

SENSURVEILEDNING

Emnekode og navn: PSY2022	Semester / År / Eksamenstype: Vår 2020 / Hjemmeeksamen
Oppgave: Det er fem oppgaver, som er beskrevet hver for seg under Eksamenskrav.	
Relevant pensumlitteratur: Se under hver enkelt oppgave.	
Eksamenskrav: Oppgave 1 (Konstruksjon av spørreskjema): Tom og Tale skal lage et spørreskjema som blant annet måler bruk av og holdninger til bruk av sosiale medier. De har hørt hvor viktig det er å ha kunnskap om hvordan respondentene tolker spørsmål, og ikke minst om ulike kontekstuelle forhold som kan påvirke svarene. De ønsker å unngå de vanligste feilene som gjøres ved konstruksjon av spørreskjema fordi kritiske røster har sagt at man ikke kan stole på svar fra spørreskjema fordi forskeren kan formulere spørsmål på måter som gjør at svarene blir slik man ønsker! <u>Relevant pensum:</u> Schwarz, N., Groves, R. M., & Schuman, H. (1998). Survey methods. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), <i>The handbook of social psychology</i> (4th ed., Vol. 1, pp. 143-179). Boston, MA: The McGraw-Hill Companies, Inc. Schwarz, N., Knäuper, B., Oyserman, D., & Stich, C. (2008). The psychology of asking questions. In E. de Leeuw & J. Hox (Eds.), <i>International handbook of survey methodology</i> (pp. 18-34): Taylor & Francis. Oppgaven omfatter måling av både egen atferd og ens holdninger til bruk av sosiale medier. Det er flere ulike kontekstuelle faktorer i et spørreskjema som kan påvirke svarene på atferds- og holdningsmål. Det er ikke grunn til å anta at sosiale medier er et spesielt følsomt tema. Det forventes at kandidaten kjenner til og kort kan gjengi Grices (1975) prinsipper om samarbeid og hva disse innebærer. Disse prinsippene gjelder for spørreskjema som ellers i samfunnet (dvs. all informasjon skal være informativ, relevant, sannferdig og tydelig). Vår oppgave som forskere er å forstå hvilke utfordringer den som fyller ut spørreskjemaet står overfor, og legge til rette for sann rapportering. Ved estimering av frekvens kan lukkede svaralternativer fungerer som mentale «rammer» for deltakerne og mange foretrekke åpne svar alternativer hvor man selv angir frekvens. Man kan drøfte fordeler og lempes ved begge tilnærmingene.	

Deltakere som besvarer spørreskjema kan også bli påvirket av hvordan spørsmål og påstander er formulert. Påstandene kan være av positiv eller negativ valør, og være bekreftende eller benektende. Folk har en tendens å være mer enige enn uenige i påstander, svare «midt på» og utstrakt monotont ved manglende motivasjon. Dette kan motvirkes ved at flere påstander etter hverandre presenteres vekselvis med ulik valør (e.g., «sosial medier styrker ungdoms sosiale nettverk» og «sosiale medier er negativt for ungdoms selvfølelse»)

Spørsmål som stilles kan i noen tilfeller påvirke svar på senere spørsmål (rekkefølge-effekter) da informasjon fra tidlige spørsmål kan assimileres inn i de senere. Dette gjelder imidlertid ikke generelt, men i tilfeller hvor man har spørsmål om spesifikke og generelle saksforhold. Det er gjennomført studier som viser måter å unngå assimileringseffekter gjennom ulike instruksjoner og endring av rekkefølge. Det viktigste er at kandidaten forstår hvordan respondenten tenker og trekker slutninger om spørsmål i spørreskjema under ulike betingelser. ##ikke replikert##

Oppgave 2 (Sensitive spørsmål):

Tom og Tale er begge psykologistudenter og ønsker å gjøre en studie på temaet «hooking up». De har lest at mer tilfeldige seksuelle relasjoner er blitt mer vanlig blant unge voksne sammenlignet med tidligere generasjoner, og forskere snakker nå om en såkalt «hook-up culture». Tom og Tale ønsker å se om bruk av datingapper (Tinder, etc.) er assosiert med at man har hatt flere tilfeldige seksualpartnere, og om det er økt forekomst av kjønnssykdommer blant de som har hatt mange tilfeldige seksualpartnere. De skal spesielt se på kjønnsforskjeller. De er imidlertid bekymret for gyldigheten av de svarene de får på spørsmål om slike følsomme tema, og diskuterer hvordan de best kan unngå sosialt ønskelig rapportering.

