

Bilaga 1 Verktyg för att granska risk för bias

WWC standards för SCED primärstudier

WWC har erkänt single-case design som en av de tre huvudsakliga studiedesignerna (vid sidan av RCT och regressionsdiskontinuitetsdesign) för att utvärdera insatsers effekter. What Works Clearinghouse (WWC) standards [1] för SCED-primärstudier innehåller metodologiska riktlinjer för genomförande av primärstudier med SCED.

Granskningsprocessen består av tre steg.

Steg 1 – Kriterier för att betraktas som en SCED primärstudie

Som ett första steg i granskningsprocessen fastställs att studien uppfyller allmänna krav samt specifika kriterier för en experimentell SCED-studie.

Allmänna kriterier

Studien ska omfatta relevant population, insats och utfall¹, vara offentligt tillgänglig på engelska och publicerats de senaste 20 åren.

Specifika kriterier

Studien ska

- vara baserad på individuella fall
 - oftast en enskild deltagare (n=1), men kan vara en grupp av deltagare
- för jämförelseändamål fungerar fallet som sin egen kontroll.
 - fallets utfallsmätningar före insatsen jämförs med utfallsmätningar under och efter insatsen (före-, under- och efterstudie, ABA)
 - utfallsvariabeln mäts upprepade gånger inom faser och tillräckligt ofta för att fånga förändring över tid
 - regelbundna och frekventa mätningar i varje fas
 - baslinjefas (A); ska representera stabilitet över flera mätningar innan insats introduceras, minst tre datapunkter krävs för att säkerställa att en stabil trend har etablerats (AAA).
 - insatsfas (B): minst tre datapunkter för att möjliggöra identifiering av en tydlig förändring från baslinjen (BBB)
- för att fastställa Insatsens möjliga effekt ska utfallsförändring observeras över flera faser
 - inom, under och efter faser
 - exempelvis första baslinjefas, första insatsfas, andra baslinjefas (At1Bt2At3), samt en andra insatsfas (At1Bt2At3Bt4), eller första baslinjefas, första insatsfas, följt av en fas med ny insats (At1Bt2Ct3)

¹ I standarderna anses populationen relevant om den består av personal och elever/studenter inom alla delar av utbildningsväsendet, samt mycket små barn till vilka tidiga insatser riktas. Populationen skall vara bosatt i USA, Australien, Canada, Irland, Nya Zeeland eller Storbritannien. Insatserna ska vara utbildningsbaserade och utfallen relevanta i utbildningssammanhang.

- genomföras i enlighet med någon av designtyperna
 - Withdrawal/Reversal-design
 - Multiple Baseline Design
 - Alternating Treatments Design
 - Changing Criterion Design, eller variationer av dessa som
 - Multiple Probe Design.

När studien bedöms motsvara kriterierna för att betraktas som en SCED primärstudie, antingen utan reservationer (Meets WWC Standards Without Reservations) eller med reservationer (Meets WWC Standards With Reservations) görs en fortsatt bedömning i enlighet med steg två. I de fall studien inte bedöms motsvara kriterierna (Does Not Meet WWC Standards) och därmed inte betraktas som en SCED primärstudie görs ingen ytterligare bedömning.

Steg 2 - Utfallsmätning och förväxlingsfaktorer

I detta steg görs en bedömning i förhållande till standarder som avser utfallsmätning och förväxlingsfaktorer. Utfall avser här observation av beteende.

Mätning av utfall

Standard 1: Valida utfallsmått

- Tydlig definition av utfallet
- Utfallet måste mätas med hjälp av en kvantitativ indikator som speglar den förväntade förändringen (till exempel antal beteenden per tidsenhet, eller frekvens).
- Utfallet ska vara direkt relaterat till insatsens mål

Standard 2: Reliabilitet i mätningen

- Utfallsvariabeln måste mätas systematiskt över tid av mer än en bedömare för varje fall
- Studien måste för minst 20 procent av datapunkterna samla in data om överensstämmelse mellan bedömare i varje fas (Inter-observer agreement (IOA))
 - minimalt acceptabla IOA är 80 % för varje utfall över alla faser och fall (dock behöver värdena för överensstämmelsen inte uppnå minimivärden separat för varje fall eller fas)
 - rådata från den sekundära bedömaren som samlats in för att mäta överensstämmelse mellan bedömare behöver inte rapporteras. Det räcker att rapportera sammanfattande mått på överensstämmelse mellan bedömare

