Det systematiske review

Erik Gahner Larsen

Kausalanalyse i offentlig politik

Dagsorden

- Litteraturgennemgang
 - ▶ Mellem narrative og systematiske reviews
 - Søgning efter litteratur
 - Behandling af studier
 - Konklusioner
- Læsning af studier
 - Hvad skal man være opmærksom på?

Seminaropgave

- Deadlines, det korte perspektiv
 - ▶ 8. februar: $\approx \frac{1}{2}$ -1 side. Brainstorm.
 - ▶ 29. februar: \approx 5 sider. Idé og fokus.
- ► Alle, der har sendt mig noget, har fået skriftlig feedback
- Hvis du/l ikke allerede har sendt mig noget: gør det så hurtigt som muligt
- Spørgsmål? Spørg eller send en mail

Hvad lærte vi sidste uge?

- Offentlig politik er fyldt med kausale spørgsmål
- Verden er kompleks
 - Vores opgave er at forsimple den
 - Hvordan?
 - ▶ Studere kausale relationer mellem klart definerede variable
- I dag: Litteraturreview

Fordele ved at lave et litteraturreview (cf. Knopf 2006)

- ► Fem (relaterede) fordele
 - 1. Giver et bredt kendskab til en litteratur
 - 2. Giver en forståelse af, hvad vi allerede ved
 - 3. Giver ideer til, hvad man kan og bør gøre
 - 4. Giver en forståelse af, hvilke problemer og udfordringer der er
 - 5. Giver en kontekst hvori man kan placere sit eget spørgsmål

Hvorfor flere studier?

- ► Et studie siger ikke ret meget
- ► Enkelte studier har ofte ikke meget diversitet (land, mål osv.)
- Enkelte studier kan være ekstreme (outliers)
 - ▶ Pas på ekstreme studier (også selvom journalister elsker dem)
- Vi vil gerne generalisere på baggrund af flere studier

Afgrænsning af litteratur

- Litteratur dumper ikke ned fra himlen (eller "læseplanen")
- Forskningsspørgsmål skal passe ind i en litteratur
 - ▶ Hvilken litteratur er vi interesseret i?
- ► I meget statskundskabslitteratur er litteraturgennemgangen arbitrær og ustruktureret
 - Narrativ gennemgang af vigtige/interessante studier
 - Vi skal efterstræbe mere systematiske reviews
- Narrativ og systematisk litteraturgennemgang
 - Tænk på et kontinium

Narrative reviews

- Introduktion af overordnet problemstilling
- Narrativ gennemgang af udvalgte studier
 - Hovedfund
 - Metodiske tilgange
 - Resultater
 - Illustrative eksempler
 - Begrænsninger
- ▶ Illustrativt eksempel: de fleste videnskabelige artikler

Styrker og svagheder ved narrative reviews

Styrker:

- Grundig beskrivelse af en litteratur
- Formulering af hypoteser
- Teoriudvikling med kvalitative kategorier

Svagheder:

- Ustruktureret
- Arbitrær udvælgelse, målefejl
- Mange forskellige relationer mellem variable
- Dårligt framework til at diskutere åbenlyse udfordringer (publikationsbias osv.)

Systematiske reviews

- ▶ Jo mindre dit review ligner en slavisk gennemgang af 'the usual suspects', desto bedre
- Overvej: Hvis der skal være en tabel eller figur i din litteraturgennemgang, hvordan skulle den så se ud?
- Vi vil især fokusere på metaanalyser
 - ► Hav altid metaanalyser i baghovedet
 - ► Tommelfingerregel: Jo tættere dit review ligger op ad en metaanalyse, desto bedre