Relevant pensum:

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133, 859-883.

Følsomme spørsmål omfatter en rekke fenomener som spenner fra inntekt og skatt, personlig tro, seksuell legning/orientering, illegal atferd og lovbrudd, rusmiddelbruk og seksuell atferd, og andre ting man av ulike grunner ønsker å holde skjult. Følsomme spørsmål kan (1) oppleves påtrengende, (2) utløse frykt for reaksjoner, og (3) utløse svar som er sosialt uakseptable eller uønskede og som dermed kan bli gjenstand for motivert «justering» av responsen dersom de blir stilt i en survey (reduisert responskvalitet). Det er dette vi vanligvis omtaler som sosialt ønskerverdige svar eller inntrykksforvaltning (eng. impression management). I regelen snakker man om at deltakere i en survey underrapporter sosialt uønskede tanker og handlinger, men man kan også se for seg at noen deltakere «legger på» litt og overdriver atferd som de tror kan gi de status (unge menn som rapporterer om antall seksualpartnere). Andre potensielle problemer knyttet til følsomme spørsmål er manglende deltakelse i surveyen, og manglende svar på de spørsmål som er av følsom natur.

Effektene av sosialt ønskerdig rapportering kan reduseres gjennom å benytte modaliteter som gir deltakerne følelse av økt grad av anonymitet. Metoder med nettbaserte løsninger (datamaskinassistert personlig intervju) og spesielt selvadministrert synes øke rapportering av illegal rusmiddelbruk og minske kjønnsforskjeller i antall seksualpartnere. Intervjusetting med tilstedeværelse av andre vil i regelen medføre underrapportering av sosial følsomme tema. Indirekte metoder for å lokke frem sensitiv informasjon kan nevnes (Random Response Technique og Bogus pipeline), men slike er relativt ressurskrevende å benytte. Man kan også benytte seg av såkalte

«Forgiving words» (signalisere at noe er vanlig eller uproblematisk) og blande sensitive og ikke-sensitive spørsmål i et spørreskjema/intervjuguide. Man kan også inkludere mål på individuelle forskjeller i *tilbøyelighet* til å svare sosialt ønskelig (e.g., Impression management eller Social Desirability Scale) med tanke på å kontrollere statistisk for slike tendenser i analysene. Noen mener man da vil kunne få et «renere» estimat av effekter og sammenhenger når det er tatt hensyn til sosialt ønskelig rapportering. En god besvarelse vil inneholde beskrivelser av i minst to av disse metodene, og avveining av kostnader og gevinster, gjerne understøttet av empiriske funn.

Oppgave 3 (Eksperimentelle design):

Tor Kåre og Leif Egil, som er stipendiater i psykologi, sitter en dag i kantinen på Dragvoll og diskuterer sine prosjekter. Tor Kåre studerer for tidlig fødsel og er interessert i om det er mulig å hjelpe tenåringer som er født for tidlig til å unngå kognitive svekkelse og lærevansker, som er en vanlig langtidsfølge av for tidlig fødsel. Han har en idé om at det går an å trene opp eksekutivfunksjoner, som denne gruppen ofte har trøbbel med. Leif Egil driver med kognitiv psykologi og har utviklet et opplegg hvor folk løser oppgaver som belaster arbeidsminne og oppmerksomhetsfunksjoner ved bruk av et dataprogram. Han foreslår at Tor Kåre kan bruke hans program til et intensivt treningsopplegg for folk som er født for tidlig. Han mener at hvis Tor Kåre lar sine for tidlig fødte tenåringer løse oppgaver fra hans program hver dag over fem uker, så vil det bringe deres kognitive evner opp til normalt nivå. Men du må ha en kontrollgruppe, sier han, ellers kan du ikke være sikker på at en eventuell bedring skyldes treningsprogrammet. Tor Kåre liker idéen, men tenker at det viktigste vil være å teste tenåringene før treningsprogrammet starter, slik at han kan dokumentere en forbedring. Liv Silje, som er stipendiat i statsvitenskap, sitter ved nabobordet, og hun skyter inn: «Jeg kunne ikke unngå å høre hva dere snakker om. Dette er utrolig spennende! Men hvordan vet du at hvis de blir bedre på disse testene, så vil det ha noe å si for hvordan de fungerer på skole og ellers i livet? Og dessuten: Hva om det ikke er selve testen som er årsaken til en eventuell forbedring?»

