

Coincidence Analysis: A New Method for Causal Inference in Implementation Science

Rebecca Garr Whitaker

Duke-Margolis Center for Health Policy

Nina Sperber

Duke University School of Medicine

Michael Baumgartner

Universitetet i Bergen

Alrik Thiem

Universitat Luzern

Deborah Cragun

University of South Florida

Laura Damschroder

VA Center for Clinical Management Research

Edward J. Miech

Regenstrief Institute Inc

Alecia Slade

Avalere Health

Sarah A. Birken (✉ sbirken@wakehealth.edu)

University of North Carolina at Chapel Hill <https://orcid.org/0000-0002-0591-4800>

Methodology

Keywords: Coincidence Analysis, Configurational Comparative Methods, causal inference, comparative analysis

Posted Date: August 17th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-58815/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on December 11th, 2020. See the published version at <https://doi.org/10.1186/s13012-020-01070-3>.

Abstract

Background

Implementation of multifaceted interventions typically involves many diverse elements working together in interrelated ways, including intervention components, implementation strategies, and features of local context. Given this real-world complexity, implementation researchers may be interested in a new mathematical, cross-case method called Coincidence Analysis (CNA) that has been designed explicitly to support causal inference, answer research questions about combinations of conditions that are minimally necessary or sufficient for an outcome, and identify the possible presence of multiple causal paths to an outcome. CNA can be applied as a standalone method or in conjunction with other approaches, and can reveal new empirical findings related to implementation that might otherwise have gone undetected.

Methods

We applied CNA to a publicly available dataset from Sweden with county-level data on human papillomavirus (HPV) vaccination campaigns and vaccination uptake in 2012 and 2014 and then compared CNA results to the published regression findings.

Results

The original regression analysis found vaccination uptake was positively associated only with the availability of vaccines in schools. CNA produced different findings and uncovered an additional solution path: high vaccination rates were achieved by either (1) offering the vaccine in all schools or (2) a combination of offering the vaccine in some schools and media coverage.

Conclusions

CNA offers a new comparative approach for researchers seeking to understand how implementation conditions work together and link to outcomes.

Contributions To The Literature

- Coincidence Analysis (CNA) represents a new mathematical, cross-case method for researchers evaluating the implementation of complex interventions in dynamic settings.
- CNA can address multiple dimensions of real-world complexity, including conjunctivity (where several conditions must be jointly present to bring about an outcome) and equifinality (where different paths can lead to the same outcome). CNA can also detect causal chains, where conditions lead to an intermediary outcome, which then leads to the final outcome.
- Intentionally designed to investigate different hypotheses and uncover different properties of causal structures than more traditional approaches, CNA can identify implementation-related findings that

might otherwise go undetected.

Background

One of the basic analytic challenges within implementation science is to study and understand implementation within real-world, dynamic settings. Implementation of multifaceted interventions typically involves many diverse elements working together in interrelated ways, including intervention components, implementation strategies, and features of local context. Furthermore, boundaries between an intervention, its implementation, and its contextual features can prove difficult to discern in practice. [1, 2]

For researchers seeking to explain these complex relationships encountered in real-world settings, causal inference can play an important role. Since the mid-1980s, Configurational Comparative Methods (CCMs) have increasingly been recognized as effective methods for causal inference, especially in the social sciences. As Fig. 1 belows shows, the cumulative number of CCM-related publications listed in the core collection of the Web of Science [3] has dramatically escalated in recent years, with more total publications appearing during the three-year period from 2017–2019 than in the entire preceding 22-year period between 1995–2016.

CCMs have also started to make prominent appearances within the health services research and implementation science literatures. CCMs, for example, were used in a recent Cochrane Review to identify conditions directly linked with successful implementation of school-based interventions for asthma self-management [4]; featured as an innovative member of the mixed-methods repertoire in in a major methodological review in public health [5]; and applied to determine different pathways for federally-qualified health centers to achieve patient-centered medical home status [6].

CCMs are designed to investigate different hypotheses and uncover different properties of causal structures than traditional regression analytical methods (RAMs).[7, 8] Qualitative Comparative Analysis (QCA) is one kind of CCM that, to date, has been most frequently applied in implementation science and health services research. The purpose of this article is to introduce a new CCM to the implementation research community: Coincidence Analysis. Coincidence Analysis (CNA) is a mathematical, cross-case approach that can be applied as a standalone method or in conjunction with other methods (including RAMs) to support causal inference, and is available via the R-package `cna`. [9, 10, 11]

CNA offers a new cross-case method for implementation and health services researchers exploring causality when evaluating or implementing multifaceted interventions in complex contexts. Investigators applying CNA can conduct analyses across entire datasets to identify specific combinations of components and conditions that consistently lead to outcomes, and can be applied to large-n as well as small-n studies. Peer-reviewed, implementation-related work involving CNA has started to emerge, including podium presentations at major implementation conferences [12, 13]; methods workshops dedicated specifically to CNA [14]; published protocols [15]; and full-length articles in established journals [16].

CNA is a new comparative approach that can be used by the implementation research community to support causal inference, answer research questions about conditions that are minimally necessary or sufficient, and identify multiple causal paths to an outcome. We present this article in three parts. In part 1, we establish the theoretical foundation for CCMs, define CNA as a method within the CCM family and describe what CNA (and CCMs) uniquely offer. In part 2, we illustrate CNA by applying the method to a publicly available dataset that was originally analyzed using RAMs. In part 3, we offer guidance for reporting CNA design and results, and we discuss the limitations and challenges of CNA. In additional files accompanying this article we provide detailed descriptions of the steps and coding used to conduct the analysis [see Additional file 1] and the analytic dataset used [see Additional file 2] along with the R script [see Additional file 3] to allow for independent replication and validation of results.

