

gesis

Leibniz Institute
for the Social Sciences



Ex-post Harmonisierung von Fragen zum gleichen Konstrukt mit Equating und Linear Stretching Ist fehlende Populationsinvarianz ein Problem?

Matthias Roth, Ranjit K. Singh

Gemeinsame Tagung der Arbeitsgemeinschaft sozialwissenschaftlicher Institute (ASI) und der

DGS-Sektion „Methoden der empirischen Sozialforschung“

Leibniz
Association

KonsortSWD

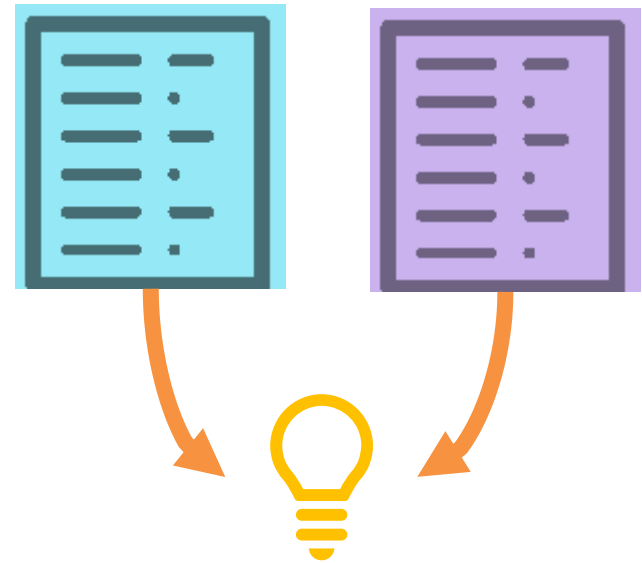


DFG

Deutsche
Forschungsgemeinschaft

Der Kontext: Harmonisierung von Umfragedaten

- Umfragen messen häufig ähnliche Konstrukte (politisches Interesse, allgemeine Lebenszufriedenheit, allgemeiner Gesundheitszustand,...)
- Die Zusammenführung von Umfragedaten unterschiedlicher Herkunft kann neue Forschung ermöglichen
- *Problem:* Messinstrumente sind oftmals nicht vergleichbar



Inhalt der Präsentation

- 1. Teil: OSE-RG & Linear Stretching
- 2. Teil: Empirische Untersuchung der Populations(in)varianz von OSE-RG und Linear Stretching

Ex-post Harmonisierung von Umfragedaten

A



Wie stark interessieren Sie sich für Politik?

- Sehr stark
- Stark
- Mittel
- Wenig
- Überhaupt nicht

B

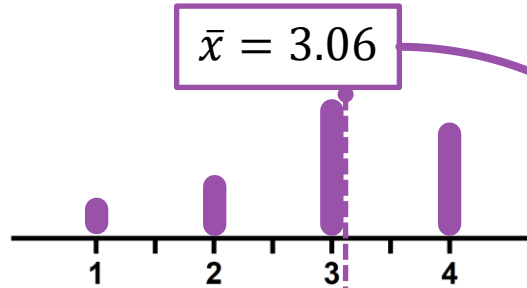
European *Values* Study 

Wie stark interessieren Sie sich für Politik? Sind Sie...

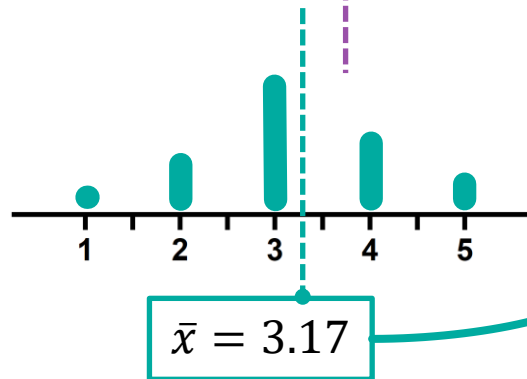
- Sehr interessiert
- Etwas interessiert
- Kaum interessiert
- Überhaupt nicht interessiert

Ex-post Harmonisierung mit OSE-RG

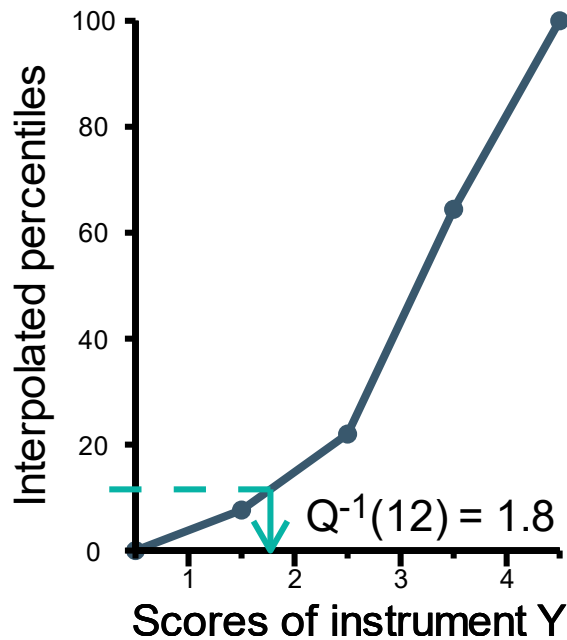
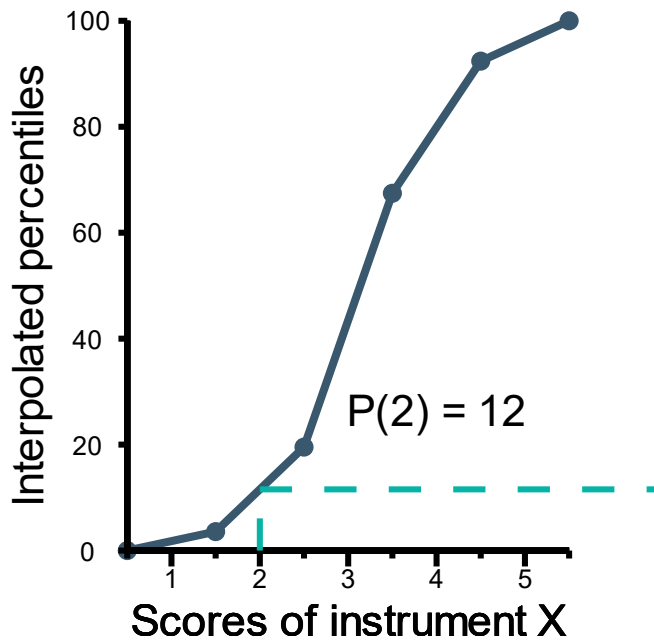
European Values Study  (B)



 Panel (A)



Ex-post Harmonisierung mit OSE-RG

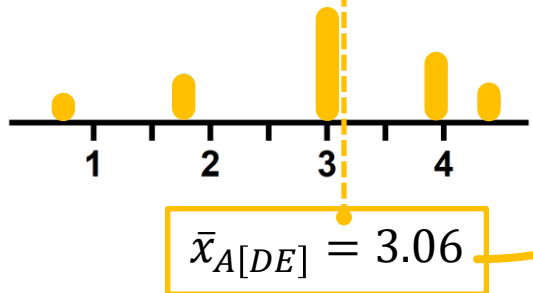
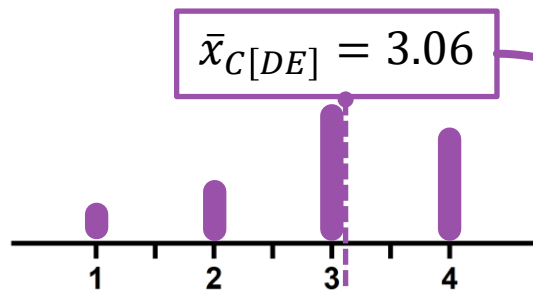


Ausgangsfrage	Zielfrage
1	0.7
2	1.8
3	3.0
4	3.9
5	4.4

Ex-post Harmonisierung mit OSE-RG

European *Values* Study  **B**

 Panel **A**

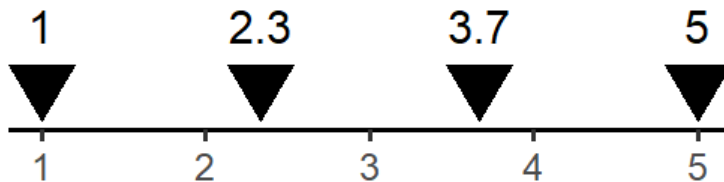


Ex-post Harmonisierung mit Linear Stretching

4 Skalenpunkte

→

5 Skalenpunkte



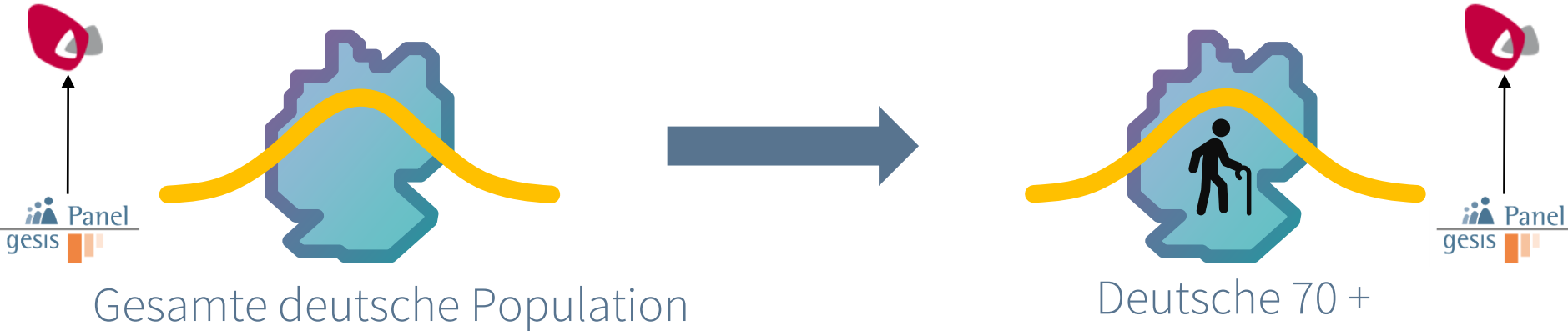
Populationsinvarianz bei **Observed Score Equating**

- Wenn OSE-RG populationsinvariant ist, sind Harmonisierungsverbindungen identisch über verschiedene Populationen.
- **Linking population** (Rekodierungstabelle wird erstellt)
- **Research population** (Rekodierungstabelle wird angewendet)