Relevant pensum:

Meltzoff, J., & Cooper, H. (2018). *Critical Thinking about Research: Psychology and Related Fields*. (Second Edition). Washington, DC. American Psychological Association. 335 sider. ISBN: 978-1-4338-2710-5. Kindle version: <https://www.amazon.com/Critical-Thinking-About-Research-Psychologyebook/dp/B075VKZDH7/>

Wilson, T. D., Aronson, E., & Carlsmith, K. (2010)

Denne oppgaven gir studentene flere hint; den henter om behov for kontrollgruppe (Leif Egil), testing før og etter intervensjonen (Tor Kåre) og problemer med ekstern validitet (Liv Silje). I likhet med læreboken, la forelesningen vekt på forskjellen mellom avhengige og uavhengige variable og begrepet intern validitet. Den eksperimentelle metode går ut på å manipulere den uavhengige variable for å *forårsake* en endring i den avhengige. Et godt design gjør at andre årsaker til slik endring enn den eksperimentelle manipulasjon, fremstår som mindre sannsynlige. Studien har intern validitet hvis endringen i den avhengige variabel skyldes den eksperimentelle manipulasjonen. «Endring» brukes her både om forskjell mellom grupper (aktiv vs. kontroll) og endring over tid (før vs. etter manipulasjon). I denne oppgaven er den foreslåtte eksperimentelle manipulasjonen en intervensjon i form av et fem ukers kognitivt treningsprogram.

Læreboken kontrasterer ulike kategorier av design:

- Et pre-eksperimentelt design har ingen kontrollgruppe og kun én post-test (etter intervensjonen). Her er det ikke mulig å trekke slutninger om årsakssammenhenger.

- Kvasiekperimentelle design kan ha pre- og posttest, eller en kontrollgruppe, eller begge deler, men har ikke *random assignment* til gruppene (intervensjon vs. kontroll).
- Et ekte eksperiment har kontrollgruppe (minst én) og random assignment, samt (minst) en post-test. Kan også ha pretest eller enda mer komplekst design.

Det er ikke et krav at studenten eksplisitt skal bruke begrepene pre- og kvasiekperimentelle design i besvarelsen, men designet som foreslås som løsning, må være et *ekte* eksperiment. Hvis det ikke er det, bør det ikke gis karakter C eller høyere.

Det finnes to forholdsvis enkle eksperimentelle design som er gode løsninger på oppgaven:

1. To grupper, posttest: Tilfeldig fordeling (random assignment) av for tidlig fødte tenåringer til to grupper; en som mottar intervensjonen og en «kontrollgruppe». Test etter intervensjon.
2. To grupper, pre- og posttest: Tilfeldig fordeling av for tidlig fødte tenåringer til intervensjon- eller kontrollgruppe. Test før og etter intervensjon.

Med design 1) kan en forskjell mellom intervensjonsgruppen og kontrollgruppen på posttest attribueres til intervensjonen, da en eventuell modningseffekt ville ha påvirket begge gruppene. NB! Det antas at gruppene i utgangspunktet er like med hensyn til alle relevante parametere (*årsaksfaktorer*) på grunn av *random assignment*. NB! Randomiseringen sannsynliggjør dette, men uten å garantere det. (Læreboken går langt i å fremstille randomisering som en universalløsning som garanterer at gruppene er like på alle andre faktorer den eksperimentelle manipulasjonen. Andre kilder er mer opptatt av å fremheve svakheter ved dette designet. Noen av disse nevnes nedenfor, og studenten bør belønnes for å drøfte disse.)

Studenten bør få ekstra poeng for å nevne at det går an statistisk å kontrollere for ulikheter mellom gruppene mht. diverse kliniske og demografiske variabler. Disse kan inkluderes i de statistiske analysene som ko-variater. Her kan man f.eks. kontrollere for alder og kjønn, samt andre variabler som angir hvor godt deltagerne fungerer kognitivt (uten å inkludere en pre-test), som skolekarakterer eller lignende.

