U.S. Department of Veterans Affairs

Configurational Data Analysis with QCA and CNA for Health Researchers

Michael Baumgartner & Alrik Thiem
University of Geneva

2 March 2017

Poll Question

What is your level of knowledge about configurational methods?

- I am an absolute beginner
- I have limited working knowledge
- I have intermediate knowledge
- I have advanced knowledge
- I consider myself an expert

Background

QCA Qualitative Comparative Analysis

CNA Coincidence Analysis

QCA and CNA are **configurational comparative methods** of causal data analysis that investigate implicational (Boolean) hypotheses of the type:

$$X = 1 \text{ is minimally sufficient/necessary for } Y = 0.$$

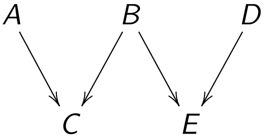
By contrast, regression-analytic methods investigate covariational (linear-algebraic) hypotheses of the type:

$$\text{The more/less of } X\text{, the more/less of } Y.$$

For more on the difference between configurational comparative methods and regression analysis see **?**.

The problem to be solved by QCA and CNA

A simple example



data generating causal structure (DGS)

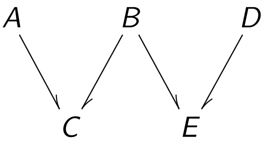
#	A	B	C	D	E
c_1	1	1	1	1	1
c_2	1	1	1	0	1
c_3	1	0	1	1	1
c_4	1	0	1	0	0
c_5	0	1	1	1	1
c_6	0	1	1	0	1
c_7	0	0	0	1	1
c_8	0	0	0	0	0

configurational data

A	painting with yellow
B	smoking
C	yellow fingers
D	having a cold
E	coughing

The problem to be solved by QCA and CNA

A simple example



data generating causal structure (DGS)

A	painting with yellow
B	smoking
C	yellow fingers
D	having a cold
E	coughing

#	A	B	C	D	E
c_1	1	1	1	1	1
c_2	1	1	1	0	1
c_3	1	0	1	1	1
c_4	1	0	1	0	0
c_5	0	1	1	1	1
c_6	0	1	1	0	1
c_7	0	0	0	1	1
c_8	0	0	0	0	0

configurational data

Problem

Infer DGS from configurational data.

Boolean dependencies

As causation is not visible in empirical data, any method of causal inference must indirectly infer causal structures **from empirically visible dependency structures**.

QCA and CNA infer causal structures from Boolean dependency structures recovered in the data.

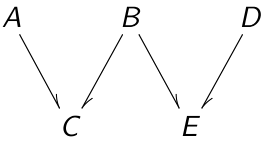
Sufficiency

A is sufficient for B if, and only if (iff), all cases featuring A also feature B, i.e. A is a subset of B.

Necessity

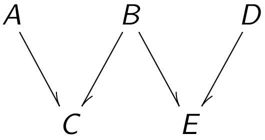
A is necessary for B iff all cases featuring B also feature A, i.e. B is a subset of A.

Back to the example



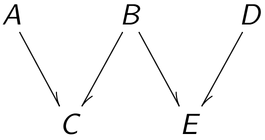
#	A	B	C	D	E
c_1	1	1	1	1	1
c_2	1	1	1	0	1
c_3	1	0	1	1	1
c_4	1	0	1	0	0
c_5	0	1	1	1	1
c_6	0	1	1	0	1
c_7	0	0	0	1	1
c_8	0	0	0	0	0

Back to the example



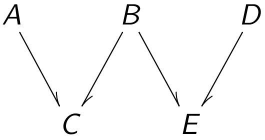
#	A	B	C	D	E
c_1	1	1	1	1	
c_2	1	1	1	0	1
c_3	1	0	1	1	1
c_4	1	0	1	0	0
c_5	0	1	1	1	1
c_6	0	1	1	0	1
c_7	0	0	0	1	1
c_8	0	0	0	0	0

Back to the example



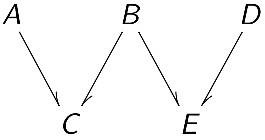
#	A	B	C	D	E
c_1	1	1	1	1	1
c_2	1	1	1	0	1
c_3	1	0	1	1	1
c_4	1	0	1	0	0
c_5	0	1	1	1	1
c_6	0	1	1	0	1
c_7	0	0	0	1	1
c_8	0	0	0	0	0

