

Kort om validering



Från innebörd till genomförande

Innehåll

- 1 Varför validitet?
- 3 Studiedesigners validitet
- 8 Metoders validitet
- 16 Praktiska råd

Författare

Örjan Ekblom och Gisela Nyberg,
Gymnastik- och Idrottshögskolan

Projektledare

Christina Wahlström och Erik Jannesson,
Skandias stiftelse Idéer för livet

Grafisk formgivare

Tomas Falk, TF Consulting

Ett stort tack till följande för deras värdefulla feedback på innehållet:

Ida Johansson, Rädda barnen
Carina Juserius, Project Playground
Arne Falkbäck, Arez Kader och
Mika Päiväniemi, Våga va' Dig Själ!!!

Om Idéer för livet

Skandias stiftelse Idéer för livet strävar efter att skapa ett samhälle med bättre hälsa och ökad trygghet. Idéer för livet agerar som en katalysator för att skapa fler och effektivare förebyggande insatser mot barn och unga. Det genom att initiera och stödja projekt, metoder och forskning. Men också via utbildning, processtöd och olika typer av böcker och guider gällande effektmätning.

Om Gymnastik- och idrottshögskolan

Gymnastik- och idrottshögskolan (GIH) är Sveriges ledande lärosäte för forskning kring idrott, fysisk aktivitet och hälsa. GIH:s målsättning är att vara en attraktiv samarbetspartner, inom liksom utanför akademien. Ambitionen är att vara en kunskapsaktör i samhällsbygget. Forskning och undervisning sker i nära samverkan med näringsliv och ideella organisationer för att kunna skapa samhällsrelevant kunskap.



Förord

Allt vi på Idéer för livet gör syftar till att skapa bättre förutsättningar för barn och unga. En god hälsa och en stark grundtrygghet bidrar till välmående och framtidstro för nästa generation som växer upp. Många hundratals organisationer bidrar till att genomföra bra verksamheter för många barn och unga, så att de kan utvecklas och tro på sig själva. Men för att veta att de gör en insats som gör skillnad är en viktig pusselbit att hjälpa organisationer att på ett bättre sätt mäta effekterna av sin verksamhet. För att ge rätt insats, till rätt individ i rätt tid är centralt för ett starkare samhälle.

Vår ambition är att bidra till utvecklingen av effektmätning i Sverige. Därför tillhandahåller vi utbildning, processledning och processtöd till framför allt våra partners. Men också till många av de organisationer som vi ger stipendier till, likväl som till andra.

Vi jobbar aktivt för att publicera böcker, guider och handledningar inom effektmätning, både med bredd och på specifika teman. Boken ”Så mäts socialt hållbart värdeskapande” (Studentlitteratur, 2022) är vår utgångspunkt, som kompletteras med guider om bland annat mätning via observation och hur vi lämpligen jobbar med enkäter riktade till barn och unga. Det vi publicerar har sin grund i en uttalad efterfrågan på mer kunskap hos de vi möter i specifika effektmättningsfrågor.

Just det är också grunden för den handledning om validering som du just nu läser. För att vi ska förstå relevansen av resultat från effektmätning är validitet en central faktor, men det behövs mer anpassad och lättillgänglig litteratur kring detta. Därför har vi tagit initiativ till att börja täppa till den luckan, och därmed förse dig som antingen jobbar med effektmätning, eller som har ett intresse av sådan, med innehåll för att förstå mer om validering.

Handledningen är framtagen av Gymnastik- och Idrottshögskolan (GIH), i samarbete med oss på

Idéer för livet. Utgångspunkten är att ge dig en bild av olika typer av validitet, och mått för sådana, tillsammans med tips om vad som är lämpliga åtgärder för stärka validiteten i en studie.

En studies validitet, det vill säga dess giltighet, beror i grunden på både vilken data som samlats in och hur. I handledningen får du som läsare därför ta del av en redogörelse av olika sätt att samla in data, så kallad studiedesign, och dess betydelse för giltigheten. Därefter beskrivs hur kvaliteten på de metoder som använts – som enkäter, intervjuer eller andra mätinstrument – kan påverka en studies giltighet. Avslutningsvis ges några praktiska tips till dig som ska planera och genomföra en studie.

Vår förhoppning är att du med detta ska få en betydligt bättre bild av innebörden av validitet, och dessutom tillgång till verktyg för att själva göra mer medvetna val som stärker validiteten i de mätningar som du kommer göra framöver.

Lycka till!



*Stina Liljekvist,
Ordf. Skandias stiftelse
Idéer för livet,
Chef Kommunikation
och Marknad, Skandia*



*Örjan Ekblom,
Professor, Institutionen för
fysisk aktivitet och hälsa,
Gymnastik- och Idrotts-
högskolan*

Varför validitet?

Tänk dig att du vill göra någon form av studie utifrån vad din verksamhet bidrar till, exempelvis när det gäller fysisk hälsa, lycka eller skolnärvaro. I en sådan är utgångspunkt alltid en eller flera specifika frågeställningar som du vill få svar på. Det kan exempelvis handla om en fråga som: "Har vårt nya närvaroprogram någon påverkan på elevernas skolnärvaro?" Utifrån detta handlar det sedan om att försöka samla information för att kunna besvara varje enskild fråga.

I det arbetet finns en rad övergripande och detaljerade aspekter att beakta, och även fatta beslut utifrån. Genomförandet av studien är dessutom beroende av en rad praktiska, ekonomiska, tidsmässiga, likväl som andra, faktorer. Detta gör att det ofta inte går att genomföra studien på exakt det sättet som vore önskvärt. Det kan exempelvis innebära att det av tidsskäl inte går att involvera så många personer som borde vara med, eller att vi av ekonomiska skäl behöver nöja oss med den näst bästa metoden i sammanhanget. Nedanstående exempel om kopplingen mellan fysisk aktiviteter och välbefinnande illustrerar detta.

Den här typen av utmaningar och vägval gör att i princip alla studier har sina svagheter när det gäller att besvara den aktuella frågeställningen. I grunden handlar detta om vilken giltighet studiens resultat har, exempelvis i vilken utsträckning resultaten går att använda för att uttala sig om hela den grupp som vi vill kunna uttala oss om.

Ett annat ord för resultatens giltighet är *validitet*. Och validitet är i sin tur nära kopplat till ett annat viktigt ord i sammanhanget – *reliabilitet*.

