Informative hypotheses evaluation

Multiple studies

R. M. Kuiper

Department of Methodology & Statistics
Utrecht University

Possibilities multiple studies

- Update GORIC(A) values & weights.
  More data collected: (re-)calculate.
- Update hypotheses.
  First data set (or a part of it) generates one or more hypotheses.
  Other data set (or part) used to determine evidence / support.
  See this html tutorial and/or this R script tutorial.
- Aggregate evidence for hypotheses.
  Aggregate the support for theories (diverse designs allowed).
  Bear in mind: Meta-analysis aggregates parameter estimates or effect sizes which need to be comparable (often same designs required).
  See this html tutorial and/or this R script tutorial.

Table of Contents

Multiple Studies: Updating hypotheses

Update Hypotheses (go from exploration to confirmation)

1. 1st study: Explore & Obtain informative hypothesis(-es).
2. Replicated study: Evaluate updated, informative hypothesis(-es).

Example:

1. 1st study: Monin, Sawyer, and Marquez (2008)
2. Replicated study: Holubar (2015).

investigate the attraction to "moral rebels", that is, persons that take an unpopular morally laudable stand.

Imagine that you are in a group (all others in group are actors) and that the atmosphere in the group is that criminal behavior is linked to having an African American background.

- You publicly have to rate your attraction to a person in a video.
- This is repeated using the same group of actors with you replaced by another person, that is, there are more participants in the experiment that have to rate the attraction to a person in a video.
- There are three experimental conditions (see the next slide).

Example Monin and Holubar: Conditions

Three conditions:

1. Condition 1: participants rate the attraction to a person that is 'obedient' and selects an African American person from a police line up of three.
2. Condition 2: participants rate a moral rebel (a person not selecting the African American person) after executing a self-affirmation task intended to boost their self-confidence.
3. Condition 3: participants rate a moral rebel after executing a bogus writing task.

Example Monin and Holubar: Explore in 1st study

Hypotheses evaluated for the Monin data

$$H_0 : \mu_1 = \mu_2 = \mu_3$$
$$H_{a1} : \mu_1 = \mu_2,\ \mu_3$$
$$H_{a2} : \mu_1 = \mu_3,\ \mu_2$$
$$H_{a3} : \mu_2 = \mu_3,\ \mu_1$$
$$H_u : \mu_1,\ \mu_2,\ \mu_3,$$

Example Monin and Holubar: Explore in 1st study

Using GORIC

	model	loglik	penalty	goric	goric.weights
1	H0	-149.907	2.000	303.815	0.000
2	Ha1	-141.191	3.000	288.383	0.610
3	Ha2	-145.404	3.000	296.809	0.009
4	Ha3	-148.907	3.000	303.815	0.000
5	unconstrained	-140.665	4.000	289.330	0.380

Conclusion: $H_{a1}: \mu_1 = \mu_2,\ \mu_3$ is best.

Descriptives obtained for the Monin data:

group	n	mean	sd
1	19	1.88	1.38
2	19	2.54	1.95
3	29	0.02	2.38

So, $\hat{\mu}_1$ and $\hat{\mu}_2$ are larger than $\hat{\mu}_3$.

Updated hypothesis: $H_1: \mu_1 = \mu_2 > \mu_3$
This will be evaluated in Holubar data.

New set of hypotheses:

- H_1 against its complement (or unconstrained hypothesis H_a).
- H_1 with another updated hypothesis, based on support in exploratory phase, and H_a.
  e.g., could also choose to update $H_u : \mu_1,\ \mu_2,\ \mu_3$ (using $\hat{\mu}_2 > \hat{\mu}_1 > \hat{\mu}_3$), leading to $H_2 : \mu_2 > \mu_1 > \mu_3$.
- H_0, H_1, and H_a.

I will show the results of the first set choice.

$$H_1 : \mu_1 = \mu_2 > \mu_3$$

$$H_a : \mu_1, \mu_2, \mu_3$$

Replicating Monin, Sawyer, and Marquez (2008) using the Holubar data

```
Results:
        model    loglik  penalty    goric  goric.weights
1          H1  -144.981    2.500  294.962          0.280
2  complement  -143.038    3.500  293.076          0.720
---
The order-restricted hypothesis 'H1' has  0.390 times
more support than its complement.
```

Hence, the results of Monin are not replicated.