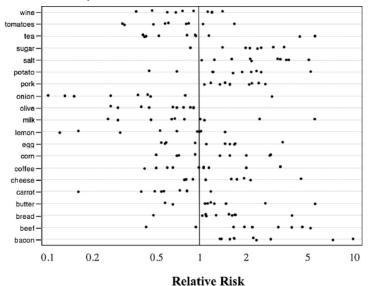
Metaanalyser

- Metaanalyse: Analyse af analyser
- Guldstandarden for litteraturreviews
- Udfordringer for metaanalyser (og andre litteraturreviews):
 - Manglende reliabitilitet i de inkluderede studier
 - Publikationsbias (og andre biases)
 - Manglende information (effektstørrelser, stikprøve etc.)
 - Bias i kodning, bias i udregning af effektstørrelser
 - Statistisk power

Hvorfor metaanalyser? Kogebogseksempel

- ► Eksempel: Forårsager bestemte typer af mad og drikke kræft?
- Studie: Is everything we eat associated with cancer? A systematic cookbook review (Schoenfeld og Ioannidis 2012)
- Metode: "We selected 50 common ingredients from random recipes in a cookbook. PubMed queries identified recent studies that evaluated the relation of each ingredient to cancer risk. Information regarding author conclusions and relevant effect estimates were extracted."

Hvorfor metaanalyser? Kogebogseksempel (Schoenfeld og loannidis 2012)



Metaanalyser: Procedure

- 1. Problemformulering
- 2. Indsamling af studier
 - ► Publicerede *og* ikke-publicerede (hvorfor?)
- 3. Vurder de enkelte studier
 - ▶ Passer det ind i rammen? Er der nok information?
- 4. Vælg et passende effektmål
 - Udregn effekterne for alle studierne
- 5. Lav et datasæt med så mange informationer som muligt
 - ► Effekter, stikprøvestørrelse, design, moderatorer, land osv.
- 6. (Meta)analyser litteraturen
 - ► Er der konsistente effekter i litteraturen? Hvad forklarer dem?
- 7. Lav robusthedstests
 - Sandsynlighed for Type I og Type II fejl



Fem principper for metaanalyser og kausal inferens

- 1. Sammenlignelighed
 - Er det fair at sammenligne de respektive studier?
- 2. Udelukkelse af irrelevante faktorer
 - Kan uobserverede forhold forklare forskelle?
- Diskrimination
 - Hvordan skal studierne vægtes? Alle studier er lige, men...
- 4. Interpolation og extrapolation
 - Hvor ville andre effekter være givet de data vi har?
- 5. Kausal forklaring
 - ▶ Hvilke *kausale* effekter er der?

Kodning af studier

- ▶ Informationer, der kan kodes:
 - ► Studie: publiceringsår, citationer
 - ► Metode: design, uafhængig og afhængig variabel, moderatorer
 - ▶ Resultater: effekter, standardfejl, observationer
 - Kontekst: land, tid
- Stort problem i mange studier: Manglende information!

Effektstørrelser

- Det vi gerne vil sammenligne er effektstørrelser
- Hvad er en effektstørrelse?
 - ► Et standardiseret mål
- Kvantificerer forskellen mellem to grupper
- Gør det muligt at sammenligne forskellige studier

Effektstørrelser

- ► Forskellige mål (Cohens d, Hedges g, odds ratio etc.)
- ► Standardiseseret gennemsnitsforskel (Cohen's d, d-index, d):

$$d_i = \frac{\overline{X_i^t} - \overline{X_i^c}}{s_i}$$

- $ightharpoonup \overline{X_i^t}$: gennemsnit for stimuligruppen i studie i
- ▶ $\overline{X_i^c}$: gennemsnit for kontrolgruppen i studie i
- ▶ s_i: standardafvigelsen for de to grupper
- For studier med flere grupper er udfordringen større
 - Standardisering af koefficienter m.v.

