

Arbetsrapport 1: Primärstudier med Single Case Experimental Design

Metodik, analys, risk för bias-bedömning, överförbarhet och rapporteringsstandarder

Innehåll

Arbetsrapport: Single Case Experimental Design	1
Inledning.....	2
Vad är SCED?	2
Designtyper.....	4
Utfall och mätning.....	11
Analys av SCED-data	12
Effektmått.....	14
Risk för bias	15
Överförbarhet	24
Rapporteringsstandarder för SCED	28
Sammanfattande slutsatser	29
Medverkande	29
Referenser	30

Inledning

Denna rapport syftar till en överblick över Single Case Experimental Design (SCED) som forskningsmetod.

Rapporten täcker följande huvudområden:

- Grundläggande principer och designtyper: Definition av SCED, terminologi, samt beskrivning av fyra huvudsakliga designtyper med illustrationer.
- Utfallsmått och analysmetoder: Beskrivning av lämpliga utfallsmått samt visuella och statistiska analysmetoder.
- Risk för bias: Genomgång av olika biasrisker och strategier för att hantera dessa.
- Verktyg för risk för bias-bedömning: Presentation av standardiserade verktyg för att granska potentiella risker för snedvridning.
- Överförbarhet: Diskussion om extern validitet och förutsättningar för generalisering av resultat.
- Rapporteringsstandarder: Riktlinjer för transparent rapportering av SCED-studier.

Vad är SCED?

SCED är en experimentell studiedesign där ett enskilt fall [1] (till exempel en individ) prospektivt studeras över tid genom upprepade mätningar, och där samma fall fungerar som sin egen kontroll. Till skillnad från en traditionell fallstudie – som vanligtvis är en okontrollerad, deskriptiv beskrivning av en individ eller händelse – kännetecknas SCED av ett aktivt införande av en oberoende variabel (en insats eller behandling) under förutbestämda faser för att pröva dess effekt på ett utfall.

SCED kan ses som ett alternativ till stora gruppstudier, antingen som en inledande studie som leder till specifika hypoteser att testa i en gruppstudie, eller som en självständig forskningsstudie för att utvärdera eventuell insatseffektivitet vid heterogena och lågfrekventa tillstånd [2] (se Tabell 1).

Tabell 1 När är SCED lämplig?

När är SCED särskilt lämplig?
För lågfrekventa eller heterogena tillstånd
När individualiserad bedömning är central
Vid klinisk eller praktisk utvärdering av insats för specifik individ
När etiska hänsyn kräver att varje deltagare får insats
För att studera behandlings- och insatseffekter i detalj
Vid hypotesprövning/hypotesgenerering

Den grundläggande analysenheten, det enskilda fallet, kan vara en enskild person, en organisatorisk eller samhällelig enhet, liksom en enskild händelse. Det finns dock metodologiska skillnader mellan att ha en analysenhet som representerar en individ ($n=1$) och exempelvis en samhällelig enhet som en stad ($n=1$). Skillnaderna berör främst den

underliggande datainsamlingen, variabiliteten inom analysenheten, och hur möjliga effekter tolkas. Denna rapport belyser single case-metodologi med individen som analysenhet.

Det finns experimentella och icke-experimentella¹ Single case-studier [3, 4]. *Experimentella* Single case-studier, som denna rapport belyser, baseras på upprepade faser av mätningar och i vissa designer införande respektive borttagande av en insats (exempelvis ABA, ABAB). Det finns olika benämningar på experimentella Single case-studier [1, 4] [3, 6], härfter används benämningen Single Case Experimental Design (SCED) eftersom den tydliggör att det är ett eller ett fåtal (single) fall (case) som studeras, samt att det är en studie där en planerad insats aktivt införs (experimental) under flera faser (minst ABA).

Centrala principer

SCED-design bygger på följande centrala principer:

- Upprepade mätningar: Förändringar i ett specifikt utfall mäts systematiskt före, under och efter, där individen fungerar som sin egen kontroll.
- Datadensitet: Minst tre, men helst fem, datapunkter per fas för att etablera stabila mönster (At1At1At1At1At1Bt2Bt2Bt2Bt2Bt2At3At3At3At3At3Bt5Bt5Bt5Bt5) [7].
- Baslinjestabilitet: Baslinjefasen (At1At1At1At1At1) måste vara stabil innan insats introduceras (Bt2Bt2Bt2Bt2Bt2).
- Experimentell kontroll: en aktiv, medveten och kontrollerad introduktion av en insats genom att forskaren kontrollerar när insatsen införs och eventuellt tas bort.
- Reproducering: Insatseffekter behöver replikeras över tid eller mellan olika kontexter/personer. Förändring, eller avsaknad av sådan, säkerställs genom att studien innehåller flera faser (till exempel ABA, ABAB), och kan påvisas vid minst tre olika fasövergångar (At1Bt2At3Bt4).

Termer och begrepp

Vanliga termer och begrepp på engelska och svenska i SCED-studier presenteras i Tabell 2.

¹ *Icke-experimentella* Single case-studier (3. Epstein L. H and Dallery J, *The Family of Single-Case Experimental Designs*. Harvard Data Science Review, 2022(Special Issue 3), 4. Tate R. L, et al., *The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016 Statement*. Journal of Clinical Epidemiology, 2016. **73**(1): p. 142–52.) består av en bakgrundsmätning av utfallet i fråga (baslinjen, A), följt av insats och mätning av utfall under tiden insatsen pågår (insatsfas, B). Detta är den enklaste studiedesignen (AB) och kallas på engelska “phase change without reversal”. Den skiljer sig från en före- och eftermätningstudie genom att mätning av utfallet sker under tiden insats pågår (4. Ibid.). AB-designen kan vara användbar för att pröva om en insats är genomförbar (5. Manolov R, et al., *Single-case experimental designs: Reflections on conduct and analysis*. Neuropsychological Rehabilitation, 2014.).

Tabell 2 Svenska och engelska termer och begrepp som används i SCED-studier.

Svensk term	Engelsk term	Förklaring
Fas	Phase	En definierad observationsperiod där en deltagares beteende mäts upprepade gånger
Baslinjefas (A)	Baseline phase	Fasen före insats, fungerar som jämförelsegrund
Insatsfas (B, C, D)	Intervention/treatment phase	Aktiva faser där olika insatser införs
Fassekvens	Phase sequence	Ordningen av faser, till exempel A1-B1-A2-B2
Primärt utfall	Target behavior/dependent variable	Det utfall som primärt mäts och utvärderas
Insats	Independent variable/intervention	Den behandling eller insats som utvärderas
Experimentell kontroll	Experimental control	Aktiv, avsiktlig och systematisk manipulation av insats
Intern validitet	Internal validity	Att insatsen faktiskt orsakar förändringar i utfallet
Extern validitet	External validity	Att resultaten kan generaliseras till andra sammanhang
Nivå	Level	Datavärdet (medelvärde/median) inom en fas
Trend	Trend	Lutningen på linje inom en fas
Variabilitet	Variability	Dataspridningen inom en fas

Designtyper

Det finns flera sätt att designa en SCED-primärstudie och olika typer av design svarar på olika forskningsfrågor. Nedan beskrivs fyra sådana; Withdrawal/Reversal Design (WRD), Multiple Baseline Design (MBD), Alternating Treatments Design (ATD) och Changing Criterion Design (CCD) [4]. Gemensamt är att en person studeras under olika faser där baslinjen (A) utgör den utgångspunkt mot vilken uppföljande mätning under och efter introduktion av en insats (B) jämförs, men de varierar i vilken situation de är bäst lämpade. En översikt finns i Tabell 3.

Tabell 3 Översikt designtyper.

Designtyp	Grundstruktur	Bäst lämpad för	Begränsningar
WRD	ABA, ABAB	Reversibla effekter	Ej för permanenta förändringar
MBD/MPD	AB över personer/sammanhang/beteenden	Irreversibla effekter	Kräver flera baslinjer
ATD	Snabbt alternerade insatser	Jämföra insatser	Kräver snabb effekt
CCD	Stegvisa kriterieförändringar	Stegvis förändring	Kräver successiv förändring

WRD (Withdrawal/Reversal Design)

WRD², även kallad införande/tillbakadragande design, används när en insats förväntas vara upptäckbar i närtid och ha en reversibel effekt på utfallet [1] [5]. Designen innebär att man först etablerar en baslinje (A), därefter inför insatsen (B), sedan drar tillbaka insatsen (återgår till A), och i en ABAB-design återinför insatsen en sista gång (B) (Faktaruta 1 och Figur 1). Reproducering (ABAB) anses ge starkare stöd för insatsens eventuella effekt [5].

² N-of-1-Trials är ett exempel på WRD-design [4].

Förutsättningar:

- Insatsens effekt ska vara reversibel
- Effekten ska vara upptäckbar i närtid
- Det ska vara etiskt rimligt att dra tillbaka insatsen

Exempel på tillämpning: En studie undersöker effekten av en uppmärksamhetsträning för en elev med koncentrationssvårigheter. Efter att ha etablerat en stabil baslinje av koncentrationstid (A), införs träningen (B), därefter återgår man till ordinarie undervisning (A), och slutligen återinförs träningen (B). Koncentrationstiden mäts systematiskt genom alla faser.

Styrkor

- Demonstration av möjlig kausalitet genom att visa att utfallet följer införandet/borttagandet av insatsen
- Stärkt intern validitet genom replikation inom samma individ.

Faktaruta 1 WRD.

ABA

Nivåer och trender jämförs mellan baslinje och insatsfaser för att fastställa om förändring är kopplad till insatsen.

En (1) person, en (1) insats utan reproducering

Steg-för-steg-process:

1. Baslinjemätning - minst tre mätningar av primärt utfall (At1At1At1).
2. Stabil baslinje - värdena i baslinjefasen uppvisar stabil nivå.
3. Insats införs - minst tre mätningar av primärt utfall (Bt2Bt2Bt2).
4. Tillbakadragande av insats - återgång till baslinje
5. Baslinjemätning igen - minst tre mätningar av primärt utfall (At3At3At3).

Sammanlagd mätserie: At1At1At1, Bt2Bt2Bt2, At3At3At3

ABAB

Nivåer och trender jämförs mellan baslinje, insatsfaser och reproduktionsfaser för att fastställa om förändring upprepas och är kopplad till insatsen.

En (1) person, en (1) insats med reproducering

Steg-för-steg-process:

1. Baslinjemätning - minst tre mätningar av primärt utfall (At1At1At1).
2. Stabil baslinje - värdena i baslinjefasen uppvisar stabil nivå.
3. Insats införs - minst tre mätningar av primärt utfall (Bt2Bt2Bt2).
4. Tillbakadragande av insats - återgång till baslinje.
5. Baslinjemätning igen - minst tre mätningar av primärt utfall (At3At3At3).
6. Insats införs igen (reproducering) - minst tre mätningar av primärt utfall (Bt4Bt4Bt4).
7. Baslinjemätning igen - minst tre mätningar av primärt utfall (At5At5At5).