Validitet handlar om både hur väl vi lyckats samla in data, och vilken data vi faktiskt samlar in. Giltigheten handlar alltså om hur giltiga resultaten är för att besvara frågeställningen. Om vi utgår ifrån en skolas mätning av elevernas längd, handlar validiteten bland annat om i vilken utsträckning den mätsticka som används faktiskt mäter just längden, det vill säga att varje angiven centimeter på mätstickan är just en centimeter (och inte någon annan längd).

Reliabilitet handlar om i vilken utsträckning vi mätt saker på ett stabilt sätt, det vill säga hur mycket ett mått varierar, antingen mellan upprepade mätningar eller mellan olika bedömare. Utifrån exemplet med mätningen av elevernas längd handlar reliabiliteten om huruvida flera mätningar av ett barns längd, exempelvis under en och samma dag, ger samma längd (vilket de borde göra givet att barnen inte växer märkbart under en enskild dag). Ger den samma utfall över ett antal sådana mätningar är reliabiliteten hög. Reliabilitet är dessutom en förutsättning för validitet. Det betyder att om en mätning varierar mycket mellan mätningar (när det som ska mätas inte ändras) kan validiteten knappast vara hög.

Exempel: Studie om kopplingen mellan fysisk aktivitet och välbefinnande

I en studie ville några forskare studera om en insats i form av stöd för ökad fysisk aktivitet bidrar till ökat välbefinnande. Till studien rekryterades 600 anställda på ett tjänsteföretag, där dessa delades in i två grupper. En grupp fick träningskort och motiverande samtal, medan den andra gruppen inte fick något stöd alls (det vill säga de utgjorde en kontrollgrupp). Forskarna valde att i sina mätningar använda en deltagarenkät som fylldes i före respektive efter insatsen för att mäta den eventuella förändringen i välbefinnande. Ett mer precist sätt att fånga den förändringen hade varit att använda en psykolog som djupintervjuat deltagarna. Ett sådant tillvägagångssätt valdes dock bort av ekonomiska och tidsmässiga skäl.

När studien startade visade det sig att det var cirka 20 % av de 600 som rekryterats till studien som faktiskt ville delta, varav mer än hälften redan var väldigt fysiskt aktiva. På grund av hög arbetsbelastning var det dock få i den grupp som fick motiverande samtal och träningskort som faktiskt hann nyttja dessa. Resultaten från studien visade på liten effekt av insatsen. För att kunna tolka dessa resultat blev forskarna tvungna att ta hänsyn till det stora bortfallet, och det faktum att väldigt få genomförde insatsen. Slutsatsen blev att vi inte vet om insatsen saknade betydande effekt, eller om resultaten berodde på att så få deltog.

I den här handledningen fokuserar vi på validitet, framför allt genom att ge dig som läsare en bild av olika typer av validitet. Syftet är att öka dina insikter om vad validitet kan vara, så att du kan nyttja dessa insikter när du själv genomför, eller tolkar resultatet av, någon form av studie. Att ha god inblick i giltigheten av en studies resultat är nämligen av vikt för att kunna avgöra, och kommunicera, dess användbarhet.

Som vi redan indikerat finns det utmaningar i att säkerställa en god validitet. Men genom att välja ett lämpligt sätt att organisera sin studie (så kallad studiedesign), samt välja en lämplig metod att samla in data (så kallad datainsamlingsmetod), är det möjligt att medvetet påverka sin studies validitet så att den blir så hög som möjligt givet förutsättningarna.

Studiedesignen är hur du organiserar din datainsamling, det vill säga om du träffar dina deltagare en eller flera gånger, om du observerar spontana händelser, eller om du tränar/behandlar dina deltagare. I studiedesignen ingår även en beskrivning av den aktuella populationen, det vill säga den totala grupp som är föremål för studien. Sättet på vilket du samlar in data kan påverka hur pålitliga dina värden blir (så kallad intern validitet), och hur lätt det blir att generalisera dina fynd till andra situationer eller grupper (så kallad extern validitet).

Datainsamlingsmetod utgörs av instrument, enkät, intervju, observation, testmetod eller ett register som används för att inhämta information. Det finns en stor mängd olika metoder, långt fler än antalet designar, och vi kommer att diskutera hur bra dessa metoder är på att mäta det de ska (metodens validitet). Vi kommer dock uteslutande fokusera på sådana metoder som används för att kvantifiera det som ska mätas, så som omfattning, förekomst eller mängd av något. Här finns en rad framtagna metoder för att mäta distinkta egenskaper som kognition, fysisk aktivitet, medicinska riskmarkörer eller miljöfaktorer.

En viktig fråga här är då vad "lämpliga" val är? Vad som är lämpligt i en viss studie beror mest på hur väl anpassat valet är i relation till den fråga som ska besvaras. Det finns därför inte några generellt sett bra eller dåliga studiedesigner och metoder, utan det är i den enskilda situationen som det finns bättre och sämre val.

Den som exempelvis vill studera hur effektiv en insats i en skola är för att stärka elevers självkänsla, bör kanske undersöka ett mindre antal elever där en mätning görs både före och efter insatsen. Den som i stället vill undersöka vilka faktorer som hänger ihop med en känsla av otrygghet, kanske snarare ska välja att studera många personer vid ett enskilt tillfälle.

En av många fördelar med en god validitet är att det normalt krävs färre deltagare i en studie för att finna ett samband av viss styrka. Omvänt innebär en låg validitet att många fler deltagare behöver inkluderas för att finna samma samband (eller en förändring).

Orsaken är att det med god validitet huvudsakligen menas att det som ska mätas också är det som mäts. Vid lägre validitet kan det lätt bli större "brus" i utfallen, det vill säga annan störande information, vilket gör att det blir svårare att finna samband i materialet.

Resterande del av den här handledningen innefattar tre kapitel. I kapitel 2 behandlar vi studiedesigners validitet genom att fördjupa oss i innebörden av intern och extern validitet, samt mått på sådant. I kapitel 3 lägger vi i stället fokus på metoders validitet, och mer specifikt det som kallas kriterievaliditet, prediktiv validitet och samtida validitet. I det fjärde och avslutande kapitlet presenterar vi några praktiska råd som vi tänker är bra att du får med dig.



Studiedesigners validitet

När en undersökning ska genomföras har den vanligen en av två huvudsakliga typer en frågeställningar. En av dessa handlar om att beskriva ett samband eller ett naturligt förekommande förlopp. Exempel på sådana kan vara om det finns ett samband mellan utbildning och upplevd hälsa hos unga vuxna, eller om de som var fysiskt aktiva som barn mår bättre som vuxna. Vid den här typen av frågeställningar används studiedesigner som är icke-experimentella, det vill säga i själva genomförandet vill vi påverka deltagare så lite som möjligt (primärt inte alls) genom att enbart observera det som förekommer.