Back to the example



#	A	B	C	D	E
c_1	1	1	1	1	1
c_2	1	1	1	0	1
c_3	1	0	1	1	1
c_4	1	0	1	0	0
c_5	0	1	1	1	1
c_6	0	1	1	0	1
c_7	0	0	0	1	1
c_8	0	0	0	0	0

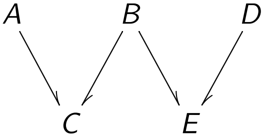
Back to the example



#	A	B	C	D	E
c_1	1	1	1	1	
c_2	1	1	1	0	1
c_3	1	0	1	1	1
c_4	1	0	1	0	0
c_5	0	1	1	1	1
c_6	0	1	1	0	1
c_7	0	0	0	1	1
c_8	0	0	0	0	0

$$ABCD + ABCd + AbCD + aBCD + aBCd + abcD \leftrightarrow E \qquad (1)$$

Back to the example

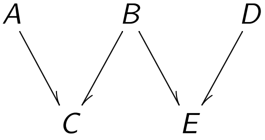


#	A	B	C	D	E
c_1	1	1	1	1	
c_2	1	1	1	0	1
c_3	1	0	1	1	1
c_4	1	0	1	0	0
c_5	0	1	1	1	1
c_6	0	1	1	0	1
c_7	0	0	0	1	1
c_8	0	0	0	0	0

$$ABCD + ABCd + AbCD + aBCD + aBCd + abcD \leftrightarrow E \qquad (1)$$

$\rightarrow$ **Most relations of sufficiency and necessity have nothing to do with causation!**

Back to the example



#	A	B	C	D	E
c_1	1	1	1	1	
c_2	1	1	1	0	1
c_3	1	0	1	1	1
c_4	1	0	1	0	0
c_5	0	1	1	1	1
c_6	0	1	1	0	1
c_7	0	0	0	1	1
c_8	0	0	0	0	0

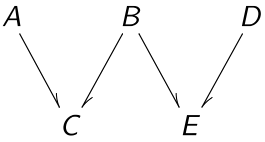
$$ABCD + ABCd + AbCD + aBCD + aBCd + abcD \leftrightarrow E \qquad (1)$$

→ **Most relations of sufficiency and necessity have nothing to do with causation!**

$$\textbf{Target:} \quad B + D \leftrightarrow E \qquad (2)$$

→ Question: How do we get from (1) to (2)?

Back to the example



#	A	B	C	D	E
c_1	1	1	1	1	1
c_2	1	1	1	0	1
c_3	1	0	1	1	1
c_4	1	0	1	0	0
c_5	0	1	1	1	1
c_6	0	1	1	0	1
c_7	0	0	0	1	1
c_8	0	0	0	0	0

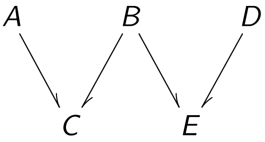
$$ABCD + ABCd + AbCD + aBCD + aBCd + abcD \leftrightarrow E \tag{1}$$

$\rightarrow$ **Most relations of sufficiency and necessity have nothing to do with causation!**

$$\textbf{Target:} \quad B + D \leftrightarrow E \tag{2}$$

$\rightarrow$ Question: How do we get from (1) to (2)? Answer: **via redundancy elimination.**

Back to the example



#	A	B	C	D	E
c_1	1	1	1	1	1
c_2	1	1	1	0	1
c_3	1	0	1	1	1
c_4	1	0	1	0	0
c_5	0	1	1	1	1
c_6	0	1	1	0	1
c_7	0	0	0	1	1
c_8	0	0	0	0	0

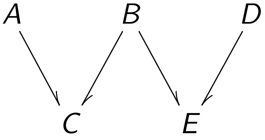
$$ABCD + ABCd + AbCD + aBCD + aBCd + abcD \leftrightarrow E \quad (1)$$

