

# Analysing data on evaluation use using EvalC3

An illustrated example using Ledermann's 2012 data on evaluation use  
by the Swiss Development Cooperation

# What is EvalC3

- A user friendly Excel package of tools for exploring and evaluating models of complex causal configurations, combining
  - A QCA view of causality (Necessary and/or Sufficient causes)
  - A choice of predictive analytics algorithms, plus
  - Manual hypothesis testing, plus
  - Case selection aids for follow up within-case investigations, and
  - Tools for optimising selection of case attributes to investigate

# A (simple) example of a data set

Case	PC	LC	NV	EQ	CD
A	0	1	1	1	0
B	1	1	1	1	1
C	0	0	1	1	1
D	1	1	0	0	1
E	0	0	1	1	1
F	1	0	0	1	1
G	0	0	0	1	0
H	1	1	0	1	1
I	0	1	0	0	0
J	0	0	0	0	0
K	1	0	0	0	0

Ledermann, S. (2012). Exploring the Necessary Conditions for Evaluation Use in Program Change. *American Journal of Evaluation*, 33(2), 159–178.

Cases = evaluations by Swiss Development Cooperation

PC = Pressure for Change

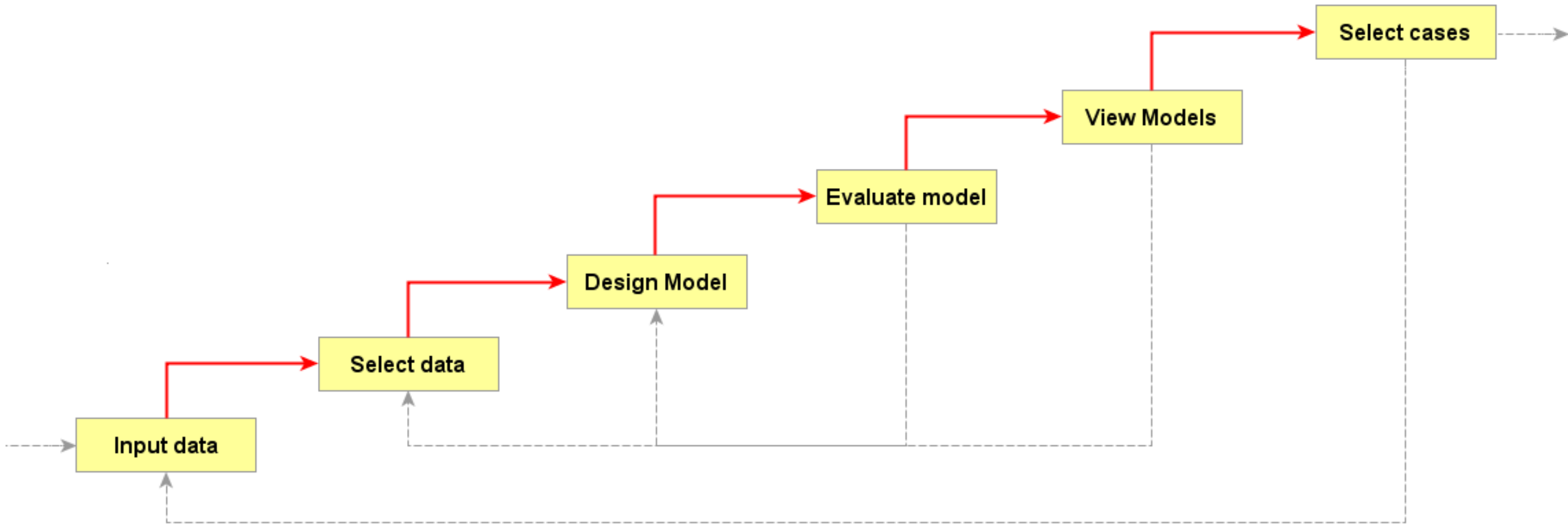
LC = Level of conflict

NV = Novelty Value

EQ = Evaluation Quality

CD = Change Decision – the outcome of interest

# The EvalC3 analysis workflow



Save model

Trigger2

## Design

Outcome	Status
Change Decision	Present

If these attributes have this status

attribute	status
Pressure for Change	1
Level of Conflict	n/a
Novelty Value	n/a
Evaluation Quality	n/a

Design menu

Performance measures

## Evaluate

Total cases: 11

Data says outcome is...

Model attributes are...	Data says outcome is...		cases	
	Present	Absent		
Present	TP = 4	FP = 1	45%	
Absent	FN = 2	TN = 4	55%	
	cases	55%	45%	100%

## Model status

The current set of attributes is:

The current set of attributes is:	
Not Necessary	Not Sufficient
for the outcome to be: Present	
Simplicity	25%
Support	45%

## Model performance

Overall	Metric	Value
	Accuracy	73%
	Balanced accuracy	73%
	F1-score	73%
	Mathews Correlation Coefficient	47%
	Gini index	61%
Specific	True positive rate	67%
	Positive Predictive Value	80%
Relative	Lift	147%
	Null error rate	45%
	Likelihood ratio (positive)	333%

## Explore

Find new models

Algorithm choices

### EvalC3: Find New Models

Find the configuration of attributes that best predicts the presence/absence of the outcome according to a selected performance indicator.