Den andra huvudtypen av frågeställningar handlar om huruvida effekter av någon form av insats, behandling eller intervention har uppstått. I sådana fall kan frågorna handla om effekten av ett behandlingsprogram mot självskadebeteende, eller utifrån förändringar i skolans undervisning. I dessa fall har vi ofta en grupp som genomgår behandlingen, eller får insatsen, och en grupp som lever vidare som vanligt. Mätningar görs sedan både före och efter insatsen/behandlingen för att kunna belysa skillnader över tid inom och mellan grupperna.

De två viktigaste aspekterna av validitet kopplat till en studies design är intern respektive extern validitet. De är lite varandras motsatser, där den förstnämnda handlar om ordning och reda, medan den andra handlar om i vilken utsträckning resultat från en studie gäller andra än de som undersökts (det vill säga hur generaliserbara resultaten är). I resterande del av detta avsnitt fördjupar vi oss något i dessa två aspekter av validitet, med betoning på extern validitet.

Intern validitet – ordning och reda

I grunden borde alla former av studier kunna ha god intern validitet, det vill säga att vi har ordning och reda kopplat till datainsamlingen. Eller att vi åtminstone har kunskap om vilken data som blev insamlad på ett korrekt sätt utifrån planen, samt vilken data som samlats in där det finns avvikelser från plan, exempelvis där det blev något tekniskt fel, något misstag skedde eller där det kan ha blivit ett missförstånd. Viktiga typer av frågor att kunna svara på är bland annat:

- Har vi gett en standardiserad och tydlig instruktion till alla som medverkade i datainsamlingen?
- Kan vi redogöra för alla detaljer kring hur data samlades in?

- Vet vi vilken data som gäller före respektive efter insatsen?
- Var instrumentet vi använde kalibrerat (exempelvis en våg), så att det fungerar på ett korrekt sätt?
- Har enkäten vi använde ett sådant språk att misstolkningar är osannolika?
- Är enkäten vi använde översatt till svenska genom en tillförlitlig metod, exempelvis genom att frågor i originalenkäten översätts till svenska, och därefter från svenska tillbaka till originalspråket (så kallad back translation).

Intern validitet handlar alltså om att ha bra kontroll på sina metoder och det material som samlats in, allt från att maskiner och instrument är väl underhållna, enkäter och intervjuguiden tydliga, till att det framgår varifrån data kommer och hur den samlats in, vem som gjort vad och vilka eventuella misstag som kan ha uppstått. Med sådan information blir insamlad data transparent, och resultaten störs i mindre utsträckning av metodfel och avvikelser. Det är lätt för en utomstående att se vad som gjorts, och för en ny person enklare att komma in i studien och delta i datainsamlingen.

Genom att genomföra så kallade pilottester, det vill säga små förstudier, kan vi lära oss mycket om hur deltagare uppfattar undersökningen, samt om alla små detaljer som är så viktiga för att vi ska kunna lita på informationen som samlas in. Gör därför gärna några sådana pilottester innan du startar en studie för att testa rutiner och för att lära dig mer om hur andra uppfattar dina frågor eller mätningar. Samtidigt lär du dig om hur du säkert och korrekt kan samla in data. Vi återkommer till denna typ av pilottester i kapitel 4.

När studier genomförs som innefattar många deltagare, där deltagarna exempelvis fyller i enkäter via webben eller där de ska rapportera själva under längre tid, är det svårt att bibehålla den interna validiteten. Det blir enklare när det handlar om en mindre grupp där du som person genomför alla mätningar själv. Men viktigast är så klart att välja en volym deltagare som är lämplig utifrån vad du vill åstadkomma med studien.

Extern validitet – generaliserbarhet

De allra flesta studier görs för att kunna vägleda andra. Om en insats för ökad trygghet ledde till framgång i en bostadsort vill vi ofta kunna säga något om hur sannolikt det är att liknande resultat skulle uppkomma vid samma insats på en annan ort.

Det kallas *generalisering*. Men det finns ofta begränsningar i en studies generalisering. En sådan begränsning är att själva genomförandet av en studie tenderar att påverka de människor som deltar. Deltagare kan mycket väl agera eller reagera annorlunda utifrån det faktum att de medverkar i studien, det vill säga att de är "under luppen". Den externa validiteten (generaliserbarheten) påverkas helt enkelt av att deltagarna vet att de är med i en studie. När vi efter en studie försöker generalisera resultaten måste vi därför vara medveten om detta och anpassa våra tolkningar därefter.

Det kan dessutom finnas begränsningar i hur lika de undersökta deltagarna är i förhållande till de som vi i framtiden vill kunna uttala oss om. Exempelvis i vilken utsträckning det finns en matchande andel gällande kön, mående, utbildningsnivå, inkomst och liknande. I vissa fall spelar detta mindre roll, men det kan vara en viktig faktor. Insatsens karaktär, och vilka som inkluderats, avgör hur generaliserbar en studie är. Det är därför viktigt att resonera sig fram till en ståndpunkt gällande generaliserbarheten i varje enskilt fall.

En speciell del av validitetsbegreppet är kopplad till de som arbetar med själva studien. Det är exempelvis viktigt att vi undviker att låta egna förhoppningar och förväntningar skina igenom vid kontakten med deltagare. Om en viss insats i skolan för bättre mående presenteras som fantastisk, och omges av en allmänt positiv inställning, är det rimligt att tro att det kan påverka deltagarnas uppfattningar om insatsen (och därmed också hur de väljer att agera). Något som därmed påverkar generaliserbarheten.

En avslutande potentiell begränsning i generaliserbarheten som vi vill lyfta fram här är den kontext, eller situation, som en studie är genomförd i. En insats som visat sig effektiv i en viss kontext kan inte utan vidare överföras till en annan kontext och ge liknande effekt. Exempelvis kan en insats i en resursstark skola inte nödvändigtvis överföras till en mindre resursstark skola, om delar av insatsens effekt beror på att en rad andra miljöfaktorer behöver finnas på plats. Här kan vi ta en insats med utomhuspedagogik som exempel, där den sannolikt är olika framgångsrik beroende på hur den fysiska miljön kring den aktuella skolan ser ut.