$\rightarrow$ **Most relations of sufficiency and necessity have nothing to do with causation!**

$$\textbf{Target:} \quad B + D \leftrightarrow E \quad (2)$$

$\rightarrow$ Question: How do we get from (1) to (2)? Answer: **via redundancy elimination.**

Back to the example



#	A	B	C	D	E
c_1	~~1~~	1	1	1	1
c_2	1	1	1	0	1
c_3	1	0	1	1	1
c_4	1	0	1	0	0
c_5	0	1	1	1	1
c_6	0	1	1	0	1
c_7	0	0	0	1	1
c_8	0	0	0	0	0

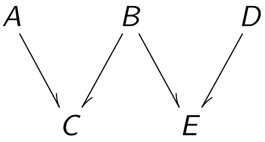
$$ABCD + ABCd + AbCD + aBCD + aBCd + abcD \leftrightarrow E \qquad (1)$$

$\rightarrow$ **Most relations of sufficiency and necessity have nothing to do with causation!**

$$\textbf{Target:} \quad B + D \leftrightarrow E \qquad (2)$$

$\rightarrow$ Question: How do we get from (1) to (2)? Answer: **via redundancy elimination.**

Back to the example



#	A	B	C	D	E
c_1	1	1	1	1	1
c_2	1	1	1	0	1
c_3	1	0	1	1	1
c_4	1	0	1	0	0
c_5	0	1	1	1	1
c_6	0	1	1	0	1
c_7	0	0	0	1	1
c_8	0	0	0	0	0

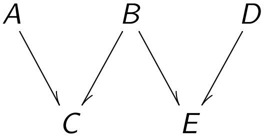
$$ABCD + ABCd + AbCD + aBCD + aBCd + abcD \leftrightarrow E \qquad (1)$$

$\rightarrow$ **Most relations of sufficiency and necessity have nothing to do with causation!**

$$\textbf{Target:} \quad B + D \leftrightarrow E \qquad (2)$$

$\rightarrow$ Question: How do we get from (1) to (2)? Answer: **via redundancy elimination.**

Back to the example



#	A	B	C	D	E
c_1	1	1	1	1	1
c_2	1	1	1	0	1
c_3	1	0	1	1	1
c_4	1	0	1	0	0
c_5	0	1	1	1	1
c_6	0	1	1	0	1
c_7	0	0	0	1	1
c_8	0	0	0	0	0

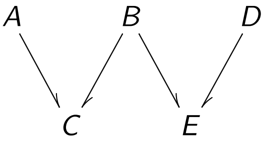
$$ABCD + ABCd + AbCD + aBCD + aBCd + abcD \leftrightarrow E \quad (1)$$

$\rightarrow$ **Most relations of sufficiency and necessity have nothing to do with causation!**

$$\textbf{Target:} \quad B + D \leftrightarrow E \quad (2)$$

$\rightarrow$ Question: How do we get from (1) to (2)? Answer: **via redundancy elimination.**

Back to the example



#	A	B	C	D	E
c_1	1	1	1	1	1
c_2	1	1	1	0	1
c_3	1	0	1	1	1
c_4	1	0	1	0	0
c_5	0	1	1	1	1
c_6	0	1	1	0	1
c_7	0	0	0	1	1
c_8	0	0	0	0	0

$$ABCD + ABCd + AbCD + aBCD + aBCd + abcD \leftrightarrow E \qquad (1)$$

$\rightarrow$ **Most relations of sufficiency and necessity have nothing to do with causation!**

$$\textbf{Target:} \quad B + D \leftrightarrow E \qquad (2)$$

$\rightarrow$ Question: How do we get from (1) to (2)? Answer: **via redundancy elimination.**

Redundancy elimination

The procedural core of QCA and CNA consists in algorithms that **rigorously eliminate all redundancies** from relations of sufficiency and necessity.

Minimal sufficiency

AX is minimally sufficient for B iff AX is sufficient for B, and AX does not contain a sufficient proper part.

Minimal necessity

$A + X$ is minimally necessary for B iff $A + X$ is necessary for B, and $A + X$ does not contain a necessary proper part.

Output of QCA and CNA

QCA and CNA output minimally necessary disjunctions of minimally sufficient conditions of modeled outcomes.