Analyse og præsentation af metaanalysedata

- Kan analyseres som andre typer af data
- Deskriptiv statistik, OLS, hierarkiske modeller (effekter indlejret i studier)
- Heterogenitet mellem studier
 - Kan studeres systematisk
 - ► Fixed og random effects
- Grafisk præsentation
 - Histogram med effekter
 - ► Funnel plot

Opbygning af metaanalyse (Moser og Schmidt 2014)

- 1. Baggrundsinformation
- 2. Hypoteser/forskningsspørgsmål
- 3. Metodesektion/fremgangsmåde
- 4. Detaljer om inkluderede studier
- 5. Detaljer om ekskluderede studier
- 6. Resultater fra metaanalyse
- 7. Resultater fra robusthedsanalyser
- 8. Diskussion, implikationer

Hvor finder vi studier?

- Søgning efter litteratur
 - Brug Google Scholar
- Søg efter:
 - Emneord
 - Bestemt artikel
- ▶ Se hvilke artikler, der citerer de pågældende artikler
- Angiv evt. periode (find de nyeste artikler)

Hvornår siger vi, at der er evidens for noget?

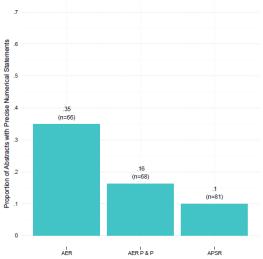
- Det er svært at konkludere, at der er systematisk evidens for et fænomen
 - 'Absence of evidence is not evidence of absence'
- ► Hvad er forskellen på evidens og ikke-evidens?
 - ► Hvor går grænsen?

Hypotetisk scenarie: Hvornår viser to studier noget forskelligt?

- Et studie viser en effekt
- Et andet studie viser, at der ikke er en effekt
- Viser de to studier noget forskelligt? Ikke nødvendigvis
 - Forskellen mellem signifikant og ikke-signifikant er ikke altid signifikant
- Vi vil gerne have numeriske resultater (og standardfejl/konfidensintervaller)

Hvor mange artikler har præcise numeriske resultater (i deres abstract)? (Sood og Guess 2015)

Figure 1: Proportion of Precise Numerical Statements in Abstracts of Empirical Papers in the APSR, AER, and the AER P & P



Hvorfor er det svært at finde sammenlignelige studier?

- ▶ Der vil *altid* være forskelle
 - ► Kontekst, tid
- Nyhedsskævhed ("novelty bias")
 - Fokus på at lave noget nyt
- ► Forskellige rapporteringsstrategier
 - Svært at identificere sammenlignelige mål

Hvad skal man være opmærksom på?

- ► Når man finder et studie, skal man ikke gå direkte til resultaterne
- Ikke alle studier er lige gode
 - Alle studier skal læses med kritiske øjne
- Ingen udtømmende liste, men...

- 1. Der er mange kilder til variation
 - Opgaven er for os at isolere fænomener
 - Udelukke andre kilder til variation
- 2. Intet mål er eksakt
 - Vi skal forholde os til målefejl
 - Hvilke potentielle målefejl er der ved vores variable?
- 3. Der er systematiske skævheder
 - ▶ Mange forhold kan føre til bias i vores resultater
 - ▶ lkke-tilfældige målefejl

- 4. Jo flere observationer, desto bedre
 - More is more
 - Mange studier har for lav statistisk styrke til at finde den effekt, de ønsker at finde
- 5. Korrelation er ikke lig kausalitet
 - ► En pointe der skal nævnes igen og igen
- 6. Regression mod gennemsnittet
 - ► Jo længere væk en observation ligger fra gennemsnittet ved første observation, desto større er sandsynligheden for at den vil falde tættere på gennemsnittet ved næste observation
 - ▶ Tendenser kan være en naturlig udvikling mod gennemsnittet

- 7. Pas på med at ekstrapolere
 - ▶ Vi har observationer der falder i et bestemt rum
 - Kræver stærk teori at sige noget om ikke-observationer
- 8. Husk sandsynligheder og stikprøvestørrelse ("base rate fallacy")
 - Ikke alle tests er perfekte (Type I og Type II-fejl)
- 9. Kontrolgrupper er vigtige
 - ▶ Vi skal have observationer, der ikke påvirkes af vores stimuli

10. Randomisering fjerner skævheder

- ▶ Uden randomisering ved vi ikke, om vores stimuli har en effekt
- Problem med effekten af offentlige politikker: Ingen randomisering

11. Søg replikationer

- Uden replikationer giver det ingen mening at tale om videnskab
- Jo flere selvstændige studier der tester et givent fænomen, desto bedre

12. Forskere er også mennesker(!)

▶ Forskere har egne interesser; er ikke altid rationelle

- 13. Signifikant er signifikant
 - ► (Se bort fra dette)
 - Diskuter altid statistisk signifikans i forhold til substantiel signifikans
- 14. Insignifikant betyder ikke, at der ikke er en effekt
- 15. Effektstørrelsen er afgørende
- 16. Der kan være et trade-off mellem relevans og generaliserbarhed

- 17. Vi er dårlige til at forholde os til risici/sandsynligheder
- 18. Når begivenheder er relateret, kan det påvirke sandsynlighederne drastisk for, at de finder sted
- 19. Data kan manipuleres
 - Selektiv præsentation af resultater
- 20. Ekstreme mål kan være misvisende

Hvor mange observationer, skal vi have med?

- Ofte har vi ikke ret mange observationer
- Mange studier er underpowered.
- Hvad er problemet?

Hvor store effekter? (Simmons et al. 2013)

Independent Variable	Dependent Variable	<u>d</u>	Required n
Gender	Height in inches [open-ended]	1.85	6
Age [median split]	How many years from now do you think that you will retire [open-ended]	1.49	9
If you were to vote today, would you plan on voting for Mitt Romney or Barack Obama?	How much do you like Michelle Obama? [1 = do not like at all; 7 = like very much]	1.47	9
Gender	How many pairs of shoes do you own? [open-ended]	1.07	15
Do you like spicy food?	How much do you like Indian food? [1 = strongly dislike; 7 = strongly like]	0.80	26
Which word best describes your political orientation: liberal or conservative?	How important is social equality? [1 = not at all important; 7 = very important]	0.69	34
Gender	Weight in pounds [open-ended]	0.59	46
Do you like to watch soccer? [yes vs. no]	How many languages do you speak? [open-ended]	0.59	46
Do you like eggs? [yes vs. no]	How often do you eat egg salad? [1 = very rarely; 7 = very often]	0.58	48
Are you a smoker? [yes or no]	What is the likelihood of a smoker dying from a smoking- related illness? [open-ended]	0.33	144
Do you prefer science or art?	How many planets can you name correctly? [open-ended]	0.07	3,669
Do you prefer apples or oranges?	How much would you be willing to pay for a pound of oranges? [open-ended, in dollars]	-0.00	∞

"Winner's curse"

- ► Forestil jer:
 - ▶ ingen effekt af stimuli (en effekt på 0)
 - publikationsbias (kun publikation af studier med effekter \neq 0)
- Hvem bliver publiceret? Den 'heldige' forsker(gruppe)
- ► Forudgående forsøg finder vej til arkivet (bliver ikke publiceret)
- ► Kan forklare Proteus-fænomenet: Det første studie vil som regel have større effekter end efterfølgende forsøg

"Winner's curse" (Button 2013)

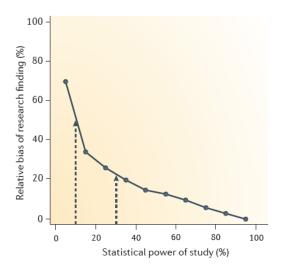


Figure 5 | The winner's curse: effect size inflation as a function of statistical power.

