimiles & Reprints; William and Robert Chambers. https://doi.org/10.1037/11993-000

Ragin, C. (1987). *The comparative method: Moving beyond qualitative and quantitative strategies*. University of California Press.
Ragin, C. (2009). Qualitative Comparative Analysis Using Fuzzy Sets (fsQCA). In B. Rihoux & C. Ragin (Eds.), *Configurational Comparative Methods* (pp. 87–122). Sage.
Ragin, C. (2014). *The comparative method: Moving beyond qualitative and quantitative strategies* (*2nd Edition* ed.). The University of California Press.
Ragin, C. (2017). *User's guide to fuzzy-set/ qualitative comparative analysis*. University of California.
Ragin, C., & Becker, H. (1992). *What is a case? Exploring the foundations of social inquiry*. Cambridge University Press.
Rihoux, B., & De Meur, G. (2009). Crisp-Set Qualitative Comparative Analysis (csQCA). In B. Rihoux & C. Ragin (Eds.), *Configurational comparative methods* (pp. 33–68). Sage.
Rihoux, B., & Ragin, C. (2009). *Configurational comparative methods: Qualitative Comparative Methods (QCA) and related techniques*. Sage.
Rihoux, B. R, Rezohazy, I Bol, D. (2011). Qualitative comparative analysis (QCA) in public policy analysis: An extensive review. *German Policy Studies*, *7*(3), 9–82. http://epa-journal.eu/download/2.-qualitative-comparative-analysis-qca-in-public-policy-analysis-an-extensive-review.pdf.
Romesburg, H. (2004). *Cluster Analysis for Researchers*. Lulu Press.
Schneider, C. Q., & Wageman, C. (2010). Standards of good practice in qualitative comparative Analysis (QCA) and Fuzzy Sets. *Comparative Sociology*, *9*(3), 397–418. https://doi.org/10.1163/156913210X12493538729793 .
Schneider, C. Q., & Wagemann, C. (2012). *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis*. Cambridge University Press.
Sneath, P., & Sokal, R. (1973). *Numerical taxonomy. The principles and practice of numerical classification*. W H Freeman and Company.
Stewart, I. (2013). *17 equations that changed the world*. Profile Books.
Tashakkori, A., & Creswell, J. W. (2007). The new era of mixed methods. *The Journal of Mixed Methods Reseach*, *1*(1), 3–7. https://doi.org/10.1177/2345678906293042
Tashiro, A., & Shaw, R. (2020). COVID-19 pandemic response in Japan: What is behind the initial flattening of the curve? *Sustainability*, *12*(13), 5250. https://doi.org/10.3390/su12135250.
Taylor, L., Haynes, P., & Darking, M. (2021). English local government finance in transition: Towards the 'marketization of income'. *Public Management Review 23*(7) 1081–1106. https://doi.org/10.1080/14719037.2020.1743343
Thomann, E. (2020). Qualitative Comparative Analysis for comparative policy analysis. In B. Peter & G. Fontaine (Eds.), *Handbook of research methods and applications in comparative policy analysis* (pp. 254–276). Edward Elgar.
Ward, J. (1963). Hierarchical grouping to optimize and objective function. *Journal of the American Statistics Association*, *58*(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845
Warnes, G., Bolker, B., Lumley, T., & Johnson, R. (2018). *gmodels: Various R Programming Tools for Model Fitting (R package version 2.18.1)*. https://cran.rstudio.com/web/packages/gmodels/gmodels.pdf
World Bank (2020). *International tourism, number of arrivals data*. Retrieved 3 May 2022, from https://data.worldbank.org/indicator/ST.INT.ARVL
Yanow, D., & Schwartz-Shea, P. (2015). *Interpretation and method: Empirical research methods and the interpretive turn*. Routledge.