Standard 3: Överanpassning av mätinstrument/vad som observeras

- Mätinstrumentet förändras på ett sätt som gynnar personen under vissa mätningar eller faser. Exempelvis inkluderas innehåll som personen bara har exponerats för under insatsfas. Sådan partiskhet göra det svårt att avgöra om förändringarna i resultat beror på insats effekt eller förändringar i sätt att mäta

Standard 4: Konsekventa datainsamlingsprocedurer

- Utfall skall mätas på samma sätt i alla faser
- Datainsamlingsprocedurerna var liknande mellan olika förhållanden för ett givet fall. Granskare bör leta efter detaljer som tyder på att data samlades in med olika metoder, vid olika tidpunkter eller av olika personer mellan förhållandena. En stegvis introduktion

av en insats, som är en avsiktlig del av designen, representerar inte ett problem med inkonsekvent datainsamling.

- Utfallsmått oberoende av insats: det som mäts för att bedöma resultatet (utfallet) ska inte påverkas direkt av själva insatsen. I single-case designs (SCD) används ofta väldigt specifika och praktiska utfallsmått, som att mäta om en person kan följa en checklista eller utföra en viss uppgift. Här behöver granskaren kontrollera att utfallsmåttet inte bara är en direkt återspeglning av något som deltagaren lärt sig mekaniskt under insatsen, utan att det faktiskt bedömer en förbättring i färdighet eller funktion som är relevant och användbar utanför själva insatsens kontext.

Fastställ förväxlingsfaktorer

Standard 5: Insatsgivare

- Personen som ger insatsen (till exempel en lärare, förälder eller kamrat) kan vara en förväxlingsfaktor.
- Om olika personer ansvarar för olika faser, till exempel baslinje och insatsfas, kan detta påverka resultaten.

Standard 6: Förändringar i procedurer eller kontext

- Ändringar mellan olika förhållanden som inte är en del av insatsen kan snedvridera resultaten.
- Exempel på sådana förändringar:
 - Sessionens längd: Skillnader i hur lång tid sessioner pågår.
 - Förstärkning: Ändringar i mängden eller kvaliteten på belöningar eller förstärkning som ges till deltagaren.
 - Miljö: Platsen där faserna sker, exempelvis klassrum i en fas och lekplats i en annan.
 - Social kontext: Förändringar i den sociala miljön som påverkar deltagarens möjligheter att agera eller svara.

Steg 3 - Kriterier för att betrakta resultat som bevis för potentiell insatseffekt

WWC beskriver Single-case-designs som experimentella designar med potential att visa kausala effekter. Fyra standarder skall uppfyllas för att resultat från en SCED primärstudie kan anses utgöra sådana potentiella effekter.

Standard 7: Datatillgänglighet

- Rådata ska finnas tillgängliga i ett grafiskt format eller tabell

Standard 8: Forskarstyrd oberoende variabel

- Kontroll över tidpunkten för insats är en avgörande faktor för att SCED ska kunna betraktas som experimentell och potentiellt kvalificera sig för WWC:s högsta betyg. Forskaren skall bestämma tidpunkten när ett individuellt fall övergår mellan olika faser eller förhållanden.

- Det skall finnas underlag som stödjer att den oberoende variabeln systematiskt manipulerats. Vid samarbete med andra personer som deltar i studien, såsom föräldrar och profession, skall slutligt beslut om när förhållandena för den oberoende variabeln ändras ligga hos forskaren. Randomiserade designmodeller eller modeller med blindad visuell analys anses uppfylla detta krav, så länge forskaren fattat de slutgiltiga besluten kring procedurerna eller beslutsreglerna för fasövergångar
- Finns bevis för att någon annan än forskaren manipulerade den oberoende variabeln bedöms studien Uppfyller inte WWC-standarder. Om så stannar man här.

Standard 9: Inga kvarstående insatseffekter

- Kvarstående insatseffekter är en potentiell störande faktor i designmodeller med mer än en insats, exempelvis Alternative Treatment Design, eller designmodeller med en mellanliggande tredje betingelse (Withdrawal Reversal Design: ABA, ABAB, ABAC).
- Withdrawal Reversal Design, Multiple Baseline Design och Multiple Probe Design har dock generellt längre faser och längre tid mellan datapunkterna än Alternative Treatment Design. Mer tid passerar mellan de icke-sammanhängande faserna som ska jämföras kan göra kvarstående insatseffekter mindre betydelsefulla, även om de förekommer.
- Risk för kvarstående insatseffekter skall baseras på teoretiska och kontextuella övervägande, inte empiriska data enbart. Dessa överväganden kopplas till specifika insatser, det sammanhang som undersökningen genomförs i, tidpunkten och längden på observationssessionerna, ordningen på insatserna eller sammanhang, samt utfallen.
- Risken för kvarstående insatseffekter skall fastställas som liten eller begränsad. Detta kan göras i samråd med sakkunnig.
- Om sakkunnig och granskare är överens om att kvarstående behandlingseffekter sannolikt förekommer, får resultatet bedömningen Uppfyller inte WWC-standarder och ingen fortsatt bedömning görs.

Standard 10: Krav på forskningsdesign

- Säkerställ att studien är utformad så att den möjliggör minst tre demonstrationer av en insatseffekt över faser vid tre olika tidpunkter
- Fastställ rimlig säkerhet att observerade data är tillräckliga för att fånga viktig information om responsmönster. Responsmönster inkluderar information som medelvärde och variabilitet inom fas, liksom eventuella ökande eller minskande trender.
 - *Withdrawal/Reversal Design (WRD)*: minst fyra faser (ABAB), minst tre men gärna fem datapunkter per fas (i baslinjefas önskvärt med sex datapunkter) At1At1At1(At1At1At1) Bt2Bt2Bt1(Bt2Bt2) At3At3At3(At3At3) Bt4Bt4Bt4(Bt4Bt4)
 - *Multiple Baseline Design (MBD)/Multiple Probe Designs (MPD)*: minst sex faser (AB (tier1), AB(tier2), AB(tier3)), minst tre men gärna fem datapunkter per fas (i baslinjefas är det önskvärt med sex datapunkter). Parametern i exemplen är tre personer där samma insats utvärderas under olika dagar.

Person 1- start kl 10 dag 1

At1p1At1p1At1p1(At1p1At1p1At1p1) Bt2p1Bt2p1Bt2p1(Bt2p1Bt2p1)
At2p1-At3p1At3p1(At3p1At3p1At3p1)

Person2 – start kl 10 dag 2

At1p2At1p2At1p2(At1p2At1p2At1p2) Bt2p2Bt2p2Bt2p2(Bt2p2Bt2p2)
At3p2At3p2At3p2(At3p2At3p2At3p2)

Person 3 - start kl 10 dag 3

At1p3At1p3At1p3(At1p3At1p3At1p3) Bt2p3Bt2p3Bt2p3(Bt2p3Bt2p3)
At3p3At3p3At3p3(At3p3At3p3At3p3)

- *Alternating Treatments Design (ATD)*: minst sex faser (At1Bt2, At3Ct4, At5Dt6), minst fyra men gärna fem datapunkter per fas (exempel en person med olika insatser At1At1At1At1(At1) Bt2-Bt2-Bt2-Bt2(Bt2) At3At3At3At3(At3) Ct4Ct4Ct4Ct4(Ct4) At5At5At5At5(At5) Dt6Dt6Dt6Dt6p1Dt6p1(Dt6Dt6))
- *Changing Criterion Design (CCD)*: minst fyra faser, med minst tre datapunkter per fas. At1At1At1, Bt2k1Bt2k1Bt2k1, Bt3k2Bt3k2Bt3k2, Bt4k3Bt4k3Bt4k3, eller At1At1At1
- Replikationer över olika sammanhang och deltagare
 - För att styrka extern validitet rekommenderas att insats provas med flera personer, eller i olika miljöer.
- Flera beteendemått
 - Om möjligt ska flera typer av utfallsmått användas för att stärka slutsatsernas tillförlitlighet.

Standard 11: Effektmått

- Använd kvantitativa mått (t.ex. "icke-överlappande mått") för att bedöma trender och effekter i data. WWC har valt måttet *nonoverlap of all pairs* (NOP) att bedöma trender i baslinjen och reversibilitet. Detta skall vara minst 0,85. NOP kan anta alla värden mellan 0 (ingen skillnad mellan faser) och 1 (maximal skillnad mellan faser). Högre NOP-värden (närmare 1) indikerar större skillnad och tydligare effekt mellan faserna.

RoBiNT

RoBiNT [2, 3] är utformat för att användas för de flesta SCED-designer och omfattar 15 punkter fördelade på två delfaktorer: intern validitet (7 punkter) och extern validitet/tolkning (8 punkter) (se Avsnitten Intern reliabilitet och Extern validitet). Dessa överensstämmer i delar med WWC:s standarder [1] för SCED primärstudier. Varje punkt i RoBiNT bedöms på en 3-gradig skala (0 = ej uppfyllt, 1 = delvis, 2 = ja/uppfyllt) i stället för enbart ja/nej, vilket ger bättre diskriminering enligt moderna standarder.

RoBiNT-skalan har visat goda psykometriska egenskaper – studier rapporterar hög interbedömarreliabilitet med intraclass-korrelationer ~0,88–0,95 för summaindex, både bland erfarna granskare och tränade nybörjare [2].

Intern validitet

Design

- Designen ska vara experimentell och lämplig för att undersöka (orsaks)samband mellan insats och utfall.
 - Replikationer: Insatseffekter ska observeras vid minst två, men helst tre tidpunkter, vilket kräver tre till sex faser beroende på design (eng. intra-subject replikation).
 - För att design skall fungera, det vill säga identifiera förändring mellan faser, behöver baslinjefas och insatsfasernas respektive innehåll och skillnader tydligt definieras och beskrivas.
 - Withdrawal/Reversal Design (WRD): Minst tre faser (ABA) som ger möjlighet att påvisa förändring vid två tillfällen (At1-Bt2, Bt2-At3) (1 p), men gärna fyra faser (ABAB) som ger möjlighet att påvisa förändring vid tre tillfällen (At1-Bt2, Bt2-At3, At3-Bt4) (2 p).
 - Multiple Baseline Design (MBD) samt Multiple Probe Design (MPD): Minst tre eller fyra faser (tiers) (2*AB) som ger möjlighet att påvisa förändring vid två tillfällen (At1-Bt2, At1-Bt1) (1 p), men gärna minst sex faser (tiers) (3*AB) som ger möjlighet att påvisa förändring vid tre tillfällen (At1-Bt2, At1-Bt2, At1-Bt2) (2 p).
 - Alternating Treatments Design (ATD): Minst tre alternerande sekvenser (exempelvis B, C, D eller B, C, B) som ger möjlighet att påvisa förändring vid två tillfällen (1p), men gärna fyra alternerande sekvenser (B, C, D, E) som ger tre tillfällen att påvisa förändring (2p) (AtBt2, At3Ct4, At5Dt6, At7Et7).
 - Changing Criterion Design (CCD): tre förändringar över kriterier.
 - Rådata ska helst presenteras grafiskt för att tydliggöra om designkriterierna uppfylls.
1. Randomisering:
 - Tidpunkt för fasstart ska randomiseras (WRD, MBD, CCD).
 - Fasordning (WRD, ATD), insatsordning (ATD) ska randomiseras.
 - Om randomisering inte är möjlig, ska systematisk och transparent allokering användas.
 2. Upprepad mätning:
 - Minst tre, men helst fem, datapunkter per fas krävs för tillförlitlig analys.
 4. Blindning av deltagare och insatsgivare:
 - Blindning deltagare (ovanligt i beteendeenriktade insatser) och/eller insatsgivare vad gäller vad som utgör baslinje- respektive insatsfas.
 5. Blindning av observatör vid mätning av utfall:
 - Blindning observatör/person som samlar in data och/eller analytiker vad gäller vilken fas som observeras.
 - Observatör och insatsgivare ska vara olika personer.
 6. Inter-rater agreement (ICC)
 - Mått på konsistens mellan de bedömningar av utfall som görs av de personer som samlar in data (observatörer) skall finnas.
 - Per utfallsmått.
 - Minst två observatörer.

- Beräkning av IRR skall göras för baslinje- respektive insatsfas (behöver inte göras för alla faser, dock skall en av varje skall vara representerad).
- Baseras på minst 20 % av datapunkterna.
- ICC skall vara minst 70 %, men gärna, 80 % i för att säkerställa tillförlitligheten i mätningarna.

7. **Konsekvent genomförande av insats (adherence):**

- Genomförande av insats enligt plan/protokoll, till minst 70, men gärna 80%.
- Skall helst fastställas genom mätning som omfattar minst 20% av datapunkterna.
- Bedömningen av adherence skall göras av annan person än insatsgivaren eller deltagare.
- Bedömningskriterierna skall vara tydliga och rapporterade.

Extern validitet

8. **Baslinjekaraktäristika deltagare:**

- Relevanta deltagar-karaktäristika och/eller tillstånd väl beskrivna. Ska minst innehålla information om ålder, kön, beskrivning av aktuellt tillstånd som insatsen är tänkt att förändra, bedömning av allvarlighetsgrad av tillståndet (baseras på någon form av test).
- Karaktäristika analyseras med avseende på dess eventuella inverkan på det beteende som studeras (om de påverkar exv. förändringsbenägenhet/möjlighet till förändring).

9. **Sammanhang**

- Sammanhang i vilket utvärderingen genomfördes ska vara väl beskrivna.
- Detta gäller både den bredare omgivningen (exv. sjukhus, skola, hemmet) som mer detaljerad beskrivning av var i den omgivningen (exv. lekterapi, klassrum eller egna rummet).

10. **Utfall (beroende variabel):**

- Tydligt beskrivna och operationaliserade.
- Datainsamlings/Mätmetod skall anges, inklusive namngivna instrument och metoder för detta.

11. **Insats och jämförelsealternativ**

- Insatsen innehåll och komponenter ska beskrivas i detalj.
- Procedur för genomförandet beskriven, inkludera hur många gånger (faser) insatsen kommer att utvärderas, hur lång insatsperioden är och frekvens av insatsen.
- Utrustning/manualer som används skall beskrivas.

12. **Rådata:**

- All data skall rapporteras över alla faser och mätpunkter, om inte designen i sig innebär att enbart vissa datapunkter mäts (Multiple Probe Design), eller att undersökningsgruppen är större än 10 personer (då skall alla data samlas in för två till tre av dem). Aggregerade data kan accepteras med sänkt trovärdighet som följd.

13. **Analys av data:**

- Beskriv typ av analys som genomförs (visuell, statistisk).
- Om statistisk analys är genomförd ska det motiveras varför en viss analys genomförts eller inte genomförts.

- Systematisk visuell analys ska ske i enlighet med protokoll. Mönster studeras i form av nivåförändringar, trender, variationer, överlapp mellan olika faser, konsistens.
- Qvasi-statistiska analyser som används för att stödja visuell analys kan också genomföras.
- Alla resultat från analyser skall rapporteras.

14. Replikation:

- Insatseffekter skall replikeras. Detta kan ske genom att flera personer erbjuds samma insats, en insats för en person genomförs i flera olika sammanhang, fler än ett utfall studeras när en insats genomförs, insats genomförs av flera olika insatsutförare mm (eng. systematic inter-subject replikation).
- Bör testas vid två eller tre tillfällen men gärna fler än tre tillfällen.

15. Generaliserbarhet:

- Ett kompletterande utfallsmått, vid sidan av primärfallet, skall identifieras och mätas för att få ytterligare underlag för att förändring skett.
- Det skall fastställas a priori.
- Det kompletterande utfallsmåttet skall helst mätas över alla datapunkter, men kan också mätas enbart före och efter insats.

Referenser

1. What Works Clearinghouse. What Works Clearinghouse Standards Handbook, Version 4.1.: Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Washington, DC: U.S. ; 2020. Available from: <https://ies.ed.gov/ncee/wwc/handbooks>
2. Tate R.L , Perdices M, Rosenkoetter U, Wakim D, Godbee K, Togher L, et al. Revision of a method quality rating scale for single-case experimental designs and N-of-1 trials: The 15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale. Neuropsychological Rehabilitation. 2013;23:619-38. Available from: <https://doi.org/10.1080/09602011.2013.824383>
3. Tate RL, Rosenkoetter, U., Wakim, D., Sigmundsdottir, L., Doubleday, J., Togher, L., McDonald, S., Perdices, M. The Risk of Bias in N-of-1 Trials (RoBiNT) Scale: An Expanded Manual for the Critical Appraisal of Single-Case Reports. Sydney, Australia: The PsycBITE Group, John Walsh Centre for Rehabilitation Research, The University of Sydney, Sydney Medical School Level 9, Kolling Institute of Medical Research, Royal North Shore Hospital, St Leonards NSW 2065 Australia; 2015.