Qualitative Comparative Analysis (QCA)

Areas of Application of QCA

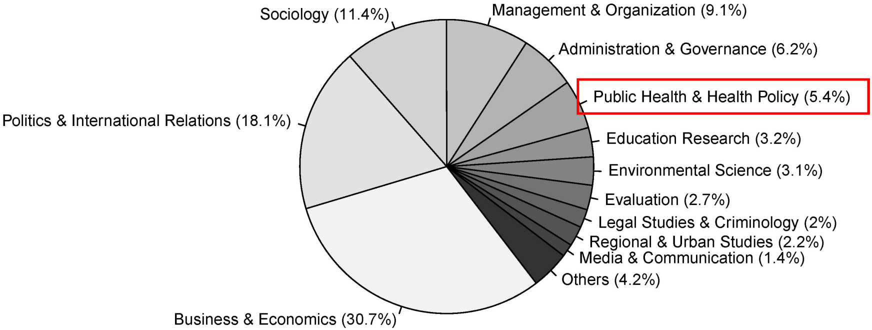


Figure: Area Distribution of 762 Applied QCA Articles

Example Study

Kahwati, Leila C., Megan A. Lewis, Heather Kane, Pamela A. Williams, Patrick Nerz, Kenneth R. Jones, Trang X. Lance, Stephen Vaisey, and Linda S. Kinsinger. 2011. "Best Practices in the Veterans Health Administration's MOVE! Weight Management Program." *American Journal of Preventive Medicine* **41** (5):457-64.

Background	Obesity substantial problem in Veterans Health Administration (VHA); VHA developed and disseminated MOVE! Weight Management Program for Veterans to medical facilities in 2006; program implementation has been variable
Purpose	Explore variation in MOVE! program implementation to identify facility structure, policies, processes associated with larger patient weight-loss outcomes
Methods	Qualitative Comparative Analysis (QCA); used to identify (combinations of) conditions associated with larger 6-month patient weight-loss outcomes

The Data

Table E-1. Summary truth table

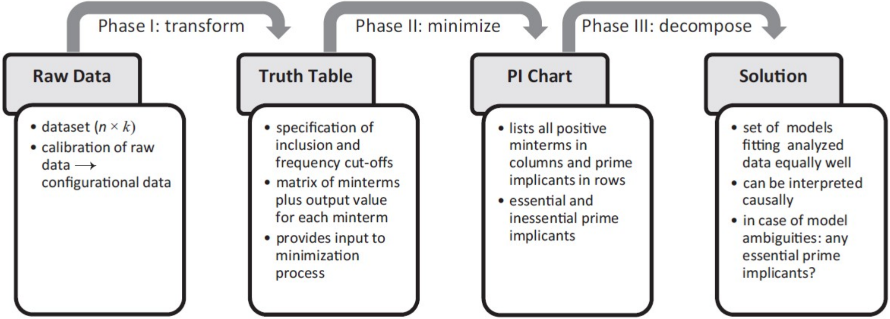
Facility ID	Larger numbers of patient weight loss outcomes	1. High interface between screening and treatment	2. Use of standard curriculum	3. Use of multidisciplinary team approach involving a dietitian and at least one other discipline	4. High program complexity	5. Use of weight loss maintenance component	6. Used group care-delivery format	7. High use of structured dietary plans
1	1	0	1	0	1	0	1	0
4	1	1	1	1	1	1	1	1
6	1	0	1	0	1	1	1	0
9	1	0	1	0	0	1	1	0
11	1	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1	1
14	1	1	1	1	1	0	1	1
15	1	1	1	1	1	1	1	0
19	1	1	1	1	1	1	1	1
20	1	1	1	1	1	1	1	1
21	1	1	1	1	1	1	1	1
2	0	0	1	0	1	1	1	1
3	0	0	1	0	1	1	0	0
5	0	1	1	1	1	1	1	0
7	0	1	1	1	0	1	1	1
8	0	1	1	1	1	0	1	0
10	0	1	0	0	0	0	0	1
13	0	1	1	1	0	1	1	0
16	0	0	0	0	0	0	0	0
17	0	1	1	0	0	0	0	1
18	0	0	0	0	0	1	0	1
22	0	1	1	0	1	1	1	0

The Data

Facility ID	Larger numbers of patient weight loss outcomes	8. High use of structured physical activity plans	9. High use of multiple behavioral strategies	10. High staff involvement	11. No use of wait list	12. High facility complexity	13. High data tracking and analysis capacity	14. Active physician involvement	15. Use of quality improvement (QI) for enhancing program and resolving challenges	16. High program accountability to facility leadership and internal reporting requirements	17. High program accountability to regional leadership and external reporting requirements
1	1	1	1	1	1	0	1	1	0	0	0
4	1	1	1	1	1	1	1	0	0	1	1
6	1	0	0	0	0	0	1	1	0	0	1
9	1	1	1	1	1	0	0	0	1	1	1
11	1	0	1	1	0	0	0	1	1	0	0
12	1	1	1	1	1	1	0	0	1	1	0
14	1	1	1	0	0	1	1	1	1	1	1
15	1	1	1	0	1	1	1	0	0	0	0
19	1	1	1	0	1	1	1	1	1	0	1
20	1	0	1	1	1	1	1	0	0	1	1
21	1	0	1	0	1	0	1	1	0	0	1
2	0	1	1	0	1	1	0	1	0	1	0
3	0	1	1	1	1	1	0	1	1	1	1
5	0	1	1	0	0	1	1	1	1	1	1
7	0	1	1	1	1	1	1	1	0	1	1
8	0	0	1	0	1	0	1	0	0	0	1
10	0	1	1	1	0	0	1	0	0	1	1
13	0	1	1	1	1	1	1	1	0	1	1
16	0	0	1	0	0	0	0	0	0	0	0
17	0	1	1	1	1	0	0	0	0	0	0
18	0	0	0	0	1	1	1	1	1	1	0
22	0	0	0	0	1	0	0	0	0	0	0

Note: The truth table summarizes the results of condition calibration. The presence and/or high levels (fully in) of the condition are indicated by a 1. The absence and/or low levels (fully out) of the condition are indicated by a 0. Consult Appendix D for more details on calibration.

The Work Flow of QCA



The Concept of Consistency/Inclusion

sometimes, cases that instantiate the same minterm show different values on the endogenous factor

for minterm 32, which has 5 cases, 3 cases show $LNP = 0$

the degree to which the hypothesis that the minterm is sufficient for the outcome is true is thus not perfect any more

we need the concept of *inclusion*

$$\mathrm{Incl}(\mathbf{X}\{\cdot\} \Rightarrow \mathbf{O}\{\cdot\}) = \frac{\sum_{i=1}^{n} \mathbf{O}\{\cdot\} = 1 | \mathbf{X}\{\cdot\} = 1}{\sum_{i=1}^{n} \mathbf{X}\{\cdot\} = 1}$$

a value often suggested in the literature is 0.75

however, theoretically speaking, any value above 0.5 would do (more evidence in favour than against)

The Concept of Coverage

usually, some minterms will be instantiated many times in the data, others only once, and some others never

minterm 28 has 1 case, minterm 29 has 2 cases

result: some causal paths will often be more prevalent than others

we need the concept of (*raw*) *coverage*

$$\mathrm{Cov_r}(\mathbf{X}\{\cdot\} \Rightarrow \mathbf{O}\{\cdot\}) = \frac{\sum_{i=1}^{n} \mathbf{O}\{\cdot\} = 1 | \mathbf{X}\{\cdot\} = 1}{\sum_{i=1}^{n} \mathbf{O}\{\cdot\} = 1}$$

empirical relevance does not equate with theoretical relevance

rarely observed paths may be the more interesting ones

Coincidence Analysis (CNA)

Input of CNA

The data input of CNA is the same as the data processed by QCA, *viz.* **tables listing configurations**.

	A	B	C	D
c_1	1	1	1	1
c_2	1	1	1	0
c_3	1	0	1	1
c_4	1	0	1	0
c_5	0	1	1	1
c_6	0	1	1	0
c_7	0	0	0	1
c_8	0	0	0	0

crisp-set data

	A	B	C	D
c_1	2	3	3	1
c_2	3	3	2	2
c_3	1	2	3	1
c_4	4	2	3	1
c_5	1	1	2	2
c_6	3	1	2	2
c_7	1	4	2	3
c_8	2	1	1	3

multi-value data

	A	B	C	D
c_1	0.1	0.6	0.7	1
c_2	1	1	0.3	0.4
c_3	0.4	0.9	0.3	1
c_4	0.7	0.6	0.5	0
c_5	0.3	0.1	0.8	1
c_6	0.9	0	1	0.4
c_7	1	1	0	0.9
c_8	0	0.8	0	1

fuzzy-set data

Input of CNA

CNA takes a consistency (*con*) and a coverage (*cov*) threshold (between 0 and 1) as input and only issues models meeting the given thresholds. Default $con = cov = 1$; should not be lowered below $con = cov = 0.75$.

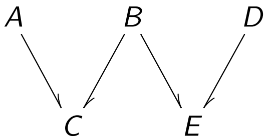
CNA does not require the identification of a factor as outcome in the data. It infers which factors can be modeled as outcome(s) from the data.

Prior causal information about which factor can or cannot cause which other factor(s) may be given to CNA in the form of an optional argument called a **causal ordering**.

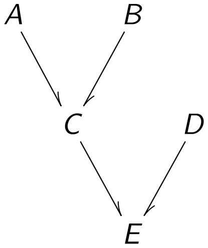
$$A, B < C, D < E$$

Complex causal structures

CNA not only searches for atomic causal models with one designated outcome, but also for **complex causal structures** with multiple outcomes.



$(A + B \leftrightarrow C) * (B + D \leftrightarrow E)$



$(A + B \leftrightarrow C) * (C + D \leftrightarrow E)$

Further Resources on QCA

Thiem, Alrik. 2016. *QCApro: Professional functionality for performing and evaluating Qualitative Comparative Analysis*. R package version 1.1-1. URL: `http://www.alrik-thiem.net/software/`.

- R extension package for performing QCA

Thiem, Alrik. 2016. "Conducting configurational comparative research with Qualitative Comparative Analysis: A hands-on tutorial for applied evaluation scholars and practitioners." *American Journal of Evaluation*. DOI: 10.1177/1098214016673902.

- provides an article-length tutorial for QCA using a recent evaluation of patient follow-through effectiveness in Lynch syndrome tumor-screening programs; includes a replication file for the QCApro package

Homepage: `http://www.alrik-thiem.net`

- publications, replication files, blog texts, video tutorials, etc. on QCA

ResearchGate: `https://www.researchgate.net/profile/Alrik_Thiem`

- publications, conference papers, presentation slides on QCA

Further Resources on CNA

Homepage: `http://www.unige.ch/lettres/baumgartner/`

pre-publication texts for download

Ambühl, Mathias, Michael Baumgartner, Ruedi Epple, Alexis Kauffmann, and Alrik Thiem. 2015. *cna: A Package for Coincidence Analysis*. R package version 1.0-3. URL: `http://cran.r-project.org/package=cna`.

R extension package for performing CNA

Baumgartner, Michael. 2009. "Inferring Causal Complexity." *Sociological Methods & Research* 38 (1):71-101.

introduces theoretical background and technicalities of CNA

Baumgartner, Michael, and Alrik Thiem. 2015. "Identifying complex causal dependencies in configurational data with Coincidence Analysis." *The R Journal* 7 (1):176-84.

introduces cna package and its major functions

General Resources on Configurational Methods

Thiem, Alrik, Michael Baumgartner, and Damien Bol. 2016. "Still lost in translation! A correction of three misunderstandings between configurational comparativists and regressional analysts." *Comparative Political Studies* 49 (6):742-74.

explains differences between configurational and regressional methods

Baumgartner, Michael. 2008. "Regularity theories reassessed." *Philosophia* 36 (3):327-54.

article on theoretical framework of QCA and CNA

COMPASSS website: `http://www.compasss.org/`

provides working paper series and bibliography on configurational methods

Contact Information

Michael Baumgartner
michael.baumgartner@unige.ch
http://www.unige.ch/lettres/baumgartner/

Alrik Thiem
alrik.thiem@gmail.com
http://www.alrik-thiem.net/