Et annet moment som studenten bør belønnes for å inkludere, er behovet for å kontrollere om deltagerne *gjør noe annet utenom det vanlige* mens eksperimentet pågår, som kan tenkes å påvirke utfallet. Dette kan være at noen deltagere gjennomfører et annet trenings- eller behandlingsopplegg samtidig med intervensjonen. Hvis dette skjer usystematisk, f.eks. at noen deltagere i begge grupper begynner å spille sjakk, så vil dette øke *within-group* varians og således gjøre det vanskeligere å oppdage en reell effekt av intervensjonen. Randomisering løser ikke dette problemet. (Dersom det skjer systematisk, altså at bare den ene gruppen utsettes for slik ekstern påvirkning, vil det være katastrofalt for studiens interne validitet. Randomisering reduserer sannsynligheten for et slik scenario.)

Med design 2) kan man også sammenligne gruppens skårer på pre-testen for å se om randomiseringen gjorde dem like, slik man bare antar under (1). Man kan direkte undersøke om begge gruppene forbedrer sin prestasjon fra pre- til post-test, og en forbedring i kontrollgruppen kan ikke tilskrives intervensjonen. En ulempe med dette designet er at det kan være testeffekter, altså at pretesten fungerer som trening for posttesten (hvis de er tilstrekkelig like). Hvis begge grupper blir bedre til å ta testen, og derfor gjør det bedre på post-test, så kan det maskere en ekte effekt av intervensjonen (falsk negativ).

Lokale og generelle historieeffekter:

Læreboken nevner lokale historieeffekter, som kan oppstå hvis gruppene for eksempel går på ulike skoler, eller i ulike klasser, og blir eksponert for ulike former for pedagogikk som gjør at den ene gruppens kognitive ferdigheter øker mer enn den andres. Dette motvirkes ved randomisering (*random assignment*). Generelle historieeffekter er slike som påvirker hele samfunnet, som naturkatastrofer, krig og pandemier(!). For å unngå generelle effekter bør gruppene testes *samtidig*. Merk at randomisering ikke vil hjelpe mot generelle historieeffekter dersom gruppene ikke testes samtidig. Dette er et veldig teoretisk poeng, men det vil kunne være noen studenter som nevner det, da det er omtalt i læreboken.

Andre problemer:

Intern validitet er spørsmålet om andre årsaker kan forklare våre resultat enn den eksperimentelle manipulasjonen. Det er en til tider subtil grenseoppgang mot begrepsvaliditet (*construct validity*), og det har ikke vært tema for denne delen av kurset. For eksempel: Hvis deltagerne gjør det bedre på posttest fordi å delta i intervensjonen gir økt motivasjon, mer selvtillit eller lignende, kan det argumenteres for at det er en trussel mot begrepsvaliditet snarere enn intern validitet (det er intervensjonen som forårsaker en bedre prestasjon, men vi er interessert i om det er en spesifikk kognitiv treningseffekt av opplegget). Dette berører spørsmålet om hva kontrollgruppen skal gjøre. Hvis de bare testes (posttest eller pre- og posttest), kan en forbedring hos intervensjonsgruppen skyldes en slik *Hawthorn effekt*, men hvis kontrollgruppen får et «placebo»-opplegg med en annen form for aktivitet på datamaskin, for eksempel å spille spill eller noe lignende som ikke antas å ha en spesifikk positiv virkning på eksekutivfunksjon, kan man kontrollere for en slik effekt.

Kontaminasjon: De to gruppene kan *i teorien* påvirke hverandre hvis de befinner seg innenfor samme geografisk område, fordi de går på samme skole eller deltar i samme fritidsaktivitet. (Hvis de ikke går på samme skole, øker risiko for lokale historieeffekter.) Dette er nok et større problem for forskning på ulike undervisningsmetoder (pedagogikk), hvor elevene kan «compare notes» og lære av hverandre. Dette virker mindre sannsynlig i forbindelse med et intenst kognitiv treningsopplegg som vi diskuterer her.