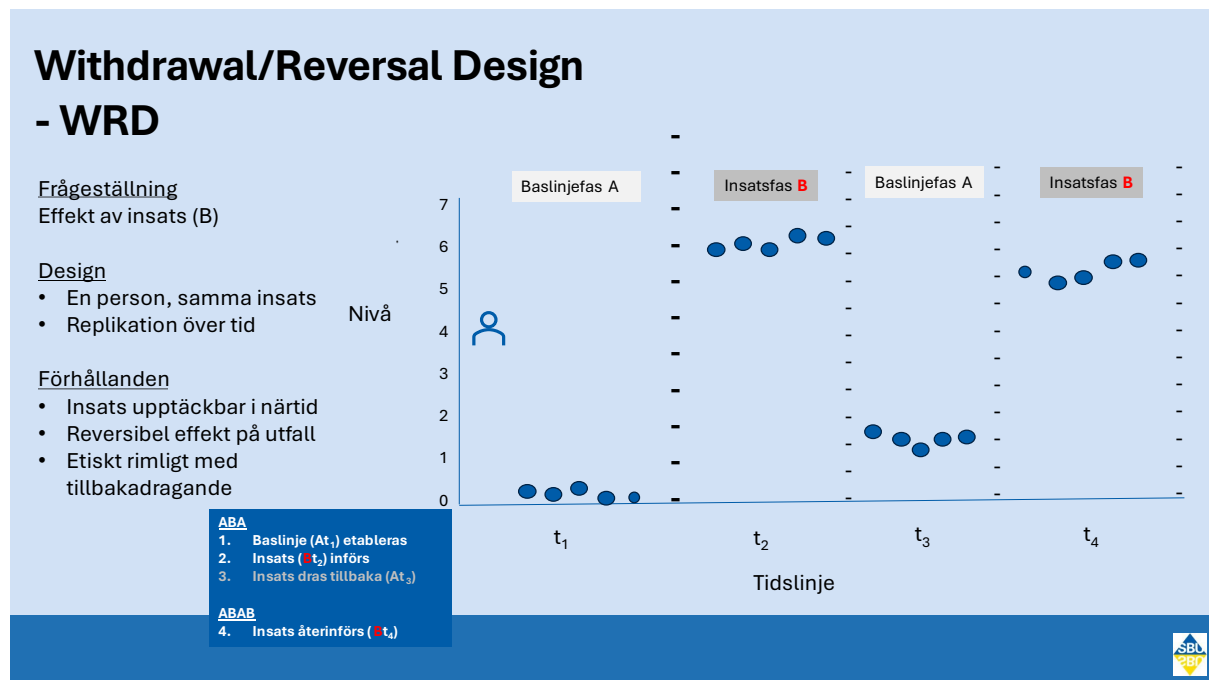
Sammanlagd mätserie: At1At1At1, Bt2Bt2Bt2, At3At3At3, Bt4Bt4Bt4, At5At5At5

Svagheter

- Olämplig för permanenta förändringar
- Etiska överväganden kring att ta bort en effektiv insats
- Kvarstående effekter kan försvåra tolkning av andra A-fasen

I Figur 1 ses en schematisk bild av en WRD studie med ABA/ABAB-design.

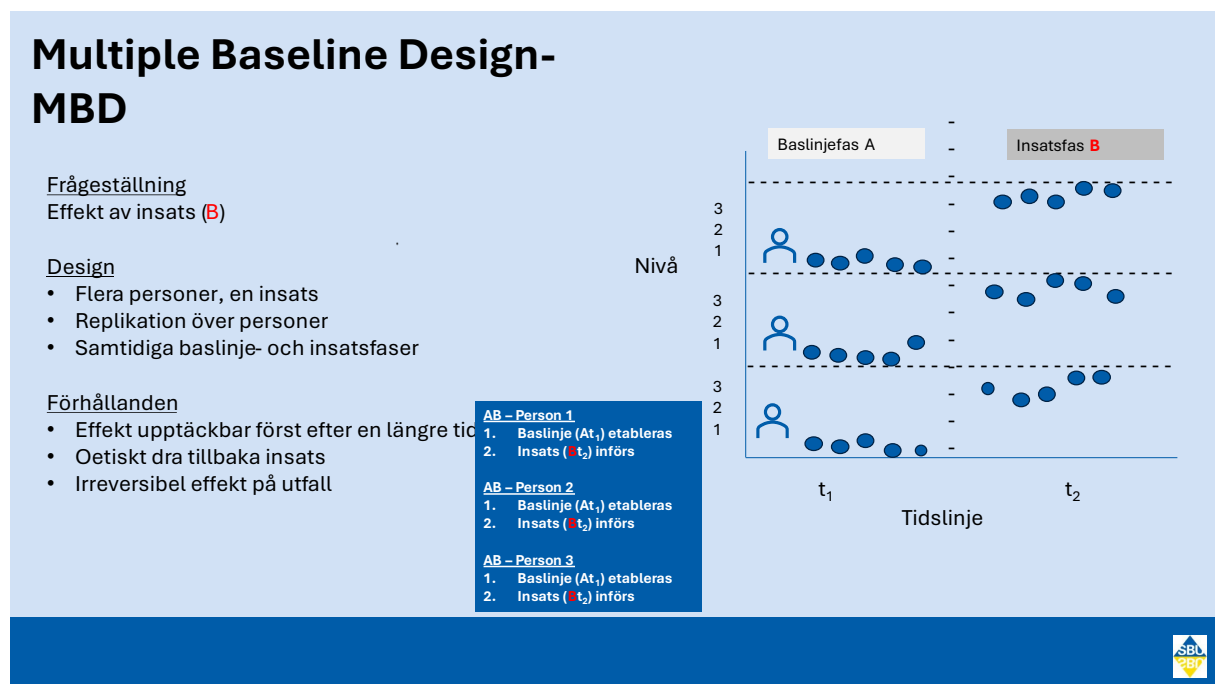
Figur 1 Schematisk bild av WRD.



MBD (Multiple Baseline Design)

När en insatseffekt inte bedöms vara reversibel och där insatsens eventuella effekt på utfallet är synlig först på sikt [1], eller om insatsen inte bedöms som etiskt rimlig att ta bort efter ett första införande kan MBD (svenska. multipel baslinjedesign) vara lämplig [5]. MBD:s flera samtidiga baslinjedesigner (AB, AB, AB) där fler än en person får samma insats eliminerar behovet av att återgå till baslinjen (ABA). Det gör designen lämpad för utvärdering av insatser som inte är reversibla eller har effekter på lång sikt, som till exempel vissa rehabiliteringseffekter (Faktaruta 2 samt Figur 2). Designen avser att påvisa att förändring endast sker när insatsen är riktad till personen, genomförs i ett visst sammanhang eller gäller ett visst eller flera specifika beteenden.

Figur 2 Schematisk bild av MBD.



Faktaruta 2 MBD.

Jämförelse av förändringar mellan olika enheter.

Flera (2 eller fler) personer, beteenden eller sammanhang med sekventiell insats

Sekvens av mätningar:

- Person 1: (At1, Bt2)
- Person 2: (At1, At2, Bt3)
- Person 3: (At1, At2, At3, Bt4)

Steg-för-steg-process:

1. Baslinjemätning - minst tre mätningar av primärt utfall för varje enhet (At1At1At1).
2. Stabil baslinje - värdena i baslinjefasen uppvisar stabil nivå.
3. Insats införs sekventiellt – enhet för enhet (Bt2, Bt3, Bt4), medan de andra fortsätter i baslinje (At2, At3).

Sammanlagd mätserie:

- At1At1At1, Bt2Bt2Bt2
- At1At1At1, At2At2At2, Bt3Bt3Bt3
- At1At1At1, At2At2At2, At3At3At3, Bt4Bt4Bt4

Den mest använda formen av MBD är den med individ, men kan även vara sammanhang och målbeteenden eller liknande, som parameter [1]. Minst tre personer skall ingå i en sådan studie. Ordningen av personer, kan randomiseras, I Faktaruta 2 preciseras detta närmare.

MPD (Multiple Probe Design)

MPD (multipel baslinje design med utvalda mätpunkter) är en variant av MBD, men med färre datapunkter i baslinjen [8]. I stället för kontinuerlig datainsamling mäts utfallet vid specifika "probpunkter" (mätpunkter), inklusive en mätning direkt före insatsens införande. Med hjälp av proberna jämförs vad som händer för de olika deltagarna. Om person 1 förbättras direkt efter

insatsen medan person 2 och 3 fortfarande är på samma nivå (insatsen har ännu inte introducerats) kan insatsen antas göra skillnad (Faktaruta 3).

Exempel på tillämpning: För tre elever samlas data in dag 1, 5 och 10 (i stället för alla dagar 1–10). Insatsen införs vid olika tidpunkter (dag 11 för elev 1, dag 16 för elev 2, dag 21 för elev 3), med en mätning precis före insatsstart för varje elev.

Styrkor:

- Mindre resurskrävande än kontinuerlig datainsamling
- Minskar risken för test effekter genom upprepad mätning
- Behåller fördelarna med Multiple Baseline Design

Faktaruta 3 MPD.

Förändringar relateras till insatsen vid kritiska tidpunkter.

En (1) person, en (1) insats med periodiska mätningar

Steg-för-steg-process:

1. Baslinjemätning - minst tre mätningar av primärt utfall (At1At1At1).
2. Stabil baslinje - värdena i baslinjefasen uppvisar stabil nivå.
3. Insats införs - korta mätperioder efter varje interventionsfas (Bt2, Bt4).