En studies validitet påverkas också delvis av den aktuella studiedesignen. Som vi nämnt tidigare finns det huvudsakligen två olika design, experimentella och icke-experimentella. I de två följande avsnitten resonerar vi utifrån ett validitetsperspektiv kopplat till dessa två design.

Experimentella studier – mäta effekt av en insats

Experimentella studier handlar om att mäta effekt av en specifik insats, exempelvis om ett program för ungdomar med fokus på ökat välmående leder till en minskad psykisk ohälsa, eller om en föräldrautbildning leder till en minskad skolfrånvaro för relaterade barn. Denna typ av studier ger goda möjligheter att beskriva en eventuell effekt, det vill säga i vilken utsträckning insatsen verkligen skapar förändring.

Idealt har deltagarna i den här typen av studie delats upp i en grupp som får insatsen, och en kontrollgrupp som inte får insatsen (eller får vänta lite). Samma mätningar genomförs sedan för båda grupperna för att se om det är någon skillnad i utfallet för deltagarna. På så sätt försöker vi skapa en situation som ger oss möjlighet att isolera effekten av en specifik insats. Om grupperna är lika i de viktigaste avseendena, och inte har för mycket kontakt (eller känner till) varandras deltagande, har studien goda möjligheter att lyckas med det.

Det finns ett flertal saker som är värda att lyfta fram när det gäller den externa validiteten kopplat till experimentella studier. En sådan handlar om att försöka undvika att en del av de uppmätta effekterna av insatser beror på deltagarnas förväntningar. Det allra bästa är därför att inte låta deltagarna själva få välja om de ska ingå i insats- eller kontrollgruppen. Framför allt då det minskar risken att deltagarna väljer utifrån vad de själva gillar, och att resultaten därmed påverkas av deras förväntningar (eller motsatsen). Randomisering, det vill säga slumpmässig indelning i grupperna, är därför en lämplig del i en experimentell studie.

Ett annan viktig aspekt handlar om att den experimentella studiedesignen (det finns lite variationer på dem) förvisso ofta låter väldigt imponerande och "vetenskapliga", men den har en stor svaghet. Deltagarna vet ofta om att de är med i studien, och beter sig därför inte så som de brukar göra. Ett illustrativt exempel här är studier om människors matvanor, där det kan antas att matvanorna påverkas av att deltagarna vet att deras val av mat och dryck registreras noga. Det kan exempelvis innebära att de väljer lite nyttigare kost än de normalt sett gör, eller att de under kortare perioder väljer att "stå ut" med en viss kost, fast de normalt sett inte skulle göra det. Konsekvensen blir att vi måste ställa oss frågan om hur överförbara, eller externt valida, resultaten från en sådan studien är. Den externa validiteten kan av samma skäl bli ännu svagare för studier som genomförs med deltagare i ett laboratorium, i och med att den situationen kan anses vara än mer konstlad. Något som mycket väl kan få tydligare konsekvenser för deltagarnas beteenden.

En tredje sak att lyfta fram utgår ifrån att alla som är med i inledningen av studien vanligen inte fullföljer den. Det kan bero på många saker, där vissa är slumpmässiga (sjukdom eller flytt), medan andra är systematiska och mer kopplade till själva studien. Det kan exempelvis handla om att deltagare inte tycker om att vara med på grund av att det är ansträngande eller tråkigt. Om det i en sådan situation till exempel är väldigt många män som väljer att hoppa av studien, blir det kanske så få kvar att vi inte kan mäta någon effekt för dessa. Och dessutom kan det vara så att de som stannar kvar har egenskaper som gör att de som grupp skiljer sig för mycket jämfört med de som avbröt. Dessa bortfall, om de är systematiska på det viset, kan påverka den externa validiteten.

Självklart finns det många exempel på experimentella studier som genomförts där den externa validiteten kan anses vara god. Nedan är ett exempel på detta vid valideringen av en insats för att öka nyanlända flyktingars trygghet.

Icke-experimentella studier – kartlägga samband

När vi gör en icke-experimentell studie (ibland kallad observationsstudie) lägger vi fokus på aspekter som hur vanligt något är, eller vad som normalt sett händer i vissa situationer. En viktig utgångspunkt är att försöka minimera påverkan på deltagarna. Det kan handla om studier som kartlägger om det är vanligare med psykisk ohälsa hos barn till ensamstående, jämfört med barn till sammanboende föräldrar, eller hur den fysiska aktiviteten är hos ungdomar med olika grad av upplevd trygghet.

Det som skiljer icke-experimentella studier från de experimentella är att alltså att det i de förstnämnda inte finns någon insats- och kontrollgrupp, utan studierna fokuserar på att undersöka deltagarna vid ett eller flera tillfällen. Syftet är därmed inte att studera effekter, utan att ta en ögonblicksbild av hur olika egenskaper och förhållanden är relaterade till varandra.

Exempel: Validering av insats för att öka nyanlända flyktingars trygghet

Ett exempel på en experimentell studie där de lyckats med validiteten utgår från en insats för ökad trygghet hos nyanlända barn och föräldrar i ett visst område. Mer specifikt innefattade insatsen tolv fysiska träffar med information, samt studiebesök.

För att rekrytera deltagare satte forskarna upp anslag i området, både vid mataffärer och vid vårdcentraler. Detta ledde till att totalt 48 familjer anmälde sig. Dessa slumpades sedan in i en insatsgrupp och en kontrollgrupp. Den förstnämnda innefattade 25 familjer, medan de resterande 23 familjerna utgjorde kontrollgruppen. Randomiseringen skedde med hjälp av en slumpgenerator i Excel, och det visade sig att upplevelse av trygghet var mycket lika mellan grupperna, och grupperna var snarlika i en rad andra viktiga avseenden.

Innan de tolv träffarna genomfördes svarade alla deltagare (både i insats- och kontrollgrupp) på samma frågor om trygghet, levnadsvanor och bakgrund. Dessutom deltog samtliga i gruppintervjuer. Forskarna var på förhand tydliga med vilket mått på trygghet som skulle användas som huvudmått, och hur data skulle analyseras. Under studien bad de deltagarna i insatsgruppen att inte diskutera studien med de som fanns i kontrollgruppen. Under genomförandet noterar de noga vilka familjer som deltog på träffarna och studiebesöken, och om något oväntat hände.

Studien visade sig vara mer omfattande än planerat, och det krävdes att ytterligare två personer anställdes för att arbeta med genomförandet. Eftersom det fanns en noggrann beskrivning av hur träffarna och mätningarna skulle gå till, var det enkelt för dessa nya personer att börja arbeta. Samtidigt påverkades data inte av vem som genomförde träffar, intervjuer och annan datainsamling. All data fördes in i tydliga protokoll som sparades systematiskt.