Populationsinvarianz bei **Observed Score Equating**

Linking population

Research population

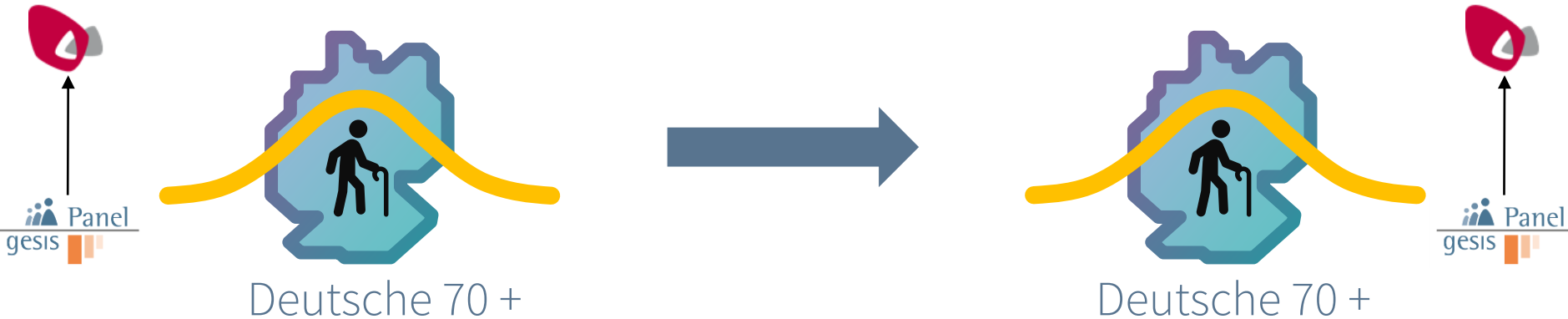


Mittelwertsdifferenz harmonisierte Daten zu Zieldaten
Cohen's $d = 0.25$

Populationsinvarianz bei **Observed Score Equating**

Linking population

Research population




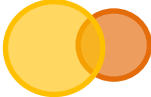

Mittelwertsdifferenz harmonisierte Daten zu Zieldaten
Cohen's $d = 0.001$

Empirische Untersuchung der Populationsinvarianz von OSE-RG & LS

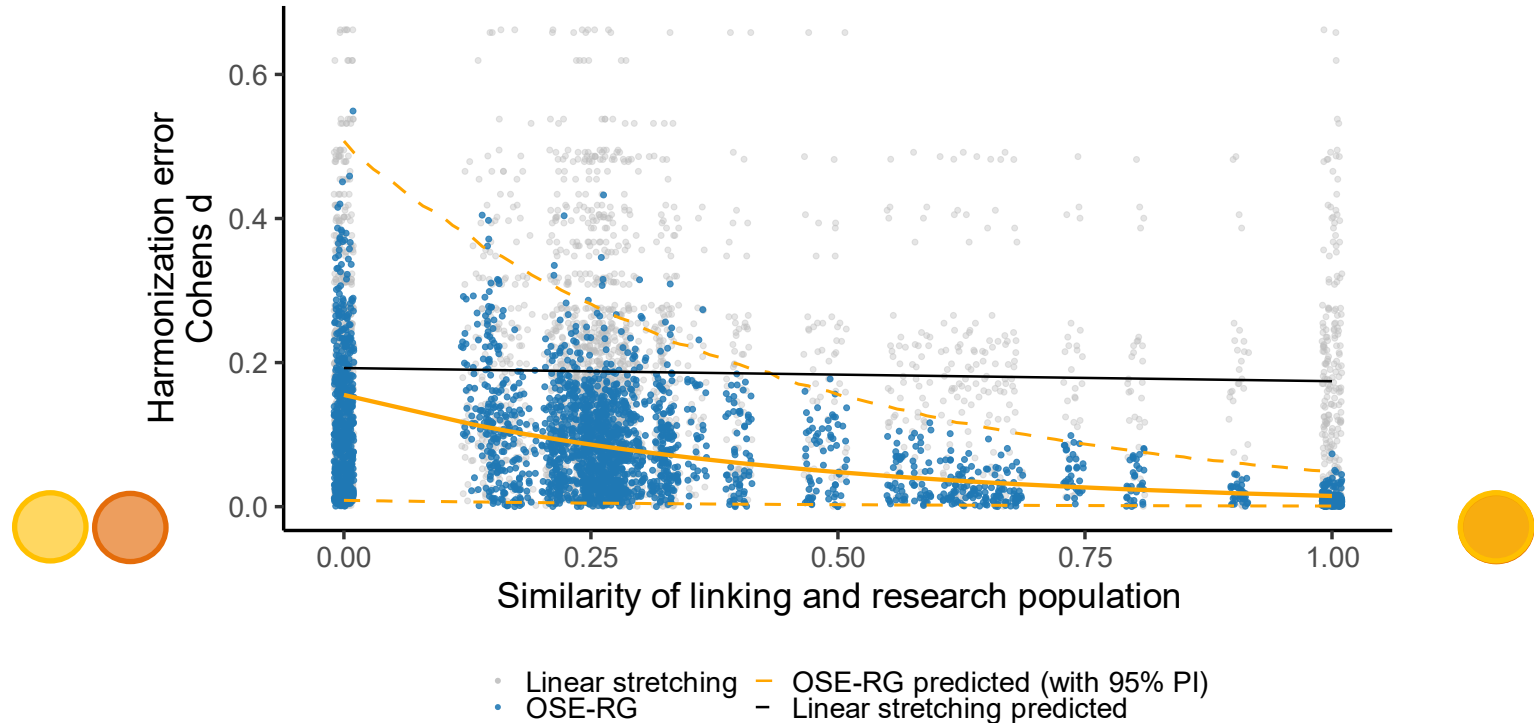
- 1 Zeigt OSE-RG einen geringeren Harmonisierungsfehler als LS?
- 2 Sind Harmonisierungen von OSE-RG populationsinvariant?
- 3 Sind die Harmonisierungsfehler von OSE-RG und LS abhängig von den harmonisierten Fragen?



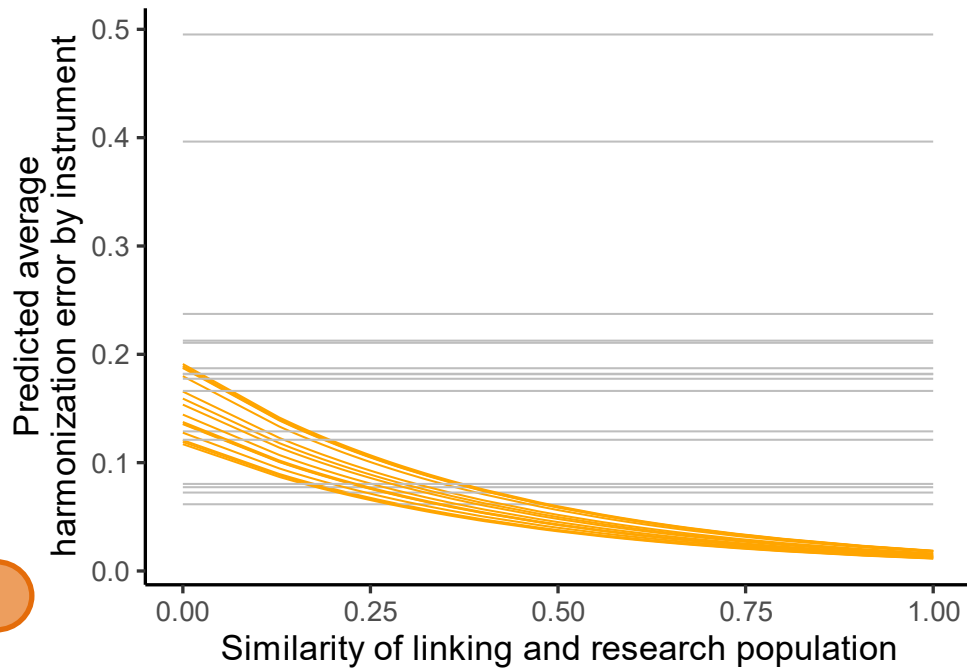
Research design

Fragen Paar	Linking population	Research population	Überschneidung (Sørensen–Dice Koeffizient)	Cohen`s d
Politisches Interesse	Ganze Deu. Population	Ganze Deu. Population	1 	0.001
Politisches Interesse	Ganze Deu. Population	70+ German Population	0.32 	0.25
Links-Rechts Orientierung	20-29 Deu. Population	Männ. Deu. Population	0.21 	0.04
...
16 Fragenpaare	13 Sozio-demo. Gruppen	13 Sozio-demo. Gruppen	=> 2074 Harmonisierungen	

Ergebnisse: RQ1 & RQ2



Ergebnisse: RQ 3



Model	R2 marginal	R2 conditional
OSE-RG	0.44	0.46
Linear stretching	0.00	0.66

Harmonization method — Linear stretching — OSE-RG

Zusammenfassung

- ① OSE-RG harmonisiert besser als LS, solange linking und research population gleich sind.
- ② Der erwartete Fehler von OSE-RG steigt je ungleicher linking und research population werden.
- ③ OSE-RG Fehler sind ähnlich über verschiedene Fragenpaare. LS Fehler sind stark abhängig von dem jeweiligen Fragenpaar.