Search type

- Exhaustively test all configurations of attributes
- Evolutionary search for best configuration of attributes (using Solver)
- Find one additional attribute that gives best performance
- Build a decision tree with maximum depth: 4

Value to optimise

Number of attributes

Number of attributes in model configuration: any

Constraints

Buttons: Add, Change, Delete, Cancel, OK

# Ledermann's theory-based model of conditions for evaluation use

Context Conditions		Assumed Necessary Actor Conditions		
Pressure for Change	Level of Conflict	Mechanism	Novelty Value	Evaluation Quality
Low	Low	Awakener	High	High
High	Low	Trigger	Irrelevant	High
High	High	Referee	Irrelevant	Irrelevant
Low	High	Conciliator	High	High

“...this article examines the interplay between the policy and decision setting on the one hand and specific characteristics of evaluation implementation on the other hand”

- Awakener

### Design ?

Outcome	Status
Change Decision	Present
If these attributes have this status	
attribute	status
Pressure for Change	0
Level of Conflict	0
Novelty Value	1
Evaluation Quality	1

### Evaluate ?

Total cases: 11

		Data says outcome is...		
		Present	Absent	cases
Model attributes are...	Present	TP = 2	FP = 0	18%
	Absent	FN = 4	TN = 5	82%
		cases	55%	45%
				100%

### Model status ?

The current set of attributes is:

Not Necessary ?	Sufficient ?
for the outcome to be:	
Present	

- Trigger

### Design ?

Outcome	Status
Change Decision	Present
If these attributes have this status	
attribute	status
Pressure for Change	1
Level of Conflict	0
Novelty Value	n/a
Evaluation Quality	1

### Evaluate ?

Total cases: 11

		Data says outcome is...		
		Present	Absent	cases
Model attributes are...	Present	TP = 1	FP = 0	9%
	Absent	FN = 5	TN = 5	91%
		cases	55%	45%
				100%

### Model status ?

The current set of attributes is:

Not Necessary ?	Sufficient ?
-----------------	--------------

- Referee

### Design ?

Outcome	Status
Change Decision	Present

If these attributes have this status

attribute	status
Pressure for Change	1
Level of Conflict	1
Novelty Value	n/a
Evaluation Quality	n/a

### Evaluate ?

Total cases: 11

Data says outcome is...

Model attributes are...	Present	Absent	cases
Present	TP = 3	FP = 0	27%
Absent	FN = 3	TN = 5	73%
	cases 55%	cases 45%	100%

### Model status ?

The current set of attributes is:

Not Necessary ?      Sufficient ?

- Conciliator

### Design ?

Outcome	Status
Change Decision	Present

If these attributes have this status

attribute	status
Pressure for Change	0
Level of Conflict	1
Novelty Value	1
Evaluation Quality	1

### Evaluate ?

Total cases: 11

Data says outcome is...

Model attributes are...	Present	Absent	cases
Present	TP = 0	FP = 1	9%
Absent	FN = 6	TN = 4	91%
	cases 55%	cases 45%	100%

### Model status ?

The current set of attributes is:

Not Necessary ?      Not Sufficient ?

for the outcome to be: Present



- Referee versions (based on examination of case details)

- Endorser

### Design ?

Outcome	Status
Change Decision	Present

If these attributes have this status

attribute	status
Pressure for Change	1
Level of Conflict	1
Novelty Value	0
Evaluation Quality	n/a

### Evaluate ?

Total cases: 11

		Data says outcome is...		
		Present	Absent	cases
Model attributes are...	Present	TP = 2	FP = 0	18%
	Absent	FN = 4	TN = 5	82%
		cases	55%	45%
				100%

### Model status ?

The current set of attributes is:

Not Necessary ?      Sufficient ?

for the outcome to be: **Present**

- Reviser

### Design ?

Outcome	Status
Change Decision	Present

If these attributes have this status

attribute	status
Pressure for Change	n/a
Level of Conflict	n/a
Novelty Value	1
Evaluation Quality	1

### Evaluate ?

Total cases: 11

		Data says outcome is...		
		Present	Absent	cases
Model attributes are...	Present	TP = 3	FP = 1	36%
	Absent	FN = 3	TN = 4	64%
		cases	55%	45%
				100%

### Model status ?

The current set of attributes is:

Not Necessary ?      Not Sufficient ?

# How many models do we need to test?

- With four attributes there are  $2^4$  possible models = 16.
  - Six tested so far, based on theory and case examination
- The number of possibly useful models rises exponentially with number of attributes. So if 5 attributes = 32 possible models
- Search algorithms can help us explore “outside the box” of our existing theory(ies) e.g. Pressure for Change + Evaluation Quality

**Design** ?

Outcome	Status
Change Decision	Present

If these attributes have this status

attribute	status
Pressure for Change	1
Level of Conflict	n/a
Novelty Value	n/a
Evaluation Quality	1

**Evaluate** ?

Total cases: 11

Data says outcome is...

	Present	Absent	cases
Model attributes are... Present	TP = 3	FP = 0	27%
Absent	FN = 3	TN = 5	73%
	cases 55%	45%	100%

**Model status** ?

The current set of attributes is:

Not Necessary	Sufficient
---------------	------------

for the outcome to be: Present

# How do we choose between models?

- Parsimony -
  - % of available attributes in the model &
    - # of models to cover all cases with outcome of interest
- Coverage - % of cases a model applies to
  - Simpler models can have wider coverage ← Share traders
- Consistency - % of True Positive cases
  - Complex models can have more consistency ← surgeons
- Within-case mechanism details – but which cases to spend time on?

Index	Cases	Awakener	Trigger	Referee	Conciliator	Endorser	Reviser	PC & EQ	# of models
1	A	0	0	0	0	0	0	0	0
2	B	0	0	1	0	0	1	1	3
3	C	1	0	0	0	0	★ 1	0	2
4	D	0	0	1	0	1	0	0	2
5	E	1	0	0	0	0	★ 1	0	2
6	F	0	1	0	0	0	0	1	2
7	G	0	0	0	0	0	0	0	0
8	H	0	0	1	0	1	0	1	3
9	I	0	0	0	0	0	0	0	0
10	J	0	0	0	0	0	0	0	0
11	K	0	0	0	0	0	0	0	0
<b># True positives:</b>		2	1	3	0	2	3	3	
<b>% Unique coverage:</b>		0%	0%	0%	0%	0%	0%	0%	
<b>Aggregate coverage:</b>									<b>100%</b>

- Look for cases with the most conflicting models
  - Especially models with wider coverage?

# Balthasar, A. (2006). The Effects of Institutional Design on the Utilization of Evaluation. Evaluation, 12(3), 353–71.

Index	Cases	PurpS0&PurpForm0	Dist0&PurpoBoth1	Dist1&PurpSum1	Dist0PurpSum0	# of models
1	PCA	0	0	0	0	0
2	SFAO	0	0	0	0	0
3	SDC	1	0	0	1	2
4	SFOPH	1	0	0	0	1
5	FOM	1	0	0	1	2
6	DDPS	0	0	1	0	1
7	FOP	0	0	0	0	0
8	SFVO	0	0	0	0	0
9	FOT	0	1	0	1	2
10	SFOE	0	0	0	0	0
<b># True positives:</b>		<b>3</b>	<b>1</b>	<b>1</b>	<b>3</b>	
<b>% Unique coverage:</b>		<b>20%</b>	<b>0%</b>	<b>20%</b>	<b>0%</b>	
					<b>Aggregate coverage:</b>	<b>100%</b>

4 models of Sufficient conditions for Outcome (=Instrumental use of evaluations)

5 conditions = 32 possible models

QCA result

Exhaustive search



Calculate Hamming Distances

Model:

Index	Status	Hamming Distance	Cases	Pressure for Change	Level of Conflict	Novelty Value	Evaluation Quality	Outcome
3	FN	33%	C	0	0	1	1	1
4	FN	67%	D	1	1	0	0	1
5	FN	33%	E	0	0	1	1	1
1	TN	55%	A	0	1	1	1	0
7	TN	35%	G	0	0	0	1	0
9	TN	35%	I	0	1	0	0	0
10	TN	30%	J	0	0	0	0	0
11	TN	45%	K	1	0	0	0	0
2	TP	25%	B	1	1	1	1	1
6	TP	25%	F	1	0	0	1	1
8	TP	17%	H	1	1	0	1	1

- Look for modal and extreme cases within True Positives, to see if the same causal mechanism can be found at work

# Where to next?

- Development of more case-selection aids, especially for case comparisons
  - Goertz, G. (2017). Multimethod research, causal mechanisms, and case studies: an integrated approach. Princeton University Press.
- Accumulation of data sets on evaluation use
  - E.g. ICAI already does an annual review, but is not yet systematically gathering analysable data
- Encourage wider use of EvalC3
  - Free software & free tech support via skype 😊

# Wider implications: ToC design & evaluation

What are we doing to bring change?	What will change in the short term?
Investing in capacity building	Capable and adaptable organisations
Nurturing and supporting the growth of not-for profit sector	Enhanced leadership and resilience
Collaborative learning	Understanding impact and value
Enabling innovation	Increased potential for scaling and replication
Sustaining engaged and collaborative relationships	Embedded learning and enhanced communication
Supporting philanthropic innovation	Collaborative investment

6 outputs leading to 6 outcomes

=

36 possible causal connections between Output and Outcome

=

$2^{36}$  possible combinations of these connections

Having a theory of how things work will help us identify connections worth testing

But there may be more than a few other possibilities worth exploring

The more cases we have the more we can explore these

But the rate of discovery of these diverse cases may diminish

Centre for Social Impact, NZ.