Hvor store korrelationer? (Richard et al. 2003)

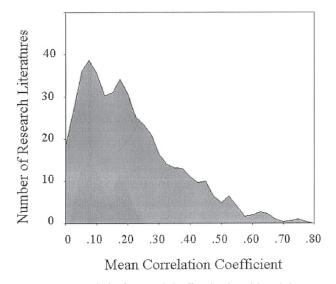


Figure 1. Magnitude of meta-analytic effect sizes in social psychology.

Eksempel: Den offentlige opinion og offentlige politikker (Burstein 2003)

- ► Spørgsmål: Er der signifikante effekter af den offentlige opinion på offentlige politikker?
- ► Metode: Kig på koefficienter i forskellige studier
- Analyse: Antallet af signifikante koefficienter

Eksempel: Den offentlige opinion og offentlige politikker (Burstein 2003)

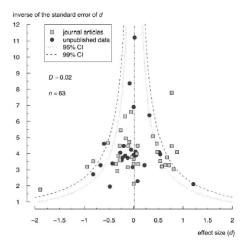
■ TABLE 2
IMPACT OF PUBLIC OPINION ON POLICY

	%
None	25
Ambiguous	2
Statistically significant, policy importance not	
discussed	35
Statistically significant, little policy importance	4
Statistically significant, substantial policy	
importance	35
Total number	52

Note: Total may be greater than 100 percent due to rounding.

Eksempel: Effekten af flere valgmuligheder (Scheibehenne et al. 2010)

FUNNEL PLOT OF ALL DATA



NOTE.—The dotted and dashed lines indicate confidence intervals under the assumption that the data set is homogeneous and normally distributed around the mean effect size.

Eksempel: Politikernes brug af twitter (Jungherr 2016)

Topic	Studies				
Which kind of parties and candidates tend to adopt Twitter?	Chi & Yang, 2011; Evans et al., 2014; Gilmore, 2012; Jackson & Lilleker, 2011; Larsson & Kalsnes, 2014; Lassen & Brown, 2011; Peterson, 2012; Shogan, 2010; Straus et al., 2013; Verqueer & Hermans, 2013; Verqueer & La, 2011, 2013				
How do parties and candidates tend to use Twitter?	Adams & McCorkindale, 2013; Adi et al., 2014; Aharony, 2012; Ahmed & Skoric, 2014; Amirullah et al., 2013; Aragón et al., 2013; Baxter & Marcella, 2012, 2013; Brossma & Grahan 2012; Bruns & Highfield, 2013; Christensen, 2013; Conway et al., 2013; D'heer & Verdegem, 2014; DiGrazia et al., 2013; Enli & Skogerbo, 2013; Evans et al., 2014; Golbeck et al., 2010; Graham et al., 2014; Graham et al., 2013; Graham et al., 2016; Grabam et al., 2014; Jackson & Lilleke 2011; Jaidka & Ahmed, 2015; Klinger, 2013; Kreiss, 2014; Kreiss et al., 2014; Jackson & Lilleke 2011; Jaidka & Ahmed, 2015; Klinger, 2013; Kreiss, 2014; Kreiss et al., 2014; Kruikemeler, 2011; Lalamre & Suzuki-Lambrecht, 2013; Lietz et al., 2014; Uner et al., 2011; Macmamara, 2011; Margaretten & Gaber, 2014; Mirer & Bode, 2013; Nilsson & Carlsson, 2014; Plotkowiak & Stanoevska-Slabeva, 2013; Raynauld & Greenberg, 2014; Small, 2010; Suiter, 2015; Vaccari & Valeriani, 2013				
What are the functions and effects of Twitter use by parties and candidates?	Conway et al., 2015; Hong, 2013; Hong & Nadler, 2011, 2012; Kobayashi & Ichifuji, 2015; Lec 2013; Lee & Jang, 2013; Lee & Oh, 2012, 2013; Lee & Shin, 2014; Parmelee, 2014; Verweij, 2012; Zhang et al., 2013				

Eksempel: Mediernes formidling af meningsmålinger (Larsen og Straubinger 2012)