Det finnes mer komplekse eksperimentelle design, som alle er ekte eksperiment og derfor godkjennes. For eksempel:

- Et pre-post test/ fire-gruppe Solomodesign kontrollerer for det meste. Her er det to intervensjonsgrupper og to kontrollgrupper. En intervensjon og en kontrollgruppe får pre- og posttest, de gjenværende gruppene får bare posttest. Her kan man blant annet kontrollere for en eventuell test-effekt. Dette er et ressurskrevende studiedesign.
- Det er også mulig å ha et to-gruppe design med to pre-tester som har like stor tidsmessig avstand mellom hverandre som mellom andre pre-test og post-testen. Dette vil også kontrollere for effekten av testing *per se*. Hvis det ikke er en signifikant læringseffekt fra første til andre pretest, antar man at det heller ikke er det mellom andre og tredje.

Post-testen:

Hvordan teste «forbedring av kognitiv funksjon»? Skal (pre- og) post-testen inneholde oppgaver som ligner på selve treningsopplegget? Eller skal den være tilstrekkelig annerledes til at det ikke oppstår en direkte treningseffekt? Man kan for eksempel bruke veletablerte kognitive tester for eksekutivfunksjon, som Stroop, Wisconsin card sorting test eller et generelt nevropsykologisk testbatteri (som D-KEFS) (se *Flak et al. (2019) Front Psychol. Apr 12;10:807*). Dette er en avansert diskusjon, som krever kjennskap til konkrete nevropsykologiske eller kognitive tester, så vi kan ikke forvente en detaljert drøfting her. Men studenten bør belønnes for å være klar over problemstillingen.

Et viktig moment er at administrasjon av test og vurdering av testprestasjon må foregå «blindt», dvs. uten kjennskap til gruppestatus. (PS! Hvis alle testene er *computerized*, vil dette kanskje kunne gjøre automatisk.)

Kontrollgruppen:

Hvem inngår i kontrollgruppen? Er det for tidlig fødte kontroller eller en terminfødt kontrollgruppe Leif Egil snakker om? Riktig svar er at det må være en gruppe av for tidlig fødte, som blir tilfeldig fordelt (*assigned*) til en intervensjons- og en kontrollgruppe.

(Spørsmål fra studenter i Diskusjonsforum på Blackboard indikerer at de hersker noe forvirring her, da noen studenter synes å tro at de for tidlig fødte er en selektert gruppe. Det er en misforståelse. Vi samler fra populasjonen av for tidlig fødte (innenfor et visst geografisk område) og så fordeler vi deltagerne tilfeldig på intervensjon eller kontroll/placebo. For ordens skyld nevnes det at koordinator svarte følgende på Blackboard, på spørsmål om de for tidlig fødte er selektert, og altså ikke tilfeldig utvalgt, og at løsningen derfor må bli et kvasiekperimentelt design: *Jeg kan ikke svare spesifikt for denne oppgaven, men generelt så har du et kvasiekperimentelt design dersom det ikke er tilfeldig hvilken gruppe deltagerne havner i. Men så er spørsmålet hvem er deltagerne og hvilken populasjon kommer de fra? PS! En bakgrunnsopplysning om for tidlig fødsel: Det finnes et norsk fødselsregister, hvor gestasjonsalder (om du ble født for tidlig eller til termin) er registrert, som forskere kan søke om å få adgang til.*)

Studenter som har skjönt at vi samler fra en populasjon av for tidlig fødte, som vi har god oversikt over pga. fødselsregisteret, bør belønnes for dette. Man kan enten invitere alle som ble født innenfor en gitt tidsrom og et gitt geografisk område, for eksempel Trondheim, eller et tilfeldig trukket utvalg av dem, til å delta i studien. Man vil naturligvis ikke få 100% samtykke.

Disse tingene ble ikke nevnt på forelesning, og det står ikke om det i læreboken, så selv om koordinator har hintet om det på Blackboard, hvor han riktignok forsøkte å unngå å si for mye, kan det ikke forventes at studentene har inngående kjennskap til dette. Dette berører spørsmål om generaliserbarhet (ekstern validitet). Til hvilken populasjon generaliserer vi?

Hva om studenten foreslår å inkludere en *kontrollgruppe av terminfødte*? Et design hvor alle for tidlig fødte får intervensjonen og en kontrollgruppe av terminfødte får placebo, er ikke et godkjent design på denne oppgaven.