Update Hypotheses: TRAILS studies

using GORICA

1. Explore:
Use results from study Nederhof, Ormel, and Oldehinkel (2014)
Use theory from Nederhof and Schmidt (2012)
Discuss with authors Nederhof and Oldehinkel.
Result: Two informative hypotheses.

2. Evaluate informative hypotheses in replication.

Reference:
Altınışık, Y., Nederhof, E., Hoijtink, H., Oldehinkel, A.J., and Kuiper, R.M. (accepted 2021). Evaluation of Inequality Constrained Hypotheses Using a Generalization of the AIC. *Psychological Methods*.

Update Hypotheses: TRAILS studies

using GORICA

- 11 years old participants are divided into three groups:
  $1 =$ Sustainers, $2 =$ Shifters, and $3 =$ Comparison group,
  based on their performance on a sustained-attention task and on a shifting-set task.
- Outcome: depressive episode
  (D: $0 =$ no depressive episode, $1 =$ endorsed an episode)
- Predictors: early life stress (ES: $0 =$ low, $1 =$ high),
  recent stress (RS, continuous), and
  their interaction.
- RS is standardized to improve interpretation of main effects when interactions exist.

Update Hypotheses: TRAILS studies

using GORICA

- Outcome is dichotomous, so logistic regression model:

$$f(\hat{D}_{ji}) = \begin{cases} \beta_{j0} + \beta_{j1}\mathrm{RS_{ji}} & \text{if } \mathrm{ES} = 0 \text{ (low)} \\ (\beta_{j0} + \beta_{j2}) + (\beta_{j1} + \beta_{j3})\mathrm{RS_{ji}} & \text{if } \mathrm{ES} = 1 \text{ (high)}. \end{cases}$$

- Note: We only have parameter estimates and their covariance matrix.
- Thus: Use gorica.
  For the goric, we need the model / (g)lm object in R and thus the full data set.

Update Hypotheses: TRAILS studies

using GORICA

$$f(\hat{D}_{ji}) = \begin{cases} \beta_{j0} + \beta_{j1}\mathrm{RS}_{\mathrm{ji}} & \text{if ES} = 0 \text{ (low)} \\ (\beta_{j0} + \beta_{j2}) + (\beta_{j1} + \beta_{j3})\mathrm{RS}_{\mathrm{ji}} & \text{if ES} = 1 \text{ (high)}. \end{cases}$$

mismatch expectation states that the risk of depression for adolescents with low levels of early life stress ($ES = 0$) increases with high recent stress levels (i.e., $\beta_{j1} > 0$), while adolescents with high levels of early life stress ($ES = 1$) are not affected by high recent stress levels (i.e., $\beta_{j1} + \beta_{j3} = 0$).

cumulative stress expectation states that there is no interaction between early and recent life stress (i.e., $\beta_{j3} = 0$), that is, only the main effect of recent stress predicts depression; and, furthermore, that this relation is positive (i.e., $\beta_{j1} > 0$).

In the hypotheses, one or none of these expectations apply to each of the three groups.

Update Hypotheses: TRAILS studies

using GORICA

H_1 (theory in Nederhof and Schmidt (2012))

- mismatch expectation applies to sustainers ($j = 1$) and shifters ($j = 2$).
- cumulative stress expectation applies to comparison groups ($j = 3$).

H_2 (based on results in Nederhof et al. (2014, p. 689))

- mismatch expectation applies to sustainers ($j = 1$).
- none of them apply to shifters ($j = 2$).
- cumulative stress expectation applies to comparison groups ($j = 3$).

H_u

no restrictions on parameters.

Included as safeguard.