Efter de tolv träffarna bad de deltagarna i båda grupperna att svara på samma frågor som tidigare. Genom att jämföra svaren i för- respektive eftermätningen för de två grupperna kunde forskarna skapa sig en bild av huruvida insatsen varit effektiva eller inte när det gäller upplevd trygghet.

Eftersom de noga antecknat hur genomförandet gått till, och vilka som kunnat delta på alla träffar, kunde de dessutom undersöka om det fanns skillnad i effekt mellan de i insatsgruppen som närvarat på alla träffar och de som missat några. De visste också vem som samlat in vilken data och kunde kontrollera informationen om något avvikande resultat dykt upp i de protokoll som sparats.

Viktigt att poängtera här är att en väl genomförd studie som denna inte är någon garanti för att den visar att den aktuella insatsen är effektiv, det vill säga att den skapar effekt. Det kan ju helt enkelt vara så att insatsen i sig är för svag för att bidra till förändring. Att genomföra en studie med god extern och intern validitet handlar därmed om att skapa förutsättningar för att få en tydlig bild av insatsens faktiska betydelse. I många fall kan vi också få en uppfattning om varför en insats är, eller inte är, effektiv så att eventuella förbättringar kan göras.

Den externa validiteten blir ofta starkare i sådana studier, eftersom vi i mindre utsträckning försätter deltagarna i en konstlad situation. Visst kan det tänkas att svaren till del kan bero på att deltagare kan känna en förväntan att svara på visst sätt, men det kan till viss del hanteras i de olika analyserna. Om en studie genomförs av exempelvis Gymnastik- och idrottshögskolan, kanske många deltagare vill visa att de är lite mer fysiskt aktiva än de verkligen är. Vilket gör att medelvärdena för aktivitetsnivå kan bli lite för höga. Men genom att undersöka skillnader mellan grupper, snarare än absoluta nivåer, kan vi ofta få en bra bild av hur aktiviteten varierar inom gruppen. Antalet deltagare i denna typ av studier är ofta större än i de experimentella studierna, vilket gör att resultaten ger en bild av en större del av hela den grupp vi vill kunna uttala oss om. Något som stärker den externa validiteten.

**Mått på intern och extern validitet
– vad är alternativen?**

Det finns få, eller inga, bra mått på intern eller extern validitet. I stället är det viktigt att resonera sig fram till en slutsats kring grad av validitet efter att studien genomförts. För att ha ett bra underlag för detta är det viktigt att vara uppmärksam på avvikelser i genomförandet som påverkar validiteten.

Ett sätt att få en uppfattning om den interna och externa validiteten är att jämföra sina ursprungliga planer med hur genomförandet blev i praktiken. Kunde vi inkludera lika många individer som planerat, och var det få som tackade nej? Det blir i så fall ett argument för att vi har en lämplig extern validitet (givet att planen i sig är bra i sammanhanget). Var det dessutom få som avbröt sin medverkan i studien är det ytterligare ett argument för att generaliserbarheten är god. Hade vi i stället svårt att få deltagare, och vet vi baserat på observationer eller andra erfarenheter att de som deltog nog var särskilt friska, eller skiljde sig på annat sätt från den grupp vi vill undersöka, kan vi anta att generaliserbarheten är lite lägre. Har vi inte möjlighet att spåra varifrån data i studien kommer, samt vem som lämnat svaren, finns det risk för att delar av resultaten kan vara påverkade av misstag som inte kan rättas till. Då kan den interna validiteten vara hotad. Det första steget är alltså att avgöra hur genomförandet avviker från det tänkta. Det andra steget är att fundera på dessa avvikelserns konsekvenser för resultaten. Men kom ihåg att det är viktigt att hitta en bra balans mellan att säga att "allt gick bra" och att vara alltför självkritisk.



Sammanfattning – Studiedesigners validitet

Intern validitet

- Intern validitet handlar om i vilken utsträckning det är ordning och reda kopplat till vår datainsamling.
- Den påverkas av om vi bland annat har maskiner och instrument som är väl underhållna, enkäter och intervjuguiden som är tydliga, samt att det framgår varifrån data kommer och hur den samlats in, vem som gjort vad och vilka eventuella misstag som kan ha uppstått.
- Testa gärna genomförandet av datainsamlingen för att kontrollera att det fungerar som planerat.
- Ju fler som deltar i undersökningen, desto större utmaning att säkerställa en god intern validitet.

Extern validitet

- Extern validitet handlar om generalisering, det vill säga hur sannolikt det är att resultatet av en insats i en kontext ger samma resultat om insatsen genomförs i en annan kontext.
- Det finns flera saker som minskar generaliserbarheten, bland annat det här:
 - Att personer som vet att de är med i en studie har en tendens att bete sig på andra sätt än de vanligen gör.
 - Att det finns betydande skillnader mellan den grupp som medverkar i studien och den grupp som vi vill kunna uttala oss om.
 - Att de som genomför studien har tydliga förhoppningar och förväntningar som skiner igenom vid kontakten med deltagare, och därmed sannolikt påverkar deltagarnas agerande.
 - Att viktiga förutsättningar för förändring finns i den kontext som studien genomförs, men inte i den som resultaten ska generaliseras till.

Experimentella och icke-experimentella studier

- Den externa validiteten är viktig vid både både experimentella och icke-experimentella studier.
- Experimentella studier används för att mäta effekt av en specifik insats där både finns en grupp individer som deltagit i en insats och en kontrollgrupp som inte gjort det (gärna baserat på slumpen).
- Icke-experimentella studier, så kallade observationsstudier, fokuserar på att antingen fånga ögonblicksbilder av hur olika egenskaper och förhållanden är relaterade till varandra, eller följa ett antal personer över tid för att kunna studera vad som naturligt sker hos dem.
- Utmaningar för den externa validiteten i experimentella studier är att deltagarna ofta vet om att de är med i studien och därmed agerar annorlunda än vad de vanligen gör, samt att avhopp kan leda till att det blir för stora skillnader mellan insats- och kontrollgruppen, vilket gör resultaten osäkra.
- I icke-experimentella studier blir den externa validiteten ofta starkare eftersom vi i mindre utsträckning försätter deltagarna i en konstlad situation (även om det inte behöver vara så); många gånger vet deltagarna inte ens om att de deltar i studien. Dessutom har dessa studier ofta en större volym av deltagare som vi kan uttala oss om, vilket kan ge ett mer träffsäkert resultat.