I tillegg er det noen andre viktige poeng studentene kan/bør drøfte:

- Hvor langvarig er en eventuell virkning av intervensjonen? Studenter som foreslår å inkludere flere post-tester over en viss periode (f.eks. etter seks måneder, ett år eller mer), for å undersøke om virkningen er langvarig, bør belønnes. Dette er et vesentlig poeng.
- Studenter som foreslår å kontrollere om deltagerne faktisk gjennomfører treningen, bør også belønnes. Kan dette registreres på datamaskinen som brukes i intervensjonen? Kan man intervjuer tenåring og foreldre, for eksempel en gang i uken, for å følge opp?
- Oppgaveteksten henter om ekstern (økologisk) validitet: Har det noe å si *i praksis* om de for tidlig fødte blir bedre på testen som brukes i studien? Fungerer de bedre i hverdagen, gjør de det bedre på skolen? For å kontrollere for dette, kan man undersøke om de viser en forbedring i skoleprestasjoner for eksempel. Man kan også bruke noen tester fra WISC/WAIS, f.eks. tester av arbeidsminne og eksekutivfunksjoner. Igjen, husk at studentene neppe har særlig god kjennskap til utvalget av nevropsykologiske tester.

Oppgave 4 (Replikasjoner og power):

Ragnhild og Rune har kommet over flere artikler som omhandler problemer med replikasjoner i psykologien. De prøver å forstå hva denne «reproduksjonskrisen» handler om, hvor stort dette problemet egentlig er, og om den gjelder hele eller kun deler av psykologien.

De har også lest at bare 8% av publiserte studier innenfor psykologien har akseptabel power = .80 eller høyere, og at dette kan være noe av grunnen til den lave replikasjonsraten. Som fremtidige forskere tenker de at de skal sikre høy grad av power i egne studies. De forsøker å få rede på hvilke faktorer som har betydning for power, og hva som er problemene knyttet til for lav og for høy power.

Relevant pensum:

Open Science Collaboration, *Science* 349, aac4716 (2015). doi: 10.1126/science.aac4716

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on “Estimating the reproducibility of psychological science”. *Science*, 351(6277), 1037-1037.
doi:10.1126/science.aad7243

Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, 6(621). doi:10.3389/fpsyg.2015.00621

G-power (Statistical Power Analyses for Window and Mac); download from:

<http://www.gpower.hhu.de/>

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.

Replikasjoner utgjør selve «skillemynten» i vitenskapen og er et kjernepunkt innen empirisk vitenskap. Om noe er replikerbart utgjør kriteriet for skillelinjen mellom vitenskap og ikke-vitenskap. En direkte replikasjon innebærer at alle metodiske aspekter er reproduert (utvelgelse, prosedyre, manipulasjoner, materiale/måle-instrumenter, m.m.). Dette sikres best ved at originalforfatteren gjør dette tilgjengelig og kan kommentere på evt. justeringer. Det er ikke mulig å 100% nøyaktig replikere en studie (utvalget vil alltid være et annet selv om alt annet er likt). Man ser oftere enn tidligere at replikasjonsstudier preregistreres. En konseptuell replikasjon er et forsøk på å validere den underliggende teorien, fenomenet eller mekanismen med metoder som ikke er helt identiske med det opprinnelige forsøket. Det er viktig å skille mellom direkte og konseptuell replikasjon og diskutere implikasjonene av disse.

Reproduksjonskrisen i psykologisk forskning handler om at funnene fra flere av de mer kjente studiene innenfor sosial- og kognitiv psykologi ikke er blitt reproduert i senere studier. Open Science Collaboration (OSC) er et forskningsprosjekt hvor man systematisk forsøkte å reproducere funnene fra 100 studier fra tre høyt rangerte tidsskrifter (første nummer i 2008 ble tilfeldig valgt). Kun 35 av de 97 signifikante effektene (36%) ble reproduert. Innenfor sosialpsykologien var det kun 25% av funnene som lot seg reproducere (50% innen kognitiv psykologi). Effektstørrelsene i replikasjonene var gjennomgående mye lavere enn i originalstudiene.

