

# **A** approach to factor selection in CNA



**bottom-up**

Edward J. Miech

Coincidence Analysis intensive  
Prague

16 May 2023

# Overview

- ▶ **brief description of bottom-up approach to factor selection**
- ▶ **optional routine - for your consideration/FYI**
- ▶ **still work in progress**
- ▶ **originally developed out of necessity for personal use**
- ▶ **fully functional R scripts provided for you in Dropbox folder**
- ▶ **you could apply to your own data tonight**

# Overview

- ▶ **exploratory data analysis phase that occurs before modeling**
- ▶ **routine (quickly) produces detailed table for your review**
- ▶ **“wide-angle” view of conjuncts with strongest apparent connections to an outcome of interest**
- ▶ **mathematical output still needs to be interpreted by analyst using theory, background knowledge, case familiarity, relevance to research question, logic, common sense**

# The Problem

- ▶ **original dataset contains large number of factors**

# The Problem

- ▶ **original dataset contains large number of factors**
- ▶ **all factors have plausible connection to outcome**

# The Problem

- ▶ **original dataset contains large number of factors**
- ▶ **all factors have plausible connection to outcome**
- ▶ **too many factors to model at once**

# The Problem

- ▶ **original dataset contains large number of factors**
- ▶ **all factors have plausible connection to outcome**
- ▶ **too many factors to model at once**
- ▶ **need way to select subset of factors to use in modeling**

# Prior Guidance

- ▶ **QCA: select factors based on “theoretical grounds alone”**



# Prior Guidance

- ▶ **QCA: select factors based on “theoretical grounds alone”**
- ▶ **deductive, top-down strategy**

# Prior Guidance

- ▶ **QCA: select factors based on “theoretical grounds alone”**
- ▶ **deductive, top-down strategy**
- ▶ **entirely theoretical**

# Prior Guidance

- ▶ **QCA: select factors based on “theoretical grounds alone”**
- ▶ **deductive, top-down strategy**
- ▶ **entirely theoretical**
- ▶ **no mathematical output to inform factor selection process**

# Issues with Prior Guidance

- ▶ **throwing away potentially valuable information**

# Issues with Prior Guidance

- ▶ **throwing away potentially valuable information**
- ▶ **replicability: would other researchers make same choice?**

# Issues with Prior Guidance

- ▶ **throwing away potentially valuable information**
- ▶ **replicability: would other researchers make same choice?**
- ▶ **justifying decisions about factor selection**

# Issues with Prior Guidance

- ▶ **throwing out information**
- ▶ **replicability**
- ▶ **justifying decisions about factor selection**
- ▶ **considering factors in isolation, not in “bundles”**

# Factor Selection: Other Methods

- ▶ **Factor Analysis**
- ▶ **Principal Component Analysis**
- ▶ **Random Forest**
- ▶ **Blinder-Oaxaca Decomposition**
- ▶ **Conditional Inference Trees**
- ▶ **etc**



# Factor Selection: Other Methods

- ▶ **Factor Analysis**
- ▶ **Principal Component Analysis**
- ▶ **Random Forest**
- ▶ **Blinder-Oaxaca Decomposition**
- ▶ **Conditional Inference Trees**
- ▶ **based on correlation and regression**
- ▶ **products of “net effects” thinking, interventionist framework**

# The Goal

**A configurational approach to factor selection in CNA that is:**

- ▶ **inductive and bottom-up**
- ▶ **systematic and comprehensive**
- ▶ **replicable**
- ▶ **transparent**
- ▶ **case-based**
- ▶ **pragmatic**

# The Goal

**A configurational approach to factor selection in CNA that:**

- ▶ **operates within a regularity framework**
- ▶ **uses all the information in the original dataset**
- ▶ **generates mathematical output for analyst to review**
- ▶ **considers factors in bundles, not just in isolation**
- ▶ **works with crisp-set, multi-value and fuzzy-set datasets**

**original approach: applied example**

# May 2020 Article in Medical Care

**e** APPLIED METHODS

OPEN

## Strategy Configurations Directly Linked to Higher Hepatitis C Virus Treatment Starts

*An Applied Use of Configurational Comparative Methods*

Vera Yakovchenko, MPH, MS,\* Edward J. Miech, EdD,†‡ Matthew J. Chinman, PhD,§||  
Maggie Chartier, PsyD, MPH,¶ Rachel Gonzalez, MPH,# JoAnn E. Kirchner, MD,\*\*  
Timothy R. Morgan, MD,# Angela Park, PharmD,†† Byron J. Powell, PhD,‡‡  
Enola K. Proctor, PhD,‡‡ David Ross, MD, PhD, MBI,¶ Thomas J. Waltz, PhD,§§|||  
and Shari S. Rogal, MD, MPH§¶¶##

**Background:** The Department of Veterans Affairs (VA) cares for more patients with hepatitis C virus (HCV) than any other US health care system. We tracked the implementation strategies that VA sites used to implement highly effective new treatments for HCV with the aim of uncovering how combinations of implementation strategies influenced the uptake of the HCV treatment innovation. We applied

**Results:** From the 73 possible implementation strategies, CCMs identified 5 distinct strategy configurations, or “solution paths.” These were comprised of 10 individual strategies that collectively explained 80% of the sites with higher HCV treatment starts with 100% consistency. Using any one of the following 5 solution paths was sufficient to produce higher treatment starts: (1) technical assistance; (2) engaging in a learning collaborative AND designating

# 73 Implementation Strategies (!!)

Powell et al. *Implementation Science* (2015) 10:21  
DOI 10.1186/s13012-015-0209-1



RESEARCH

Open Access

## A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project

Byron J Powell<sup>1\*</sup>, Thomas J Waltz<sup>2</sup>, Matthew J Chinman<sup>3,4</sup>, Laura J Damschroder<sup>5</sup>, Jeffrey L Smith<sup>6</sup>, Monica M Matthieu<sup>6,7</sup>, Enola K Proctor<sup>8</sup> and JoAnn E Kirchner<sup>6,9</sup>

a final compilation of 73

### Abstract

**Background:** Identifying, developing, and testing implementation strategies are important goals of implementation science. However, these efforts have been complicated by the use of inconsistent language and inadequate descriptions of implementation strategies in the literature. The Expert Recommendations for Implementing Change (ERIC) study aimed to refine a published compilation of implementation strategy terms and definitions by systematically gathering input from a wide range of stakeholders with expertise in implementation science and clinical practice.

**Methods:** Purposive sampling was used to recruit a panel of experts in implementation and clinical practice who engaged in three rounds of a modified Delphi process to generate consensus on implementation strategies and definitions. The first and second rounds involved Web-based surveys soliciting comments on implementation strategy terms and definitions. After each round, iterative refinements were made based upon participant feedback. The third round involved a live polling and consensus process via a Web-based platform and conference call.

**Results:** Participants identified substantial concerns with 31% of the terms and/or definitions and suggested five additional strategies. Seventy-five percent of definitions from the originally published compilation of strategies were retained after voting. Ultimately, the expert panel reached consensus on a final compilation of 73 implementation strategies.

# VA Project

- ▶ **80 different VA medical centers around the United States**
- ▶ **data for each VA facility for 70 implementation strategies**
- ▶ **was each specific implementation strategy used as part of implementation of the local Hepatitis C virus treatment program at that VA facility? (1=YES; 0=NO)**
- ▶ **NOTE: implementation strategies often weak by themselves**
- ▶ **most effective when combined with other implementation strategies in a larger “bundle”**

# VA Project

- ▶ **joined project very late after data collection was complete**
- ▶ **goal: conduct CNA analysis to identify bundles of implementation strategies linked to outcome of interest (higher HCV treatment starts)**
- ▶ **original data matrix: 80 rows by 70 columns**





# Specific Problem: Factor Selection

- ▶ **no compelling a priori theoretical reasons to select certain implementation strategies over others**
- ▶ **conditional formatting (e.g. “heat maps”) of limited use**
- ▶ **strong theoretical and practical sense that “strategy bundles” somehow at work**





OUT	S24	S34	S45	S18	S47	S70	S22	S61	S56	S71
1	1	1	1	1	1	1	1	1	1	1
1	0	1	1	1	1	1	1	1	1	1
1	0	1	1	1	0	1	1	1	1	1
1	1	1	1	1	1	1	1	0	1	1
1	1	1	0	1	1	1	0	0	1	1
1	0	1	1	1	0	1	1	0	1	1
1	1	1	1	1	0	1	1	1	1	0
1	1	1	1	1	1	1	1	1	0	1
1	0	1	0	0	0	1	1	1	0	1
1	0	1	1	1	1	1	0	1	0	1
1	0	1	1	1	1	1	0	0	0	1
1	0	1	0	1	1	1	1	0	0	1
1	0	1	0	1	1	1	0	0	0	1
1	1	1	1	1	0	1	0	1	0	1
1	0	1	1	1	0	1	0	0	0	1
1	1	1	1	0	1	1	1	0	0	1
1	0	1	1	1	1	1	0	1	0	0
1	0	0	0	0	0	1	0	0	1	1
1	0	0	1	1	1	1	0	0	0	1
1	0	0	0	1	1	1	0	0	0	1
1	0	0	0	1	1	1	0	0	0	1
1	1	0	0	0	1	1	1	1	0	0
1	0	0	0	0	1	0	0	0	1	1
1	0	0	0	0	0	0	0	0	0	1
1	0	0	0	0	0	0	0	0	0	1
1	0	0	0	0	0	0	0	0	0	1
1	0	0	0	0	0	0	0	0	0	1
1	1	1	1	0	1	1	1	1	1	1
1	1	1	1	1	1	1	1	0	0	1
1	0	1	1	1	1	1	0	1	0	1
1	0	1	0	1	0	1	0	0	0	1
1	0	1	0	1	0	0	1	1	0	1
1	1	1	0	0	0	1	0	1	0	1
1	0	1	0	0	0	1	0	0	0	1
1	0	0	0	0	1	1	0	0	1	1
1	1	0	0	0	0	1	0	0	0	1
1	0	0	0	0	1	1	1	1	0	0
1	0	0	0	0	0	1	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	1	0	1	0	1
0	0	1	0	1	0	0	0	0	0	0