Mått på intern och extern validitet

- Det finns få, eller inga, bra mått på intern och extern validitet.
- Därför är det viktigt att resonera sig fram till en slutsats kring grad av validitet efter att studien genomförts, där en central utgångspunkt är de avvikelser i genomförandet som påverkar validiteten.
- Ett sätt att få en uppfattning om den interna och externa validiteten är att jämföra sina ursprungliga planer med hur genomförandet blev i praktiken.

Metoders validitet

Hittills har vi utgått ifrån innebörden av intern och extern validitet utifrån ett studiedesignsperspektiv. Det här tredje kapitlet kommer vi i stället viga åt metoders validitet, vilket är något annat än de två förstnämnda typerna av validitet. Med metoder avses exempelvis den enkät, det instrument eller det observationsschema som vi använder i en studie för att samla in data. Dessa metoders validitet handlar i enkla ordalag om i vilken utsträckning de mäter det som ska mätas.

För att sätta in detta i en något större kontext är följande stegmodell en bra utgångspunkt:



Den *företeelse* vi vill mäta utgörs av det som är i fokus i en studie, exempelvis psykisk hälsa, trygghet, depression, kost, fysisk aktivitet, kroppsvikt, självkänsla, framtidstro, kunskap eller demokrati. För att mäta företeelsen behöver ett, eller ofta flera, *mått* tas fram (även kallat indikator). Det handlar om mått som hjälper oss att fånga företeelsen. För företeelsen självkänsla kan måttet exempelvis utgöras av en totalpoäng som deltagarna får utifrån en självskattning av sin självkänsla. Avslutningsvis handlar *mätinstrument* om att göra företeelsen mätbar genom att bestämma vad i den aktuella företeelsen som ska vara i fokus. Ett exempel för självkänsla är Rosenbergs mätinstrument som utgörs av en enkät med totalt tio frågor (och en standardiserad fyrgradig skala) som beaktar olika nyanser av självkänsla.

Den här processen, från företeelse att mäta till ett konkret mätinstrument (exempelvis med specificerade frågor och skalor), kallas *operationalisering*. I studier där vi vill mäta andelar, nivåer eller andra kvantitativa mått är detta ibland en ganska enkel process. Framför allt då måtten ofta är mer specifika och lättare att finna lämpliga mätmetoder för. I studier som vill mäta mer svårfångade företeelser som psykisk hälsa, trygghet och självkänsla kan operationaliseringen vara svårare, eftersom det finns många sätt att fånga sådana företeelser på. Oavsett handlar metoders validitet alltså om i vilken utsträckning mätinstrumentet faktiskt mäter den aktuella företeelsen.

Alla metoder har en inbyggd osäkerhet kopplat till detta, som gör att dess utfall inte fullt ut speglar det vi vill mäta. Det är därför viktigt att känna till dessa osäkerheter, och att kunna förstå hur de kan

påverka resultat och de slutsatser som kan dras. Det kan därför finnas fördelar med att använda mätinstrument som sedan tidigare testats avseende validitet. Framför allt då det gör det möjligt att på ett resurseffektivt sätt beskriva för andra hur validiteten är, men också då det gör det möjligt att på ett bra sätt jämföra sina resultat med resultat från andra studier där samma mätinstrument använts. Vanligt är dock att använda mätinstrument bestående av flera enkäter sammanfogade med egna frågor, och ibland behöver vi konstruera helt egna frågor för att verkligen undersöka det vi vill. I sådana fall är det viktigt att beskriva vilka delar som är egenutvecklade, och vilka delar som tidigare använts eller testats. Baserat på det kan ett resonemang om den sammantagna validiteten föras.

En viktig utgångspunkt är att vald metod ska passa studiens syfte och frågeställning. På så sätt finns det generellt sett inga bra eller dåliga metoder, utan bara metoder som passar mer eller mindre bra i en specifik studie. Exempelvis finns det många enkäter som mäter psykisk hälsa, där dessa är ungefär lika bra. Men språket och antalet frågor i en specifik enkät kan göra att den passar utmärkt för den tänkta gruppen i en specifik studie, och blir då ett bra metodval. Ett annat exempel är att det finns separata stegräknare som är bättre än stegräknarna i en telefon, men då en stegräknande telefonapp är tillgänglig för många kan den utgöra en bra metod i fall då det är viktigare att nå många individer än att mäta det exakta antalet steg.

Viktigt att poängtera är att metoders validitet kan diskuteras utifrån olika perspektiv, framför allt i form av att olika typer av validitet belyser olika saker. Här följer därför tre avsnitt som behandlar tre sådana former av validitet – kriterievaliditet, prediktiv validitet och samtida validitet.

Kriterievaliditet – grad av träffsäkerhet

I de flesta studier tvingas vi använda metoder som inte är de allra mest exakta. Det beror ofta på ekonomi, praktiska möjligheter eller belastningen på deltagarna. I stället för dyra och exakta rörelsemätare för att mäta fysisk aktivitet, skulle vi kunna använda enkätfrågor eller enklare stegräknare. Resultaten av en sådan studien kan då bli lite annorlunda jämfört med det bästa alternativet, och det gör att vi måste veta mer om hur mycket den enklare ersättningsmetoden avviker från den mest exakta. Det går i allra högsta grad få fram viktig och relevant information med andra mer oprecisa metoder, men det är då viktigt att känna till den valda metodens begränsningar.

Ett sätt att undersöka en metods träffsäkerhet är att jämföra dess resultat med resultat från bästa möjliga metod, kallad *kriteriemetod* eller golden standard. En metods överensstämmelse med den bästa metoden kallas *kriterievaliditet*. Viktigt är att en sådan jämförelse görs som innebär att metoderna används tillräckligt nära varandra i tiden för att det som mäts inte hinner ändra sig. Men inte så nära att de två testerna påverkar varandra (för då mäter dem inte exakt det vi vill att de ska mäta, utan påverkas av andra faktorer). I vissa fall kan de två testen göras samtidigt. Oavsett är det intressanta att undersöka hur resultaten från de två metoderna skiljer sig från varandra, även om de helst ska ge samma resultat. Om ersättningsmetoden ger resultat som är väldigt lika de som kriteriemetoden ger, kan vi säga att ersättningsmetoden har hög kriterievaliditet. Är det större skillnader är kriterievaliditeten lägre. Ett exempel på en studie som synliggör den här typen av validitet är följande:

Exempel: Kriterievaliditet för cykeltest

Uppfinnarna av ett nytt cykeltest ville ta reda på metodens kriterievaliditet gällande individers kondition. De valde därför att jämföra resultatet av cykeltestet med kriteriemetoden, som innebar mätning av individers maximal syreupptagningsförmåga i ett laboratorium. I det sistnämnda fallet fick deltagarna springa på ett löpband till maximal förmåga, samtidigt som syreupptagningen (konditionen) mättes mycket noggrant med speciell metodik. Cykeltestet å andra sidan gjordes på deltagarnas arbetsplats, genom att deltagarna fick cykla en viss tid samtidigt som deras puls noterades vid två tillfällen under testet. Konditionen beräknas sedan med hjälp av en formel.

Resultatet från de två metoderna visade att cykeltestet inte är ett perfekt sätt att fånga konditionen; resultatet avviker lite från det "rätta". Men skillnaden mellan individernas två värden var liten, så sambandet (eller korrelationen som det ofta kallas) mellan resultaten är ganska högt. Vilket innebär att det enklare cykeltestet på ett helt okej sätt fångar individers kondition. Kriterievaliditeten är därmed rätt god. För att sammanfatta så är det alltså så att ju mindre avvikelser det finns i resultaten från kriterie- respektive ersättningsmetoden, desto högre är kriterievaliditeten. Det vill säga en hög kriterievaliditet innebär att det knappt är någon skillnad i resultatet från de två metoderna.

I studier som mäter företeelser som självskattade hälsa, grad av stress, eller självkänsla används ofta enkäter som metod, där dessa validerats

mot kriteriemetoden psykologintervju eller annan mer ingående psykiatrisk undersökning. Mätning av fysisk aktivitet, matvanor, alkoholbruk och tobaksvanor görs också ofta med enkätinstrument som ersättningsmått för mer specifika mätningar. De sistnämnda kan handla om att fysisk aktivitet mäts med rörelsemätare, matvanor med hjälp av kostregistrering, samt alkohol- och tobaksbruk via blodprover som testar för leverskada respektive nedbrytningsprodukter av nikotin.

Resultaten från denna typ av tester har ofta en ganska hög korrelation (det vill säga ett ganska starkt samband), men korrelationer är ibland svåra att överföra till någon praktisk nytta. Ett cykeltests resultat kan visa ett starkt samband till resultaten från kriteriemetoden, vilket rent statistiskt innebär att vi exempelvis kan läsa att det finns ett samband på 0.8 – där 1.0 är högst möjliga värde, och -1.0 är det lägsta. Värdet på skalan mellan 1.0 och -1.0 kallas korrelationskoefficient.

En korrelationskoefficient som är nära 1.0 betyder i detta fall att både kriteriemetodens test och cykeltestet visar höga värden för en enskild individ, och vice versa. Men att det finns en del skillnader. Att kunna visa på ett sånt samband kan ju kännas betryggande, men lite svårt att använda praktiskt. Ett mer tillämpbart mått är hur mycket kriteriemetoden och ersättningsmetodens resultat skiljer sig uttryckt i de enheter vi är intresserade av. Ett sådant mått gör att det till exempel går att säga: "cykeltestet visade att personen hade ett konditionsvärde på 45, men eftersom vi använt ett enklare test är det korrekta värdet någonstans mellan 42 och 48."

För de flesta metoder finns det redan genomförda mätningar av kriterievaliditeten, vilket ju är till stor hjälp då det gör att vi inte behöver testa metoden i varje studie. Det går i dessa fall att läsa sig till vilka enkäter, instrument eller andra metoder som lämpar sig för en speciell grupp och hur mycket olika metoder skiljer sig från kriteriemetoden.

Ofta rapporteras kriterievaliditeten i statistiska mått (så som i exemplet med cykeltestet), där dessa beskriver hur stora skillnaderna är mellan metoderna. Sådana mått belyser både hur lika metoderna är (med någon form av mått på överensstämmelse, till exempel en korrelation) och hur stor skillnaden mellan metoderna är (med något mått på hur stort medelfelet är). Det finns en uppsjö sådana mått, där den här handledningen inte rymmer en meningsfull genomgång av dessa.

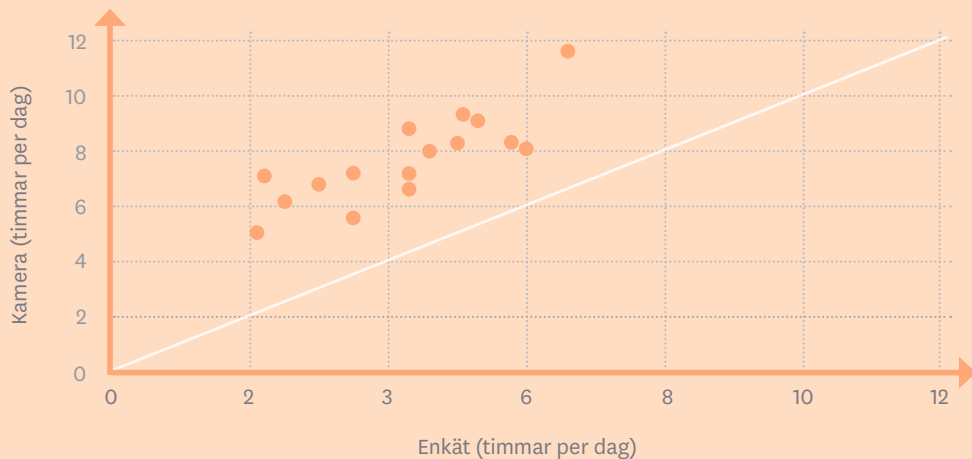
Värt att nämna gällande korrelationers styrka är dock att värden under 0.3 brukar klassas som låga, 0.3–0.8 klassas som "måttliga" och värden över 0.8 som höga. Men för att kunna tolka och använda informationen om en metods kriterievaliditet krävs

ofta att vi läser mer om just den metodens kvalitet och validitet, och att vi förstår mer om kontexten för att kunna avgöra vad som är god respektive

mindre god validitet. Vi avslutar detta avsnitt med ytterligare två exempel på studier gällande kriterievaliditet.

Exempel: Kriterievaliditet i studie om stillasittande

I en stor studie av hur lång tid individer sitter ner användes en enkät. De svarande uppgav hur många timmar och minuter som de satt stilla under en dag. Enkäten hade tidigare testats för kriterievaliditet, genom att en mindre grupp både svarat på enkätfrågorna (ersättningsmetod) och registrerats avseende verklig tid sittande med direkta observationer via en kamera (kriteriemetod). När resultaten från den mindre gruppens enkätsvar jämfördes med mätarnas registreringar blev resultatet det som presenteras i diagrammet nedan.