OSC-studien er blitt kritisert av flere, blant annet Gilbert et al.(2016), som blant annet viser til ManyLabs-prosjektet som rapporterer langt høyere reproduksjon av funn enn OSC. Gilbert finner

dessuten flere mangler ved OSC-studien som sår litt tvil om funnene. Artikkelen til OSC og artikkelen til Earp & Trafimow nevner flere mulige grunner til de mange manglende replikasjonene og forslår tiltak for å redusere sjansen for publisering av falske positive funn. En god besvarelse vil drøfte noe av dette og diskutere hva som kan være implikasjonene av denne forskningen for psykologi som vitenskap.

Power handler om sjansen for korrekt å forkaste H_0 . Hypoteser handler om populasjoner. Testene vi gjør er på bakgrunn av utvalg. Power er noe alle forskere bør beregne forut for at man gjennomfører en studie og trekker utvalg. Faktorer som har betydning for power er først og fremst antall deltakere og størrelsen på effekten man avdekker. Selv om man får signifikante funn er studier ofte «underpowered» fordi effektene man finner er svakere enn forventet med det antall deltakere man gjorde testen på. Resultatet blir da at den statistiske testen forblir uavklart (og forskningen er bortkastet). Før høy power skyldes gjerne at man har mange deltakere. Dette ansees å være bortkastede ressurser (man kunne klart seg med færre).

Oppgave 5 (Etikk):

Ragnhild og Rune ønsker å studere voksne studenters erfaringer med voldelige konflikter i barndommen og mulige virkninger av disse. De vil bruke et elektronisk (nettbasert) spørreskjema med mulighet for helt anonym rapportering. De ønsker å rekruttere fra forelesninger. Hvilke forskningsetiske utfordringer står de overfor, hva må de gjøre, hvor må de eventuelt søke eller melde inn prosjektet og hvilket lovverk kommer til anvendelse for denne studien?

Mens de holder på å planlegge studien får de høre om et nasjonalt prosjekt som skal kartlegge omfanget av vold og seksuelle overgrep blant elever i ungdomsskolen (alderen 12-16 år). Dette er ikke tidligere gjort i Norge, og flere departementer mener det er viktig å få denne kunnskapen. I denne studien gir elevene selv samtykke til deltakelse, mens foreldrene kun er informert om studien. Det sies i informasjonsskrivet at alle som svarer vil forbli anonyme. Elevene benytter sin egen skole-PC ved besvarelse og logger på en nettportal med en unik kode som er koblet til den enkeltes navn. Ragnhild og Rune lurer på hvilke forskningsetiske vurderinger som er gjort i dette prosjektet, og om det er noe kritikkverdige ved prosjektet.

Relevant pensum:

https://helseforskning.etikkom.no/reglerogrutiner/loverogregler?p_dim=34770&_ikbLanguageCode=n

De viktigste lovene som regulerer forskning på mennesker er helseforskningsloven og personopplysningsloven. Følgende prinsipper gjelder:

1. Forskningsetiske prinsipper – hensyn til personer

- Barn og unge som deltar i forskning, har særlige krav på beskyttelse
- Forskeren skal vise respekt for individets privatliv og familieliv. Informanter har krav på å kunne kontrollere om sensitiv informasjon om dem selv skal gjøres tilgjengelig for andre
- Forskeren skal ikke tillegge informanter og aktører irrasjonelle eller lite aktverdige motiver uten å kunne gi overbevisende dokumentasjon og begrunnelse. Forskeren skal vise respekt

for verdier og holdninger hos forskningsdeltakerne, ikke minst når disse avviker fra det som er mest vanlig i storsamfunnet

- Identifiserbare personopplysninger innsamlet for ett bestemt forskningsformål kan ikke uten videre benyttes til annen forskning.
- Opplysninger om identifiserbare enkeltpersoner skal lagres forsvarlig. Slike opplysninger skal ikke lagres lenger enn det som er nødvendig for å gjennomføre formålet med behandlingen.
- Forskeren har et ansvar for å unngå at forskningsdeltakerne blir utsatt for alvorlig fysisk skade eller andre alvorlige eller urimelige belastninger.
- Forskeren bør vurdere og foregripe virkninger på tredjepart som ikke er direkte inkludert i forskningen