Update Hypotheses: TRAILS studies

using GORICA

	(Sustainers)	**(Shifters)**	**(Comparison)**
H_1 :	$\beta_{11} + \beta_{13} = 0, \beta_{11} > 0,$	$\beta_{21} + \beta_{23} = 0, \beta_{21} > 0,$	$\beta_{33} = 0, \beta_{31} > 0,$
H_2 :	$\beta_{11} + \beta_{13} = 0, \beta_{11} > 0,$	$\beta_{21} = \beta_{23} = 0,$	$\beta_{33} = 0, \beta_{31} > 0,$
H_u :	$\beta_{11}, \beta_{13},$	$\beta_{21}, \beta_{23},$	$\beta_{31}, \beta_{33}.$

TRAILS studies: Results

using GORICA

	model	loglik	penalty	gorica	gorica.weights
1	H1	-1.373	1.500	5.746	0.776
2	H2	-3.168	1.000	8.335	0.212
3	unconstrained	-0.045	7.000	14.089	0.012

Notes

H_2 is more specific and thus it has a lower penalty.
H_1 fits data better and fit difference outweighs penalty difference.

Conclusion

Hypothesis H_1 has $0.776/0.212 = 3.65$ times more support than hypothesis H_2.
That is, mismatch expectation applies to both sustainers and shifters, and cumulative stress expectation applies to comparison groups.

Table of Contents

BMS & GORIC(A) for Multiple Studies: Aggregating support (= evidence synthesis)

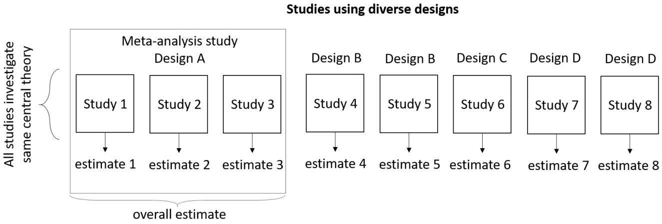
Motivation

In science, the gold standard for evidence is an empirical result that is consistent across multiple studies.

- **Replicability/Replication crisis** in social science.
- Political scientists call for meta-scientific introspection.

Therefore, need for aggregating results.

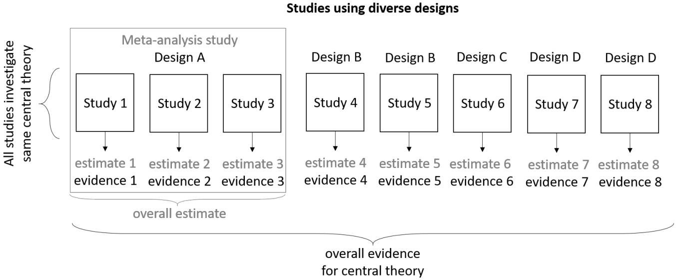
Current best practice



Current best practice is meta-analysis and Bayesian updating.

- Not applicable for diverse research designs.
- Not applicable for incomparable estimates.

Need for new methodology: Evidence Synthesis



Note: All studies do investigate the same theory (using diverse designs).

Trust Example: Meta-Analysis versus Evidence Synthesis

Study	Type of model
1	univariate regression
2	univariate regression
3	probit regression
4	three-level logistic regression
	Same design? e.g., same set of predictors?

Conceptual replications!

	Meta-Analysis	Evidence Synthesis
Effect size not required		✓
Deal with diverse designs		✓
Main results	Estimate of effect size	Evidence for hypotheses
Check:		same theoretical relationships?

Reference:

Kuiper, R.M., Buskens, V.W., Raub, W., and Hoijtink, H. (2013). Combining statistical evidence from several studies: A method using Bayesian updating and an example from research on trust problems in social and economic exchange. *Sociological Methods and Research*, 42 (1), (pp. 60-81) (22 p.).

Trust Example: Meta-Analysis versus Evidence Synthesis

Study	Type of model
1	univariate regression
2	univariate regression
3	probit regression
4	three-level logistic regression
	Same design? e.g., same set of predictors?

Conceptual replications!

	Meta-Analysis	Evidence Synthesis
Effect size not required		✓
Deal with diverse designs		✓
Main results	Estimate of effect size	Evidence for hypotheses
Check:		same theoretical relationships?

Reference:

Kuiper, R.M., Buskens, V.W., Raub, W., and Hoijtink, H. (2013). Combining statistical evidence from several studies: A method using Bayesian updating and an example from research on trust problems in social and economic exchange. *Sociological Methods and Research*, 42 (1), (pp. 60-81) (22 p.).

Trust Example: Meta-Analysis versus Evidence Synthesis

Study	Type of model
1	univariate regression
2	univariate regression
3	probit regression
4	three-level logistic regression
	Same design? e.g., same set of predictors?

Conceptual replications!

	Meta-Analysis	Evidence Synthesis
Effect size not required		✓
Deal with diverse designs		✓
Main results	Estimate of effect size	Evidence for hypotheses
Check:		same theoretical relationships?

Reference:

Kuiper, R.M., Buskens, V.W., Raub, W., and Hoijtink, H. (2013). Combining statistical evidence from several studies: A method using Bayesian updating and an example from research on trust problems in social and economic exchange. *Sociological Methods and Research*, 42 (1), (pp. 60-81) (22 p.).

Trust Example: Meta-Analysis versus Evidence Synthesis

Study	Type of model
1	univariate regression
2	univariate regression
3	probit regression
4	three-level logistic regression
	Same design? e.g., same set of predictors?

Conceptual replications!

	Meta-Analysis	Evidence Synthesis
Effect size not required		✓
Deal with diverse designs		✓
Main results	Estimate of effect size	Evidence for hypotheses
Check:		**same theoretical relationships?**

Reference:

Kuiper, R.M., Buskens, V.W., Raub, W., and Hoijtink, H. (2013). Combining statistical evidence from several studies: A method using Bayesian updating and an example from research on trust problems in social and economic exchange. *Sociological Methods and Research*, 42 (1), (pp. 60-81) (22 p.).

Example: 4 studies regarding one concept

Study	Type of study	Number of observations n	Type of model
1	survey	895 transactions	univariate regression
2	experiment	348 decisions by 40 subjects	univariate regression
3	experiment	1249 decisions by 125 subjects	probit regression
4	experiment	2160 decisions by 144 subjects	three-level logistic regression
Study	**Outcome y (trust)**		**scale y**
1	effort invested in management		ratio
2	effort invested in management		ratio
3	choice of vignettes		dummy
4	trustfulness		dummy
Study	**Predictor x_1 (past / previous experience)**		**scale x_1**
1	existence relationship with supplier		dummy
2	type of relationship with supplier		interval
3	bought a car from The Autoshop before		dummy
4	number of times a trustee honored trust in the past		ratio
Study	**some of the other predictors**		
1	transaction characteristics, expected future transactions, network embeddedness		
2	transaction characteristics, expected future transactions, network embeddedness		
3	expected future transactions, network embeddedness		
4	future interactions, network embeddedness		

One-Parameter Example: Hypotheses of interest

Parameter of interest in each study

parameter corresponding to x_1 = previous experience; i.e., β_1.
For simplicity, only one here, could have been more.

Expectation in each study

x_1 = previous experience has a positive effect on y = trust; i.e, $\beta_1 > 0$.

Set of central theories

$$H_0 : \textit{no effect},$$
$$H_> : \textit{positive effect},$$
$$H_< : \textit{negative effect}.$$

Note 1: These are hypotheses for the effect in all studies,
and thus not regarding the average parameter.
In each data set, the hypotheses reflecting the theories may differ (e.g., $\beta > 0$ versus $OR > 1$). Note 2: In practice, I would not include H_0...

One-Parameter Example: Hypotheses of interest

Parameter of interest in each study

parameter corresponding to x_1 = previous experience; i.e., β_1.
For simplicity, only one here, could have been more.

Expectation in each study

x_1 = previous experience has a positive effect on y = trust; i.e, $\beta_1 > 0$.

Set of central theories

$$H_0 : \textit{no effect},$$
$$H_> : \textit{positive effect},$$
$$H_< : \textit{negative effect}.$$

Note 1: These are hypotheses for the effect in all studies,
and thus not regarding the average parameter.
In each data set, the hypotheses reflecting the theories may differ (e.g.,
$\beta > 0$ versus $OR > 1$). Note 2: In practice, I would not include H_0...

One-Parameter Example: Hypotheses of interest

Parameter of interest in each study

parameter corresponding to x_1 = previous experience; i.e., β_1.
For simplicity, only one here, could have been more.

Expectation in each study

x_1 = previous experience has a positive effect on y = trust; i.e, $\beta_1 > 0$.

Set of central theories

$$H_0 : \textit{no effect},$$
$$H_> : \textit{positive effect},$$
$$H_< : \textit{negative effect}.$$

Note 1: These are hypotheses for the effect in all studies,
and thus not regarding the average parameter.
In each data set, the hypotheses reflecting the theories may differ (e.g., $\beta > 0$ versus $OR > 1$). Note 2: In practice, I would not include H_0...

Example: Trust (y) & previous experience (x_1)

Not full data set (and probit regression), so use

- GORICA (not GORIC) using *goric* function in R package *restriktor*

Input:

- parameter estimates and their covariance matrix

t	$\hat{\beta}_1$	$\hat{\sigma}_{\beta_1}$
1	0.090	0.029
2	0.140	0.054
3	1.090	0.093
4	1.781	0.179

Note: Here, one parameter (β_1); thus, cov. matrix $\hat{\beta}_1$ = variance $\hat{\beta}_1 = \hat{\sigma}^2_{\beta_1}$ (not $\hat{\sigma}_{\beta_1}$)

One-Parameter Example: results per study

using GORICA

Results per study (not aggregated yet)!

Table: GORICA weights ($w_{t,m}$) for Hypothesis H_m in Study t

	$w_{t,m}$			
m / t	1	2	3	4
0	0.013	0.052	0.000	0.000
$>$	**0.979**	**0.916**	**1.000**	**1.000**
$<$	0.008	0.032	0.000	0.000

Note: Weight is at max 1.

So, now on forehand already clear.... but no quantification yet.

One-Parameter Example: Results & Conclusions

using GORICA

Table: Overall GORICA weights ($w^1_{t,m}$) for Hypothesis H_m in Study t

	$w^1_{t,m}$			
m / t	1	2	3	4
0	0.013	0.001	0.000	0.000
$>$	**0.979**	**0.999**	**1.000**	**1.000**
$<$	0.008	0.000	0.000	0.000

- $w^1_{4,>} = 1$ => full support for $H_>$
  $w^1_{4,0} = w^1_{4,<} = 0$ => no support for H_0 and $H_<$
- Support for $H_>$ ($w^1_{4,1}$) is highest: favor $H_>$ over H_0 and $H_<$.

One-Parameter Example: Results & Conclusions

using GORICA

Table: Overall GORICA weights ($w^1_{t,m}$) for Hypothesis H_m in Study t

	$w^1_{t,m}$			
m / t	1	2	3	4
0	0.013	0.001	0.000	0.000
$>$	**0.979**	**0.999**	**1.000**	**1.000**
$<$	0.008	0.000	0.000	0.000

- $w^1_{4,>} = 1$ => full support for $H_>$
  $w^1_{4,0} = w^1_{4,<} = 0$ => no support for H_0 and $H_<$
- Support for $H_>$ ($w^1_{4,1}$) is highest: favor $H_>$ over H_0 and $H_<$.

One-Parameter Example: Results & Conclusions

using GORICA

Table: Overall GORICA weights ($w^1_{t,m}$) for Hypothesis H_m in Study t

	$w^1_{t,m}$			
m / t	1	2	3	4
0	0.013	0.001	0.000	0.000
$>$	**0.979**	**0.999**	**1.000**	**1.000**
$<$	0.008	0.000	0.000	0.000

- $w^1_{4,>} = 1$ => full support for $H_>$
  $w^1_{4,0} = w^1_{4,<} = 0$ => no support for H_0 and $H_<$
- Support for $H_>$ ($w^1_{4,1}$) is highest: favor $H_>$ over H_0 and $H_<$.

One-Parameter Example: Results & Conclusions

using GORICA

Table: Overall GORICA weights ($w^1_{t,m}$) for Hypothesis H_m in Study t

	$w^1_{t,m}$			
m / t	1	2	3	4
0	0.013	0.001	0.000	0.000
$>$	**0.979**	**0.999**	**1.000**	**1.000**
$<$	0.008	0.000	0.000	0.000

- $w^1_{4,>} = 1$ => full support for $H_>$
  $w^1_{4,0} = w^1_{4,<} = 0$ => no support for H_0 and $H_<$
- Support for $H_>$ ($w^1_{4,1}$) is highest: favor $H_>$ over H_0 and $H_<$.