- ► Spørgsmål: Formidler medierne metodiske informationer fra meningsmålinger?
- ▶ Metode: Indsamling af numeriske resultater fra andre studier
- Analyse: Deskriptiv statistik med vægtede gennemsnit

Eksempel: Mediernes formidling af meningsmålinger (Larsen og Straubinger 2012)

Tabel 1: Litteraturgennemgang

	Enheder	Institut	Medie	Stikprøve	Ordlyd	Usikkerhed	Indsamling	Metode
Gennemsnit	354	85,6	69,7	55,5	33,9	29,1	56,6	35,3
Standardafvigelse	384	10,1	13,6	27,2	24	21,2	13,4	13,9
Minimum	23	71,5	47	0	6,4	0	39,1	11
Maksimum	1335	99,4	96	98	87	84	87	84
N	5303	4549	3803	5303	4775	5303	5044	3804
Studier	15	11	12	15	13	15	14	12

Note: De metodiske variable er angivet i procent vejet efter frekvensen af analyserede enheder i studierne. N angiver antallet af analyserede enheder i litteraturen på hver parameter.

Data: Andersen (2000), Brettschneider (1997), Chang (1999), de Vreese & Semelko (2002), Ferguson & de Clercy (2005), Gosselin & Pétry (2009), Hardmeier (1999), Lee (2006), Marton & Stephens (2001), Miller & Hurd (1982), Salwen (1985), Strömbäck (2009), Weaver & Kim (2002), Welch (2002) og Willnat et al. (2012).

Eksempel: Internetbrug og politisk engagement (Boulianne 2009)

► Spørgsmål: Påvirker brugen af internettet borgerens politiske engagement?

Metode: Effektstørrelser

Analyse: Distribution af effekter

Eksempel: Internetbrug og politisk engagement (Boulianne 2009)

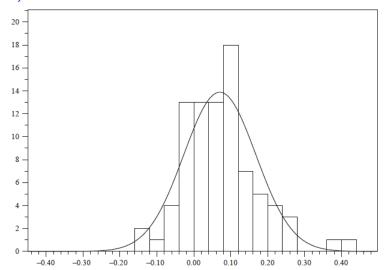


Figure 2. Distribution of effects of Internet use and political engagement.

- Spørgsmål: Er der responsivitet til den offentlige opinion online?
- Fremgangsmåde

► Teori: Analytisk framework

► Metode: Indsamling af litteratur

Resultater: Kategorisering af studier

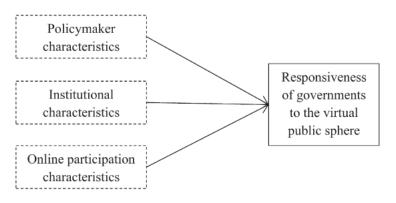


Fig. 1. Analytical framework.

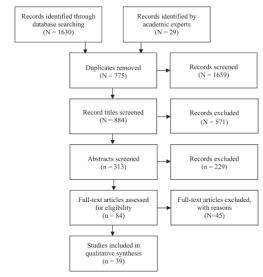


Fig. 2. PRISMA flow diagram.

 Table 3

 Policymaker explanations of governments' responsiveness to the virtual public sphere.

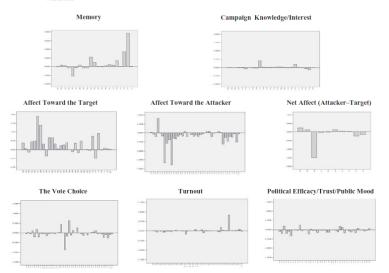
Category	Explanations	Articles	N (%)	
Policymaker characteristics	Competencies and skills	Carlitz & Gunn, 2002 5 (13.6%) Soon & Soh, 2014 Seltzer & Mahmoudi, 2013 Valtysson, 2014 Evans-Cowley & Hollander, 2010		
	Tensions professional role	Klang & Nolin, 2011 Soon & Soh, 2014	2 (5.3%)	
	Position of policymaker in policy networks	Klang & Nolin, 2011 Soon & Soh, 2014 Bekkers et al., 2011	4 (10.5%)	

Eksempel: Effekten af negative kampagner (Lau et al. 2007)

- ► Spørgsmål: Påvirker negative kampagner vælgerne?
- Metode: Kodning af effekter
- ► Analyse: En lang række af afhængige variable

Eksempel: Effekten af negative kampagner (Lau et al. 2007)

FIGURE 1 Effect Sizes Adjusted for Sampling Error and Measurement Unreliability, by Type of Dependent Variable



Eksempel: Personlighedstræk og prosocial adfærd (Kline et al. 2015)

- Spørgsmål: Påvirker personlighedstræk borgerens sociale adfærd?
- ▶ Metode: Indsamling af data med de relevante variable
- Analyse: Multilevel metaanalyse

Eksempel: Personlighedstræk og prosocial adfærd (Kline et al. 2015)

Table 1: Studies Included in Meta-Analysis

Table 1: Studies included in Meta-Ariatysis							
Study	Game	Personality Measure	Significant Predictors	Incentivized	Sample Size		
Ben-Ner and Kramer (2011)	Dictator	Big Five	(+)N;(-/+)E;(-)C	no	198		
Ben-Ner et al. (2004)	Dictator	Big Five	(+)A;(+)N	yes	322		
Brocklebank et al. (2011)	Dictator	Big Five	(-)N;(-)E;(+)O	yes	67		
Hilbig and Zettler (2009)	Public Goods	HEXACO	No test	no	134		
Hilbig et al. (2012a)	Public Goods	HEXACO	(+)O	no	424		
Hilbig et al. (2012b)	Public Goods	HEXACO	(+)O;(+)A	no	531		
Hirsh and Peterson (2009)	Prisoner's Dilemma	Big Five	(+)N	no	52		
Koole et al. (2001)	Common Pool Resource	Big Five	(-)E;(+)A	yes	71		
Kurzban and Houser (2001)	Public Goods	Big Five	None	yes	91		
Pothos et al. (2011)	Prisoner's Dilemma	Big Five	(+)A	no	113		
Schmitt et al. (2004)	Ultimatum Game	MBTI	None	yes	360		
Swope et al. (2008)	UG, DG, TG, PD	MBTI	(+)E	yes	276		
Artinger et al. (2010)	Dictator	N/A	N/A	yes	116		
Gunnthorsdottir et al. (2002)	Trust Game	N/A	N/A	yes	67		
Fischbacher et al. (2001)	Public Goods	N/A	N/A	yes	44		

Note: N = "Neuroticism"; E = "extroversion"; O = "Openness"; A = "Agreeableness"; C = "Conscientiousness";

^{(-) =} statistically significant, negative association; (+) = statistically significant, positive association

Dit studie og andres litteraturreview

- ▶ I den bedste af alle verdener kan andre bruge dit studie
- Rapporter tilstrækkelige informationer
 - Andre skal kunne bruge det i metaanalyser
 - Nøgleordet er transparens

Dagens øvelse

- ► Find en review artikel, der belyser den afhængige variabel eller det spørgsmål, du ønsker at beskæftige dig med
- ▶ Google Scholar
 - ► Konkrete emneord ('review', 'overview', 'agenda')
 - Bestemte tidsskrifter ('Annual Review of Political Science')
- ► Alene eller sammen med en anden
- 10 minutter

Næste gang

- Udvikle teorier
 - Definere kausaleffekt
 - Opbygge kausalmodeller
 - Eksplicitere antagelser
 - Validitet
- Senere: Teste og evaluere teorier
- ▶ Bemærk: næste gang (mandag) er vi tilbage
 - ► (Se altid seneste version af læseplanen for tid og sted)