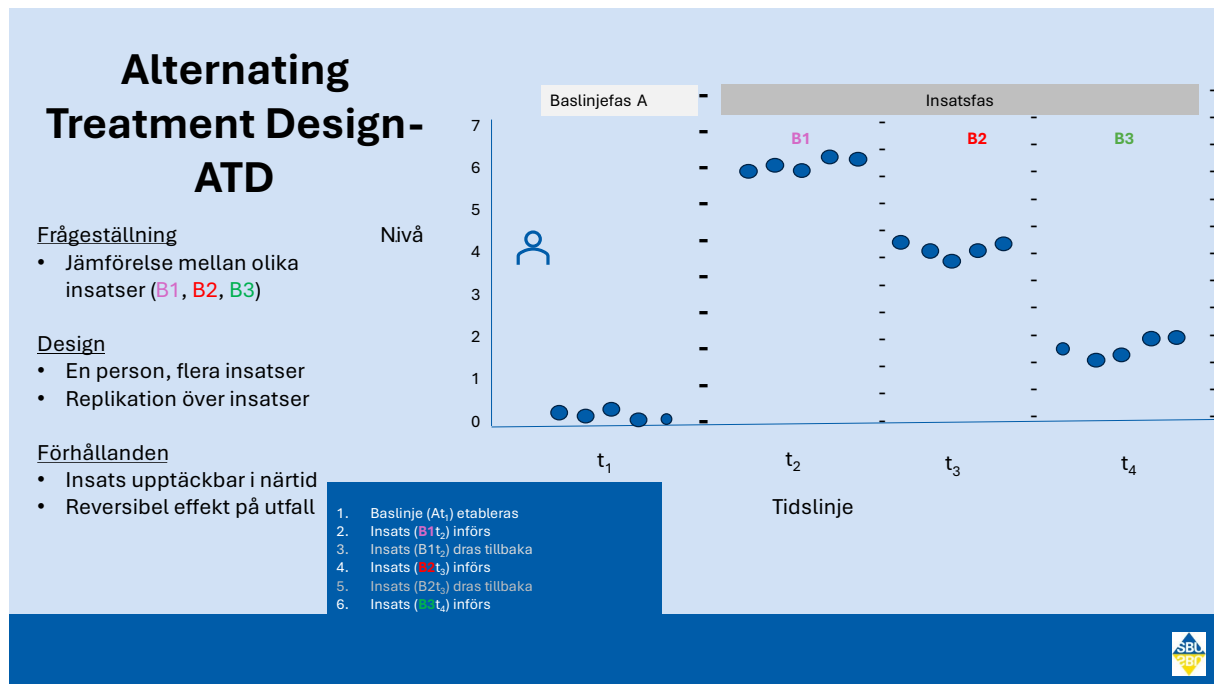
Sammanlagd mätserie:

- At1, Bt2, At3, Bt4, At5

ATD (Alternating Treatments Design)

ATD (alternerande och samtidiga insatser) innebär att flera insatser (B, C, D) utvärderas genom att snabbt alternera dem för samma person. De olika insatsfaserna kan även alterneras med baslinjefaser [1, 4]. Ordningen av insatser kan randomiseras. Se även Figur 3 och Faktaruta 4.

Figur 3 Schematisk bild av ATD.



Förutsättningar:

- Insatseffekten måste vara snabbt synlig [1].
- Effekten måste vara reversibel [5].
- Minimal risk för överföringseffekter mellan insatser

Exempel på tillämpning: En arbetsterapeut jämför tre olika typer av fotleds-ortoser och deras respektive effekt på en persons gångkapacitet vid ett och samma besök, genom att personen får prova varje ortos under korta perioder i slumpmässig ordning.

Styrkor:

- Effektiv för att jämföra olika insatser
- Kräver kortare tid än andra designer
- Kan randomisera ordningen för insatser

Svagheter:

- Risk för överföringseffekter mellan insatser
- Endast lämplig för snabbt verkande insatser

- Kan vara uttröttande för deltagaren

Faktaruta 4 ATD.

Nivåer och trender jämförs för varje insats.

En (1) person, två (2) eller flera insatser som alterneras utan återgång till baslinje.

Steg-för-steg-process:

1. Baslinjemätning - minst tre mätningar av primärt utfall (At1At1At1).
2. Stabil baslinje - värdena i baslinjefasen uppvisar stabil nivå.
3. Insatser alterneras - minst tre mätningar för varje insats (At1Bt2Ct3).

Sammanlagd mätserie:

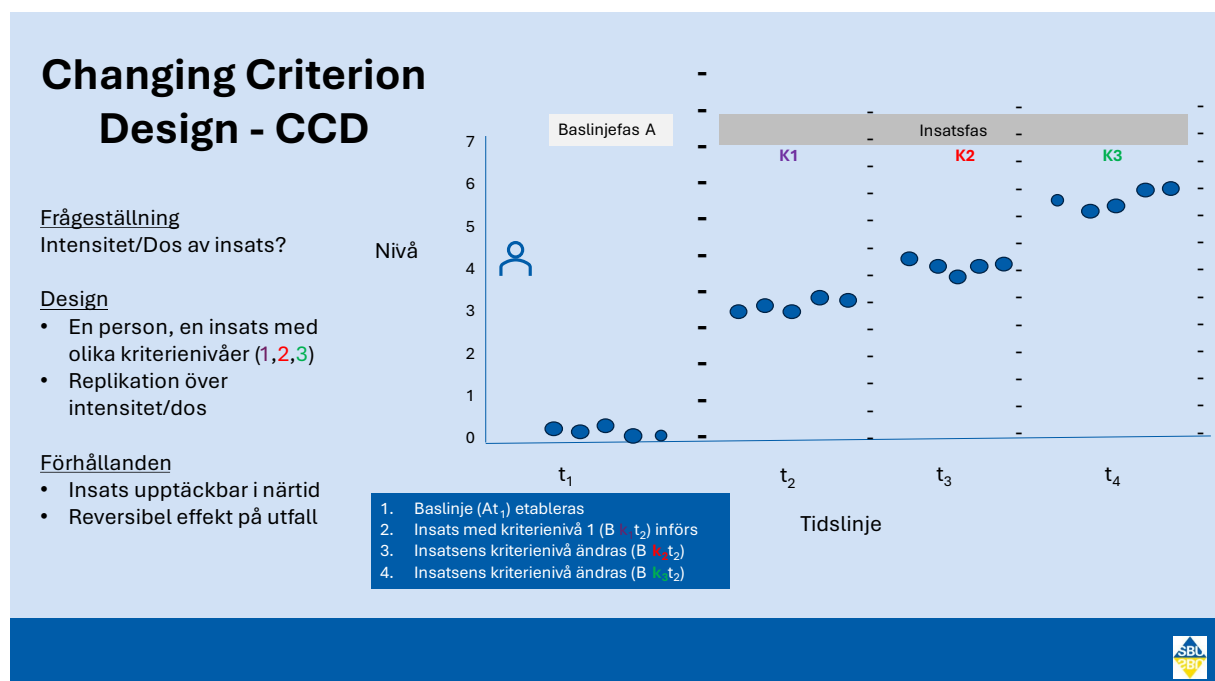
- At1At1At1, Bt2Bt2Bt2, Ct3Ct3Ct3, Dt4Dt4Dt4
- At1At1At1, Bt2Bt2Bt2, At3At3At3, Ct4Ct4Ct4, At5At5At5, Dt6Dt6Dt6, At7At7At7

CCD (Changing Criterion Design)

CCD (förändrad kriterienivå design) utgår från ett antal hierarkiskt baserade kriterienivåer som implementeras på ett sekventiellt sätt. Kriterienivån kan avse olika aspekter av insatsen, som intensitet eller längd. Genom sådana kriterienivåer kan ökning eller minskning av ett beteende mätas. Logiken bygger på att möjlig effekt på utfall varierar i takt med fas-längd eller [9].

Kriteriernas ordning kan randomiseras. Se Figur 4 samt Faktaruta 5.

Figur 4 Schematisk bild av CCD.



Förutsättningar:

- Lämplig för beteenden som kan förändras gradvis
- Kräver att deltagaren kan anpassa sig till förändrade kriterier
- Behöver stabil prestation på varje kriterienivå innan nästa införs

Styrkor:

- Visar kontroll över beteende genom stegvis förändring
- Behöver inte ta bort en effektiv insats
- Användbar för inläring av nya färdigheter

Svagheter:

- Begränsad till beteenden som kan förändras stegvis
- Kräver stabil prestation på varje nivå
- Kan vara tidskrävande

Faktaruta 5 CCD.

Jämförelse av förändringar mellan kriterienivåer.

En (1) person, en (1) insats med successiva kriterieförändringar.

Steg-för-steg-process:

- Baslinjemätning - minst tre mätningar av primärt utfall (At1At1At1).
- Stabil baslinje - värdena i baslinjefasen uppvisar stabil nivå.
- Insats införs med första kriterienivån - minst tre mätningar av primärt utfall (Bt2Bt2Bt2).
- Förändring av kriterium - successivt ändrade kriterier med minst tre mätningar per nivå (Ct3Ct3Ct3, Dt4Dt4Dt4).

Sammanlagd mätserie:

- At1At1At1, Bt2Bt2Bt2, Ct3Ct3Ct3, Dt4Dt4Dt4

Utfall och mätning

Typer av utfallsmått

Det primära utfallet i en SCED-studie, knutet till den person som deltar i utvärderingen [1], handlar ofta om någon form av beteendeförändring.

Huvudtyper av utfallsmått [1]:

- Primärutfall: Det centrala beteende eller utfall som insatsen avser att påverka
- Jämförelsevariabler: Utfall som inte förväntas förändras av insatsen
- Generaliserande variabler: Utfall som inte är primära men kan påverkas indirekt
- Implementeringsdata: Data om metodtrohet och genomförande
- Standardiserade testresultat: Kompletterande standardiserade mått

Krav på utfallsmätning

För att säkerställa tillförlitliga resultat bör utfallsmätning uppfylla följande krav:

- Operationell definition: Tydlig beskrivning av vad som mäts
- Frekventa mätningar: Regelbundna mätningar inom varje fas
- Interbedömarreliabilitet: Minst två oberoende bedömare för minst 20% av mätpunkterna [1, 7]. Samstämmigheten skall uppnå en nivå som satts på förhand, till exempel Cohens κ på minst 0,6 [7].

- Konsistenta metoder: Samma mätmetod används genom alla faser
- Objektivitet: Mätningen bör vara så objektiv som möjligt

Exempel på operationell definition:

Koncentrationsförmåga definieras som antal minuter eleven kan sitta stilla vid sin bänk och arbeta med tilldelad uppgift utan att lämna platsen eller avbryta arbetet i mer än 10 sekunder. Mätningen görs genom direkt observation under 30-minuterspass, tre gånger per vecka.

Faktaruta 6 Huvudkomponenter i visuell analys av rådata.

Inom faser

- Variabilitet/Stabilitet: Grad av fluktuation eller konsistens i data
- Nivå inom fas: Bedömning av nivå inom respektive fas

Mellan faser

- Trend: Systematisk ökning eller minskning mellan faser
- Nivå: Bedömning av förändring exv. i medelvärde mellan faser
- Latens/omedelbarhet: Hur snabbt en förändring uppträder efter fasbyte, där kort latens innebär en omedelbar effekt efter insatsens start
- Överlapp: I vilken utsträckning data från insatsfasen överlappar med baslinjens data
- Konsistens: Hur konsekvent ett förändringsmönster återkommer över flera liknande faser

Analys av SCED-data

Visuell analys är den traditionella metoden för att analysera SCED-data [10] [11] [12]. Den fokuserar på att identifiera systematiska mönster i rådata [13] genom att granska mätpunkter och grafer inom och mellan faser. För att komplettera visuell analys och öka objektiviteten har flera statistiska metoder utvecklats specifikt för SCED-data [7, 10]. Dessa tar hänsyn till de särskilda utmaningar som finns med SCED-data som små urval, klustrade data, autokorrelation och inom-individ-variation [13]. Nedan presenteras komponenter i visuell analys och olika statistiska metoder. Det finns ännu ingen konsensus vilken eller vilka metoder som bör användas [13], utan skall ses som exempel på analysmetoder. För en uttömmande beskrivning hänvisas till [14]. Oavsett analysmetod krävs vanligtvis att en liknande förändring observeras vid minst tre separata tillfällen eller i tre oberoende fall för att en pålitlig effekt ska anses föreligga [11] [15].

Visuell analys

Visuell analys bör göras i förhållande till sex aspekter [7, 15]; nivå, trend, variabilitet/stabilitet, överlapp, latens/omedelbarhet, samt konsistens. Proceduren för visuell analys startar med bedömning av baslinjestabilitet följt av bedömning av varje fas separat med avseende på

trendriktning, datastabilitet/variabilitet och nivå. Därefter jämförs trender, nivåer och överlappning mellan faser (Faktaruta 6). Styrkor med visuell analys är att det är en intuitiv och direkt metod som kan fånga kliniskt/praktiskt relevanta förändringsmönster [10] liksom mönster som statistisk analys kan missa. Svagheter är av naturliga skäl den subjektivitet som bedömning baseras på, vilken inför risk för över- eller underskattningar av faktiskt insatseffekt.

Statistisk analys

Exempel på statistiska metoder presenteras nedan.

Bootstrapping [1] [2] [3]

- Beräkningsmetod för att uppskatta fördelningen av en parameter genom resampling
- Bra vid små urval och när normalfördelning inte kan antas
- Bootstrap ökar reliabiliteten i dataanalys
- Bootstrap-tekniker har anpassats för att öka tillförlitligheten i effekttestimat och konfidensintervall i single-case data
- En reservation är att bootstrap kräver relativt stabil data eller många mätpunkter för att ge robusta resultat

Randomization Test (Permutation Test) [19] [20] [21]

- Testar signifikans genom att slumpmässigt omfördela datapunkter mellan faser
- Relevant i SCED-designer där faser eller insatsordning med flera faktiskt randomiserats
- Kan kombineras med Tau-U för p-värdes bedömning.

Trendanalys

Trendanalys kan handla om att identifiera trender [22], justera för trender, eller utnyttja trendinformation i analysen. Styrkan med trendanalys är att undvika att tillskriva insats en effekt som beror på en redan pågående trend (till exempel en förbättring som hade skett ändå). Svårigheten är att tester för trend i väldigt korta baslinjer har låg styrka och att trend kan vara icke-linjär.

- Split-middle trend line: Grafisk metod för att analysera trendförändringar [23]
- Celeration Line Analysis: Bedömning av förändringar i nivå och trend inom en fas [24].
- Piecewise Regression/Segmenterad Regression: Analyserar förändring i trend före och efter insats [15].

Tidsserieanalyser

- C-statistik: Testar förändringar i nivå och trend mellan faser [25].
- Interrupted Time-Series Analysis (ITSA)
 - Statistisk modellering av nivå- och trendförändringar [26, 27]
 - Analyserar förändringar i seriebunden data
 - Kräver längre tidsserier och möjlighet att kontrollera för externa trender.
- Auto-regressive Integrated Moving Average (ARIMA) [28]
 - Modell för autokorrelation och tidsberoende effekter.
 - Används vid "säsongseffekter" eller starka trender i data.
- Time-Varying Effect Modeling (TVEM) [29]
 - Modellerar hur insatseffekten förändras över tid [30].

- Användbart när effekten inte är konstant utan varierar mellan faser.

TWEM skulle potentiellt kunna anpassas till SCED-data

Multilevel och autokorrelationsmodeller

- Generalized Estimating Equations (GEE). GEE är en metod för att analysera klustrade data (till exempel upprepade mätningar inom en individ). Analyserar förändringar i single-case data och hanterar autokorrelation [31].
- Hierarkisk Regressionsanalys/Multilevel Modeling (MLM): Hanterar hierarkiska data och används för SCED med flera deltagare exempelvis [32].
- Robust Variance Estimation (RVE): Justerar osäkerheter i variansskattningar i hierarkiska modeller [33]. Användbar när aggregade data från flera SCED-studier analyseras.

Bayesianska metoder [34]

Med få observationer kan bayesianska metoder med informativa priorer stabilisera skattningar där frekventistiska metoder fallerar eller ger mycket osäkra resultat.

- Beräknar sannolikheter och osäkerhet i effektstorlekar [2].
- Används vid små urval och när det finns hög osäkerhet i data.
- Tillåter uppskattning av effektstorlek även vid begränsade data.
- Kräver specifikation av prior-fördelningar; beräkningsmässigt komplex
- Prior är ett subjektivt mått, om prior är fel kan resultatet bli snedvridet.

Effektmått

Det finns olika effektmått som kan användas vid statistisk analys av SCED-studiedata. Manolov och kollegor erbjuder en guide för att välja sådana [35].

Mått på procentuell förändring

Dessa mått fokuserar på att mäta förändringen från baslinjen i procent eller relativa skillnader.

- Percent of Goal Obtained (PoGO) [36]: Speglar måluppfyllelse. $PoGO = 100 \rightarrow$ mål uppnått, $>100 \rightarrow$ överskridit, $<0 \rightarrow$ negativ effekt.
- Improvement Rate Difference (IRD) [10, 20] [37]: Skillnad i förbättringsfrekvens mellan baslinje och insats.

Mått på överlappning mellan baslinje och insats

Dessa mått är alla besläktade genom att de kvantifierar i vilken grad insatsfasens data överlappar (eller inte överlappar) med baslinjens data.

För en bred översikt av överlap-baserade effektmått i single-case-forskning se [38].

- Percentage of Non-Overlapping Data (PND) [10] [39]: Mäter andel datapunkter i insatsfasen som överstiger den högsta datapunkten/datavärdet i baslinjen. $PND \geq 90\%$ anses som stark effekt, 70–89% som måttlig, osv. [18]. En nackdel med PND är att det endast använder ett datapunktsvärde som referens och ignorerar resterande variation i baslinjen [40].

- Percentage of Data Points Exceeding the Median Difference (PND-Median/PEM) [10] [17] [18]: Mäter andelen datapunkter i insatsfas som överstiger baslinje-median
- Percentage of All Non-Overlapping Data (PAND) [10] [40]: En förbättrad version av PND där alla datapunkter används, inte bara den högsta från baseline. PAND innebär att man försöker minimera antalet överlappande datapunkter genom att eventuellt att ta bort några datapunkter, och sedan räkna procenten av de datapunkter som inte överlappar.
- Nonoverlap of All Pairs (NAP) [8][41]: NAP är ett senare utvecklat effektmått som anses mer precist än andra. Mäter andelen par av datapunkter, en från baslinjen, en från insatsfas, som är ordnade som förväntat (dvs insats > baslinje om förbättring är uppnådd). Varje insatspunkt jämförs med varje baslinjepunkt och proportionen av dessa par där insatsvärdet är större än baslinjevärdet beräknas [42], mäter andel datapunktspar som inte överlappar mellan faser. Indikerar hur mycket de två faserna skiljer sig åt, men tar inte hänsyn till riktningen av förändringen.

Trendjusterande mått

- Tau (AB): Rangkorrelationsmått mellan baslinje (fas A) och insats (fas B) som visar andelen förbättrade parvisa jämförelser mellan faserna. Värdet varierar från -1 till 1; ett positivt Tau (AB) indikerar förbättring av insats, ett negativt indikerar försämring, och 0 innebär ingen nettoeffekt [43].
- Tau (U): Rangkorrelation för monoton trend inom en enskild fas, oftast använd för att mäta och justera oönskad trend i baslinjen (fas A) innan insats. Värdet varierar från -1 till 1; ett positivt Tau (U) betyder att det finns en uppåtgående trend (förbättring) inom fasen, ett negativt betyder en nedåtgående trend (försämring), medan 0 indikerar ingen trend [44].
- Tau-U: Ett effektstorleksmått som kombinerar förbättringen mellan fas A och B (Tau AB) med eventuell trend under insatsens fas, med möjlighet att korrigera för oönskad baslinjetrend. Tau-U varierar från -1 till 1; positiva värden tyder på en total förbättring av beteendet efter justering för baslinjetrend, negativa värden tyder på total försämring, och 0 innebär ingen nettovariation [37] [43, 44].

Förändring i medelvärden

- Log Response Ratio (LRR) [45]: Log-transformerad proportionell förändring i medelvärde). Positiv = ökning, negativ = minskning, 0 = ingen förändring. Om baslinjedata är nära noll kan logaritmttransformation ge missvisande värden
- Mean Difference (MD): Medelvärdesskillnader mellan faser.
- Standardised Mean Difference (SMD) [46]: Medelvärdesskillnad mellan faser satt i relation till standardavvikelsen. Hedges' g, en korrigerad version av Cohen's d, anpassad för små urval [47].

Risk för bias

SCED-studier, som kännetecknas av inom-individ-jämförelser över tid, ställer specifika krav på hur risk för bias hanteras jämfört med gruppbaseade studier [4] [5] [6] [7]. En översikt av identifierade bias-risker och hur de kan hanteras ses i Tabell 5.

Tabell 4 Huvudkategorier av risk för bias i SCED-studier.

Biasrisk	Typiska problem	Viktiga hanteringsstrategier
Confounding	Mognadseffekter, externa händelser	Stabil baslinje, replikation, randomisering
Förväntningsbias	Observatörsförväntningar	Blindning av utfallsbedömare, objektiva mått
Behandlingsbias	Variation i insatsgenomförande	Standardiserad procedur, fidelitymätning
Bedömningsbias	Inkonsekvent mätning	Standardiserade mätmetoder, interbedömarreliabilitet
Analysbias	Selektiv/subjektiv analys	Förspecificerad analys, blindning, rådata
Rapporteringsbias	Selektiv rapportering	Fullständig rapportering, följa standarder

Confounding

Confounding avser systematiska felkällor som uppstår när en tredje variabel (confounder) påverkar både den oberoende variabeln (exponeringen/insatsen) och den beroende variabeln (mål beteendet/utfallet). Detta leder till en snedvridning av den verkliga sambandsuppskattningen och kan skapa ett skensamband mellan insatsen och utfallet, eller dölja ett verkligt samband [51].

Confounders hör ihop med förändringar inom en individ mellan mättillfällena, samt förändringar mellan mättillfällena i förhållanden runt individen [51]. De kan uppstå både vid baslinjen och över tid. Tidskopplade confounders handlar om externa faktorer som konstant påverkar insatsens effekt (tidsstabil confounding), samt om förändringar i exempelvis deltagarens tillstånd som uppstår under studiens gång (tidsvarierande confounding) [52]. De olika typerna av confounders presenteras nedan inklusive en kortfattad beskrivning av hur de kan hanteras. Det senare utvecklas i avsnittet Strategier för hantering av confounding.

Baslinje-relaterad confounding

Regression mot medelvärdet (Regression to the mean)

- Problem: Om deltagaren valts ut baserat på extrema värden kan förbättring bero på statistisk variation [53].
- Exempel: Ett barn med kraftigt utagerande beteende testas när det är som värst. Om det testas igen en vecka senare kan beteendet ha förbättrats, även utan någon insats.
- Hantering: Lång och stabil baslinje, replikation av effekt.

Instabil baslinjefas

- Problem: Variabel baslinje försvårar bedömning av insatseffekt [14].
- Hantering: Säkerställ stabil baslinje innan insats introduceras, helst minst 5 mätpunkter

Tidsstabil confounding

Egenskaper hos insatsgivare

- Problem: Personliga egenskaper hos insatsgivaren kan påverka resultatet [54].
- Hantering: Dokumentera insatsgivare, samma insatsgivare genom alla faser.

Grad av följsamhet till planerad insats i förhållande till protokoll

- Problem: Skillnader mellan planerat och faktiskt genomförande skapar osäkerhet kring vilken del av proceduren eller insatsen som eventuellt orsakar observerade förändringar [55]. Skillnader kan exempelvis avse hur lång tid fasen pågår, vid vilken tid på dagen en insats påbörjas, eller intensitet av insats.
- Hantering: Dokumentera genomförande, mät följsamhet till protokoll (fidelity).

Tidsvarierande confounding

Maturation/Mognadseffekter

- Problem: Naturlig utveckling över tid misstas för insatseffekt [14].
- Exempel: Språkförbättring hos barn kan bero på naturlig utveckling
- Hantering: Multiple Baseline Design (MBD) med förskjuten start av insats, kontrollmått som inte bör påverkas av insatsen

Externa händelser/Livshändelser som uppstår över tid

- Problem: Andra händelser som inträffar samtidigt med insatsen påverkar utfallet [51].
- Exempel: En elev får en ny beteendeterapi, men samtidigt byter hen klass och får en ny lärare som hanterar beteende annorlunda.
- Hantering: Dokumentera externa händelser, Multiple Baseline Design (MBD), replikation

Test-repetitionseffekter

- Problem: Upprepade mätningar leder till förbättring genom övning [54].
- Exempel: Förbättring på kognitiva tester p.g.a. träningseffekt
- Hantering: Parallella versioner av test, Multiple Baseline Design (MBD), kontrollmått.

Skillnader i insatsprocedur eller innehåll mellan faser/mätpunkter

- Problem: Variationer i hur insatsen genomförs i olika faser [55].
- Exempel: Förändrat sammanhang, såsom platsen där faserna sker (klassrum i en fas och lekplats i en annan) eller om olika personer ger insatsen i olika faser.
- Hantering: Standardiserade protokoll, övervakning av insatsgenomförande

Skillnader i utfallsmätningens procedur eller innehåll mellan faser/mätpunkter

- Problem: Variationer i sättet mäta utfallet över tid, gör det svårt att avgöra om en förändring beror på insatsen eller på ändringen i mätproceduren [56].
- Exempel: Ett barn får en lästräning och testas med ett lättare läs-test i insatsfasen jämfört med baslinjen, förbättringen kan då tillskrivas insatsen fast den egentligen beror på det enklare testet.
- Hantering: Standardiserade protokoll, övervakning av insatsgenomförande.

Kvarstående insatseffekter

- Problem: Effekter från en fas kvarstår in i nästa fas (gäller WRD)
- Exempel: En insats förbättrar impulsivitet under en At1-Bt2-At2-design, men förbättringen kvarstår även i den andra At2-fasen. Då utgör baslinjen inte längre en neutral jämförelsepunkt, vilket gör det svårt att dra slutsatser om insatsens effekt.
- Hantering: Washout-perioder, statistisk kontroll, alternativa designar.

Strategier för att hantera confounding

I SCED primärstudier identifieras och hanteras confounders enligt nedan metoder. De är bland annat baserade på aktuella riktlinjer från What Works Clearinghouse (WWC) [11] och Risk of Bias in N-of-1 Trials (ROBiNT) Scale [48, 50].

Protokoll finns och följs

Ett välutvecklat protokoll som följs noga är grundläggande för att minska risken för bias. Protokollet bör innehålla [13]:

- Kriterier för urval av deltagare, målbeteenden, sammanhang eller insats.
- Detaljerad beskrivning av utfall och hur det ska mätas, det skall tydligt framgå vad som skiljer baslinjefas från insatsfas.
- Tydlig redogörelse för insatsens innehåll och genomförande, om det är fler insatser som ges skall det tydligt gå att särskilja dem emellan.
- Plan för fasernas längd och när fasövergångar ska ske
- Specificering av mätfrekvens och mätmetoder

Baslinjekaraktäristika skall beskrivas

Baslinjekaraktäristika skall noggrant beskrivas för att möjliggöra korrekt tolkning av förändringar över tid [49].

Observation och dokumentation av avsteg eller förändringar från protokoll

Under hela studiens gång bör observation och dokumentation av potentiella avsteg eller förändringar från designkrav och planer, liksom andra förändringar över tid i exempelvis deltagarens miljö, hälsa eller dagliga rutiner göras för att identifiera möjliga confounders [13].

Upprepad mätning och replikation

Oavsett designtyp skall följande kriterier uppfyllas [5] [6] [7].

- Utfallsvariabeln ska ha mätts upprepade gånger inom faser och tillräckligt ofta för att fånga förändring över tid
- Mätningarna ska vara regelbundna och frekventa i varje fas
- Baslinjefas (A) ska visa stabilitet över flera mätningar innan insats introduceras, och minst tre datapunkter krävs för att påbörja bedömning av stabilitet, men fem eller fler datapunkter rekommenderas för att säkrare kunna fastställa att en stabil trend har etablerats (AAAAA) [49].

- Insatsfas (B) ska mätas med minst tre datapunkter (BBB) för att möjliggöra identifiering av en tydlig förändring från baslinjen (A)
- För att fastställa insatsens möjliga effekt ska utfallsförändring observeras över flera faser (replikation) (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).

Ovanstående är centrala krav i litteraturen om SCED [5] [6] [7], och det är avgörande att genomförandet av en studie sker i enlighet med de designkrav som finns i Withdrawal/Reversal-design, Multiple Baseline Design, Alternating Treatments Design, Changing Criterion Design, eller variationer av dessa som Multiple Probe Design.

Randomisering

För att minska risk för bias kan randomisering användas på olika sätt beroende av designtyp (Tabell 5) [11]. Det är viktigt att notera att randomisering i SCED-studier kan ha praktiska begränsningar, särskilt i kliniska eller andra miljöer där etiska hänsyn, praktiska begränsningar eller patientens tillstånd kan göra det svårt att implementera randomisering [57].

Tabell 5 Översikt randomisering per designtyp.

Design typ	Vad kan randomiseras	Vilken bias minskas
WRD	Tidpunkt för fasstart	• Tidsmässig bias • Förväntningsbias
MBD	Ordning av personer/sammanhang/målbeteenden	• Selektionsbias • Förväntningsbias
ATD	Ordning av insatser	• Exponeringsbias • Bedömningsbias
CCD	Ordning/nivå av kriterier	• Tidsmässig bias • Behandlingsbias

Withdrawal/Reversal Design (WRD eller N-of-1)

- Tidpunkten för fasstart kan randomiseras, men inte ordningen av baslinjefas och insatsfas [54].
- Biasområden:
 - Tidsmässig bias: Slumpmässig starttidpunkt minskar risken för att tidsrelaterade faktorer påverkar resultaten.
 - Förväntningsbias: Randomisering av tidpunkter kan minska påverkan från förväntningar om faser.

Multiple Baseline Design (MBD)

- Ordningen av personer, sammanhang (t.ex. miljöer) och målbeteenden kan randomiseras [58].
- Biasområden:
 - Confounding faktorer: Randomisering minskar risken för att externa faktorer som individuella skillnader och miljövariationer påverkar resultaten.
 - Bedömningsbias: Slumpmässig ordning av personer, sammanhang och målbeteenden minskar risken för att bedömare påverkas av förväntningar eller förutfattade meningar om för vilken person, i vilket sammanhang eller för vilket målbeteende en insats är mest effektiv.

Alternating Treatments Design (ATD)

- Ordningen av flera insatser som testas samtidigt kan randomiseras [59].
- Biasområden:
 - Exponeringsbias: Randomisering minskar risken för att en insats alltid testas före en annan och därmed påverkar utfallet på grund av sekvensen.
 - Bedömningsbias: Slumpmässig insatsordning minskar risken för att bedömare påverkas av förväntningar eller förutfattade meningar om vilken insats som är mest effektiv.

Changing Criterion Design (CCD)

- Ordningen av kriterier som längd, grad, styrka eller intensitet/dos av insatsen kan randomiseras [59].
- Biasområden:
 - Tidsmässig bias: Randomisering hjälper till att särskilja insatsens effekt från gradvisa förändringar över tid.
 - Behandlingsbias: Randomisering säkerställer att variationer i dos eller intensitet inte snedvrider resultaten.

Washout-perioder

I det fall det finns risk för kvarstående insatseffekter kan en paus mellan faser läggas in för att säkerställa att effekten från en fas inte "spiller över" till nästa [59].

Analys

Hur har studien hanterat identifierade confounders i analys och tolkning.

- Dokumentation av confounders:
 - Studien bör lista kända potentiella confounders, till exempel komorbiditet, samtidiga insatser, kontextuella förändringar.
- Analysstrategier:

- Visuell analys: kontrollera om författarna kommenterar påverkan från confounders.
- Statistisk analys: I mer avancerade analyser kan confounding hanteras genom kovariat justering, segmenterad regression, etcetera.
- Transparens: Studien bör tydligt redovisa vilka överväganden som gjorts gällande möjliga störfaktorer.

Övriga former av bias och dess hantering

Övriga former av bias handlar om hur studien genomförts.

Förväntningsbias

Blindning

Blindning kan användas för att hantera förväntansbias. Att blinda utvärderad insats för den enskilda deltagaren eller insatsgivaren är svårt, men skulle kunna göras genom att maskera vilken fas som är baslinje- respektive insatsfas [48, 50]. Den person som observerar utfallet (målbetaendet) kan blindas för vilken fas deltagaren befinner sig i genom att exempelvis visa video- eller ljudinspelningar i slumpmässig ordning [59].

- Problem: Förväntningar på insatsens effekt påverkar bedömningen
- Hantering:
 - Blindning av enskilda deltagaren eller insatsgivaren för vilken fas deltagaren befinner sig igenom maskering.
 - Objektiva mätmetoder när möjligt.