One-Parameter Example: Results & Conclusions

using GORICA

Table: Overall GORICA weights ($w^1_{t,m}$) for Hypothesis H_m in Study t

	$w^1_{t,m}$			
m / t	1	2	3	4
0	0.013	0.001	0.000	0.000
$>$	**0.979**	**0.999**	**1.000**	**1.000**
$<$	0.008	0.000	0.000	0.000

- $w^1_{4,>} = 1$ => full support for $H_>$
  $w^1_{4,0} = w^1_{4,<} = 0$ => no support for H_0 and $H_<$
- Support for $H_>$ ($w^1_{4,1}$) is highest: favor $H_>$ over H_0 and $H_<$.

One-Parameter Example: Results & Conclusions

using GORICA

Table: Overall GORICA weights ($w^1_{t,m}$) for Hypothesis H_m in Study t

	$w^1_{t,m}$			
m / t	1	2	3	4
0	0.013	0.001	0.000	0.000
$>$	**0.979**	**0.999**	**1.000**	**1.000**
$<$	0.008	0.000	0.000	0.000

- $w^1_{4,>} = 1$ => full support for $H_>$
  $w^1_{4,0} = w^1_{4,<} = 0$ => no support for H_0 and $H_<$
- Support for $H_>$ ($w^1_{4,1}$) is highest: favor $H_>$ over H_0 and $H_<$.

Table of Contents

Multiple (Conceptual) Replication Studies: Updating hypotheses & Evidence synthesis

Example

using GORICA

Example based on results in Zondervan-Zwijnenburg et al. (2020):

RQ: Can age of the mother predict externalizing problem behavior of children around the age of 11.
(rated by the mother using the CBCL child behavior checklist)

Studied by 3 cohort studies in the Netherlands:
TRAILS (N=1955), NTR (N=21921), and GEN-R (N=4549).

Reference:
Zondervan-Zwijnenburg et al. (2020). Parental Age and Offspring Childhood Mental Health: A Multi-Cohort, Population-Based Investigation. *Child Development*. 91(3), 964-982.

Example: Notes

using GORICA

Each of the cohorts measured the variables in their own way:
so, different operationalisation of same constructs.
Hence, cannot use meta-analysis nor Bayesian updating.

They did not want evidence for pattern on average, but evidence that pattern exist in each of the three studies.

Updating hypotheses & Evidence synthesis

using GORICA

Steps:

1. Randomly divide the data of each cohort into an exploratory and confirmatory part.
2. Use the exploratory data to construct informative hypotheses.
3. Use the confirmatory data to evaluate the informative hypotheses.
4. Evidence synthesis: Combine the results obtained for the three cohorts into one overall conclusion.

Updating hypotheses & Evidence synthesis: Example

Step 1

After randomly choosing 50% of each data set (the exploration set), the following results were obtained for each cohort:

Cohort	β_1	p-val	β_2	p-val	R^2
Gen-R	-.10	$<.001$	.02	$<.001$	.02
NTR	-.11	$<.001$	.06	$<.001$	.02
TRAILS	-.13	$<.001$	.06	.06	.02

where the model was:

$$\text{CBCL} = \beta_0 + \beta_1 \text{age} + \beta_2 \text{age}^2 + \text{error} \quad (1)$$

Updating hypotheses & Evidence synthesis: Example

Step 1

Cohort	β_1	p-val	β_2	p-val	R^2
Gen-R	-.10	$<$.001	.02	$<$.001	.02
NTR	-.11	$<$.001	.06	$<$.001	.02
TRAILS	-.13	$<$.001	.06	.06	.02

Updated hypothesis:
- Significance and sign imply: $\beta_1 < 0$ & $\beta_2 > 0$.

Competing hypotheses:
- Because effects seem small: $\beta_1 = 0$ & $\beta_2 = 0$.
- Because second one not always significant: $\beta_1 < 0$ & $\beta_2 = 0$.