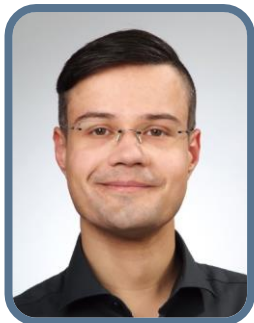
Contact



Matthias Roth (M.Sc.)



matthias.roth@gesis.org



Dr. Ranjit K. Singh



ranjit.singh@gesis.org



@_R_K_Singh

Vielen Dank für Ihre Aufmerksamkeit!

gesis

Leibniz-Institut
für Sozialwissenschaften

Mitglied der
Leibniz
Leibniz-Gemeinschaft

KonsortSWD



DFG

Deutsche
Forschungsgemeinschaft

Literature

- **OSE-RG:** Kolen, Michael J., und Robert L. Brennan. 2014. Test Equating, Scaling, and Linking. New York, NY: Springer New York.
- **Linear Stretching:** Jonge, Tineke de, Ruut Veenhoven, and Lidia Arends. 2014. “Homogenizing Responses to Different Survey Questions on the Same Topic: Proposal of a Scale Homogenization Method Using a Reference Distribution.” *Social Indicators Research* 117:275–300. doi:10.1007/s11205-013-0335-6.
- **Overlap:** Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskab*, 5(4), 1–34.
- **Gamma-GLM:** Dunn, P. K., & Smyth, G. K. (2018). Chapter 11: Positive Continuous Data: Gamma and Inverse Gaussian GLMs. In P. K. Dunn & G. K. Smyth (Eds.), *Springer Texts in Statistics. Generalized linear models with examples in R* (pp. 425–456). Springer. https://doi.org/10.1007/978-1-4419-0118-7_11
- **Prediction interval for Gamma-GLM:** Gelman, A., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. *Analytical methods for social research*. Cambridge University Press.

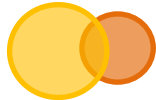
Kombination aus Linking und Research Population



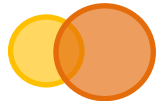
Linking und Research Population sind getrennt



Linking und Research Population sind Subpopulationen



Linking Population ist volle Population



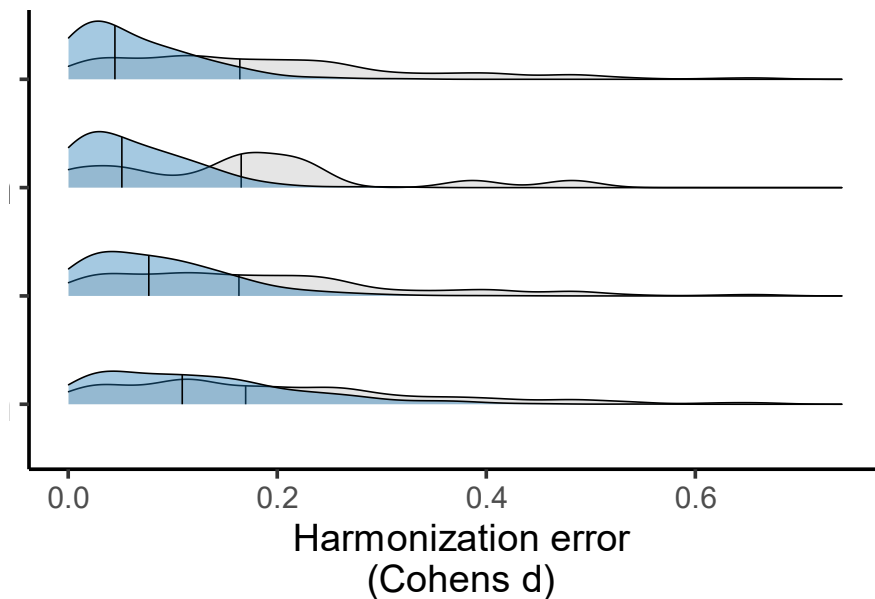
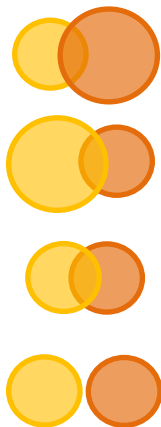
Linking und Research Population sind getrennt



Linking und Research Population sind getrennt

Linking and research Population sind nicht gleich

Linking Research



**OSE-RG Fehler
kleiner als LS Fehler**

85%

75%

74%

65%

Harmonization method Linear stretching OSE-RG