Behandlingsbias

Genomförande av insats

Standardiserad procedur

- Problem: Variationer i hur insatsen genomförs.
- Hantering [55]:
 - Tydligt definierade insatskomponenter
 - Träning av insatsgivare
 - Kontroll av följsamhet till protokoll (fidelity)
 - Oberoende observatörer som övervakar insatsgenomförande.

Säkerställande av experimentell kontroll

Insatsproceduren skall omfatta krav på forskarstyrd kontroll (experimentell kontroll) över tidpunkten för införandet av insats, det vill säga forskaren beslutar när insatsfasen ska påbörjas. Även om randomisering kan användas för att bestämma fasbyten, är det vanligare att forskare väljer att inleda insatsfasen baserat på grafisk inspektion av baslinjedata, känt som responsstyrd experimentering [58]. Detta tillvägagångssätt kan dock leda till överskattade

behandlingseffekter på grund av systematisk underskattning av varians om den används utan kompletterande randomiserade procedurer [60] [30].

- Problem: Okontrollerade insatsförändringar.
- Hantering:
 - Forskarstyrd kontroll över tidpunkter för fasbyten
 - Dokumentation av alla avvikelser från protokoll
 - Responsstyrd experimentering endast med tydliga kriterier.

Följsamhet

Två eller fler bedömare skall övervaka och bedöma om insatsen genomförs som planerat för att säkerställa att insatsen har hög procedurtrohet och konsekvent innehåll och därmed minska risken för behandlingsbias [4].

Bedömningsbias

Bortfall

Till skillnad från bortfall i gruppbaseade studiedesigner handlar det i SCED primärstudier huvudsakligen om bortfall av data från hela faser, eller av vissa mätpunkter inom faser, det vill säga dataförlust som uppstår genom att utfallsmätning inte sker enligt plan eller inte följer designkrav [55]. Bortfall knutet till individer kan förekomma (vid $n > 1$), även om det är aktuellt mer sällan. Om mätpunkter eller faser saknas ska detta rapporteras och analyseras för att avgöra om det påverkar resultaten.

Utfallsmätning

Standardiserad mätprocedur

Procedur för utfallsmätning skall vara i förväg preciserad och konsekvent genomförd vid varje måttillfälle [4].

- Problem: Inkonsekvent mätning över tid, mätinstrumentet som förändras över tid, kan gynna personen under vissa mätningar eller faser [56].
- Hantering:
 - Förspecificerade mätprocedurer
 - Samma mätinstrument genom alla faser
 - Träning av bedömare
 - Interbedömarreliabilitet $\geq 80\%$ för minst 20% av mätpunkterna [49]
 - De som observerar eller kodar resultaten kan blindas för vilken fas deltagaren befinner sig i för att undvika förväntansbias [4].

Analysbias

Selektiv rapportering och analys

- Problem: Selektiv rapportering eller analys av data i primärstudierna.

- *Hantering:*
 - Förspecificerad analysplan [4].
 - Presentation av all rådata [56].
 - Blindning av analytiker genom att avkoda information om vilken fas data representerar [59].
 - Kombination av visuell [49] och statistisk analys.
 - Hantering av autokorrelation i statistiska modeller [28]
 - Nyare metoder som Tau-U, piecewise regression och Bayesianska metoder bör övervägas [56].
 - Valda effektmått för att kvantifiera effekten skall vara lämpliga och inkludera konfidensintervall eller liknande osäkerhetsintervall [49].

Rapporteringsbias

Fullständig och transparent rapportering [4].

- *Problem:* Selektiv rapportering av resultat.
- *Hantering:*
 - Rapportera samtliga planerade utfallsmått
 - Presentera data från alla faser och mätpunkter
 - Följa standardiserade rapporteringsriktlinjer (till exempel SCRIBE)
 - Rapportera även negativa eller oväntade resultat.

Verktyg för att granska risk för bias

Även om det saknas ett instrument som anses utgöra huvudstandard vid granskning av risk för bias i primärstudier med SCED [15] har flera mallar och skalor [6] [48, 50] [61] [62] [63] liksom kriteriesystem [7] [11] utvecklats specifikt för detta ändamål. Nedan presenteras What Works Clearinghouse (WWC) standards [7] för SCED-primärstudier samt RobINT [48, 50] (se bilaga 1 för detaljerad beskrivning).

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 interventioners effekter. What Works Clearinghouse (WWC) standards [7] för SCED-primärstudier innehåller metodologiska riktlinjer för genomförande av primärstudier med SCED.

Granskningsprocessen består av tre steg:

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

I steg 2 görs en bedömning i förhållande till standarder som avser utfallsmätning (valida mått, mätreliabilitet, konsekventa mätprocedurer) och förväxlingsfaktorer (förändringar kopplade till insatsgivare, insatsprocedurer/innehåll eller kontextuella förändringar). Utfall avser här observation av beteende.

I det tredje steget listas kriterier som skall vara uppfyllda för att betrakta ett resultat som bevis för potentiell insatseffekt. Dessa handlar om datatillgänglighet (tillgång till rådata), forskarstyrd kontroll över tidpunkt för införande av insats, hantering av eventuella kvarstående insatseffekter, krav på forskningsdesign (minst tre demonstrationer av en insatseffekt över faser vid tre olika tidpunkter, att observerade data är tillräckliga för att fånga viktig information om responsmönster, replikationer över olika sammanhang och deltagare, användning av flera utfallsmått), samt krav på användning av kvantitativa effektmått.

RoBiNT

RoBiNT [16] [48, 50] är utformat för att användas för de flesta SCED-designer och omfattar 15 faktorer fördelade på två delfaktorer:

- intern validitet (7 punkter), och
- extern validitet/tolkning (8 punkter).

Dessa överensstämmer i delar med WWC:s standarderna för SCED primärstudier [7]. Varje punkt i RoBiNT bedöms på en 3-gradig skala (0 = ej uppfyllt, 1 = delvis, 2 = ja/uppfyllt).

Intern validitet handlar om designfaktorer (experimentell design, antal replikationer, tydlig baselinjefas, presentation av rådata), krav på randomisering och upprepad mätning (minst tre, men helst fem, datapunkter per fas), blindning (deltagare, insatsgivare, observatör, analytiker), krav på konsistens mellan bedömningar av utfall som görs av de personer (minst två) som samlar in data (observatörer), samt krav på konsekvent genomförande av insats i enlighet med protokoll.

Extern validitet avser krav på tydlig beskrivning av baslinjekaraktäristika, insats- och jämförelsealternativ och de sammanhang i vilket utvärderingen genomfördes. Utfall skall vara tydligt beskrivna och mätmetoder anges. Rådata skall rapporteras för alla faser och mätpunkter och sättet de analyserades på skall anges och motiveras. Extern validitet omfattar också krav på replikation av insatseffekter (två till tre tillfällen, gärna fler än tre tillfällen), samt kompletterande utfallsmått.

Mall för RoB-granskning

Baserat på genomgången av designtyper, risk för bias, kvalitetsbedömningsverktyg och rapporteringsstandarder presenteras i bilaga 2 en mall för att bedöma risk för bias i SCED-studier.

Överförbarhet

I SCED-studier är intern validitet (fastställande av funktionell relation mellan insats och utfall [14, 64]) ofta huvudfokus, medan extern validitet (överförbarhet av resultat till andra individer

eller sammanhang) är mer utmanande. Klassiska statistiska generaliseringsmetoder som används i gruppstudier är inte direkt tillämpbara på SCED.

Ur klinisk och praktisk synvinkel kan det dock vara av intresse att bedöma i vilken mån ett resultat är tillämpbart i andra kliniska eller praktiska situationer [49].

Överförbarheten av resultat från SCED-studier bygger primärt på replikation [65] och detaljerade beskrivningar av de förhållanden under vilka studien genomförts [4]. Genom att noggrant beskriva deltagare, insats, sammanhang och utfall, samt genomföra systematiska replikationer, kan resultaten från SCED-studier få bredare relevans utöver den specifika studiesituationen. Nyare metodologiska utvecklingar har ytterligare förbättrat möjligheterna att generalisera resultat från dessa studier.

Systematisk replikation som grund för överförbarhet

Extern validitet (överförbarhet/generaliserbarhet) i SCED uppnås genom systematisk replikation [64, 66]. För att resultaten ska kunna generaliseras till andra individer, miljöer, insatser eller målbeteenden krävs att effekt upprepat kan visas inom en studie (intern validitet) och genom upprepade SCED-studier [11]. Överförbarhet är inte en binär egenskap där resultat antingen gäller eller inte gäller för andra deltagare eller sammanhang. I stället bör det ses som en gradient, där graden av generaliserbarhet ökar genom systematisk replikering på olika sätt [51]. Om flera oberoende SCED-studier alla pekar mot en liknande effekt av insatsen ökar tilltron till att fyndet har bredare giltighet och inte är unikt för ett enskilt fall [10].

Typer av replikation genom upprepade studier [59, 66]

- Direkt replikation: Samma forskare upprepar studien med samma deltagare, under samma förhållanden.
- Systematisk replikation: Samma forskare upprepar studien men varierar en aspekt (till exempel annan deltagare, annan miljö, eller modifierad insats).
- Klinisk/Praktisk replikation: Studien upprepas av andra forskare, ofta i en annan miljö eller med modifierad metodik.

Gradvis uppbyggnad av evidens

Generaliserbarhet i SCED är en gradvis process som bygger på en kombination av [64]:

- Antal lyckade replikationer
- Variation i deltagare, miljöer och insatsgivare
- Kvalitet på primärstudierna

- Konsekventa resultat över replikationer.

Exempel på evidensuppbyggnad: "Efter initiala positiva resultat för insatsen med en deltagare (studie 1), replikerades effekten med fem ytterligare deltagare med liknande problematik (studie 2). Därefter visade två oberoende forskargrupper liknande effekt i skolmiljö (studie 3) respektive hemmiljö (studie 4), vilket stärker evidensen för insatsens generaliserbarhet till barn med liknande förutsättningar i olika miljöer.

Faktorer som påverkar överförbarhet

Deltagarfaktorer

Är deltagaren representativ?

För att bedöma om resultaten kan vara relevanta för andra individer behövs detaljerad beskrivning av deltagarkaraktäristika [48] [49].

- Demografiska faktorer: Ålder, kön, och andra relevanta faktorer hos deltagarna [4].
- Kliniska och praktiska faktorer: Beskrivning av tillståndet eller problemet som insatsen syftar till att förändra, inklusive diagnoskriterier när tillämpligt [64], allvarlighetsgrad, bedömd genom validerade testinstrument [11], relevant historik och tidigare insatser som kan påverka resultaten [55], aktuell medicinering eller samtidiga behandlingar [4].
- Individuella faktorer: Kognitiva förutsättningar, specifika styrkor och svårigheter.

Insatsfaktorer

Är insatsen relevant och realistisk för klinisk/praktisk användning?

Överförbarhet av insats till andra kontexter kräver tydlig beskrivning [4] av:

- Insatsens komponenter: Innehåll och komponenter med tydlig operationell definition [64], insatsens teoretiska grund och verkningsmekanism [11].
- Genomförandeaspekter: Procedur för genomförande (antal faser, insatslängd och frekvens) [49], utrustning och manualer som används [48] [59], vem som genomför insatsen och deras utbildning/kompetens [55], grad av anpassning av insatsen till den individuella deltagaren [54].
- Praktisk genomförbarhet: resurskrav, kostnader, tillgänglighet.

Kontextuella faktorer

Är sammanhanget representativt?

För att kunna överföra resultaten till andra sammanhang måste det vara tydligt var studien genomfördes [64]. Sammanhangsbeskrivning bör inkludera både det bredare sammanhanget (exempelvis sjukhus, skola, hemmet, arbetsplats) och specifik miljö inom sammanhanget (exempelvis klassrum, terapi, individuellt samtal) [55].

- Fysisk miljö:
 - Typ av miljö (klinik, skola, hem)
 - Specifika miljöegenskaper.
 - Grad av kontrollerande faktorer i miljön (artificiell vs. naturlig miljö) [59].
- Social miljö:
 - Stöd från omgivningen, exv. närvaro av andra personer under insatsen [49].
 - Kulturella och socioekonomiska aspekter [10].
 - Sociala interaktionsmönster.
- Systemfaktorer:
 - Organisatoriska förutsättningar.
 - Resurstillgång.
 - Policys och riktlinjer.

Utfall

Är utfallsmåttet relevant för den praktiska frågan?

Utfallsmåttets relevans är avgörande för överförbarheten [4].

- Är måttet direkt kopplat till den kliniska eller praktiska frågeställningen? [64].
- Fångar måttet förändring som är meningsfull för deltagaren? (social validitet) [67].

Utfallsmåttets egenskaper

- Validitet och reliabilitet.
- Är måttet tillräckligt känsligt för att upptäcka kliniskt relevanta förändringar? [11].

Är utfallet brett nog för att fånga insatsens fulla effekt?

Ett enstaka snävt utfallsmått kan missa viktiga aspekter av insatsens effekt [51]. För att säkerställa att insatsens fullständiga effekt fångas bör följande övervägas:

- Inkludering av både proximala och distala utfallsmått [59].
- Bedömning av både primära och sekundära effekter [49].
- Användning av flera datakällor (t.ex. direkta observationer, självrapportering, fysiologiska mått) [14].
- Bedömning av generalisering till andra beteenden, miljöer eller situationer [55].
- Uppföljningsmätningar för att bedöma långsiktiga effekter [4].

Nyare utvecklingar inom överförbarhet för SCED

Systematiserad bedömning av generaliserbarhet

Ramverk som specifikt utvecklats för att bedöma överförbarhet i SCED inkluderar:

- The Model of Generalized Outcomes [68]
 - Bedömning av överförbarhet till olika målgrupper och miljöer
 - Strukturerad analys av modererande faktorer.
- Replication Logic Framework [69]
 - Systematisk analys av vilka aspekter som replikerats
 - Bedömning av konsistens över replikationer.
- Contextual Factors Framework [70]
 - Strukturerad bedömning av kontextuella faktorer
 - Analys av hur dessa påverkar överförbarhet.

Rapporteringsstandarder för SCED

SCRIBE (Single-Case Reporting Guidelines in Behavioural Interventions) [4, 13] och CENT-CONSORT Extension for N-of-1 Trials [71] är rapporteringsstandarder särskilt utvecklade för att säkerställa en konsekvent och transparent rapportering av primärstudier med SCED. SCRIBE utvecklades 2016 genom en internationell Delphi-process med experter inom SCED-metodik och är den mest omfattande och citerade rapporteringsstandarden för beteendeorienterade SCED-studier. CENT-CONSORT är en rapporteringsstandard som fokuserar på att förbättra transparens och reproducerbarhet i rapporteringen av så kallade N-of-1-studier som används i medicinska studier och motsvarar WRD. Standarderna finns översiktligt beskrivna i Tabell 6 samt detaljerat i Bilaga 3.

Tabell 6 SCRIBE och CENT.

Aspekt	SCRIBE	CENT
Primärt fokusområde	Beteendeinsatser	Medicinska insatser
Typ av SCED	Alla SCED-typer	Främst N-of-1 trials (WRD)
Randomisering	Betonas när möjligt	Central komponent
Blindning	Beskrivs om tillämplig	Starkt betonad
Carryover-effekter	Behandlas som risk för bias	Specifika analyskrav
Ursprung	Utvecklad från SCED-metodik	Anpassning av CONSORT
Strukturnivå	Medelhög strukturering	Hög strukturering

Sammanfattande slutsatser

Single-Case Experimental Design (SCED) utgör en metodik för att studera effekten av insatser på individnivå. Centralt för SCED-studier är:

- Experimentell kontroll genom systematisk och kontrollerat införande av insats i olika faser
- Upprepad mätning av utfall över tid med tillräcklig datadensitet
- Replikation av effekter inom och mellan studier
- Systematisk hantering av potentiella biasrisker
- Transparent rapportering enligt etablerade standarder.

Genom att kombinera visuell och statistisk analys, samt säkerställa tillräcklig metodologisk stringens, kan SCED generera kunskap värdefull för heterogena tillstånd och individualiserade insatser. Systematisk replikation över personer, sammanhang och insatsgivare stärker den externa validiteten och möjliggör bedömning av för vilka och under vilka omständigheter en insats är verksam.

Styrkor med SCED

- Individualiserad bedömning av insatseffekter, särskilt värdefull för heterogena tillstånd
- Hög intern validitet genom att individen fungerar som sin egen kontroll
- Praktisk genomförbarhet med relativt små resurser jämfört med gruppstudier
- Klinisk/praktisk relevans genom direkt koppling till praktisk verksamhet
- Flexibilitet att anpassa design efter forskningsfråga och kontext
- Detaljerad beteendeanalys genom täta mätningar över tid
- Möjlighet att studera förändringsprocesser, inte bara slutresultat
- Etiska fördelar då alla deltagare får insats (om än vid olika tidpunkter).

Begränsningar med SCED

- Begränsad direkt statistisk generaliserbarhet till bredare populationer
- Kräver stabil baslinje vilket kan vara svårt att uppnå för vissa tillstånd
- Kräver intensiv datainsamling vilket kan vara resurskrävande
- Begränsad blindning kan öka risk för bias
- Krav på replikation för att stärka evidensvärdet.

Medverkande

Susanna Larsson Tholén, projektledare SBU

Jessica Dagerhamn, projektledare SBU

Rebecca Silverstein, projektledare SBU

Maria Ahlberg, projektadministratör

Anna Attergren Granath, projektadministratör

Sakgranskning: SBU:s metodstöd samt SBU:s metodgrupp

Referenser

1. Krasny-Pacini A and Evans J, *Single-case experimental designs (SCEDs) to assess intervention effectiveness in rehabilitation: a practical guide*. . Annals of Physical and Rehabilitation Medicine, 2018. **61**(3): p. pp. 164-179.
2. Miočević, M., et al., *Using Bayesian methods to test mediators of intervention outcomes in single-case experimental designs*. Evidence-Based Communication Assessment and Intervention, 2020. **14**(1-2): p. 52-68.
3. Epstein L. H and Dallery J, *The Family of Single-Case Experimental Designs*. Harvard Data Science Review, 2022(Special Issue 3).
4. Tate R. L., et al., *The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016 Statement*. Journal of Clinical Epidemiology, 2016. **73**(1): p. 142-52.
5. Manolov R, et al., *Single-case experimental designs: Reflections on conduct and analysis*. Neuropsychological Rehabilitation, 2014.
6. Tate RL, et al., *Rating the methodological quality of single-subject designs and n-of-1 trials: introducing the Single-Case Experimental Design (SCED) Scale*. Neuropsychol Rehabil, 2008. **18**(4): p. 385-401.
7. What Works Clearinghouse, *What Works Clearinghouse Standards Handbook, Version 4.1.*, I.o.E.S. Department of Education, National Center for Education Evaluation and Regional Assistance. Washington, DC: U.S. , Editor. 2020, Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Washington, DC: U.S. .
8. Horner, R.D. and D.M. Baer, *Multiple-probe technique: a variation on the multiple baseline*. J Appl Behav Anal, 1978. **11**(1): p. 189-96.
9. Klein L A, et al., *Best Practices in Utilizing the Changing Criterion Design*. Behav Anal Pract, 2015. **10**(1): p. 52-61.
10. Byiers, B.J., J. Reichle, and F.J. Symons, *Single-subject experimental design for evidence-based practice*. Am J Speech Lang Pathol, 2012. **21**(4): p. 397-414.
11. Kratochwill T. R, et al., *Single-case designs technical documentation*. 2010, What Works Clearinghouse: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.
12. Ferron, J. and P. Jones, *Tests for the Visual Analysis of Response-Guided Multiple-Baseline Data*. Journal of Experimental Education - J EXP EDUC, 2006. **75**: p. 66-81.
13. Tate R. L., et al., *The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016: Explanation and elaboration*. Archives of Scientific Psychology, 2016. **4**: p. 10-31.
14. Kazdin, A.E., *Single-case research designs: Methods for clinical and applied settings, 2nd ed*. 2011, New York, NY, US: Oxford University Press. xi, 452-xi, 452.
15. Lobo, M.A., et al., *Single-Case Design, Analysis, and Quality Assessment for Intervention Research*. Journal of Neurologic Physical Therapy, 2017. **41**(3).
16. Borckardt, J.J. and M.R. Nash, *Simulation modelling analysis for small sets of single-subject data collected over time*. Neuropsychol Rehabil, 2014. **24**(3-4): p. 492-506.
17. Parker, R.I., *Increased reliability for single-case research results: is the bootstrap the answer?* Behav Ther, 2006. **37**(4): p. 326-38.
18. Carlin, M.T. and M.S. Costello, *Statistical Decision-Making Accuracies for Some Overlap- and Distance-based Measures for Single-Case Experimental Designs*. Perspect Behav Sci, 2022. **45**(1): p. 187-207.
19. Levin, J.R., Ferron, J. M., Kratochwill, T. R. , *Nonparametric statistical tests for single-case systematic and randomized ABAB...AB and alternating treatment intervention designs: New developments, new directions*. 2012. **50**: p. 599-624.
20. Parker, R.I., K.J. Vannest, and L. Brown, *The improvement rate difference for single-case research*. Exceptional Children, 2009. **75**(2): p. 135-150.

