Necessary Condition Analysis (NCA)
A novel view on causality and on empirical data analysis

Explaining the world

Natural sciences \( Y = f(X) \)
- Pythagoras’ equation \( c = \sqrt{a^2 + b^2} \)
- Velocity equation \( v = \frac{d}{t} \)
- Force equation \( F = m \cdot a \)
- Water equation \( H_2O = 2H + O \)
- Energy equation \( E = \frac{1}{2}m \cdot v^2 \)
- Nuclear energy equation \( E = m \cdot c^2 \)

Explaining the world

Social sciences \( Y = f(X) + \epsilon \)
- Regression equation \( Y = \beta_0 + \beta_1 X_1 + \epsilon \)
- Regression equation \( Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \)
- Regression equation \( Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon \)
- Regression equation \( Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon \)
- Regression equation …

Concerns/Questions

• Are practitioners interested in models that explain only a small part of the outcome?
• Can practitioners handle many factors at once (that may also interact and moderate)?
• Can we find a few critical factors that really matter?

A different look at our logic and data
**Why NCA?**

1. Novel method
2. New theoretical insights
3. Practical relevance
4. Publication opportunities
5. For all disciplines

---

**Various logics and methods**

**Regression:**

Aditive logic: *single* determinants are *sufficient but not necessary* for producing the (average) outcome

**NCA:**

Necessity logic: *single* determinants are *necessary but not sufficient* to allow the outcome.

---

**Necessary conditions are everywhere**

- Traveling to Nijmegen is necessary but not sufficient for attending this seminar in person
- A high GMAT test score is necessary but not sufficient for admission to a PhD program
- HIV is necessary but not sufficient for AIDS

---

**Practical relevance**

- If the necessary condition is not in place, there is guaranteed failure
- This absence of a necessary condition cannot be compensated by other conditions/determinants
- The presence of the necessary condition does not guarantee success, but may increase the likelihood of success

The necessary condition works in *isolation*, independently of the context

---

**Necessary conditions in Academia**

Examples from Academy of Management Journal:

- “Emotion recognition may be a necessary but insufficient ability involved in the performance of transformational leadership behavior.” (Rubin, Munz, & Bommer, 2005)
- Social relationships are necessary but not sufficient for promoting high-performing cross-BU collaboration.” (Martin & Eisenhardt, 2010)
- “Organizational commitment may be a necessary but insufficient condition for low absenteeism” (Haasnecht et al., 2008)
- “Managerial ties is necessary, but not sufficient for business success” (Peng & Luo, 2005)

---

**Necessary condition formulations**

**Enablers**

- X is necessary for Y
- X is needed for Y
- X is critical for Y
- X is crucial for Y
- X is essential for Y
- X is indispensable for Y
- X is a prerequisite for Y
- X is a requirement for Y
- X is a condition sine qua non for Y
- X is a pre-condition for Y
- X allows Y
- X enables Y
- There must be X to have Y
- Y requires X

**Constraints**

- X constrains Y
- X limits Y
- X blocks Y
- X restricts Y
- X is a barrier for Y
- X is a bottleneck for Y
- Without X there cannot be Y

Necessary conditions in academia

Goertz’ first law: “for any research area one can find important necessary condition hypotheses.”


Basic types of necessary conditions

NCA searches for empty spaces, independent of the data structure:

Combinations of dichotomous, discrete and continuous variables are possible

Combinations of types of necessary conditions

EXAMPLE large N NCA

Sociability is necessary but not sufficient for sales performance. Hogan Personality Inventory (HPI) (Hogan, & Hogan, 2007).

CR-FDH Ceiling line

Empty zone: Ceiling zone (C)
Total zone: Scope (S)
Effect size: $d = C/S$ (0 ≤ $d$ ≤ 1)
Accuracy: percentage of observations on or below the ceiling line

Bottleneck table

<table>
<thead>
<tr>
<th>Sales performance (%)</th>
<th>Sociability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>NN</td>
</tr>
<tr>
<td>10</td>
<td>NN</td>
</tr>
<tr>
<td>20</td>
<td>NN</td>
</tr>
<tr>
<td>30</td>
<td>NN</td>
</tr>
<tr>
<td>40</td>
<td>NN</td>
</tr>
<tr>
<td>50</td>
<td>NN</td>
</tr>
<tr>
<td>60</td>
<td>NN</td>
</tr>
<tr>
<td>70</td>
<td>23</td>
</tr>
<tr>
<td>80</td>
<td>48</td>
</tr>
<tr>
<td>90</td>
<td>75</td>
</tr>
<tr>
<td>100</td>
<td>98</td>
</tr>
</tbody>
</table>
Seminar Radboud University, Nijmegen, Netherlands, 12 December 2018

More information: email Jan Dul at jdul@rsm.nl.
See NCA website www.erim.nl/nca

Bottleneck table

<table>
<thead>
<tr>
<th>Sales performance (%)</th>
<th>Sociability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>NN</td>
</tr>
<tr>
<td>10</td>
<td>NN</td>
</tr>
<tr>
<td>20</td>
<td>NN</td>
</tr>
<tr>
<td>30</td>
<td>NN</td>
</tr>
<tr>
<td>40</td>
<td>NN</td>
</tr>
<tr>
<td>50</td>
<td>NN</td>
</tr>
<tr>
<td>60</td>
<td>NN</td>
</tr>
<tr>
<td>70</td>
<td>23</td>
</tr>
<tr>
<td>80</td>
<td>48</td>
</tr>
<tr>
<td>90</td>
<td>73</td>
</tr>
<tr>
<td>100</td>
<td>98</td>
</tr>
</tbody>
</table>

Bottleneck table

<table>
<thead>
<tr>
<th>Sales performance (%)</th>
<th>Sociability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>NN</td>
</tr>
<tr>
<td>10</td>
<td>NN</td>
</tr>
<tr>
<td>20</td>
<td>NN</td>
</tr>
<tr>
<td>30</td>
<td>NN</td>
</tr>
<tr>
<td>40</td>
<td>NN</td>
</tr>
<tr>
<td>50</td>
<td>NN</td>
</tr>
<tr>
<td>60</td>
<td>NN</td>
</tr>
<tr>
<td>70</td>
<td>23</td>
</tr>
<tr>
<td>80</td>
<td>48</td>
</tr>
<tr>
<td>90</td>
<td>73</td>
</tr>
<tr>
<td>100</td>
<td>98</td>
</tr>
</tbody>
</table>

Bottleneck table

<table>
<thead>
<tr>
<th>Sales performance (%)</th>
<th>Sociability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>NN</td>
</tr>
<tr>
<td>10</td>
<td>NN</td>
</tr>
<tr>
<td>20</td>
<td>NN</td>
</tr>
<tr>
<td>30</td>
<td>NN</td>
</tr>
<tr>
<td>40</td>
<td>NN</td>
</tr>
<tr>
<td>50</td>
<td>NN</td>
</tr>
<tr>
<td>60</td>
<td>NN</td>
</tr>
<tr>
<td>70</td>
<td>23</td>
</tr>
<tr>
<td>80</td>
<td>48</td>
</tr>
<tr>
<td>90</td>
<td>73</td>
</tr>
<tr>
<td>100</td>
<td>98</td>
</tr>
</tbody>
</table>

Bottleneck table

<table>
<thead>
<tr>
<th>Sales performance (%)</th>
<th>Sociability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>NN</td>
</tr>
<tr>
<td>10</td>
<td>NN</td>
</tr>
<tr>
<td>20</td>
<td>NN</td>
</tr>
<tr>
<td>30</td>
<td>NN</td>
</tr>
<tr>
<td>40</td>
<td>NN</td>
</tr>
<tr>
<td>50</td>
<td>NN</td>
</tr>
<tr>
<td>60</td>
<td>NN</td>
</tr>
<tr>
<td>70</td>
<td>23</td>
</tr>
<tr>
<td>80</td>
<td>48</td>
</tr>
<tr>
<td>90</td>
<td>73</td>
</tr>
<tr>
<td>100</td>
<td>98</td>
</tr>
</tbody>
</table>

© Jan Dul
CAUSAL Theory: X \rightarrow Y \text{ means } “causes”

NCA:
- Necessary cause
- X allows Y
- X is necessary but not sufficient for Y
- \( Y \leq f_{\text{ceiling}}(A,S,I,L) \)

Multiple Regression (OLS)

Regression Coefficient: 0.13

Effect size: 0.18

Classification:
- \( d < 0.1 \) "small effect"
- \( 0.1 \leq d < 0.3 \) "medium effect"
- \( 0.3 \leq d < 0.5 \) "large effect"
- \( d \geq 0.5 \) "very large effect"

What is effect of omitted variables?

Effect size: 0.18
How to test for necessity?

Three necessary but not sufficient conditions for a necessary condition:

1. Theoretical support
2. Practical significance (effect size $d > 0$)
3. Statistical significance ($p$ value $< 0.05$): not random

Significance test for NCA

Null Hypothesis testing:
1. Calculate the observed necessity effect size ($d$).
2. Formulate the null hypothesis: the effect size is the result of random chance.
3. Create a large set of random samples (e.g., 10,000) under the null hypothesis, using approximate permutation.
4. Calculate the effect sizes of these samples.
5. Compare this distribution of effect sizes of these samples with the effect size of the observed sample (see step 1).

The percentage of random resamples for which the effect size is equal to, or larger than the observed effect size ($p$-value) informs us about the statistical (in)compatibility of the data with the null hypothesis.

EXAMPLE

Intelligence is necessary but not sufficient for creativity

“The basic idea behind the threshold hypothesis is that high creativity requires high or at least above-average intelligence. At this, above-average intelligence is thought to form a necessary but not a sufficient condition for high creativity (Guilford, 1967).”

EXAMPLE large N NCA

NCA versus OLS

NCA
Searches for empty space
Draws a ceiling line (maximum $Y$ for given $X$)
$Y \leq f_{\text{ceiling}}(X)$

OLS
Searches for full space
Draws a central line (average $Y$ for given $X$)
$Y = f_{\text{central}}(X) + e$
Opinion of users of NCA:

- Simple and straightforward
- Provides new theoretical insights
- Complements other methods
- Relevant for practice

NCA Community

Since publication of NCA’s core paper in 2016:

- Start of NCA community with now 700+ researchers
- 50+ authors have used NCA
- Publications appeared in top journals
- Used in many fields

NCA website: [www.erim.nl/nca](http://www.erim.nl/nca)

NCA Calculator (on NCA website)

NCA software in R

Conclusion

Why wouldn’t you try NCA?

- NCA as main logic and analysis
- NCA in combination with regression
- NCA in combination with case study research