# Factor Selection and “msc” output

- ▶ **conduct msc analysis on entire dataset**
- ▶ **look across all 80 cases and 70 factors at once**
- ▶ **consider all 1-condition, 2-condition and 3-condition configurations instantiated in dataset that meet consistency threshold**
- ▶ **examine mathematical output to identify strategy configurations with strongest apparent connections to outcome**

# Factor Selection and “msc” output

- ▶ **within 1-condition mscs, examine msc with highest coverage score**
- ▶ **look for mscs that explain sizable chunk of positive cases AND have separation**
  - separation = when top-scoring msc stands out because it has substantially higher coverage score than next-nearest neighbor
- ▶ **repeat with 2- and 3-conditions mscs**
- ▶ **iteratively lower consistency threshold as needed & repeat**

# Factor Selection and “msc” output

- ▶ **using theory and background knowledge, interpret msc output to identify initial subset of conditions to include in modeling**
- ▶ **conjuncts represent the “building blocks” of CNA models**
- ▶ **analogy: detecting “signal” within noisy context**



# Special Note

- ▶ **for theoretical reasons, project team interested only in presence of implementation strategies**
- ▶ **in this dataset, absence of implementation strategy not considered informative**
- ▶ **for this project only: focused on mscs with no negated conditions**

# Condition Table

1	condition	consistency	coverage	complexity
2	S24->OUT	1	0.3	1
3	S17->OUT	1	0.1	1
4	S07->OUT	1	0.075	1
5	S08->OUT	1	0.075	1
6	S44->OUT	1	0.075	1
7	S57->OUT	1	0.05	1
8	S34*S45->OUT	1	0.4	2
9	S28*S55->OUT	1	0.3	2
10	S55*S70->OUT	1	0.3	2
11	S22*S47->OUT	1	0.275	2
12	S22*S61->OUT	1	0.275	2
13	S09*S26->OUT	1	0.25	2
14	S22*S39->OUT	1	0.25	2
15	S27*S55->OUT	1	0.25	2
16	S37*S54->OUT	1	0.25	2
17	S43*S54->OUT	1	0.25	2
18	S54*S67->OUT	1	0.25	2
19	S54*S70->OUT	1	0.25	2
20	S56*S70->OUT	1	0.25	2
21	S56*S71->OUT	1	0.25	2

# Condition Table

1	condition	consistency	coverage	complexity
2	S24->OUT	1	0.3	1
3	S17->OUT	1	0.1	1
4	S07->OUT	1	0.075	1
5	S08->OUT	1	0.075	1
6	S44->OUT	1	0.075	1
7	S57->OUT	1	0.05	1
8	S34*S45->OUT	1	0.4	2
9	S28*S55->OUT	1	0.3	2
10	S55*S70->OUT	1	0.3	2
11	S22*S47->OUT	1	0.275	2
12	S22*S61->OUT	1	0.275	2
13	S09*S26->OUT	1	0.25	2
14	S22*S39->OUT	1	0.25	2
15	S27*S55->OUT	1	0.25	2
16	S37*S54->OUT	1	0.25	2
17	S43*S54->OUT	1	0.25	2
18	S54*S67->OUT	1	0.25	2
19	S54*S70->OUT	1	0.25	2
20	S56*S70->OUT	1	0.25	2
21	S56*S71->OUT	1	0.25	2

# Condition Table

1	condition	consistency	coverage	complexity
259	S18*S47*S70->OUT	1	0.4	3
260	S19*S47*S70->OUT	1	0.375	3
261	S20*S47*S70->OUT	1	0.375	3
262	S27*S45*S70->OUT	1	0.375	3
263	S27*S32*S45->OUT	1	0.35	3
264	S47*S70*S73->OUT	1	0.35	3
265	S01*S27*S34->OUT	1	0.325	3
266	S09*S28*S34->OUT	1	0.325	3
267	S09*S37*S70->OUT	1	0.325	3
268	S18*S34*S65->OUT	1	0.325	3
269	S18*S66*S70->OUT	1	0.325	3
270	S19*S22*S31->OUT	1	0.325	3
271	S22*S31*S34->OUT	1	0.325	3
272	S22*S31*S43->OUT	1	0.325	3
273	S25*S45*S52->OUT	1	0.325	3
274	S25*S45*S67->OUT	1	0.325	3
275	S25*S45*S70->OUT	1	0.325	3
276	S26*S27*S48->OUT	1	0.325	3
277	S26*S43*S70->OUT	1	0.325	3

# Condition Table

1	condition	consistency	coverage	complexity
259	S18*S47*S70->OUT	1	0.4	3
260	S19*S47*S70->OUT	1	0.375	3
261	S20*S47*S70->OUT	1	0.375	3
262	S27*S45*S70->OUT	1	0.375	3
263	S27*S32*S45->OUT	1	0.35	3
264	S47*S70*S73->OUT	1	0.35	3
265	S01*S27*S34->OUT	1	0.325	3
266	S09*S28*S34->OUT	1	0.325	3
267	S09*S37*S70->OUT	1	0.325	3
268	S18*S34*S65->OUT	1	0.325	3
269	S18*S66*S70->OUT	1	0.325	3
270	S19*S22*S31->OUT	1	0.325	3
271	S22*S31*S34->OUT	1	0.325	3
272	S22*S31*S43->OUT	1	0.325	3
273	S25*S45*S52->OUT	1	0.325	3
274	S25*S45*S67->OUT	1	0.325	3
275	S25*S45*S70->OUT	1	0.325	3
276	S26*S27*S48->OUT	1	0.325	3
277	S26*S43*S70->OUT	1	0.325	3

	CASE	OUT	Path 1	Path 2	Path 3	Path 4	Path 5	Covered by Solution	
			S24	S34 S45	S22 S61	S56 S71	S18 S47 S70		
Sites with Higher HCV Treatment Starts (N=40)	SITE 21			• •			• • •	•	
	SITE 01		•				• • •	•	
	SITE 68		•				• • •	•	
	SITE 24				• •			•	
	SITE 57			• •			• • •	•	
	SITE 02			• •			• • •	•	
	SITE 18		•		• •			•	
	SITE 04								•
	SITE 59						• • •	•	
	SITE 26		•		• •				•
	SITE 03								•
	SITE 34			• •	• •		• •		•
	SITE 39					• •			•
	SITE 41							• • •	•
	SITE 47			• •	• •		• •		•
	SITE 72			• •	• •	• •	• •	• • •	•
	SITE 13					• •	• •		•
	SITE 33								•
	SITE 23		•						•
	SITE 45							• • •	•
	SITE 27			• •	• •				•
	SITE 71						• •		•
	SITE 60		•	• •	• •				•
	SITE 40			• •	• •	• •	• •		•
	SITE 77		•	• •	• •		• •	• •	•
	SITE 30				• •	• •			•
	SITE 32			• •	• •			• • •	•
	SITE 17						• •		•
	SITE 55		•	• •	• •	• •	• •	• • •	•
	SITE 80							• • •	•
	SITE 10							• • •	•
	SITE 79		•	• •	• •	• •	• •	• • •	•
	SITE 76			• •	• •	• •	• •	• • •	•
	SITE 15		•	• •	• •				•
	SITE 54								•
	SITE 58					• •			•
	SITE 14		•	• •	• •		• •	• •	•
	SITE 44								•
	SITE 63								•
	SITE 78							• • •	•
Treatment Starts (N=40)	SITE 58								
	SITE 20								
	SITE 67								
	SITE 50								
	SITE 25								
	SITE 12								
	SITE 19								
	SITE 52								
	SITE 64								
	SITE 70								
	SITE 61								
	SITE 49								
	SITE 43								
	SITE 09								
	SITE 75								
	SITE 69								
SITE 74									