21. Tanious, R. and P. Onghena, *A systematic review of applied single-case research published between 2016 and 2018: Study designs, randomization, data aspects, and data analysis*. Behavior Research Methods, 2021. **53**(4): p. 1371-1384.
22. Manolov, R. and A. Solanas, *Analytical Options for Single-Case Experimental Designs: Review and Application to Brain Impairment*. Brain Impairment, 2018. **19**(1): p. 18-32.
23. Fisher, W.W., M.E. Kelley, and J.E. Lomas, *Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs*. J Appl Behav Anal, 2003. **36**(3): p. 387-406.
24. Gast, D. and A. Spriggs, *Visual Analysis of Graphic Data*. 2009. p. 199-233.
25. Manolov, R. and A. Solanas, *Comparing $N = 1$ effect size indices in presence of autocorrelation*. Behav Modif, 2008. **32**(6): p. 860-75.
26. Bernal, J.L., S. Cummins, and A. Gasparrini, *Interrupted time series regression for the evaluation of public health interventions: a tutorial*. Int J Epidemiol, 2017. **46**(1): p. 348-355.
27. Shadish, W.R., E.N. Kyse, and D.M. Rindskopf, *Analyzing data from single-case designs using multilevel models: new applications and some agenda items for future research*. Psychol Methods, 2013. **18**(3): p. 385-405.
28. Manolov, R. and M. Moeyaert, *How Can Single-Case Data Be Analyzed? Software Resources, Tutorial, and Reflections on Analysis*. Behavior Modification, 2017. **41**(2): p. 179-228.
29. Tan, X., et al., *A time-varying effect model for intensive longitudinal data*. Psychol Methods, 2012. **17**(1): p. 61-77.
30. Swan, D.M., P.J. E., and S.N. and Beretvas, *The impact of response-guided designs on count outcomes in single-case experimental design baselines*. Evidence-Based Communication Assessment and Intervention, 2020. **14**(1-2): p. 82-107.
31. Ballinger, G.A., *Using Generalized Estimating Equations for Longitudinal Data Analysis*. Organizational Research Methods, 2004. **7**(2): p. 127-150.
32. Van den Noortgate, W., Onghena, P., *Combining single-case experimental data using hierarchical linear models*. 2003. **18**: p. 325-346.
33. Pustejovsky, J.E., Hedges, L. V., & Shadish, W. R., *Design-comparable effect sizes in multiple baseline designs: A general modeling framework*. Journal of Educational and Behavioral Statistics, 2014. **39**(5): p. 368-393.
34. Rindskopf, D., *Bayesian analysis of data from single case designs*. Neuropsychol Rehabil, 2014. **24**(3-4): p. 572-89.
35. Manolov, R., M. Moeyaert, and J.E. Fingerhut, *A Priori Justification for Effect Measures in Single-Case Experimental Designs*. Perspect Behav Sci, 2022. **45**(1): p. 153-186.
36. Ferron, J., et al., *Indexing effects in single-case experimental designs by estimating the percent of goal obtained*. Evidence-Based Communication Assessment and Intervention, 2020. **14**(1-2): p. 6-27.
37. Dowdy, A., et al., *Meta-analyses and effect sizes in applied behavior analysis: A review and discussion*. J Appl Behav Anal, 2021. **54**(4): p. 1317-1340.
38. Parker, R.I., K.J. Vannest, and J.L. Davis, *Effect size in single-case research: a review of nine nonoverlap techniques*. Behav Modif, 2011. **35**(4): p. 303-22.
39. Scruggs, T.E., M.A. Mastropieri, and G. Casto, *The quantitative synthesis of single-subject research: Methodology and validation*. RASE: Remedial & Special Education, 1987. **8**(2): p. 24-33.
40. Parker, R.I., S. Hagan-Burke, and K. Vannest, *Percentage of all non-overlapping data (PAND) an alternative to PND*. The Journal of Special Education, 2007. **40**(4): p. 194-204.
41. Parker, R.I. and K. Vannest, *An improved effect size for single-case research: nonoverlap of all pairs*. Behav Ther, 2009. **40**(4): p. 357-67.

42. Chen, M., et al., *Comparing Single Case Design Overlap-Based Effect Size Metrics From Studies Examining Speech Generating Device Interventions*. Am J Intellect Dev Disabil, 2016. **121**(3): p. 169-93.
43. Parker, R.I., et al., *Combining nonoverlap and trend for single-case research: Tau-U*. Behav Ther, 2011. **42**(2): p. 284-99.
44. Lee, J.B. and L.R. Cherney, *Tau-U: A Quantitative Approach for Analysis of Single-Case Experimental Data in Aphasia*. Am J Speech Lang Pathol, 2018. **27**(1s): p. 495-503.
45. Pustejovsky, J.E., *Using response ratios for meta-analyzing single-case designs with behavioral outcomes*. Journal of School Psychology, 2018. **68**: p. 99-112.
46. Shadish, W.R., R.D. M., and L.V. and Hedges, *The state of the science in the meta-analysis of single-case experimental designs*. Evidence-Based Communication Assessment and Intervention, 2008. **2**(3): p. 188-196.
47. Hedges, L.V., J.E. Pustejovsky, and W.R. Shadish, *A standardized mean difference effect size for multiple baseline designs across individuals*. Research Synthesis Methods, 2013. **4**(4): p. 324-341.
48. Tate R.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**: p. 619-638.
49. Kratochwill, T.R., et al., *Single-case intervention research design standards*. Remedial and Special Education, 2013. **34**(1): p. 26-38.
50. Tate, R.L., 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*. 2015, 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.
51. Shadish, W.R., Cook, T. D. , Campbell, D. T., *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. 2002, University of Michigan: Houghton Mifflin.
52. Biglan, A., D. Ary, and A.C. Wagenaar, *The value of interrupted time-series experiments for community intervention research*. Prev Sci, 2000. **1**(1): p. 31-49.
53. Barnett, A.G., J.C. van der Pols, and A.J. Dobson, *Regression to the mean: what it is and how to deal with it*. Int J Epidemiol, 2005. **34**(1): p. 215-20.
54. Gast, D.L., & Ledford, J. R., *Single case research methodology: Applications in special education and behavioral sciences, 2nd ed*, ed. D.L. Gast and J.R. Ledford. 2014, New York, NY, US: Routledge/Taylor & Francis Group. xxi, 415-xxi, 415.
55. Ledford, J.R., Gast, D. L., *Single Case Research Methodology. Applications in Special Education and Behavioral Sciences*. 2018, New York: Routledge.
56. Vannest, K.J., Davis, J. L., Parker, R. I., *Single Case Research in Schools: Practical Guidelines for School-Based Professionals*. 2013, New York: Routledge.
57. Onghena, P. and E.S. Edgington, *Customization of pain treatments: single-case design and analysis*. Clin J Pain, 2005. **21**(1): p. 56-68; discussion 69-72.
58. Ferron, J.M.E.L., Joel R. (Ed), *Single-case permutation and randomization statistical tests: Present status, promising new developments*, in *Single-case intervention research: Methodological and statistical advances.*, J.M. Ferron and J.R. Levin, Editors. 2014, American Psychological Association: Washington, DC, US. p. 153-183.
59. Barlow, D.H., Nock, M., Hersen, M., *Single case experimental designs. strategies for studying behavior for change*. 2009: Pearson/Allyn and Bacon.
60. Kratochwill, T. and J. Levin, *Enhancing the Scientific Credibility of Single-Case Intervention Research: Randomization to the Rescue*. Psychological methods, 2010. **15**: p. 124-44.

61. Reichow B, Barton E. E, and Maggin D. M, *Development and applications of the single-case design risk of bias tool for evaluating single-case design research study reports*. Res Dev Disabil, 2018(79): p. 53-64.
62. Schlosser R. W, et al., *Appraisal of comparative single-case experimental designs for instructional interventions with non-reversible target behaviors: Introducing the CSCEDARS ("Cedars")*. Research in Developmental Disabilities 2018(79): p. 33-52.
63. Wendt O and Miller B, *Quality Appraisal of Single-Subject Experimental Designs: An Overview and Comparison of Different Appraisal Tools*. Education and Treatment of Children, 2012. **35**(2): p. 235-268.
64. Horner, R.H., et al., *The Use of Single-Subject Research to Identify Evidence-Based Practice in Special Education*. Exceptional Children, 2005. **71**(2): p. 165-179.
65. Smith, J.D., *Single-Case Experimental Designs: A Systematic Review of Published Research and Current Standards*. Psychological Methods, 2012.
66. Sidman, M., *Tactics of scientific research: Evaluating experimental data in psychology*. Tactics of scientific research: Evaluating experimental data in psychology. 1960, Oxford, England: Basic Books. x, 428-x, 428.
67. Wolf, M.M., *Social validity: the case for subjective measurement or how applied behavior analysis is finding its heart*. J Appl Behav Anal, 1978. **11**(2): p. 203-14.
68. Petry, K., B. Maes, and C. Vlaskamp, *Operationalizing quality of life for people with profound multiple disabilities: a Delphi study*. J Intellect Disabil Res, 2007. **51**(Pt 5): p. 334-49.
69. Barnett, S.D., et al., *Small N designs for rehabilitation research*. J Rehabil Res Dev, 2012. **49**(1): p. 175-86.
70. Onghena, P., B. Maes, and M. Heyvaert, *Mixed Methods Single Case Research: State of the Art and Future Directions*. Journal of Mixed Methods Research, 2019. **13**: p. 461-480.
71. Vohra S, et al., *CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement*. British Medical Journal, 2015. **350**: p. h1738.

1. Borckardt JJ, Nash MR. Simulation modelling analysis for small sets of single-subject data collected over time. Neuropsychol Rehabil. 2014;24(3-4):492-506. Available from: <https://doi.org/10.1080/09602011.2014.895390>
2. Parker RI. Increased reliability for single-case research results: is the bootstrap the answer? Behav Ther. 2006;37(4):326-38. Available from: <https://doi.org/10.1016/j.beth.2006.01.007>
3. Carlin MT, Costello MS. Statistical Decision-Making Accuracies for Some Overlap- and Distance-based Measures for Single-Case Experimental Designs. Perspect Behav Sci. 2022;45(1):187-207. Available from: <https://doi.org/10.1007/s40614-021-00317-8>
4. Kratochwill T. R, Hitchcock J, Horner R. H, Levin J. R, Odom S. L, Rindskopf D. M, et al. Single-case designs technical documentation. http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf: What Works Clearinghouse; 2010.
5. Tate RL, 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. Neuropsychol Rehabil. 2013;23(5):619-38. Available from: <https://doi.org/10.1080/09602011.2013.824383>
6. Kratochwill TR, Hitchcock JH, Horner RH, Levin JR, Odom SL, Rindskopf DM, et al. Single-Case Intervention Research Design Standards. Remedial and Special Education. 2012;34(1):26-38. Available from: <https://doi.org/10.1177/0741932512452794>
7. 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.