Part 1: Laying the Theoretical Foundation for CCMs

Defining causal inference in CCMs. CNA is one method within a class of CCMs used to model complex patterns of conditions hypothesized to contribute to an outcome within a set of data. CCMs search for causal relations as defined by a regularity theory of causality, according to which a cause is a “difference-maker” of its effect within a fixed set of background conditions. More specifically, X is a cause of Y if there exists a fixed configuration of background factors F such that, in F, a change in the value of X is systematically associated with a change in Y. If X does not make a difference to Y in any F, X is redundant to account for Y and, thus, not a cause of Y. The most influential theory defining causation along these lines is Mackie's INUS-theory,[17] with refinements by Graßhoff and May[18] and Baumgartner.[19] An INUS condition of an outcome Y is an Insufficient but Necessary part of a condition that is itself Unnecessary but Sufficient for Y. To use a common example for illustrating INUS conditions: not every fire is caused by a short circuit—fires can also be started by, for example, arson or lightning. However, a short circuit in combination with other conditions – e.g., presence of flammable material and absence of a suitably placed sprinkler – is sufficient for a fire. In this example, the short circuit is an INUS condition: it is a necessary part of a sufficient condition for a fire. This particular causal path to a fire includes the combination of three specific conditions: presence of a short circuit, presence of flammable material, and absence of a sprinkler. All three of these conditions are difference-makers, for if one of them is missing, the fire does not occur along this causal path.

Regularity theories can be described using Boolean properties of causation, which encompass three dimensions of complexity. The first is conjunctivity: to bring about an outcome, several conditions must be jointly present. For example, in a study of high performance work practices and front line health care worker outcomes, Chuang and colleagues[20] found that no single high performance work practice was alone sufficient to produce the outcome of high job satisfaction. Instead, a configuration consisting of creative input, supervisor support, and team-based work practices together accounted for 65 percent of highly satisfied front-line health care workers.[20] Chuang and colleagues identified a second configuration that also led to high job satisfaction: supervisor support, incentive pay, team-based work and flexible work.[20] Both configurations resulted in high job satisfaction independently of each other. These configurations illustrate disjunctivity, a second dimension of complexity in which an outcome can

alternatively result from multiple causal paths. The third dimension of complexity is sequentiality: outcomes tend to produce further outcomes, propagating causal influence along causal chains. For instance, high job satisfaction of health care workers may, in turn, promote patient satisfaction.[21]

Why use CCMs in implementation research? CCMs study different properties of causal structures than RAMs and thus are appropriate for exploring different types of hypotheses. RAMs examine statistical properties characterized by probabilistic or intervention theories of causation. In the probabilistic model, X is a cause of Y if and only if the probability of Y given X is greater than the probability of Y alone and there does not exist a further factor, Z, that explains (i.e., neutralizes) the probabilistic dependence between X and Y.[8] In the intervention model, X causes Y if intervention X changes the values of outcome Y controlling for other variables. The intervention theory of causation is counterfactual in that a case cannot simultaneously “receive” and “not receive” an intervention; instead, the intervention model maps possible values of Y onto possible values of X, focuses on how variables X and Y relate to one another, and generates average treatment effects over a population.[8]

Conversely, CCMs examine Boolean properties of the data as described by regularity theories of causation, according to which X is a cause of Y if and only if X is an INUS condition of Y (see INUS definition above).[8, 17] CCMs study implication hypotheses that link specific values of variables as “X = χ_i is (non-redundantly) sufficient/necessary for Y = γ_i .”[8, 11] In this way, CCMs, including CNA, model the effect of conditions (e.g., high degree of X) on outcomes. This is a fundamentally different vantage point than the one adopted by RAMs which examine covariation hypotheses that link variables. Further, CCMs are case-oriented methods, in which observations consist of bounded, complex entities (e.g., organizations) that are considered as a whole.[22] A case-based unit of analysis differs from the approach taken in RAMs, where cases are deconstructed into a series of variables, and estimates represent the net effect of a variable for the average case. Because CNA and other CCMs employ case-based analysis, they present opportunities for implementation and health services research questions in particular because these methods can be used to identify which interventions work in an array of contexts.

Different types of CCMs

While CCMs have a common regularity theoretic foundation, various types of CCMs rely on different a priori conceptions of outcome and causal factors and build causal models in different ways. For example, Qualitative Comparative Analysis (QCA), in its standard implementation that uses the Quine-McCluskey (QMC) algorithm,[23, 24] requires identification of exactly one factor as an endogenous outcome. It begins by identifying maximal sufficient and necessary conditions of the outcome, which are subsequently minimized using standard inference rules from Boolean algebra to arrive at a redundancy-free solution composed of INUS conditions of the outcome.[7] However, the QMC algorithm was not designed for causal inference. For instance, the absence of cases instantiating a potential causal model, also known as limited diversity, forces QMC to draw on counterfactual reasoning that goes beyond available data and sometimes requires assumptions contradicting the very causal structures under

investigation.[25] Moreover, QMC has built-in protocols for ambiguity reduction when multiple solutions fit the data equally well. Potential solutions are often eliminated to reduce ambiguity without justification, which is problematic for causal discovery.[25, 26]

Advantages of using CNA. Coincidence Analysis (CNA) is a new addition to the family of CCMs.[27, 28] It uses an algorithm specifically designed for exploring causal inference, thus avoiding the problems mentioned above. In particular, it does not build causal models by means of a top-down approach that first searches for maximally sufficient and necessary conditions and then gradually minimizes them using the QMC algorithm. Rather, CNA employs a bottom-up approach that first tests single factors for sufficiency and necessity, and then tests factor combinations of two, three, etc.[10, 11] All sufficient and necessary conditions revealed by this approach are, by definition, minimal and redundancy-free.

Additionally, CNA is designed to treat any number of variables as endogenous and is therefore capable of analyzing causal chains, or common-cause structures.[29] Analyzing causal chains may be advantageous if, for example, intervention factor A occurs as a result of other factors but is not the ultimate outcome of interest. Identifying the full causal model, including which factors produce A on the path to the ultimate outcome of interest, is valuable when seeking to understand causal complexity. CNA is the only member of the CCM family that builds and evaluates models representing causal chains.

Part 2: Demonstrating CNA using Publicly Available Data

Data source

In March 2016, Rehn and colleagues reported the impact of implementation strategies on human papillomavirus (HPV) "catch-up vaccination" uptake in Sweden among 5th and 6th grade girls.[30] The purpose of the original study was to estimate the impact of various information channels and delivery settings on county-level catch-up vaccine uptake to inform future vaccination campaigns in Sweden.

The authors obtained county-level data on catch-up vaccinations and the eligible population from administrative data. They collected implementation strategies from county health care offices via an open-ended questionnaire emailed in 2012 asking respondents to list and describe "information channels" used to reach eligible girls and the settings in which they offered the vaccine. A subsequent phone interview was conducted in 2014 to update the lists.

Rehn and colleagues used regression analysis to estimate county-level catch-up vaccine uptake as a function of information channels and delivery settings. The authors concluded that the availability of vaccines in schools explained differences in county-level vaccine uptake; no information channels were found to make a difference in uptake.

Rehn and colleagues defined the outcome and predictor variables as follows:

Outcome variable. County-level catch-up vaccine uptake was defined as the percent of eligible girls born between 1993–1998 who received at least one dose of vaccine by 2014.

Predictor variables. Ten variables represented information channels and four variables represented the delivery settings where the vaccinations were available (some schools, all schools, primary health care centers, and other health care centers). All 14 factors were dichotomized with values of 1 (present) or 0 (absent).

All county-level data on vaccine uptake, information channels and delivery settings used for the CNA illustration were reported in the article.

Methods

We re-analyzed the data using CNA. A step-by-step guide for conducting CNA, using this study as an illustration, is provided in a document accompanying this article [see Additional file 1] as well as the analytic dataset [see Additional file 2] and the R script [see Additional file 3] used in the analysis.

Step 1: Define, calibrate and select the factors (i.e., outcomes and conditions) to create a data set. Vaccination rate represented the outcome of interest and ranged from 49–84% across the 21 counties. We selected 65% as the threshold defining “high” catch-up vaccination rates after conducting sensitivity analyses in which we varied the threshold for “high-uptake,” using two different existing break points in the data: the 65% cut-off generated the greatest diversity among cases for the conditions and the outcome and yielded a sufficiently high number of cases featuring the outcome. We coded the 21 counties into a new dichotomous outcome called HI_UPTAKE where 1 = “catch-up vaccination rate of 65% or higher” and 0 = “catch-up vaccination rate less than 65%.” (See Additional file 1 for details on the rationale for each step in the analytic process [see Additional file 1]). A secondary analysis identified conditions leading to the absence of the outcome (HI_UPTAKE = 0) because conditions that prevent the outcome may differ from those that contribute to the presence of the outcome.

We prepared a dataset that included the uptake rates, delivery settings and implementation strategies as reported by Rehn and colleagues (see Additional file 2 to view the analytic dataset used in the Coincidence Analysis [see Additional file 2]). We transformed a number of factors from the original dataset for use with CNA because in the original dataset these factors had characteristics unsuitable for CCM processing. For instance, the original data set contained the factor “Primary health care centre” (PHC) that was constantly present in all 21 counties (cases). Constant factors like PHC can be automatically excluded as difference-makers. Another delivery setting, “Other health care center;” (HC) was eliminated given limited variation across cases. We combined “All schools” (“organized delivery of the vaccine in all schools in the county;” AS = 1) and “Some schools” (“organized delivery of the vaccine in schools in some of the municipalities in the county...;” SS = 1) into a new multi-value ordinal factor called “SCHOOLS,” where SCHOOLS = 0 if AS = 0 and SS = 0; SCHOOLS = 1 if SS = 1 and AS = 0; and SCHOOLS = 2 if AS = 1. The resulting dataset included 12 potential explanatory factors. These factors could be combined into 6144 logically possible configurations, which could not be covered to an informative degree by the 21 cases included. Thus, the diversity index for the original data, i.e. the ratio of observed configurations to all possible configurations, was exceedingly small. The smaller the diversity

index, the more challenging it is to draw informative configurational conclusions. To improve the diversity index, we included a subset of the 12 exogenous factors in our analysis. This is analogous to maximizing degrees of freedom in RAM.

We selected schools (SCHOOLS) and four of the ten information channels to include: school-based information (SBI), media coverage (MC), social media (SM), and Cinema commercial/YouTube (CCY). These four implementation strategies represented likely information channels for students and their parents to be informed of why, where, how, and when to access vaccinations.

In our initial analysis plan, three of the seven cases (counties) exhibiting the outcome (high-uptake) instantiated exactly the same configuration of conditions, leaving only five observed configurations featuring HI_UPTAKE = 1 out of a total of 48 logically possible configurations. As we could not justify removing one of the four selected information channels on a theoretical basis, but still wanted to increase the diversity index, we decided to assess “meta-factors” by disjunctively aggregating two of the information channels into single, new factors – a common approach in CCMs to reduce the number of conditions without eliminating either of the properties represented by these conditions from the analysis. An aggregation of two information channels yielded a disjunctive factor taking the value 1 if, and only if, at least one of the two aggregated channels was present in a county. There were six possible ways to disjunctively aggregate two of the four information channels in our analysis. We explored all of these possibilities by replacing two information channels by their corresponding aggregates which resulted in six datasets, or analytic samples, each with four exogenous factors and the outcome HI_UPTAKE. (To view this analytic dataset, please see Additional file 2 [Additional file 2].)

Step 2. Perform CNA using the cna package in R.[10] Two parameters of fit – consistency and coverage – provide insight into the strength of the dependence between conditions and the outcome. Consistency, represented by a score ranging from 0 to 1, measures how often a combination of conditions is associated with the outcome, or the degree to which the cases that share a configuration also share the same outcome. [7] Lower consistency values indicate lower confidence in the causal interpretability of the dependence between conditions and the outcome. Coverage scores range from 0 to 1 and represent the proportion of cases with the outcome that also have a particular configuration. Coverage measures a given configuration’s empirical importance based on the available data.[7] When applying CNA to the dataset, we set our minimum consistency and coverage scores to 1.0.[7]

Results

Step 3. Interpret results and refine model inputs if necessary.

CNA encourages an iterative approach where researchers can run analyses, interpret results and redefine model inputs before finalizing a solution set. As such, we discuss our interpretation of the findings as our iterative analyses progress.

Our analyses produced model ambiguity, meaning that the data were insufficient to determine exactly which causal structure was operative. Of the six analytic samples, three did not yield any models at consistency and coverage thresholds of 1.0, whereas the remaining three datasets yielded a total of five causal models for $HI_UPTAKE = 1$, all of which featured maximal consistency and coverage scores (see the R replication script for a complete list of models). All models had the following identical terms as part of their solution, where "+" symbolizes the Boolean operator OR, "*" symbolizes AND, and " \leftrightarrow " expresses sufficiency and necessity:

$$SCHOOLS = 2 + SCHOOLS = 1 * MC = 1 \leftrightarrow HI_UPTAKE = 1 \quad (1)$$

The above expression (1) translates to: "counties with high catch-up vaccination rates offered vaccination in all schools OR offered vaccination in some (but not all) schools AND used a media coverage implementation strategy." All five causal models resulting from our analysis were supersets of (1), so we concluded that the factor values contained in (1) were causally relevant for high-uptake. However, (1) was not a complete model because it only achieved a consistency score of 0.875 (i.e., only 87.5% of counties with this configuration were high-uptake counties). Expression (1) covers a county, Skåne, that exhibited the configuration $SCHOOLS = 1 * MC = 1$ but was not high-uptake. The other counties with this configuration, $SCHOOLS = 1 * MC = 1$, were always associated with high-uptake of vaccines. Thus, other factors unique to Skåne must be missing from expression (1).

While our data were insufficient to identify the factors that distinguished Skåne from other counties with the same configuration, the five complete models inferred by CNA provided five possible explanations. Interestingly, all five of these solution models combined $SCHOOLS = 1 * MC = 1$ with the absence of other information channels. Only when $SCHOOLS = 1 * MC = 1$ was accompanied by $SM = 0$ or $SBI = 0$ or $CCY = 0$ was it associated with high uptake with 100% consistency, which could indicate these other information channels produced a "backfiring" effect.

After further reviewing data for each county, we determined that the most plausible implementation strategy to backfire was cinema commercials and/or YouTube. Online media like YouTube can backfire because these open-source platforms contain unvalidated content that can appear automatically through newsfeeds or advertisements, potentially overriding legitimate health-related information. [31] Furthermore, we observed a negative relationship between the presence of CCY and the vaccination rate outcome in the underlying dataset. As Fig. 2 shows, the presence of CCY was relatively well-represented in the overall dataset: 8 of the 21 cases had $CCY = 1$. In 7 of these 8 cases, $HI_UPTAKE = 0$. The lone exception was the county of Jonkoping, where the sufficient condition of $SCHOOLS = 2$ was also present. For these reasons, we deemed the following complete model to be the most plausible:

$$SCHOOLS = 2 + SCHOOLS = 1 * MC = 1 * CCY = 0 \leftrightarrow HI_UPTAKE = 1 \quad (2)$$

Expression (2) translates to: "counties with high catch-up vaccination rates offered vaccination in all schools OR offered vaccination in some (but not all) schools AND used a media coverage implementation strategy but not cinema commercials/YouTube." Expression (2) demonstrated perfect consistency and coverage scores (1.0, respectively) and differentiated Skåne (a county without high

uptake) from high uptake counties that also had vaccinations available at some schools and used media coverage. Figure 2 highlights the configurations instantiating this model in the data.

Applying the same analytic approach to model the absence of the outcome yielded seven solution models for HI_UPTAKE = 0 such that six of these seven models had a common core that corresponded exactly to the negation of the core of the positive models:

$$\text{SCHOOLS} = 0 + \text{SCHOOLS} = 1 * \text{MC} = 0 \leftrightarrow \text{HI_UPTAKE} = 0 \quad (3)$$

Expression (3) translates to: “counties without high catch-up vaccination rates did not offer vaccination in all schools OR offered vaccination in some (but not all) schools AND did not use a media coverage implementation strategy.” Expression 3 exhibited 1.0 consistency and 0.93 coverage. Taken together, these results provide substantive evidence that media coverage is relevant for differentiating counties with and without high vaccination uptake, adding additional information to the regression results of the original study.

Discussion

The results of this CNA indicate that, under specific conditions, information channels made a difference for high vaccination uptake. This contrasted with the results of the regression analysis from the original study, which concluded that information channels made no difference in increasing vaccination uptake. Our results implied that the availability of vaccination in some schools was only sufficient for high vaccination rates if media coverage was employed and certain other communication channels were not used. In other words, the data contained enough evidence to infer that when vaccination was available at some but not all schools, availability must be complemented by media coverage to achieve high uptake. The data did not contain enough evidence, however, to have absolute certainty which communication channels should be avoided. Even so, cinema commercials and/or YouTube might be the most plausible information channel to avoid; YouTube in particular might backfire and reduce vaccination rates as a result of unsolicited content that undermines county-sanctioned media coverage on vaccines.

Closely examining data from individual cases (counties) corroborated the theory of some conditions backfiring with respect to producing high vaccination uptake. Jonkoping provided vaccination in all schools, but only achieved vaccination uptake among 65% of eligible girls, as opposed to over 80% uptake achieved in the three other counties with vaccination availability at all schools. Notably, Jongkoping used two communication channels (CCY = 1 and SM = 1) that were absent in the other three counties (CCY = 0 and SM = 0).

In sum, the CNA results indicated that whenever vaccination was available at only some schools, media coverage made a difference for high uptake. Furthermore, the common core of the resulting CNA models for HI_UPTAKE = 0 indicated that a lack of media coverage when vaccinations were provided only in some schools made a difference for lower vaccine uptake. By systematically scrutinizing the configurations of implementation strategies associated with high-uptake and low-uptake counties, CNA extended the conclusions drawn from Rehn and colleagues’ regression model.

Part 3: Tips for Practitioners

Recommendations for reporting analyses and results for CCMs

Variation in reporting study design, analysis and results exists in previously published peer-reviewed CCM literature. To advance the methodological rigor with which CCMs are applied, we offer recommendations for describing design, analysis and results for CCMs. We also provide additional material in Additional files 1 and 2 accompanying this article to allow for independent replication of these analyses and findings [Additional file 1][Additional file 2].

We suggest that future studies and publications applying CCMs: 1) describe the rationale for using the CCM (e.g., CNA); 2) describe the rationale for selecting which factors (outcomes and conditions) to include in the analysis; 3) describe the process used to assign cases to factor values [32, 33] (e.g., high vaccination uptake); 4) specify the software (and version) used for analysis; 5) describe the iterative analyses used to refine factors (e.g., different approaches to calibration) and models (e.g., adding/dropping factors); 6) list the number of solutions generated in each iteration of the analysis and identify commonalities across solutions; 7) report consistency and coverage thresholds for final models along with ranges for models not part of the final solution; and 8) describe the rationale for selecting the final model(s).

In this article, we used conceptual knowledge of HPV vaccination uptake and the existence of variability across cases to select the factors to include in the initial analysis. Given the structure of the data, we relied on the original binary factor calibration for all but one factor (SCHOOLS), which was calibrated as a multi-value ordinal factor. Moreover, to improve the diversity index, we explored all disjunctive combinations of two factors representing information channels into a single factor. We analyzed six datasets using the cna package in R.[10] Five models reached perfect consistency and coverage and, thus, fit the data equally well. All models exhibited a common core, which we reported as the final configuration supporting causal inference in our data. At the same time, we acknowledged that the common core did not amount to a complete causal model. Based on theoretical considerations, which we explicitly described, we then selected one of five viable completions as the most plausible overall model. We reported all relevant model fit scores for our preferred model and provide additional files with detailed descriptions of each step in the analytic process [see Additional file 1], the analytic dataset [see Additional file 2] and the R script [see Additional file 3] to allow for independent replication and verification of our results.

Limitations of CNA

CNA has several limitations of which implementation researchers should be aware. First, although CNA supports causal inference, there are limitations to the extent to which results may be generalized. Results can be confounded by unmeasured causes that are located on causal paths to outcomes that do not go through any measured factors. If the data cannot be assumed to be homogenous in confounder distributions—meaning that unmeasured factors do not affect all cases/configurations equally—

generalization becomes problematic given the risk of over-interpreting the data or incorrectly inferring a causal relationship. As with other CCMs, familiarity with cases helps to evaluate generalizability—by, for example, justifying that cases included in the analysis are homogeneous with respect to potential confounders—and to interpret solutions generated by mathematical modeling. Second, CCMs rarely uncover full model solutions. Configurational data analyzed in observational studies tend to be fragmented (i.e., exhibit low diversity), so most logically possible combinations of conditions are not present in the observed cases. Under these circumstances, CCMs may reveal only portions of the underlying causal structures. Thus, the fact that some factor X is not contained in a CCM solution must not be taken to mean that X is causally irrelevant. Unless there is reason to assume that the data are non-fragmented, the absence of X from a model can only mean that the data do not contain evidence for X's relevance (which is not the same as X's irrelevance).

Conclusions

CNA offers new insights of potential high interest to implementation researchers. We demonstrated the utility of CNA using data previously analyzed with RAMs. The authors of the original, RAM-based analysis indicated that offering vaccination in schools increased county-level vaccination uptake while no information channels significantly increased vaccination uptake.[30] By contrast, CNA results indicated that under specific conditions, information channels did make a difference for high vaccination uptake. Specifically, our results implied that higher vaccination rates were achieved by either (1) offering the vaccine in all schools or (2) offering the vaccine in some schools and using media coverage but not certain other communication channels.

Compared to RAMs, CCMs have fundamentally different methodological goals and search for different properties of causal structures. CCMs and RAMs answer distinct types of questions. RAMs are useful for estimating the average influence of a specific variable on an outcome while holding other variables constant. CCMs are useful for identifying combinations of specific conditions that may be on the causal path (i.e., are minimally necessary or sufficient) to an outcome. Of the two CCMs, CNA was built expressly for causal inference and can be used to uncover causal chains underlying the data.[10, 11, 29]

CNA has the potential to offer implementation researchers alternative and more nuanced knowledge about causal relationships when examining complex interventions in settings with interdependent or interrelated factors. In particular, CCMs are well-suited for implementation and health services research questions regarding the implementation of multifaceted interventions in complex, real-world settings, the dynamics of which can be influenced by many factors acting in combination.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The full dataset used in this analysis along with detailed descriptions of each step of the analysis and the R script are all available for review and replication in additional files accompanying this article [see Additional file 1][see Additional file 2][Additional file 3]. The original dataset is publicly available within the following open-access article: Rehn M, Uhnöo I, Kühlmann-Berenzon S, Wallensten A, Sparén P, Netterlid E. Highest vaccine uptake after school-based delivery - a county-level evaluation of the implementation strategies for HPV catch-up vaccination in Sweden. PLoS One. 2016;11(3):e0149857. doi:10.1371/journal.pone.0149857.

Competing interests

The authors declare that they have no competing interests.

Funding

This secondary analysis was conducted by the investigators without additional funding.

Acknowledgements

Not applicable

Authors' contributions

RGW, SB, NS, LD and AS originated the idea for the project and convened the full team of co-authors. RGW found the published dataset that we used for secondary analysis. RGW, MB, and EJM took lead roles in conducting the Coincidence Analysis. MB wrote the R script provided in Additional file 3. RGW, SB, NS, LD, AS, MS, AT and EJM participated in a series of discussions to review preliminary results from the Coincidence Analysis and provide direct feedback to inform ongoing analyses. RGW, SB, NS, LD, AS, MS, AT and EJM participated in writing sections of the main manuscript plus the step-by-step guide provided in Additional file 1, as well as editing drafts. All authors read and approved the final manuscript.

References

1. Pfadenhauer L, Gerhardus A, Mozygemba, K et al. Making sense of complexity in context and implementation: the context and implementation of complex interventions (CICI) framework.

- Implement Sci. 2013;12:21.
2. Medical Resource Council. A framework for development and evaluation of RCTs for complex interventions to improve health. 2000. <https://mrc.ukri.org/documents/pdf/rcts-for-complex-interventions-to-improve-health>. Accessed 11 Oct 2018.
 3. Web of Science Core Collection database. Clarivate Analytics, Philadelphia. 2020. <http://www.webofknowledge.com>. Accessed 12 May 2020.
 4. Palinkas LA, Mendon SJ, Hamilton AB. Innovations in mixed methods evaluations. *Annual Rev Public Health*. 2019;40:423–442.
 5. Harris K, Kneale D, Lasserson TJ, et al. School-based self-management interventions for asthma in children and adolescents: a mixed methods systematic review. *Cochrane Database Syst Rev*. 2019;1:CD011651.
 6. Mendel P, Chen EK, Green HD, Armstrong C, Timbie JW, Kress AM, Friedberg MW, Kahn KL. Pathways to medical home recognition: a Qualitative Comparative Analysis of the PCMH transformation process. *Health services research*. 2018;53:2523-46.
 7. Thiem A. Conducting configurational comparative research with Qualitative Comparative Analysis. *Am J Eval*. 2017;38:420-433. doi:10.1177/1098214016673902.
 8. Rohlfing I, Zuber CI. Check your truth conditions! Clarifying the relationship between theories of causation and social science methods for causal inference. *Sociol Methods Res*. 2019;0049124119826156.
 9. Thiem A, Baumgartner M, Bol D. Still lost in translation! A correction of three misunderstandings between configurational comparativists and regressional analysts. *Comp Polit Stud*. 2016;49(6):742-774. doi:10.1177/0010414014565892.
 10. Ambuehl M, Baumgartner M, Epple R, Thiem. *Alrik. cna: Causal modeling with Coincidence Analysis*. <https://cran.r-project.org/package=cna> (2018). Accessed 09 Oct 2018.
 11. Baumgartner M, Ambühl M. Causal modeling with multi-value and fuzzy-set Coincidence Analysis. *Polit Sci Res Method*. 2018;1-17. doi:10.1017/psrm.2018.45.
 12. Miech EJ, Rattray NA, Damush TM. S123 Necessary but not sufficient: A multimethod study of the role of champions in healthcare-related implementation. *Proceedings from the 12th Annual Conference on the Science of Dissemination and Implementation*. *Implementation Sci*. 2020;15:25. <https://doi.org/10.1186/s13012-020-00985-1>.
 13. Cragun D, Rahm AK. Coincidence analysis: A methodology to identify contextual conditions influencing implementation across multiple settings. *Proceedings from the 11th Annual Conference on the Science of Dissemination and Implementation*. *Implementation Sci*. 2019;14:27. <https://doi.org/10.1186/s13012-019-0878-2>.
 14. Birken S, Damschroder L, Miech E, Cragun D. Coincidence Analysis (CNA) workshop at the D&I Conference. In: *Agenda for the 12th Annual Conference on the Science of Dissemination and Implementation*. 2019.

<https://academyhealth.confex.com/academyhealth/2019di/meetingapp.cgi/Session/23183>.
Accessed 15 May 2020.

15. Rahm AK, Cragun D, Hunter JE, Epstein MM, Lowery J, Lu CY, Pawloski PA, Sharaf RN, Liang SY, Burnett-Hartman AN, Gudgeon JM. Implementing universal Lynch syndrome screening (IMPULSS): Protocol for a multi-site study to identify strategies to implement, adapt, and sustain genomic medicine programs in different organizational contexts. *BMC Health Serv Res*. 2018;18:824.
16. Yakovchenko V, Miech EJ, Chinman MJ, Chartier M, Gonzalez R, Kirchner JE, Morgan TR, Park A, Powell BJ, Proctor EK & Ross D. Strategy configurations directly linked to higher Hepatitis C virus treatment starts: An applied use of Configurational Comparative Methods. *Med Care*. 2020;58:e31-38. doi:10.1097/MLR.0000000000001319.
17. Mackie JL. *The Cement of the Universe: A Study of Causation*. Oxford: Clarendon Press; 1974.
18. Graßhoff G, May M. Causal Regularities. In: Spohn W, Ledwig M, Esfeld M, editors. *Current Issues in Causation*. Mentis-Verlag: Paderborn; 2001. p. 85–114.
19. Baumgartner M. Regularity Theories Reassessed. *Philosophia (Mendoza)*. 2008;36(3):327-354. doi:10.1007/s11406-007-9114-4.
20. Chuang E, Dill J, Morgan JC, Konrad TR. A Configurational Approach to the Relationship between High-Performance Work Practices and Frontline Health Care Worker Outcomes. *Health Serv Res*. 2012;47(4):1460-1481. doi:10.1111/j.1475-6773.2011.01366.x.
21. Sikorska-Simmons E. Linking Resident Satisfaction to Staff Perceptions of the Work Environment in Assisted Living: A Multilevel Analysis. *Gerontologist*. 2006;46:590-598. doi:10.1093/geront/46.5.590.
22. Rihoux B, Lobe B. The Case for Qualitative Comparative Analysis (QCA): Adding Leverage for Thick Cross-Case Comparison. In: Byrne D, Ragin C, eds. *The SAGE Handbook of Case-Based Methods*. London : SAGE Publications Ltd; 2009:222-242. doi:10.4135/9781446249413.n13.
23. Quine W V. On Cores and Prime Implicants of Truth Functions. *Am Math Mon*. 1959;66:755. doi:10.2307/2310460.
24. McCluskey EJ. *Introduction to the Theory of Switching Circuits*. McGraw-Hill; 1965.
25. Baumgartner M. Parsimony and Causality. *Qual Quant*. 2015;49:839-856. doi:10.1007/s11135-014-0026-7.
26. Baumgartner M, Thiem A. Model Ambiguities in Configurational Comparative Research. *Sociol Methods Res*. October 2015. doi:10.1177/0049124115610351.
27. Baumgartner M. Uncovering deterministic causal structures: a Boolean approach. *Synthese*. 2009;170:71-96. doi:10.1007/s11229-008-9348-0.
28. Baumgartner M. Inferring Causal Complexity. *Sociol Methods Res*. 2009;38(1):71-101. doi:10.1177/0049124109339369.
29. Baumgartner M, Epple R. A Coincidence Analysis of a causal chain. *Sociol Methods Res*. 2014;43:280-312. doi:10.1177/0049124113502948.

30. Rehn M, Uhnö I, Köhlmann-Berenzon S, Wallensten A, Sparén P, Netterlid E. Highest vaccine uptake after school-based delivery - a county-level evaluation of the implementation strategies for HPV catch-up vaccination in Sweden. *PLoS One*. 2016;11:e0149857. doi:10.1371/journal.pone.0149857.
31. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *J Med Internet Res*. 2013;15:e85. doi:10.2196/jmir.1933.
32. Thiem A. Membership function sensitivity of descriptive statistics in fuzzy-set relations. *Int J Soc Res Methodol*. 2014;17:625-642. doi:10.1080/13645579.2013.806118.
33. Verkuilen J. Assigning membership in a fuzzy set analysis. *Sociol Methods Res*. 2005;33:462-496. doi:10.1177/0049124105274498.

Figures

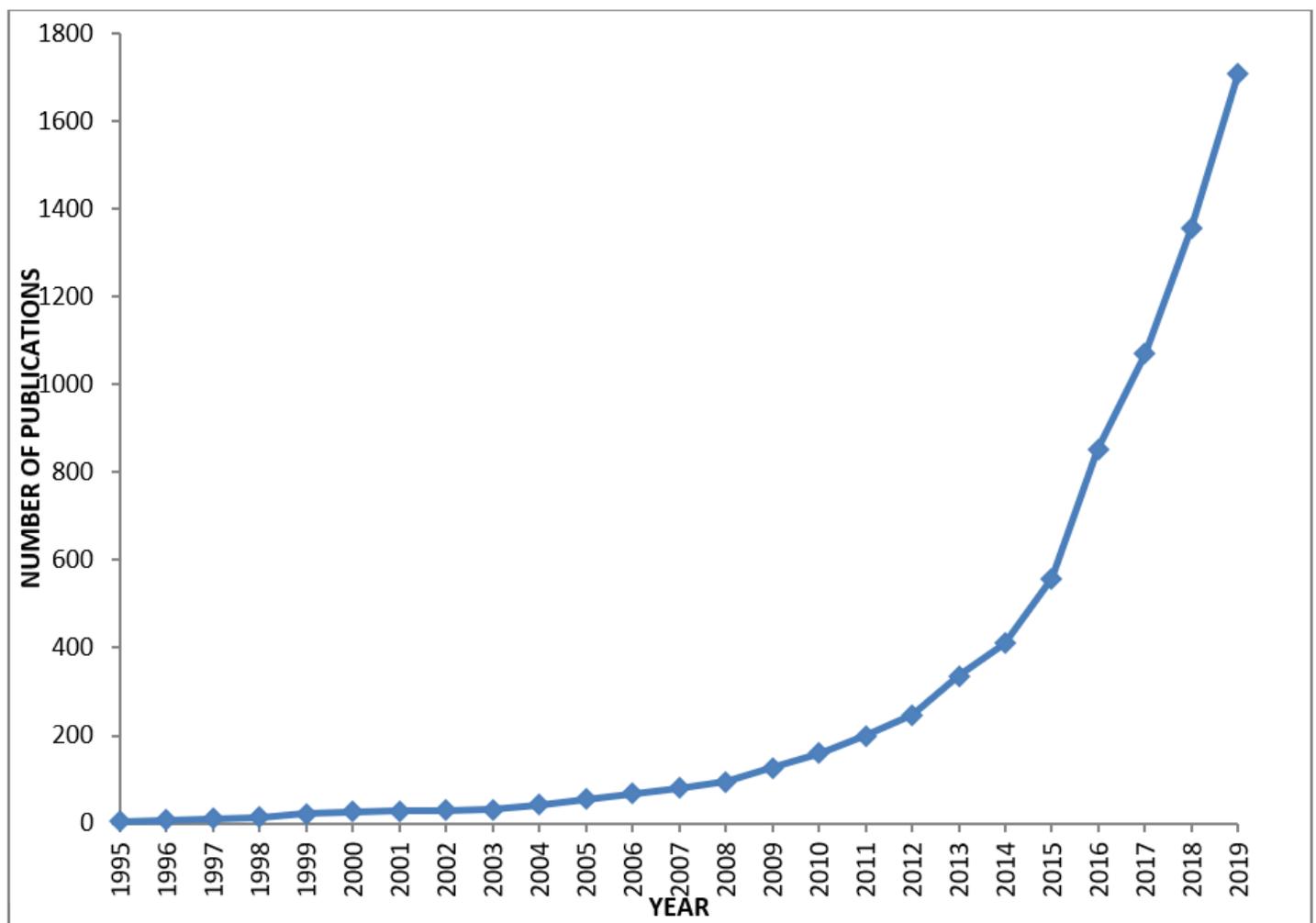


Figure 1

Running total of CCM-related publications listed by year in the core collection of the Web of Science, 1995-2019.

	HI_UPTAKE	SCHOOLS	MC	CCY	TI	SP	CW	SBI	LI	AD	SM	OHC	HC	PHC
Västernorrland	1	2	0	0	0	0	0	1	0	0	0	0	0	1
Jämtland	1	2	0	0	0	0	1	1	0	1	0	0	0	1
Värmland	1	2	0	0	0	0	0	1	1	0	0	0	0	1
Jönköping	1	2	0	1	0	0	1	1	1	1	1	0	0	1
Örebro	1	1	1	0	0	0	0	1	1	1	0	0	0	1
Dalarna	1	1	1	0	0	0	0	0	1	1	0	0	0	1
Halland	1	1	1	0	0	0	1	0	1	1	1	1	0	1
Skåne	0	1	1	1	0	0	1	1	1	1	1	1	0	1
Stockholm	0	1	0	0	0	0	0	1	1	1	1	0	1	1
Västerbotten	0	1	0	1	1	0	1	0	1	1	0	0	0	1
Västmanland	0	1	0	1	0	0	0	1	1	1	0	0	0	1
Västra Götaland	0	1	0	1	1	0	0	0	1	1	1	0	0	1
Norrbottn	0	0	1	0	0	0	0	1	1	1	1	0	0	1
Gävleborg	0	0	1	0	0	0	1	0	1	1	0	0	0	1
Kronoberg	0	0	1	0	0	0	0	0	1	0	0	0	0	1
Uppsala	0	0	1	1	0	1	1	0	1	1	0	0	0	1
Kalmar	0	0	0	0	0	0	0	0	1	0	0	0	0	1
Blekinge	0	0	0	0	0	0	0	0	1	1	0	0	0	1
Gotland	0	0	0	0	0	0	0	0	1	1	0	0	0	1
Östergötland	0	0	0	1	0	1	1	0	1	1	0	0	0	1
Södermanland	0	0	0	1	0	1	1	1	1	1	0	0	1	1

All Schools
Some Schools*Media Coverage*No Cinema Commercial/YouTube

*HI_UPTAKE = vaccination uptake >= 65%; SCHOOLS = all schools (2) some schools (1) no schools (0); MC = media coverage; CCY = Cinema commercial/YouTube; TI = Targeted Information; SP = Smart phone app; CW = county website; SBI = School-based information; LI = Letter/invitation; AD = Advertisement; SM = Social media; OHC = Online health care consulting; HC = health care center; PHC = primary health care center

Figure 2

Overall model identified by Coincidence Analysis with 100% consistency and 100% coverage.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [7CoincidenceAnalysisAdditionalfile3.docx](#)
- [6CoincidenceAnalysisAdditionalfile2.docx](#)
- [5CoincidenceAnalysisAdditionalfile1.docx](#)