2. Hovedregel om samtykke

- Det kreves samtykke fra deltakere i medisinsk og helsefaglig forskning, med mindre annet følger av lov.
- Samtykket skal være informert, frivillig, uttrykkelig og dokumenterbart (skriftlig samtykke er hovedregelen). Samtykket skal bygge på spesifikk informasjon om et konkret forskningsprosjekt med mindre det er adgang til å avgi et bredt samtykke.
- Dersom forskningsdeltakeren kan anses å være i et avhengighetsforhold til den som ber om samtykke skal det informerte samtykket innhentes av en annen.
- Barn / unge < 16 år (for invasiv forskning: < 18): stedfortredende samtykke, ofte foreldre. Barnet skal forespørres og hvis barnet motsetter seg det, kan ikke forskningen gjennomføres. Barn mottar eget informasjonskriv.

Eksplisitt henvisning til lovverket vil veie klart positivt i besvarelsen (Forskningsetikkloven, Personopplysningsloven, og Helseforskningsloven). Drøfting av forutsetninger som alder, avhengighet til forsøksansvarlig og hvorvidt det er forskning som har som «formål å skaffe til veie ny kunnskap om helse», vil styrke besvarelsen.

I en studie om voldelige konflikter i barndommen kan det være snakk om å ha en viss (dokumentert i informasjonsskrivet) beredskap ettersom studien tar opp spørsmål som kan være belastende for deltakerne og vekke vonde minner. Ved elektroniske svar blir det er spørsmål om anonymitet og hvordan personidentifiserbare opplysninger oppbevares. Dersom anonym løsning benyttes trenger man ikke melde prosjektet inn for NSD, personvernombudet. Hvis ikke anonym løsning benyttes må prosjektet meldes NSD. Da gjelder de nye EU-reglene for personvern (GDPR). En viss aktsomhet må utvises dersom man rekrutterer gjennom forelesninger, spesielt fra egne forelesninger (da man er en autoritetsperson og egne studenter kan stå i et avhengighetsforhold). Dersom studien er egnet til å fremskaffe ny kunnskap om helse, er REK riktig instans da den omfattes av helseforskningsloven. Teksten «mulige virkninger» må derfor klargjøres. Hvis virkninger er helseutfall (angst, depresjon, og lignende) skal prosjektet meldes REK. Hvis man med virkninger mener meninger, aksept, holdninger, eller måter å tolke hendelser på, er dette ikke å regne som helseforskning. En god besvarelse vil diskutere dette.

Del 2 av oppgaven handler primært om to forhold: (1) Om man kan fravike kravet om stedfortredende samtykke fra foreldre/foresatte blant barn under 16 år, og (2) Om deltakernes lovnad om anonymitet er reell og om det som er gitt av informasjon om dette er korrekt.

I tilfeller hvor man søker å kartlegge vold og seksuelle overgrep i befolkningen vil det være noen tilfeller der foreldre/foresatt selv er voldelige overfor egne barn. Selv om denne gruppen ikke antas å være stor (den er ukjent), kan det være tilstrekkelig grunnlag for at REK har fraveket hovedregelen om samtykke i denne studien. Dersom samtykke fra foreldre/foresatte hadde vært et krav, ville man risikert at barn i familier med misbruk og overgrep ikke kunne delta. Gitt at studiens formål er tilstrekkelig aktverdig vil det være tilstrekkelig at elevene selv gir sitt samtykke, og at deres foresatte

er informert. Studien omfattes selvsagt av helseforskningsloven og søknad sendes REK. Når det gjelder personvernet, kan det reises tvil om opplysningene som er gitt deltakerne er korrekte. Nettløsningen som benyttes er ikke anonym, så besvarelsene er ikke anonyme i lovens forstand, iallfall ikke før alle data er aidentifisert. En diskusjon rundt hvordan dette skal formidles til deltakerne vil trekke opp. Alle informasjonsskriv skal være alderstilpasset, og det er ikke sikkert at 12-åringer forstår begrepet «anonym». Et moment til diskusjon kan også være beredskap i undersøkelsesperioden, og om eventuelle plikter man har som forsker å varsle skole eller myndigheter dersom man oppdager lovbrudd (man plikter ikke dette, men skal informere om hvor man kan søke hjelpe ved behov).

Karakterbeskrivelse:

Faglærer / oppgavegiver:

Navn:

Sted / dato: