



Der „Zeitunterschreiter“ als Problem der Online-Befragung

Mannheim, 27. Mai 2011

Marco Meyer, Robert Greszki, Harald Schoen

Tagung der Methodensektion in der Deutschen Gesellschaft für Soziologie und
des Arbeitskreises „Methoden“ der Deutschen Vereinigung für Politische Wissenschaft



- Interviewverhalten in Online-Befragungen nicht direkt zu kontrollieren
 - unterschiedliche Datenqualität bei den Befragten möglich
 - ein Indikator für Datenqualität ist Zeit/Dauer der Befragung
 - große Unterschiede bei der Interviewdauer deuten auf Qualitätsmängel hin
- „Zeitunterschreiter“ als Problem der Online-Befragung

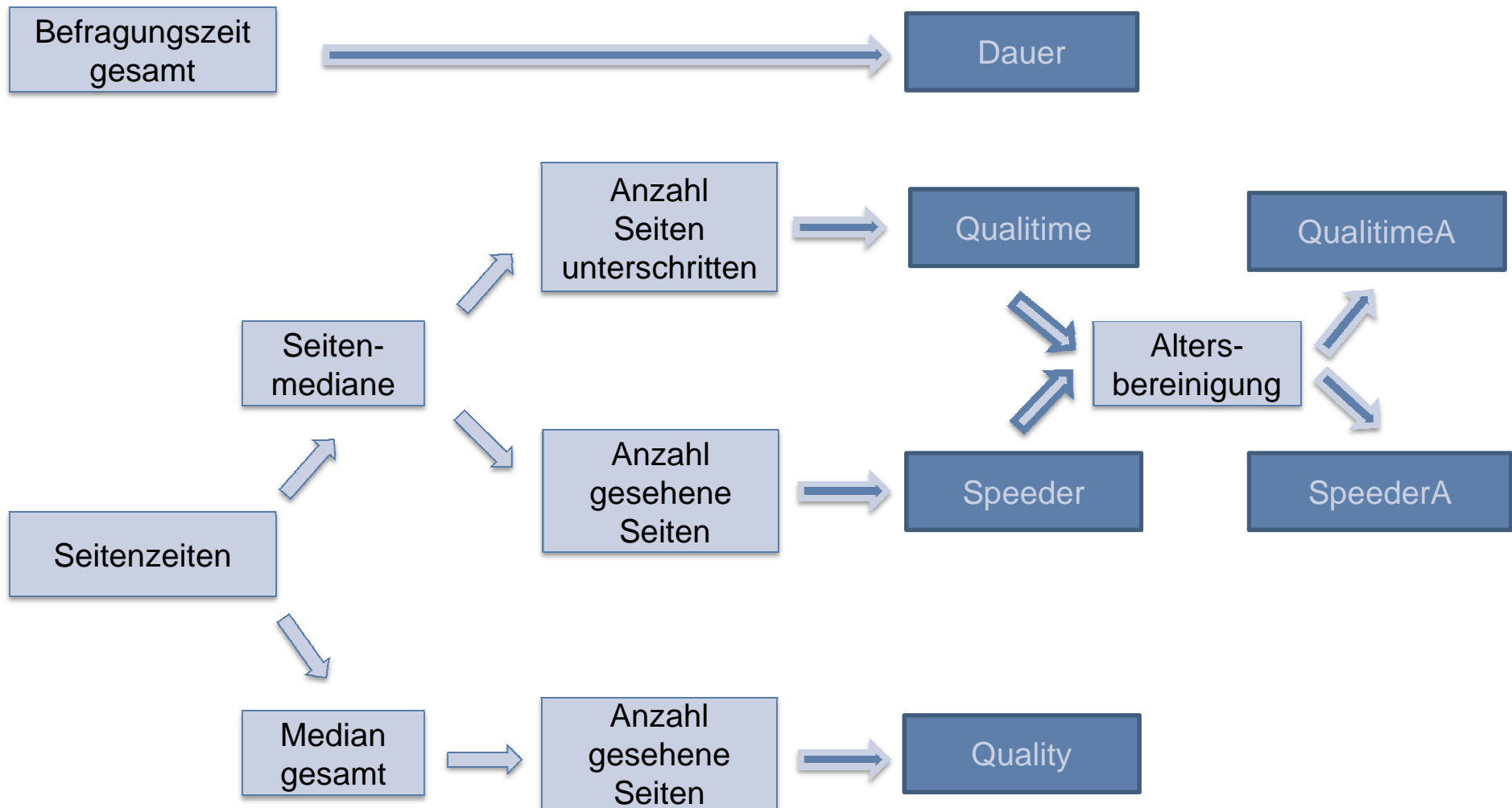


1. Was ist die beste Methode zur Abbildung der Datenqualität auf der Basis der Befragungszeit?
 - Vergleich der Güte verschiedener Indizes zur Bestimmung der Datenqualität anhand von objektiven Qualitäts-Indikatoren
2. Welcher Schwellenwert zur Identifikation von „Zeitunterschreitern“ sollte herangezogen werden?
 - Ermittlung von Schwellenwerten zur Abgrenzung von „Zeitunterschreitern“ anhand der Qualitäts-Indikatoren
3. Wie lassen sich „Zeitunterschreiter“ charakterisieren?
 - Soziodemographische Merkmale, Befragungsprozessdaten



- German Longitudinal Election Study (GLES)
- Komponente X/8: Langfrist-Online-Tracking, Umfrage 6 (ZA5339)
- Erhebungszeitraum: 18. September bis 27. September 2009
- N = 1.153
- Bearbeitungsdauer
 - Minimum: 00:02:23
 - Maximum: 02:21:33
 - Median: 00:33:01
 - Arithmetisches Mittel: 00:36:42

Indizes zur Bestimmung der Datenqualität



Korrelationen der Indizes



	Dauer	Quality	Speeder	SpeederA	Qualitime	QualitimeA
Dauer	1,000 ^c	,479 ^c	,425 ^c	,388 ^c	,349 ^c	,316 ^c
Quality		1,000 ^c	,919 ^c	,839 ^c	,780 ^c	,726 ^c
Speeder			1,000 ^c	,921 ^c	,890 ^c	,840 ^c
SpeederA				1,000 ^c	,821 ^c	,877 ^c
Qualitime					1,000 ^c	,910 ^c
QualitimeA						1,000 ^c

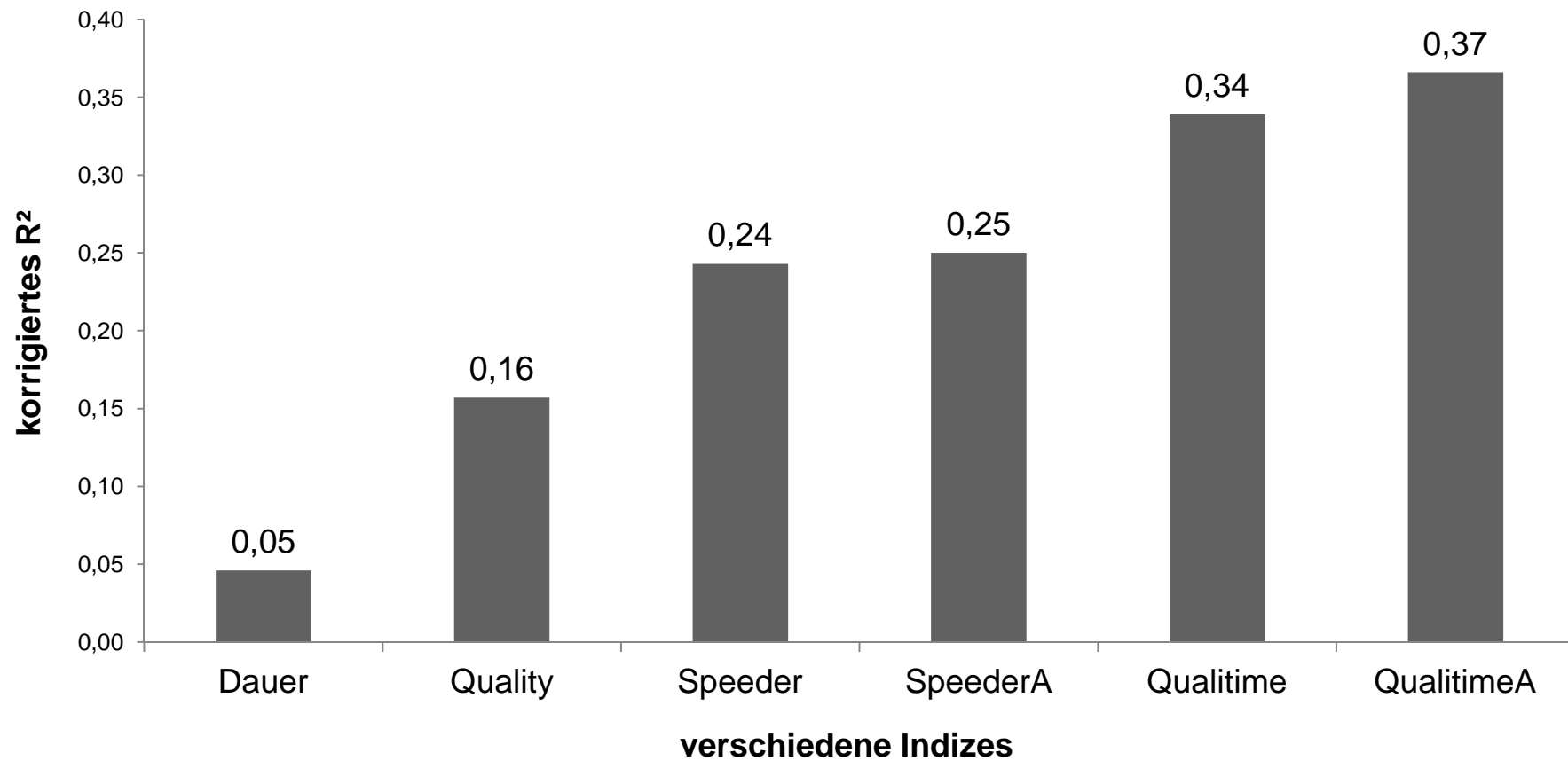
Signifikanz-Niveau: $c = p < 0,001$



- „keine Angabe“-Index
 - ohne Mehrfachnennungen
- „weiß nicht“-Index
- Mittelkategorie-Index
 - vertikale und horizontale, ordinale Skalen mit mind. fünf Ausprägungen
 - „weiß nicht“-Kategorien nicht berücksichtigt
- „Straightlining“-Index
 - nur Itembatterien mit mind. fünf Ausprägungen und vier Items
 - inhaltlich logisches „Straightlining“ wurde berücksichtigt
 - „weiß nicht“-Kategorien nicht berücksichtigt
- Wertebereiche aller Indizes: 0 bis 1

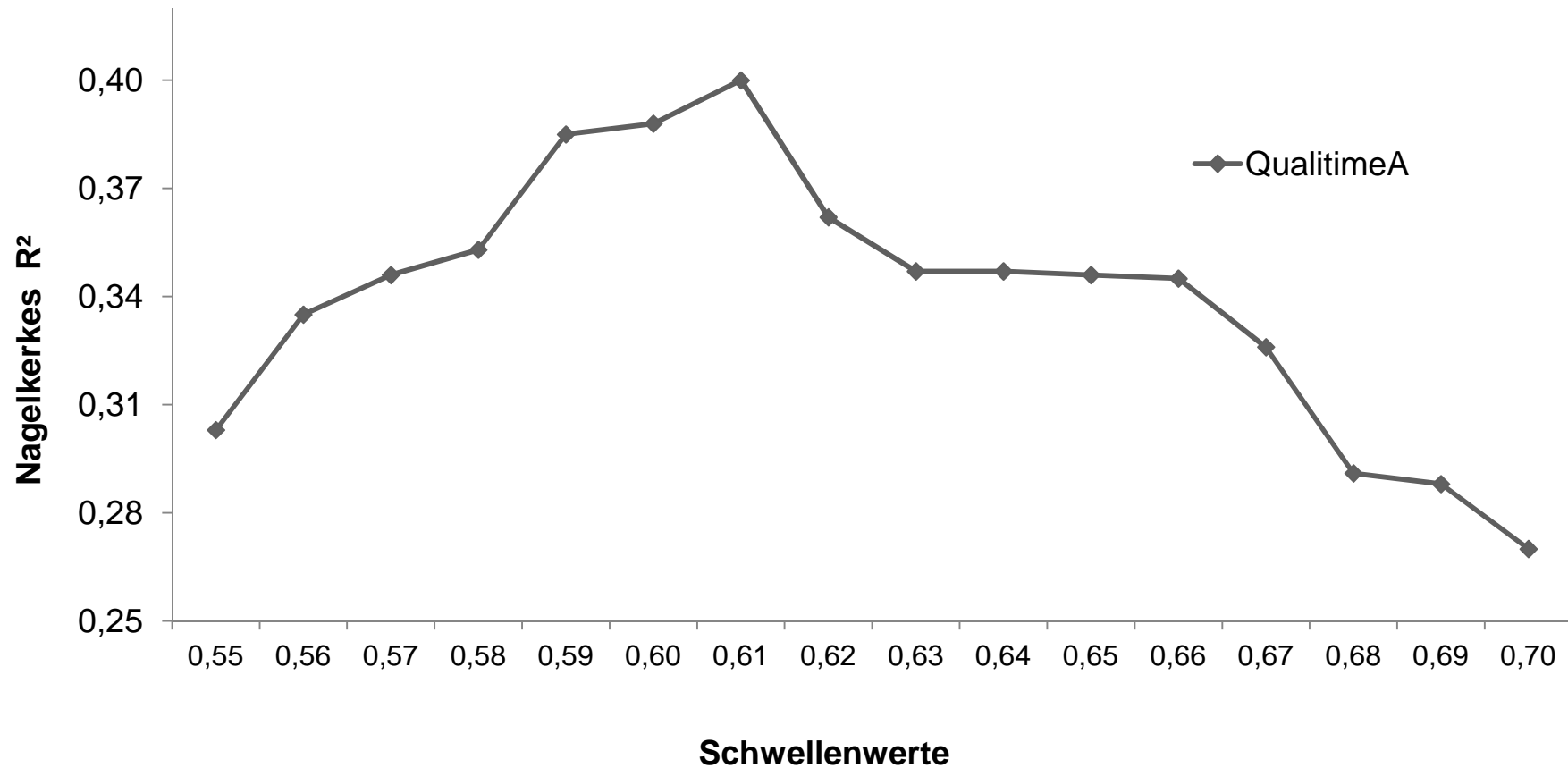


Varianzaufklärung durch Qualitäts-Indikatoren (OLS)

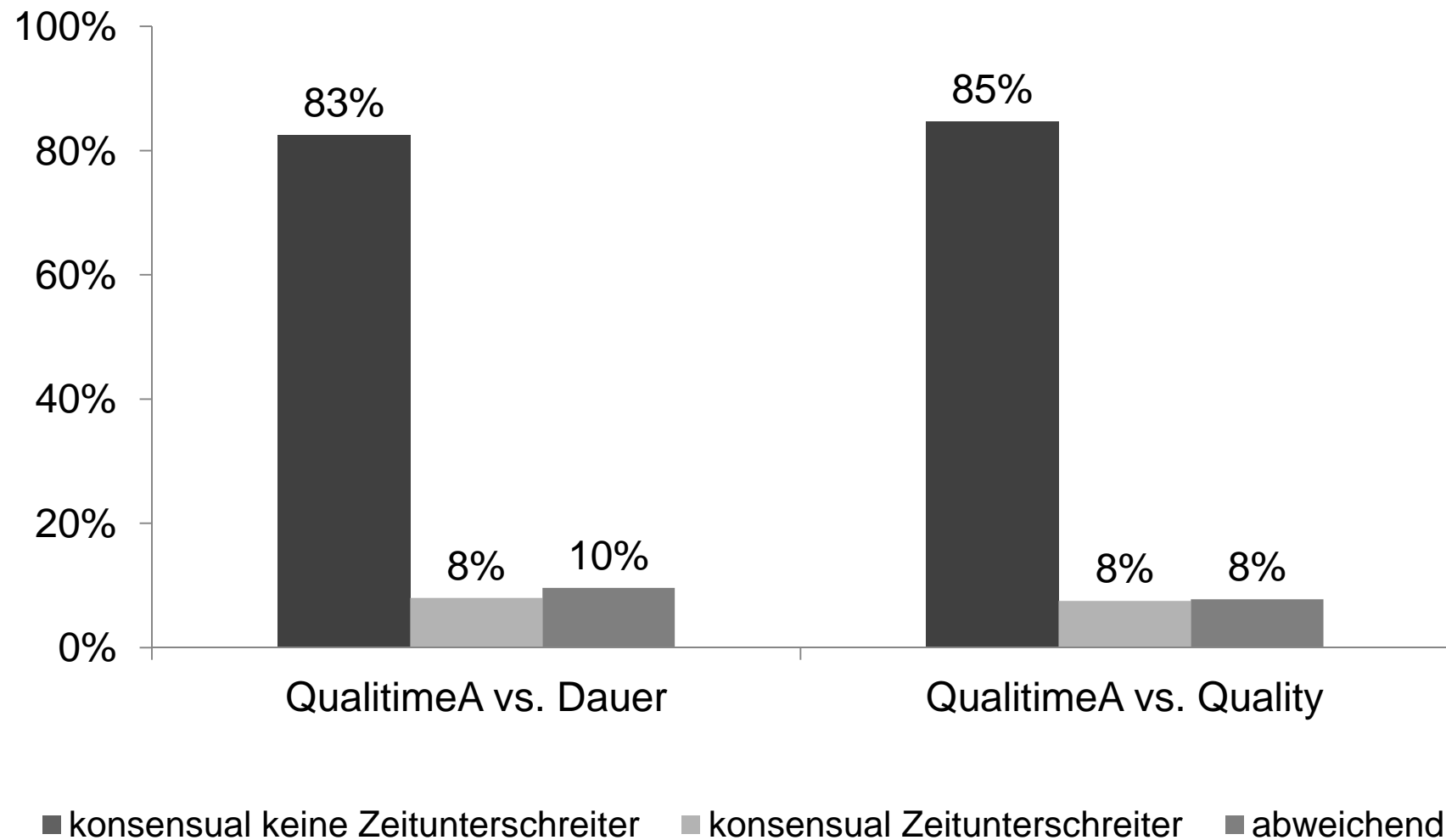




Modellgüte verschiedener Schwellenwertdefinitionen (Logit)



Vergleich der dichotomisierten Indizes





- Soziodemographie (Geschlecht, Alter, Bildung)
- Befragungsprozessdaten
 - Intern
 - Tageszeit (8-11h, 12-14h, 15-17h, 18-21h, 22-7h)
 - Reaktionszeit (ohne Einladung, 1-120m, 121-360m, 361-1200m, 1201+m)
 - Extern
 - Dauer der Panelmitgliedschaft (in Tagen)
 - Weg des Paneleintritts (Selbstrekrutierung, Sozioland, Freundschaftswerbung, extern verlinkte Seiten)
 - Anzahl beendeter Umfragen (letzte(r) Monat, 2-3 Monate, 4-12 Monate)
 - Anteil abgebrochener Umfragen (letzte(r) Monat, 2-3 Monate, 4-12 Monate)

Effekte Charakterisierung: Logit-Modell



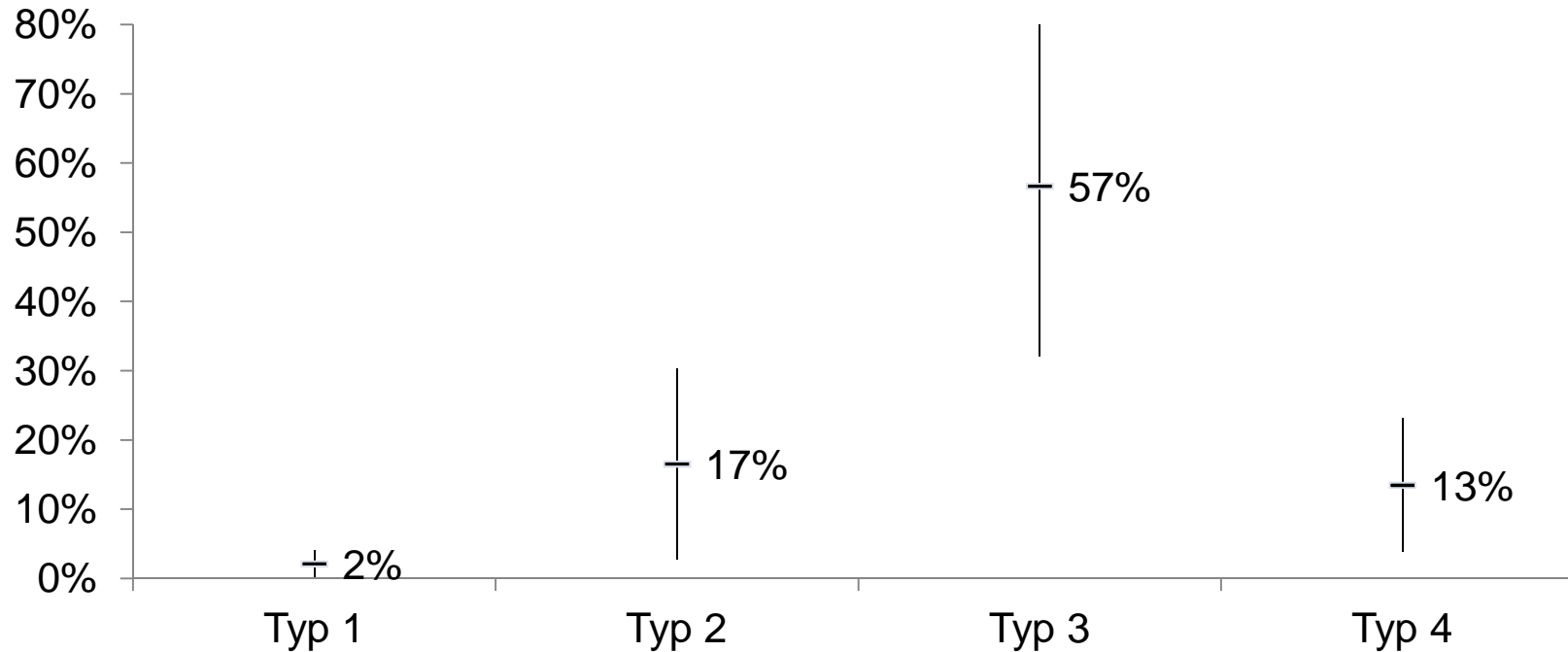
Konstrukt	Referenz	-	0	+
Alter	18-29	40-49 ^b , 50-59 ^a	30-39, 60+	
Tageszeit	22-7h	18-21h ^a	8-11h, 12-14h, 15-17h	
Reaktionszeit	1-120m		361-1200m	Ohne Einladung ^a , 121-360m ^a , 1201m+ ^a
Weg Paneleintritt	extern verlinkte Seiten		Sozioland	Selbstrekrutierung ^c Freundschaftswerbung ^c
N beendete Umfragen	-		letzter Monat, letzte 2-3 Monate	letzte 4-12 Monate ^a
% abgebrochene Umfragen	-		letzter Monat, letzte 2-3 Monate	letzte 4-12 Monate ^a

Signifikanz-Niveau: a = $p < 0,05$; b = $p < 0,01$; c = $p < 0,001$; Nagelkerkes R^2 : 0,125.

Anmerkung: Geschlecht, Bildung und Dauer der Panelmitgliedschaft sind Bestandteile des Modells, jedoch substantiell ohne Bedeutung.



Erwartete Wahrscheinlichkeiten mit 95%-Konfidenzintervall



- Typ 1: 40-49 Jahre, 18-21h, 1-120m, extern verlinkte Seiten
- Typ 2: 30-39 Jahre, 15-17h, 121-360m, Sozioland
- Typ 3: 30-39 Jahre, 22-7h, keine Einladungsmail, Freundschaftswerbung
- Typ 4: 40-49 Jahre, 8-11h, keine Einladungsmail, Selbstrekrutierung



- Präzisere Abbildung der Datenqualität durch Berücksichtigung individueller Seitenzeiten sowie der Relation zu der Anzahl unterschrittener Seiten
- Zusätzliche Erhöhung der Validität des Index durch Einbeziehung des Alters als „Proxy“-Variable für die Basisgeschwindigkeit
- Objektive Festsetzung des Schwellenwertes anhand von Qualitätsindikatoren
- Anknüpfungspunkte für weitere Forschung:
 - Prüfung der Robustheit des Ansatzes durch Übertragung auf weitere online erhobene Umfragedaten
 - Auswirkung unterschiedlicher Indizes auf inhaltliche Befunde



Vielen Dank für Ihre Aufmerksamkeit!

Kontakt:

Robert Greszki – robert.greszki@baces.uni-bamberg.de – BACES/Universität Bamberg

Marco Meyer – marco.meyer@uni-bamberg.de – Universität Bamberg

Harald Schoen – harald.schoen@uni-bamberg.de – Universität Bamberg