					CASE	OUT	Path 1 S24	Path 2 S34 S45	Path 3 S22 S61	Path 4 S56 S71	Path 5 S18 S47 S70	Covered by Solution					
					SITE 21												
					SITE 04												
					SITE 68												
					SITE 24												
					SITE 57												
					SITE 02												
					SITE 18												
					SITE 04												
					SITE 59												
					SITE 26												
					SITE 03												
					SITE 34												
					SITE 39												
					SITE 41												
					SITE 47												
					SITE 72												
					SITE 13												
					SITE 33												
					SITE 23												
					SITE 45												
					SITE 27												
					SITE 71												
					SITE 60												
					SITE 40												
					SITE 77												
					SITE 30												
					SITE 32												
					SITE 17												
					SITE 55												
					SITE 80												
					SITE 10												
					SITE 79												
					SITE 76												
					SITE 15												
					SITE 54												
					SITE 58												
					SITE 14												
					SITE 44												
					SITE 63												
					SITE 78												
1	condition	consistency	coverage	complexity													
2	S24->OUT	1	0.3	1													
3	S17->OUT	1	0.1	1													
4	S07->OUT	1	0.075	1													
5	S08->OUT	1	0.075	1													
6	S44->OUT	1	0.075	1													
7	S57->OUT	1	0.05	1													
8	S34*S45->OUT	1	0.4	2													
9	S28*S32->OUT	1	0.3	2													
10	S55*S70->OUT	1	0.3	2													
11	S22*S47->OUT	1	0.275	2													
12	S22*S61->OUT	1	0.275	2													
13	S09*S26->OUT	1	0.25	2													
14	S22*S39->OUT	1	0.25	2													
15	S27*S55->OUT	1	0.25	2													
16	S37*S54->OUT	1	0.25	2													
17	S43*S54->OUT	1	0.25	2													
18	S54*S67->OUT	1	0.25	2													
19	S54*S70->OUT	1	0.25	2													
20	S56*S70->OUT	1	0.25	2													
21	S56*S71->OUT	1	0.25	2													
					<b>Sites with Higher HCV Treatment Starts (N=40)</b>												
1	condition	consistency	coverage	complexity													
259	S18*S47*S70->OUT	1	0.4	3													
260	S19*S47*S70->OUT	1	0.375	3													
261	S20*S47*S70->OUT	1	0.375	3													
262	S27*S45*S70->OUT	1	0.375	3													
263	S27*S32*S45->OUT	1	0.35	3													
264	S47*S70*S73->OUT	1	0.35	3													
265	S01*S27*S34->OUT	1	0.325	3													
266	S09*S28*S34->OUT	1	0.325	3													

# Cross-Validation of Results

- ▶ **set out to cross-validate results using Logic Regression**
- ▶ **fundamentally different method**



# Logic Regression

- ▶ **form of regression specifically designed to identify Boolean dependencies in very large datasets in genetics**
- ▶ **developed by biostatisticians**
- ▶ **explicitly states limitations of traditional linear regression for identifying combinations of conditions linked to an outcome**
- ▶ **main article published in 2003 with >400 citations**
- ▶ **uses “simulated annealing” approach**

# Logic Regression

- ▶ **CCMs and LR both belong to a higher-order family of methods targeting Boolean conjunctivity and disjunctivity**
- ▶ **The two approaches apply distinct tools, implement distinct fit measures, are built for distinct data scenarios, and up to now have been applied in distinct, non-overlapping areas of research**
- ▶ **There are data scenarios in which both methods are applicable, creating new potential for cross-validation**

# Cross-Validation Example



- ▶ **using Logic Regression in conjunction with FR score routine, generated ~ 1000 models with no negations using exact same dataset used in Medical Care paper**
- ▶ **results directly confirmed three paths in model**
- ▶ **top-scoring model consists of three of exactly the same configurations identified in the earlier article, consisting of a one-, two- and three-condition conjunct**
- ▶ **furthermore, the top four models all share a model-submodel relationship**

	condition	FR score
403	S18*S47*S70+S24+S56*S71<->OUT	1
644	S18*S47+S24+S56*S71<->OUT	0.921899
636	S18*S47*S70+S24+S56<->OUT	0.771822
639	S18*S47*S70+S24<->OUT	0.770291
1	S14*S56+S18*S47*S70+S24<->OUT	0.765697
206	S14*S56+S18*S47+S24<->OUT	0.614089
847	S24*S70+S47*S70+S56<->OUT	0.459418
339	S18*S47*S70+S24*S70+S56*S70<->OUT	0.382848
641	S18*S47+S18*S0I+S22*S47+S22*S0I+S24+S56*S71<->OUT	0.287902
845	S24*S70+S47*S70+S56*S70<->OUT	0.283308
835	S24*S70+S43*S47*S70+S56*S70<->OUT	0.185299
200	S14*S56+S18*S47+S18*S0I+S22*S47+S22*S0I+S24<->OUT	0.163859
205	S14*S56+S18*S47+S24+S34*S0I+S55<->OUT	0.151608
967	S24+S27*S45*S70+S27*S54*S70+S56*S71<->OUT	0.073507
1018	S24+S45*S70+S54*S70+S56<->OUT	0.058193
1012	S24+S45*S70+S54*S70+S56*S71<->OUT	0.050536
1003	S24+S34*S45+S54<->OUT	0.049005
320	S18*S26+S18*S54+S24+S56*S71<->OUT	0.045942
1020	S24+S54*S70+S56*S70<->OUT	0.045942
837	S24*S70+S45*S70+S54*S70+S56*S70<->OUT	0.041348
965	S24*S70+S54*S70+S56*S70<->OUT	0.029096
829	S24*S70+S34*S45*S70+S54*S70<->OUT	0.019908
843	S24*S70+S45*S70+S54*S70<->OUT	0.019908

	CASE	OUT	Path 1	Path 2	Path 3	Path 4	Path 5	Covered by Solution	
			S24	S34 S45	S22 S61	S56 S71	S18 S47 S70		
Sites with Higher HCV Treatment Starts (N=40)	SITE 21			•				•	
	SITE 01		•					•	
	SITE 68		•					•	
	SITE 24				•	•		•	
	SITE 57			•	•			•	
	SITE 02			•	•			•	
	SITE 18		•			•	•	•	
	SITE 04							•	
	SITE 59						•	•	•
	SITE 26		•		•	•			•
	SITE 03								•
	SITE 34			•	•		•	•	•
	SITE 39					•	•		•
	SITE 41							•	•
	SITE 47				•	•			•
	SITE 72				•	•		•	•
	SITE 13						•	•	•
	SITE 33								•
	SITE 23		•						•
	SITE 45							•	•
	SITE 27				•	•			•
	SITE 71						•	•	•
	SITE 60		•		•	•			•
	SITE 40			•	•	•	•		•
	SITE 77		•		•	•		•	•
	SITE 30				•	•			•
	SITE 32				•	•		•	•
	SITE 17						•	•	•
	SITE 55		•		•	•		•	•
	SITE 80							•	•
	SITE 10							•	•
	SITE 79		•		•	•		•	•
	SITE 76				•	•		•	•
	SITE 15		•		•	•			•
SITE 54								•	
SITE 58				•	•			•	
SITE 14		•		•	•		•	•	
SITE 44								•	
SITE 63								•	
SITE 78							•	•	
ment Starts (N=40)	SITE 58								
	SITE 20								
	SITE 67								
	SITE 50								
	SITE 25								
	SITE 12								
	SITE 19								
	SITE 52								
	SITE 64								
	SITE 70								
	SITE 61								
	SITE 49								
	SITE 43								
	SITE 09								
	SITE 75								
	SITE 69								
SITE 74									

# **2<sup>nd</sup> validation study**

# BMJ Open Pairing regression and configurational analysis in health services research: modelling outcomes in an observational cohort using a split-sample design

Edward J Miech <sup>1,2</sup>, Anthony J Perkins,<sup>3</sup> Ying Zhang,<sup>4</sup> Laura J Myers,<sup>1,2</sup> Jason J Sico,<sup>5,6</sup> Joanne Daggy,<sup>3</sup> Dawn M Bravata <sup>1,2</sup>

**To cite:** Miech EJ, Perkins AJ, Zhang Y, *et al.* Pairing regression and configurational analysis in health services research: modelling outcomes in an observational cohort using a split-sample design. *BMJ Open* 2022;**12**:e061469. doi:10.1136/bmjopen-2022-061469

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2022-061469>).

Received 28 January 2022  
Accepted 12 May 2022

## ABSTRACT

**Background** Configurational methods are increasingly being used in health services research.

**Objectives** To use configurational analysis and logistic regression within a single data set to compare results from the two methods.

**Design** Secondary analysis of an observational cohort; a split-sample design involved randomly dividing patients into training and validation samples.

**Participants and setting** Patients who had a transient ischaemic attack (TIA) in US Department of Veterans Affairs hospitals.

**Measures** The patient outcome was the combined endpoint of all-cause mortality or recurrent ischaemic stroke within 1 year post-TIA. The quality-of-care outcome was the without-fail rate (proportion of patients who received all processes for which they were eligible, among seven processes).

**Results** For the recurrent stroke or death outcome, configurational analysis yielded a three-pathway model identifying a set of (validation sample) patients where the number was 15.3% (99/559), substantially

## STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Logistic regression and configurational methods (CNA) were applied to the same data to examine similarities and differences in results.
- ⇒ The split-sample approach to development and validation of models is a key methodological strength.
- ⇒ The results are based on data from the Department of Veterans Affairs and may not generalise to other healthcare systems.

## INTRODUCTION

Configurational Comparative Methods (CCMs) have been used in a wide variety of disciplines since at least the 1990s and have recently started to gain traction in the general medical research literature<sup>1–4</sup> as well as within implementation science.<sup>5 6</sup> CCMs draw on mathematical approaches that are fundamen-

# 2nd Validation Study

- ▶ **VA study with > 3000 cases**
- ▶ **Original dataset had 48 factors**
- ▶ **randomly split into training sample and validation sample**
- ▶ **derived cna model in training sample, then “tested” against validation sample**
- ▶ **modeled two different outcomes**
- ▶ **cna models for both outcomes were validated using this split-sample design**



**refined msc routine**

# refined msc routine

- ▶ **automates the process of conducting 5 msc runs**
- ▶ **75, 80, 85, 90 and 95% consistency thresholds**
- ▶ **selects the “top 10” mscs per complexity level**
- ▶ **integrates runs into a single Excel workbook**
- ▶ **display all 5 runs side-by-side on a single tab (75to95),**
- ▶ **adds additional reference tabs for each specific run**

**refined approach: applied example**

RESEARCH

Open Access



# Identifying configurations of behavior change techniques in effective medication adherence interventions: a qualitative comparative analysis

Leila Kahwati<sup>1\*</sup>, Meera Viswanathan<sup>1</sup>, Carol E. Golin<sup>2</sup>, Heather Kane<sup>1</sup>, Megan Lewis<sup>1</sup> and Sara Jacobs<sup>1</sup>

## Abstract

**Background:** Interventions to improve medication adherence are diverse and complex. Consequently, synthesizing this evidence is challenging. We aimed to extend the results from an existing systematic review of interventions to improve medication adherence by using qualitative comparative analysis (QCA) to identify necessary or sufficient configurations of behavior change techniques among effective interventions.

**Methods:** We used data from 60 studies in a completed systematic review to examine the combinations of nine behavior change techniques (increasing knowledge, increasing awareness, changing attitude, increasing self-efficacy, increasing intention formation, increasing action control, facilitation, increasing maintenance support, and motivational interviewing) among studies demonstrating improvements in adherence.

**Results:** Among the 60 studies, 34 demonstrated improved medication adherence. Among effective studies, increasing patient knowledge was a necessary but not sufficient technique. We identified seven configurations of behavior change techniques sufficient for improving adherence, which together accounted for 26 (76 %) of the effective studies. The intervention configuration that included increasing knowledge and self-efficacy was the most empirically relevant, accounting for 17 studies (50 %) and uniquely accounting for 15 (44 %).

**Conclusions:** This analysis extends the completed review findings by identifying multiple combinations of behavior change techniques that improve adherence. Our findings offer direction for policy makers, practitioners, and future comparative effectiveness research on improving adherence.

**Keywords:** Medication adherence, Qualitative comparative analysis, Behavior change, Systematic review

# Applied Example

- ▶ **2016 systematic review**
- ▶ **QCA applied in original cross-study analysis**
- ▶ **60 cases**
- ▶ **9 behavioral change techniques**
- ▶ **1 outcome factor: improved medication adherence**

# Applied Example: 9 Factors

---

Knowledge (K)

Facilitation (F)

Awareness (R)

Self-efficacy (S)

Intention formation (I)

Action control (C)

Attitude (T)

Maintenance (M)

Motivational interviewing  
(G)

---

	A	B	C	D	E	F	G	H	I	J	K	L	
1	REF_NUMBER	APPENDIXA	CITATION	KNOWL	SELF_EFF	AWARE	ATT	INT_FORM	ACT_CON	MAINT	FACIL	MOTIV_INT	OUTCOME
2	1		Pyne 2011	1	0	1	0	0	1	0	1	0	0
3	2		Wolever 2010	0	1	0	1	1	0	0	0	0	1
4	3		Bender 2010	1	0	1	0	0	0	0	0	0	1
5	4		Smith 2008	1	0	1	0	0	0	0	0	0	1
6	5		Mann 2010	1	0	1	0	0	0	0	0	0	0
7	6		Montori 2011	1	0	1	0	0	0	0	0	0	0
8	7		Wakefield 2011	1	0	1	0	0	0	0	0	0	0
9	8		Bogner 2010	1	0	0	1	0	0	0	1	0	1
10	9		Wilson 2010	1	1	1	0	1	0	0	1	1	1
11	10		Carter 2009	1	0	0	0	0	0	0	1	0	0
12	11		Okeke 2009	1	0	0	0	0	0	0	1	0	1
13	12		Waalén 2009	1	0	0	0	0	0	0	1	0	1
14	13		Rudd 2009	1	0	0	0	0	0	0	1	0	0
15	14		Lee 2006	1	0	0	0	0	0	0	1	0	1
16	15		Schnipper 2006	1	0	0	0	0	0	0	1	0	0
17	16		Taylor 2003	1	0	0	0	0	0	0	1	0	0
18	17		Nietert 2009	1	0	0	0	0	0	0	1	0	0
19	18		Solomon1998	1	0	0	0	0	0	0	1	0	1
20	19		Janson 2009	1	1	1	0	0	0	0	0	0	1
21	20		Schaffer 2004	1	1	1	0	0	0	0	0	0	1
22	21		Hunt 2008	1	0	1	0	0	0	0	1	0	0
23	22		Bogner 2008	1	0	0	1	0	0	0	1	0	1
24	23		Pearce 2008	0	1	0	0	0	0	0	1	0	0
25	24		Bosworth 2008	1	0	1	1	1	1	0	1	1	1
26	25		Weymiller 2007	1	0	0	0	0	0	0	0	0	0
27	26		Ross 2004	1	0	0	0	0	0	0	0	0	1
28	27		Powell 1995	1	0	0	0	0	0	0	0	0	0
29	28		Murray 2007	1	1	0	0	0	0	0	1	0	1
30	29		Katon 1996	1	1	0	0	0	0	0	1	0	1
31	30		Katon 1995	1	1	0	0	0	0	0	1	0	1
32	31		Katon 1999	1	1	0	0	0	0	0	1	0	1
33	32		Sledge 2006	0	0	0	0	0	0	0	1	0	0
34	33		Johnson 2006a	1	1	1	1	0	0	1	0	0	1
35	34		Simon 2006	1	0	1	0	0	0	0	1	1	0
36	35		Johnson 2006b	1	1	1	1	0	0	0	0	0	1
37	36		Wu 2012	1	1	1	1	0	0	0	0	0	1
38	37		Lin 2006	0	0	0	0	1	0	1	0	0	0
39	38		Berger 2005	0	0	0	0	0	0	0	0	1	1
40	39		Bosworth 2005	1	0	1	0	1	1	1	1	0	0
41	40		Rudd	1	1	1	0	1	0	1	1	0	1
42	41		Janson 2003	1	1	1	0	0	0	1	0	0	1
43	42		Grant 2003	1	0	0	0	0	0	1	0	0	0
44	43		Hoffman 2003	1	0	0	0	0	1	0	0	0	1
45	44		Weinberger 2002	1	0	0	0	0	1	0	0	0	0
46	45		Vivian 2002	1	0	1	0	1	0	1	1	0	0
47	46		Guthrie 2001	1	0	1	0	0	1	0	0	0	0
48	47		Katon 2001	1	1	1	0	1	1	1	0	1	1
49	48		Fulomer 1999	0	0	0	0	0	1	0	0	0	1
50	49		Berg 1997	1	1	0	0	0	0	0	0	0	1
51	50		Rich 1996	1	0	1	0	1	1	0	1	0	1
52	51		Schectman 1994	0	1	1	0	0	0	0	1	0	0
53	52		Stacy 2009	1	1	1	1	1	0	1	0	0	1
54	53		Rickles 2005	1	0	1	0	1	0	0	1	0	0
55	54		Capoccia 2004	1	0	1	0	1	0	0	1	0	0
56	55		Powers 2011	1	0	1	1	0	0	0	0	0	0
57	56		Friedman 1996	1	0	1	0	0	0	0	0	1	1
58	57		Bogner 2012	1	0	0	1	0	0	0	1	0	1
59	58		Ogedegbe 2012	1	1	0	1	1	0	0	0	1	1
60	59		Solomon 2012	1	0	1	1	0	0	0	1	1	0
61	60		Simon 2004	1	1	0	0	1	0	0	1	1	1

# Published Model

**Table 3** Intermediate solution for configurations of behavior change techniques (BCTs) used in effective interventions to improve medication adherence

## Solution parameters

Consistency<sup>a</sup>: 100 % (34 studies)

Coverage<sup>b</sup>: 76 % total (26 studies)

68 % unique (23 studies)

8 % overlapping (3 studies)

Truth table rows covered: 7.2 % (37 of 512 rows)

Number of studies with outcome not covered<sup>c</sup>: 8

Bender et al. 2010 [46], Hoffman et al. 2003 [47], Lee et al. 2006 [48],

Okeke et al. 2009 [49], Ross et al. 2004 [50], Smith et al. 2008 [51],

Solomon et al. 1998 [52], Waalen et al. 2009 [53]

Configuration <sup>d</sup>	Consistency (%)	Raw coverage % (# of cases)	Unique coverage % (# of cases)	Study
KS	100	50 (17)	44 (15)	<sup>e</sup> Berg et al. 1997 [24], <sup>e</sup> Janson et al. 2003 [18], <sup>e</sup> Janson et al. 2009 [17], <sup>e</sup> Johnson et al. 2006a [29], <sup>e</sup> Johnson et al. 2006b [28], <sup>e</sup> Katon et al. 1995 [26], <sup>e</sup> Katon et al. 1999/Katon et al. 2002 [25], <sup>e</sup> Katon et al. 1996 [21], <sup>e</sup> Katon et al. 2001/Ludman et al. 2003/Van Korff et al. 2003 [22], <sup>e</sup> Murray et al. 2007 [23], <sup>e</sup> Ogedegbe et al. 2012 [54], <sup>e</sup> Rudd et al. 2004 [19], <sup>e</sup> Schaffer et al. 2004 [33], <sup>e</sup> Simon et al. 2004 [30], <sup>e</sup> Stacy et al. 2009 [32], <sup>e</sup> Wilson et al. 2010 [20], <sup>e</sup> Wu et al. 2012 [31]
fG	100	12 (4)	6 (2)	<sup>e</sup> Berger et al. 2005 [35], <sup>e</sup> Friedman et al. 1996 [34], <sup>e</sup> Katon et al. 2001 [22], <sup>e</sup> Ogedegbe et al. 2012 [54]
rSIT	100	6 (2)	0 (0)	<sup>e</sup> Ogedegbe et al. 2012 [54], <sup>e</sup> Wolever et al. 2010 [55]
kfCm	100	3 (1)	3 (1)	<sup>e</sup> Fulmer et al. 1999 [43]
fSmIT	100	6 (2)	0 (0)	<sup>e</sup> Ogedegbe et al. 2012 [54], <sup>e</sup> Wolever et al. 2010 [55]
KRFICm	100	6 (2)	6 (2)	<sup>e</sup> Bosworth et al. 2008 [37], <sup>e</sup> Rich et al. 1996 [38]
KrFT	100	9 (3)	9 (3)	<sup>e</sup> Bogner et al. 2012 [40], <sup>e</sup> Bogner et al. 2010 [39], <sup>e</sup> Bogner et al. 2008 [41]



# Published Model

$$(K=1 * S=1)$$

# Published Model

$$(K=1 * S=1) + (F=0 * G=1)$$

# Published Model

$$(K=1 * S=1) + (F=0 * G=1) + (R=0 * S=1 * I=1 * T=1)$$

# Published Model

$$(K=1 * S=1) + (F=0 * G=1) + (R=0 * S=1 * I=1 * T=1) + \\ (K=0 * F=0 * C=1 * M=0)$$

# Published Model

$$(K=1 * S=1) + (F=0 * G=1) + (R=0 * S=1 * I=1 * T=1) + \\ (K=0 * F=0 * C=1 * M=0) + (F=0 * S=1 * M=0 * I=1 * T=1)$$

# Published Model

$$\begin{aligned} & (K=1 * S=1) + (F=0 * G=1) + (R=0 * S=1 * I=1 * T=1) + \\ & (K=0 * F=0 * C=1 * M=0) + (F=0 * S=1 * M=0 * I=1 * T=1) + \\ & (K=1 * R=1 * F=1 * I=1 * C=1 * M=0) \end{aligned}$$

# Published Model

$$\begin{aligned} & (K=1 * S=1) + (F=0 * G=1) + (R=0 * S=1 * I=1 * T=1) + \\ & (K=0 * F=0 * C=1 * M=0) + (F=0 * S=1 * M=0 * I=1 * T=1) + \\ & (K=1 * R=1 * F=1 * I=1 * C=1 * M=0) + (K=1 * R=0 * F=1 * T=1) \end{aligned}$$

# Published Model

$$\begin{aligned} & (K=1 * S=1) + (F=0 * G=1) + (R=0 * S=1 * I=1 * T=1) + \\ & (K=0 * F=0 * C=1 * M=0) + (F=0 * S=1 * M=0 * I=1 * T=1) + \\ & (K=1 * R=1 * F=1 * I=1 * C=1 * M=0) + (K=1 * R=0 * F=1 * T=1) \end{aligned}$$

Knowledge (K)

Facilitation (F)

Awareness (R)

Self-efficacy (S)

Intention formation (I)

Action control (C)

Attitude (T)


Maintenance (M)

Motivational interviewing  
(G)