Updating hypotheses & Evidence synthesis: Example

Step 2

Set of competing informative hypotheses:

$$H_3 : \beta_1 < 0 \ \& \ \beta_2 > 0,$$

that is, the older the mothers the less externalizing problems occur, and, the rate of decrease 'decreases' with age.

$$H_1 : \beta_1 = 0 \ \& \ \beta_2 = 0,$$

that is, age cannot be used to predict externalizing problems,

$$H_2 : \beta_1 < 0 \ \& \ \beta_2 = 0,$$

that is, there is only a linear effect of age, and,

$$H_a : \text{ no restrictions on the parameters}$$

Updating hypotheses & Evidence synthesis: Example

Step 3 - using GORICA

1. For each of H_1, H_2, H_3, and H_{unc}, the GORICA weights are computed; denoted w_m for H_m.

Updating hypotheses & Evidence synthesis: Example

Steps 3 and 4 - using GORICA

Using the second 50% of the data of each of the three cohorts (the confirmation set), the following GORICA weights were obtained:

Cohort	w_1	w_2	w_3	w_{unc}
Gen-R	.82	.04	.10	.05
NTR	.00	.97	.02	.01
TRAILS	.00	.88	.09	.03
All	.00	.99	.01	.00

Updating hypotheses & Evidence synthesis: Example

Steps 3 and 4 - using GORICA

Cohort	w_1	w_2	w_3	w_{unc}
Gen-R	.82	.04	.10	.05
NTR	.00	.97	.02	.01
TRAILS	.00	.88	.09	.03
All	.00	**.99**	.01	.00

Conclusion: Based on the combined evidence in the three cohorts, there is overwhelmingly support for $H_2 : \beta_1 < 0 \;\&\; \beta_2 = 0$. That is, there is

only a linear effect of age of the mother on externalizing problem behavior of children around the age of 11.

Table of Contents

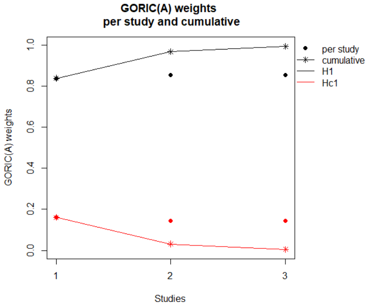
Two approaches: Added- vs Equal-evidence approach

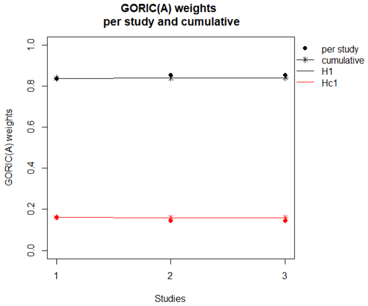
Situation A: Evidence from 5 studies with n = 100.
Situation B: Evidence from 1 study with n = 500.

Approach 1: Situation A is stronger than Situation B
Conclusion: Evidence theory true in all studies.
Then, as we did before: Added-evidence approach.

Approach 2: Situation A is equally strong as Situation B (cf. meta-analysis)
Conclusion: Evidence theory true on average.
Then, alternative method needed: Equal-evidence approach.

Added- vs Equal-evidence approach





Magnitude-hypotheses

Set of central theories regards height of effect size.
E.g., Cohen's d measured in some studies, one could evaluate in those:

$$\begin{aligned} H_1 &: d < 0, \\ H_2 &: d > 0, \\ H_3 &: d > 0.2, \\ H_4 &: d > 0.5, \\ H_5 &: d > 0.8. \end{aligned}$$

Now, overlapping hypotheses.

$$\begin{aligned} H_1 &: d < 0, \\ H_2 &: 0 < d < 0.2, \\ H_3 &: 0.2 < d < 0.5, \\ H_4 &: 0.5 < d < 0.8, \\ H_5 &: d > 0.8. \end{aligned}$$

Now, range restrictions
(complexity as if equalities).

Future research: Variation in overall evidence

1) Should look at variation measures!

2) Look at outlier studies (not to make results better):
Do evidence synthesis for all but one study.
Leave every time one out.

Software

Currently, beta versions of software:

- R function *evSyn* in R package *restriktor*
- Interactive web application (Shiny app) of *GoricEvSyn*