# refined msc approach

- ▶ **creates an Excel workbook on your computer**

- ▶  KAHWATI2016\_ConditionTable\_3\_75to95 study analysis

# refined msc approach

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	outcome	condition	consistency	coverage	complexity		outcome	condition	consistency	coverage	complexity		outcome	condition
1	OUTCOME=0	SELF_EFF=0*MAINT=1->OUTCOME=0	1	0.153846154	2		OUTCOME=0	SELF_EFF=0*MAINT=1->OUTCOME=0	1	0.153846154	2		OUTCOME=0	SELF_EFF=0*MAINT=1->OUTCOME=0
2	OUTCOME=0	KNOWL=0*FACIL=1->OUTCOME=0	1	0.115384615	2		OUTCOME=0	KNOWL=0*FACIL=1->OUTCOME=0	1	0.115384615	2		OUTCOME=0	KNOWL=0*FACIL=1->OUTCOME=0
3	OUTCOME=0	AWARE=0*MAINT=1->OUTCOME=0	1	0.076923077	2		OUTCOME=0	AWARE=0*MAINT=1->OUTCOME=0	1	0.076923077	2		OUTCOME=0	AWARE=0*MAINT=1->OUTCOME=0
4	OUTCOME=0	KNOWL=0*AWARE=1->OUTCOME=0	1	0.038461538	2		OUTCOME=0	KNOWL=0*AWARE=1->OUTCOME=0	1	0.038461538	2		OUTCOME=0	KNOWL=0*AWARE=1->OUTCOME=0
5	OUTCOME=0	KNOWL=0*MAINT=1->OUTCOME=0	1	0.038461538	2		OUTCOME=0	KNOWL=0*MAINT=1->OUTCOME=0	1	0.038461538	2		OUTCOME=0	KNOWL=0*MAINT=1->OUTCOME=0
6	OUTCOME=0	SELF_EFF=0*AWARE=1*MOTIV_INT=0->OUTCOME=0	0.785714286	0.423076923	3		OUTCOME=0	SELF_EFF=0*AWARE=1*FACIL=1->OUTCOME=0	0.8	0.307692308	3		OUTCOME=0	AWARE=1*INT_FORM=0*FACIL=1->OUTCOME=0
7	OUTCOME=0	SELF_EFF=0*AWARE=1*ACT_CON=0->OUTCOME=0	0.769230769	0.384615385	3		OUTCOME=0	AWARE=1*INT_FORM=0*FACIL=1->OUTCOME=0	1	0.192307692	3		OUTCOME=0	SELF_EFF=0*INT_FORM=1*ACT_CON=0->OUTCOME=0
8	OUTCOME=0	SELF_EFF=0*AWARE=1*INT_FORM=0->OUTCOME=0	0.75	0.346153846	3		OUTCOME=0	SELF_EFF=0*ATT=0*INT_FORM=1->OUTCOME=0	0.833333333	0.192307692	3		OUTCOME=0	AWARE=1*INT_FORM=0*ACT_CON=1->OUTCOME=0
9	OUTCOME=0	SELF_EFF=0*AWARE=1*FACIL=1->OUTCOME=0	0.8	0.307692308	3		OUTCOME=0	SELF_EFF=0*INT_FORM=1*MOTIV_INT=0->OUTCOME=0	0.833333333	0.192307692	3		OUTCOME=0	INT_FORM=0*FACIL=1*MOTIV_INT=1->OUTCOME=0
10	OUTCOME=0	AWARE=1*ACT_CON=0*FACIL=1->OUTCOME=0	0.777777778	0.269230769	3		OUTCOME=0	SELF_EFF=0*INT_FORM=1*ACT_CON=0->OUTCOME=0	1	0.153846154	3		OUTCOME=0	KNOWL=0*SELF_EFF=1*ATT=0->OUTCOME=0
11	OUTCOME=0	AWARE=1*FACIL=1*MOTIV_INT=0->OUTCOME=0	0.777777778	0.269230769	3		OUTCOME=0	KNOWL=0*ACT_CON=0*MOTIV_INT=0->OUTCOME=0	0.8	0.153846154	3		OUTCOME=0	KNOWL=0*SELF_EFF=1*INT_FORM=0->OUTCOME=0
12	OUTCOME=0	AWARE=1*INT_FORM=0*FACIL=1->OUTCOME=0	1	0.192307692	3		OUTCOME=0	KNOWL=0*ATT=0*ACT_CON=0->OUTCOME=0	0.8	0.153846154	3		OUTCOME=0	ACT_CON=1*MAINT=1*FACIL=1->OUTCOME=0
13	OUTCOME=0	SELF_EFF=0*ATT=0*INT_FORM=1->OUTCOME=0	0.833333333	0.192307692	3		OUTCOME=0	KNOWL=0*ATT=0*MOTIV_INT=0->OUTCOME=0	0.8	0.153846154	3		OUTCOME=0	ACT_CON=1*MAINT=1*MOTIV_INT=0->OUTCOME=0
14	OUTCOME=0	SELF_EFF=0*INT_FORM=1*MOTIV_INT=0->OUTCOME=0	0.833333333	0.192307692	3		OUTCOME=0	AWARE=1*INT_FORM=0*ACT_CON=1->OUTCOME=0	1	0.076923077	3		OUTCOME=0	ATT=1*INT_FORM=0*MOTIV_INT=1->OUTCOME=0
15	OUTCOME=0	SELF_EFF=0*INT_FORM=1*ACT_CON=0->OUTCOME=0	1	0.153846154	3		OUTCOME=0	INT_FORM=0*FACIL=1*MOTIV_INT=1->OUTCOME=0	1	0.076923077	3		OUTCOME=0	INT_FORM=0*ACT_CON=1*FACIL=1->OUTCOME=0
16	OUTCOME=1	SELF_EFF=1->OUTCOME=1	0.9	0.529411765	1		OUTCOME=1	SELF_EFF=1->OUTCOME=1	0.9	0.529411765	1		OUTCOME=1	SELF_EFF=1->OUTCOME=1
17	OUTCOME=1	ATT=1->OUTCOME=1	0.833333333	0.294117647	1		OUTCOME=1	ATT=1->OUTCOME=1	0.833333333	0.294117647	1		OUTCOME=1	ATT=1*MOTIV_INT=0->OUTCOME=1
18	OUTCOME=1	MOTIV_INT=1->OUTCOME=1	0.777777778	0.205882353	1		OUTCOME=1	INT_FORM=1*MOTIV_INT=1->OUTCOME=1	1	0.147058824	2		OUTCOME=1	ATT=1*FACIL=0->OUTCOME=1
19	OUTCOME=1	INT_FORM=1*MAINT=0->OUTCOME=1	0.75	0.176470588	2		OUTCOME=1	ATT=0*MOTIV_INT=1->OUTCOME=1	0.833333333	0.147058824	2		OUTCOME=1	AWARE=0*ATT=1->OUTCOME=1
20	OUTCOME=1	INT_FORM=1*FACIL=0->OUTCOME=1	0.8	0.117647059	2		OUTCOME=1	FACIL=0*MOTIV_INT=1->OUTCOME=1	1	0.117647059	2		OUTCOME=1	INT_FORM=1*MOTIV_INT=1->OUTCOME=1
21	OUTCOME=1	AWARE=0*INT_FORM=1->OUTCOME=1	0.75	0.088235294	2		OUTCOME=1	INT_FORM=1*FACIL=0->OUTCOME=1	0.8	0.117647059	2		OUTCOME=1	ATT=1*INT_FORM=1->OUTCOME=1
22	OUTCOME=1	INT_FORM=1*ACT_CON=1->OUTCOME=1	0.75	0.088235294	2		OUTCOME=1	AWARE=0*MOTIV_INT=1->OUTCOME=1	1	0.088235294	2		OUTCOME=1	FACIL=0*MOTIV_INT=1->OUTCOME=1
23	OUTCOME=1	KNOWL=0*FACIL=0->OUTCOME=1	0.75	0.088235294	2		OUTCOME=1	ACT_CON=1*MOTIV_INT=1->OUTCOME=1	1	0.058823529	2		OUTCOME=1	AWARE=0*MOTIV_INT=1->OUTCOME=1
24	OUTCOME=1	KNOWL=0*ACT_CON=1->OUTCOME=1	1	0.029411765	2		OUTCOME=1	KNOWL=0*ACT_CON=1->OUTCOME=1	1	0.029411765	2		OUTCOME=1	ACT_CON=1*MOTIV_INT=1->OUTCOME=1
25	OUTCOME=1	AWARE=1*MAINT=1*FACIL=0->OUTCOME=1	1	0.117647059	3		OUTCOME=1	KNOWL=0*MOTIV_INT=1->OUTCOME=1	1	0.029411765	2		OUTCOME=1	ATT=1*MAINT=1->OUTCOME=1
26	OUTCOME=1	AWARE=1*ACT_CON=0*MAINT=1->OUTCOME=1	0.8	0.117647059	3		OUTCOME=1	MAINT=1*MOTIV_INT=1->OUTCOME=1	1	0.029411765	2		OUTCOME=1	ATT=1*ACT_CON=1->OUTCOME=1
27	OUTCOME=1	KNOWL=1*MAINT=1*FACIL=0->OUTCOME=1	0.8	0.117647059	3		OUTCOME=1	AWARE=1*MAINT=1*FACIL=0->OUTCOME=1	1	0.117647059	3		OUTCOME=1	AWARE=1*MAINT=1*FACIL=0->OUTCOME=1
28	OUTCOME=1	AWARE=1*INT_FORM=0*MAINT=1->OUTCOME=1	1	0.058823529	3		OUTCOME=1	AWARE=1*ACT_CON=0*MAINT=1->OUTCOME=1	0.8	0.117647059	3		OUTCOME=1	AWARE=0*INT_FORM=1*MAINT=0->OUTCOME=1
29	OUTCOME=1	ACT_CON=1*MAINT=1*FACIL=0->OUTCOME=1	1	0.029411765	3		OUTCOME=1	KNOWL=1*MAINT=1*FACIL=0->OUTCOME=1	0.8	0.117647059	3		OUTCOME=1	KNOWL=0*MAINT=0*FACIL=0->OUTCOME=1
30							OUTCOME=1	AWARE=0*INT_FORM=1*MAINT=0->OUTCOME=1	1	0.088235294	3		OUTCOME=1	KNOWL=1*INT_FORM=1*FACIL=0->OUTCOME=1
31							OUTCOME=1	KNOWL=0*MAINT=0*FACIL=0->OUTCOME=1	1	0.088235294	3		OUTCOME=1	AWARE=1*INT_FORM=0*MAINT=1->OUTCOME=1
32							OUTCOME=1	AWARE=1*INT_FORM=0*MAINT=1->OUTCOME=1	1	0.058823529	3		OUTCOME=1	AWARE=1*INT_FORM=1*FACIL=0->OUTCOME=1
33							OUTCOME=1	INT_FORM=1*ACT_CON=1*MAINT=0->OUTCOME=1	1	0.058823529	3		OUTCOME=1	INT_FORM=1*ACT_CON=1*MAINT=0->OUTCOME=1
34							OUTCOME=1	KNOWL=0*INT_FORM=0*FACIL=0->OUTCOME=1	1	0.058823529	3		OUTCOME=1	INT_FORM=1*ACT_CON=1*MAINT=0->OUTCOME=1
35							OUTCOME=1	KNOWL=1*AWARE=0*INT_FORM=1->OUTCOME=1	1	0.058823529	3		OUTCOME=1	INT_FORM=1*MAINT=0*FACIL=0->OUTCOME=1
36							OUTCOME=1	KNOWL=1*ACT_CON=1*MAINT=1*FACIL=0->OUTCOME=1	1	0.029411765	3		OUTCOME=1	KNOWL=1*AWARE=0*INT_FORM=1->OUTCOME=1
37							OUTCOME=1	ACT_CON=1*MAINT=1*FACIL=0->OUTCOME=1	1	0.029411765	3		OUTCOME=1	KNOWL=1*AWARE=0*INT_FORM=1->OUTCOME=1



# Interpreting Condition Table

- ▶ **goal – separate signal from noise**
- ▶ **identify “best of class” mscs**
- ▶ **mscs with top coverage score within complexity level**
- ▶ **separation from next-nearest neighbor**
- ▶ **align with theory, background knowledge, case familiarity, logic, common sense**

	A	B	C	D	E
1	outcome	condition	consistency	coverage	complexity
2	OUTCOME=0	SELF_EFF=0*MAINT=1->OUTCOME=0	1	0.153846154	2
3	OUTCOME=0	KNOWL=0*FACIL=1->OUTCOME=0	1	0.115384615	2
4	OUTCOME=0	AWARE=0*MAINT=1->OUTCOME=0	1	0.076923077	2
5	OUTCOME=0	KNOWL=0*AWARE=1->OUTCOME=0	1	0.038461538	2
6	OUTCOME=0	KNOWL=0*MAINT=1->OUTCOME=0	1	0.038461538	2
7	OUTCOME=0	SELF_EFF=0*AWARE=1*MOTIV_INT=0->OUTCOME=0	0.785714286	0.423076923	3
8	OUTCOME=0	SELF_EFF=0*AWARE=1*ACT_CON=0->OUTCOME=0	0.769230769	0.384615385	3
9	OUTCOME=0	SELF_EFF=0*AWARE=1*INT_FORM=0->OUTCOME=0	0.75	0.346153846	3
10	OUTCOME=0	SELF_EFF=0*AWARE=1*FACIL=1->OUTCOME=0	0.8	0.307692308	3
11	OUTCOME=0	AWARE=1*ACT_CON=0*FACIL=1->OUTCOME=0	0.777777778	0.269230769	3
12	OUTCOME=0	AWARE=1*FACIL=1*MOTIV_INT=0->OUTCOME=0	0.777777778	0.269230769	3
13	OUTCOME=0	AWARE=1*INT_FORM=0*FACIL=1->OUTCOME=0	1	0.192307692	3
14	OUTCOME=0	SELF_EFF=0*ATT=0*INT_FORM=1->OUTCOME=0	0.833333333	0.192307692	3
15	OUTCOME=0	SELF_EFF=0*INT_FORM=1*MOTIV_INT=0->OUTCOME=0	0.833333333	0.192307692	3
16	OUTCOME=0	SELF_EFF=0*INT_FORM=1*ACT_CON=0->OUTCOME=0	1	0.153846154	3
17	OUTCOME=1	SELF_EFF=1->OUTCOME=1	0.9	0.529411765	1
18	OUTCOME=1	ATT=1->OUTCOME=1	0.833333333	0.294117647	1
19	OUTCOME=1	MOTIV_INT=1->OUTCOME=1	0.777777778	0.205882353	1
20	OUTCOME=1	INT_FORM=1*MAINT=0->OUTCOME=1	0.75	0.176470588	2
21	OUTCOME=1	INT_FORM=1*FACIL=0->OUTCOME=1	0.8	0.117647059	2
22	OUTCOME=1	AWARE=0*INT_FORM=1->OUTCOME=1	0.75	0.088235294	2
23	OUTCOME=1	INT_FORM=1*ACT_CON=1->OUTCOME=1	0.75	0.088235294	2
24	OUTCOME=1	KNOWL=0*FACIL=0->OUTCOME=1	0.75	0.088235294	2
25	OUTCOME=1	KNOWL=0*ACT_CON=1->OUTCOME=1	1	0.029411765	2
26	OUTCOME=1	AWARE=1*MAINT=1*FACIL=0->OUTCOME=1	1	0.117647059	3
27	OUTCOME=1	AWARE=1*ACT_CON=0*MAINT=1->OUTCOME=1	0.8	0.117647059	3
28	OUTCOME=1	KNOWL=1*MAINT=1*FACIL=0->OUTCOME=1	0.8	0.117647059	3
29	OUTCOME=1	AWARE=1*INT_FORM=0*MAINT=1->OUTCOME=1	1	0.058823529	3

	A	B	C	D	E	F	G	H	I	J	K
1	outcome	condition	consistency	coverage	complexity		outcome	condition	consistency	coverage	complexity
2	OUTCOME=0	SELF_EFF=0*MAINT=1->OUTCOME=0	1	0.153846154	2		OUTCOME=0	SELF_EFF=0*MAINT=1->OUTCOME=0	1	0.153846154	2
3	OUTCOME=0	KNOWL=0*FACIL=1->OUTCOME=0	1	0.115384615	2		OUTCOME=0	KNOWL=0*FACIL=1->OUTCOME=0	1	0.115384615	2
4	OUTCOME=0	AWARE=0*MAINT=1->OUTCOME=0	1	0.076923077	2		OUTCOME=0	AWARE=0*MAINT=1->OUTCOME=0	1	0.076923077	2
5	OUTCOME=0	KNOWL=0*AWARE=1->OUTCOME=0	1	0.038461538	2		OUTCOME=0	KNOWL=0*AWARE=1->OUTCOME=0	1	0.038461538	2
6	OUTCOME=0	KNOWL=0*MAINT=1->OUTCOME=0	1	0.038461538	2		OUTCOME=0	KNOWL=0*MAINT=1->OUTCOME=0	1	0.038461538	2
7	OUTCOME=0	SELF_EFF=0*AWARE=1*MOTIV_INT=0->OUTCOME=0	0.785714286	0.423076923	3		OUTCOME=0	SELF_EFF=0*AWARE=1*FACIL=1->OUTCOME=0	0.8	0.307692308	3
8	OUTCOME=0	SELF_EFF=0*AWARE=1*ACT_CON=0->OUTCOME=0	0.769230769	0.384615385	3		OUTCOME=0	AWARE=1*INT_FORM=0*FACIL=1->OUTCOME=0	1	0.192307692	3
9	OUTCOME=0	SELF_EFF=0*AWARE=1*INT_FORM=0->OUTCOME=0	0.75	0.346153846	3		OUTCOME=0	SELF_EFF=0*ATT=0*INT_FORM=1->OUTCOME=0	0.833333333	0.192307692	3
10	OUTCOME=0	SELF_EFF=0*AWARE=1*FACIL=1->OUTCOME=0	0.8	0.307692308	3		OUTCOME=0	SELF_EFF=0*INT_FORM=1*MOTIV_INT=0->OUTCOME=0	0.833333333	0.192307692	3
11	OUTCOME=0	AWARE=1*ACT_CON=0*FACIL=1->OUTCOME=0	0.777777778	0.269230769	3		OUTCOME=0	SELF_EFF=0*INT_FORM=1*ACT_CON=0->OUTCOME=0	1	0.153846154	3
12	OUTCOME=0	AWARE=1*FACIL=1*MOTIV_INT=0->OUTCOME=0	0.777777778	0.269230769	3		OUTCOME=0	KNOWL=0*ACT_CON=0*MOTIV_INT=0->OUTCOME=0	0.8	0.153846154	3
13	OUTCOME=0	AWARE=1*INT_FORM=0*FACIL=1->OUTCOME=0	1	0.192307692	3		OUTCOME=0	KNOWL=0*ATT=0*ACT_CON=0->OUTCOME=0	0.8	0.153846154	3
14	OUTCOME=0	SELF_EFF=0*ATT=0*INT_FORM=1->OUTCOME=0	0.833333333	0.192307692	3		OUTCOME=0	KNOWL=0*ATT=0*MOTIV_INT=0->OUTCOME=0	0.8	0.153846154	3
15	OUTCOME=0	SELF_EFF=0*INT_FORM=1*MOTIV_INT=0->OUTCOME=0	0.833333333	0.192307692	3		OUTCOME=0	AWARE=1*INT_FORM=0*ACT_CON=1->OUTCOME=0	1	0.076923077	3
16	OUTCOME=0	SELF_EFF=0*INT_FORM=1*ACT_CON=0->OUTCOME=0	1	0.153846154	3		OUTCOME=0	INT_FORM=0*FACIL=1*MOTIV_INT=1->OUTCOME=0	1	0.076923077	3
17	OUTCOME=1	SELF_EFF=1->OUTCOME=1	0.9	0.529411765	1		OUTCOME=1	SELF_EFF=1->OUTCOME=1	0.9	0.529411765	1
18	OUTCOME=1	ATT=1->OUTCOME=1	0.833333333	0.294117647	1		OUTCOME=1	ATT=1->OUTCOME=1	0.833333333	0.294117647	1
19	OUTCOME=1	MOTIV_INT=1->OUTCOME=1	0.777777778	0.205882353	1		OUTCOME=1	INT_FORM=1*MOTIV_INT=1->OUTCOME=1	1	0.147058824	2
20	OUTCOME=1	INT_FORM=1*MAINT=0->OUTCOME=1	0.75	0.176470588	2		OUTCOME=1	ATT=0*MOTIV_INT=1->OUTCOME=1	0.833333333	0.147058824	2
21	OUTCOME=1	INT_FORM=1*FACIL=0->OUTCOME=1	0.8	0.117647059	2		OUTCOME=1	FACIL=0*MOTIV_INT=1->OUTCOME=1	1	0.117647059	2
22	OUTCOME=1	AWARE=0*INT_FORM=1->OUTCOME=1	0.75	0.088235294	2		OUTCOME=1	INT_FORM=1*FACIL=0->OUTCOME=1	0.8	0.117647059	2
23	OUTCOME=1	INT_FORM=1*ACT_CON=1->OUTCOME=1	0.75	0.088235294	2		OUTCOME=1	AWARE=0*MOTIV_INT=1->OUTCOME=1	1	0.088235294	2
24	OUTCOME=1	KNOWL=0*FACIL=0->OUTCOME=1	0.75	0.088235294	2		OUTCOME=1	ACT_CON=1*MOTIV_INT=1->OUTCOME=1	1	0.058823529	2
25	OUTCOME=1	KNOWL=0*ACT_CON=1->OUTCOME=1	1	0.029411765	2		OUTCOME=1	KNOWL=0*ACT_CON=1->OUTCOME=1	1	0.029411765	2
26	OUTCOME=1	AWARE=1*MAINT=1*FACIL=0->OUTCOME=1	1	0.117647059	3		OUTCOME=1	KNOWL=0*MOTIV_INT=1->OUTCOME=1	1	0.029411765	2
27	OUTCOME=1	AWARE=1*ACT_CON=0*MAINT=1->OUTCOME=1	0.8	0.117647059	3		OUTCOME=1	MAINT=1*MOTIV_INT=1->OUTCOME=1	1	0.029411765	2
28	OUTCOME=1	KNOWL=1*MAINT=1*FACIL=0->OUTCOME=1	0.8	0.117647059	3		OUTCOME=1	AWARE=1*MAINT=1*FACIL=0->OUTCOME=1	1	0.117647059	3
29	OUTCOME=1	AWARE=1*INT_FORM=0*MAINT=1->OUTCOME=1	1	0.058823529	3		OUTCOME=1	AWARE=1*ACT_CON=0*MAINT=1->OUTCOME=1	0.8	0.117647059	3
30	OUTCOME=1	ACT_CON=1*MAINT=1*FACIL=0->OUTCOME=1	1	0.029411765	3		OUTCOME=1	KNOWL=1*MAINT=1*FACIL=0->OUTCOME=1	0.8	0.117647059	3
31							OUTCOME=1	AWARE=0*INT_FORM=1*MAINT=0->OUTCOME=1	1	0.088235294	3
32							OUTCOME=1	KNOWL=0*MAINT=0*FACIL=0->OUTCOME=1	1	0.088235294	3
33							OUTCOME=1	AWARE=1*INT_FORM=0*MAINT=1->OUTCOME=1	1	0.058823529	3
34							OUTCOME=1	INT_FORM=1*ACT_CON=1*MAINT=0->OUTCOME=1	1	0.058823529	3
35							OUTCOME=1	KNOWL=0*INT_FORM=0*FACIL=0->OUTCOME=1	1	0.058823529	3
36							OUTCOME=1	KNOWL=1*AWARE=0*INT_FORM=1->OUTCOME=1	1	0.058823529	3
37							OUTCOME=1	ACT_CON=1*MAINT=1*FACIL=0->OUTCOME=1	1	0.029411765	3

# Model informed by msc routine

**SELF\_EFF=1 + ATT=1 + MOTIV\_INT=1**

**Consistency = .83, Coverage = .71**

# Model informed by msc routine

**SELF\_EFF=1 + ATT=1 + MOTIV\_INT=1**

**Consistency = .83, Coverage = .71**

**(K=1\*S=1) + (F=0\*G=1) + (R=0\*S=1\*I=1\*T=1) +  
(K=0\*F=0\*C=1\*M=0) + (F=0\*S=1\*M=0\*I=1\*T=1) +  
(K=1\*R=1\*F=1\*I=1\*C=1\*M=0) + (K=1\*R=0\*F=1\*T=1)**

**Consistency = 1.0, Coverage = .76**

Knowledge (K)

Facilitation (F)

Awareness (R)

Self-efficacy (S)

Intention formation (I)

Action control (C)

Attitude (T)

Maintenance (M)

Motivational interviewing  
(G)



# **R script for msc routine**

```

1 library(cna)
2 library(dplyr)
3 library(openxlsx)
4
5 # Rename filepath in command below to point to where R can find the file on your computer
6
7 HK <- read.csv("C:/Users/Edward Miech/Documents/R_Work/Kahwati_2016.csv")
8
9 names(HK)
10
11 # Create your analytic dataset in the next line by subsetting the relevant factors from HK
12 |
13 SA <- HK[, c(3:12)]
14
15 names(SA)
16
17 # Below is a sample script for running the msc routine for factor selection across five different consistency thresholds, from .75 to .95
18 # This is an exploratory data phase that happens before modeling
19 # The goal in this initial phase is to analyze your entire dataset (and use all the information it contains) to inform decisions about factor selection
20 # The output generated by this routine still needs to be interpreted in light of theory, background knowledge, case familiarity and common sense
21
22 # Note that coverage threshold is not needed; the routine automatically returns all mscs that meet the consistency threshold
23 # Note that suff.only = TRUE and strict = TRUE and maxstep is set to (3,1,20)
24 # NOTE: The outcome factor in the dataset needs to be renamed "OUTCOME"
25 # The 3 in the maxstep function tells the routine to assess every possible combination of values in your dataset up to 3 conditions long
26 # WARNING: if you set the first number of the maxstep argument to 5 or higher you may crash your computer
27
28 TK <- mvcna(SA,
29             con = .75,
30             ordering = "OUTCOME" ,
31             suff.only=T,
32             strict=T, maxstep=c(3,1,20))
33
34 # The line below is the second step of the msc routine which creates the condition table
35
36 VV75 <- msc(TK)
37

```

```
36 VV75 <- msc(TK)
37
38 # VV75 is the full condition table
39
40 # Next is a data manipulation argument from dplyr to create a more compact table
41 # It prints out top 10 mscs by outcome and complexity level
42
43 # NOTE: the %>% in the line below just means "then next..."
44
45 NN_75 <- VV75 %>%
46   group_by (outcome, complexity) %>%
47   slice_max(coverage, n = 10, with_ties = FALSE)
48
49 NN_75 <- NN_75[ , c(1:5)]
50
51 # Now the first msc run at the 0.75 threshold is complete
52
53 # The next msc run below is at the 0.80 threshold
54
55 TK <- mvcna(SA,
56             con = .8,
57             ordering = "OUTCOME" ,
58             suff.only=T,
59             strict=T, maxstep=c(3,1,20))
60 VV80 <- msc(TK)
61
62 NN_80 <- VV80 %>%
63   group_by (outcome, complexity) %>%
64   slice_max(coverage, n = 10, with_ties = FALSE)
65
66 NN_80 <- NN_80[ , c(1:5)]
67
```

```
68 # The next msc run below is at the 0.85 threshold, etc.
69
70 TK <- mvcna(SA,
71             con = .85,
72             ordering = "OUTCOME"      ,
73             suff.only=T,
74             strict=T, maxstep=c(3,1,20))
75 VV85 <- msc(TK)
76
77 NN_85 <- VV85 %>%
78   group_by (outcome, complexity) %>%
79   slice_max(coverage, n = 10, with_ties = FALSE)
80
81 NN_85 <- NN_85[ , c(1:5)]
82
83 TK <- mvcna(SA,
84             con = .90,
85             ordering = "OUTCOME"      ,
86             suff.only=T,
87             strict=T, maxstep=c(3,1,20))
88
89 VV90 <- msc(TK)
90
91 NN_90 <- VV90 %>%
92   group_by (outcome, complexity) %>%
93   slice_max(coverage, n = 10, with_ties = FALSE)
94
95 NN_90 <- NN_90[ , c(1:5)]
96
97 TK <- mvcna(SA,
98             con = .95,
99             ordering = "OUTCOME"      ,
100            suff.only=T,
101            strict=T, maxstep=c(3,1,20))
102
```

```

96
97 TK <- mvcna(SA,
98             con = .95,
99             ordering = "OUTCOME"      ,
100             suff.only=T,
101             strict=T, maxstep=c(3,1,20))
102
103 VV95 <- msc(TK)
104
105 NN_95 <- VV95 %>%
106   group_by (outcome, complexity) %>%
107   slice_max(coverage, n = 10, with_ties = FALSE)
108
109 NN_95 <- NN_95[ , c(1:5)]
110
111 # The rest of the script below integrates the 5 msc runs into a single workbook
112 # It uses the R package "openxlsx"
113 # The 5 msc runs are organized by individual tabs, with the first tab listing all 5 runs side-by-side on a single worksheet
114
115 wb <- createWorkbook()
116
117 addWorksheet(wb, "75to95")
118 addWorksheet(wb, "75")
119 addWorksheet(wb, "80")
120 addWorksheet(wb, "85")
121 addWorksheet(wb, "90")
122 addWorksheet(wb, "95")
123
124 writeData(wb, "75", NN_75)
125 writeData(wb, "80", NN_80)
126 writeData(wb, "85", NN_85)
127 writeData(wb, "90", NN_90)
128 writeData(wb, "95", NN_95)
129 writeData(wb, "75to95", NN_75)
130 writeData(wb, "75to95", NN_80, startCol = 7, startRow = 1)
131 writeData(wb, "75to95", NN_85, startCol = 13, startRow = 1)
132 writeData(wb, "75to95", NN_90, startCol = 19, startRow = 1)
133 writeData(wb, "75to95", NN_95, startCol = 25, startRow = 1)
134
135 saveWorkbook(wb, "KAHWATI2016_ConditionTable_3_75to95.xlsx")

```

# The Goal

**A configurational approach to factor selection in CNA that is:**

- ▶ **inductive and bottom-up**
- ▶ **systematic and comprehensive**
- ▶ **replicable**
- ▶ **transparent**
- ▶ **case-based**
- ▶ **pragmatic**

# The Goal

**A configurational approach to factor selection in CNA that:**

- ▶ **operates within a regularity framework**
- ▶ **uses all the information in the original dataset**
- ▶ **generates mathematical output for consultation**
- ▶ **considers factors in bundles, not just in isolation**
- ▶ **works with crisp-set, multi-value and fuzzy-set datasets**

# Closing Thoughts

- ▶ **draws on bottom-up philosophy fundamental to cna**
- ▶ **akin to taking an “X-ray” of your dataset**
- ▶ **computationally intensive (but fast)**
- ▶ **provides a panoramic view of co-occurrence within your dataset**
- ▶ **allows you to apply your understanding of theory and background knowledge to mathematical output when making decisions about factor selection**
- ▶ **creates space for surprising/unexpected mscs to emerge**



# Questions

# Discussion