På x-axeln finns svaren från enkäten och på y-axeln resultaten via kameran. Varje punkt i diagrammet utgör resultatet från en individ. Om punkten ligger på det diagonala strecket är värdena från de två metoderna lika. Om de ligger över strecket är resultat från kameran högre än det självskattade. I detta exempel har samtliga deltagare en mer omfattande tid sittandes via kameran, jämfört med enkätsvaren.

Detta innebär alltså att deltagarna underskattar sitt sittande. Men, det finns en korrelation mellan resultaten; de som skattar sig högt gällande det egna sittandet är också de som sitter mest (och vice versa). I detta fall kan enkäten sägas ha låg kriterievaliditet avseende förmåga att mäta verklig tid sittande, men hyfsad kriterievaliditet när det gäller att identifiera de som sitter mest.



Exempel: Kriterievaliditet i studie om psykisk hälsa

I en studie av unga vuxnas psykiska hälsa ville de ansvariga kunna identifiera andelen som sannolikt lever med mild till måttlig depression. Studien var så omfattande att en enkätmetod var det enda rimliga. För att välja enkät gjordes först en studie där två olika enkäters kriterievaliditet jämfördes. Studien gick till så att en mindre grupp om 54 individer som genomgått en psykologisk undersökning fick svara på två enkäter om psykisk hälsa. De två enkäterna var olika långa och de ”depressionspoäng” som kunde räknas ut var därför helt olika. Båda enkäterna hade dock gränser för vad som kan klassas som mild till måttlig depression. För att få ett mått på hur väl de två enkäterna kunde identifiera personer som led av just mild till måttlig depression, jämfördes resultatet från de två enkäterna med det från psykologundersökningen. Tabellen nedan visar studiens resultat.

		Enkät A		Enkät B	
		Frisk	Mild till måttlig depression	Frisk	Mild till måttlig depression
Psykolog-intervju	Frisk	21	0	14	11
	Mild till måttlig depression	2	31	9	20

Raderna visar antalet individer som under psykologintervjuer klassats som friska respektive har mild till måttlig depression. Kolumnerna visar samma sak för enkät A och B. Gröna siffror indikerar att metoderna ger samma klassning gällande depression, medan röda siffror innebär att enkät och psykologintervju resulterat i olika klassningar.

Av tabellen framgår att enkät A identifierade alla som deprimerade (31 grön). Alla dessa var också deprimerade enligt psykologintervjun (0 röd). Av de 23 personer som enkät A identifierade som friska (21 grön + 2 röd), var det enligt psykologintervjun 2 som i realiteten hade mild till måttlig depression. Enkät B hade däremot avsevärt lägre överensstämmelse mellan enkät och psykologintervju. Exempelvis framgår det att av de totalt 31 individer som enkät B identifierade som deprimerade, var det bara 20 som psykologintervjun identifierade som det. Övriga 11 bedömdes av psykologen som friska. Eftersom resultatet visade att kriterievaliditeten för enkät A var betydligt högre än för enkät B, användes enkät A i den stora studien om unga vuxnas psykiska hälsa.

Sammanfattning – Kriterievaliditet

- Kriterievaliditet handlar om i vilken utsträckning resultatet via en viss metod motsvarar resultatet från bästa möjliga metoden. Den förstnämnda kallas ersättningsmetod och den senare kriteriemetoden (golden standard).
- Om ersättningsmetoden ger resultat som är väldigt lika de som kriteriemetoden ger, kan vi säga att ersättningsmetoden har hög kriterievaliditet (och vice versa).
- Det är en fördel att använda metoder som redan är testade för kriterievaliditeten, för då vet vi redan hur väl resultaten matchar kriteriemetodens.
- Sambandet mellan ersättningsmetoden och kriteriemetoden uttrycks ofta med hjälp av en korrelationskoefficient, där högsta möjliga värde är 1.0 och lägsta möjliga -1.0. Men korrelationen kan ibland vara svår att överföra till någon praktisk nytta.
- Ett mer tillämpligt mått är hur mycket resultatet från ersättnings- och kriteriemetoden resultat skiljer sig uttryckt i de enheter vi är intresserade av. Exempelvis: ”cykeltestet visade att personen hade ett konditionsvärde på 45, men eftersom vi använt ett enklare test är det korrekta värdet någonstans mellan 42 och 48”.

Prediktiv validitet – sambandet med andra företeelser

När vi genomför en studie fokuserar vi generellt på en eller flera aspekter som vi vill mäta, där vi som sagt väljer en lämplig metod för mätning givet omständigheterna i studien. Exempelvis en enkät om fysisk aktivitet med en hyfsad kriterievaliditet. I studien är det då just grad av fysisk aktivitet som är av intresse. Men ibland är vi inte bara intresserade av det, utan även om det finns någon koppling till helt andra saker. Svaren på den aktuella aktivitetsenkäten kanske är relaterat till en persons risk att bli sjuk, chans att få ett arbete, eller klara en utbildning. Givet att vi har tillgång till information om dessa andra företeelser kan vi med hjälp av en statistisk analys få en uppfattning om det finns några sådana samband. Finns det sådana är det i sig av stor relevans, även om vi inte alls vet eller förstår varför sambandet finns.

Det är faktiskt till och med så att även om kriterievaliditeten av aktivitetsenkäten skulle vara låg, det vill säga den mäter inte den fysiska aktiviteten särskilt bra, kan informationen om sambanden ändå vara värdefull. Men då för ett annat resultat än det om fysisk aktivitet. Den här

typen av samband mellan olika parametrar kallas *prediktiv validitet*, och betyder helt enkelt att ett utfall hänger ihop med ett annat utfall, och att vi därmed fortsättningsvis inte behöver mäta det sistnämnda (men kan uttala oss om det ändå). Ett exempel på prediktiv validitet är högskoleprovet, där höga poäng inte direkt kan kopplas till förmåga att lära sig detaljerna i till exempel en jurist- eller läkarutbildning. Men resultatet på provet kan ändå användas för att hjälpligt identifiera individer som skulle kunna klara detta. Det finns alltså en samvariation mellan höga högskolepoäng och förutsättningarna att klara en mer omfattande utbildning. Ett annat exempel kan vara enkätfrågor om hur ofta barn äter middag med sin familj. De som svarar att de ofta eller alltid gör det kanske också uppvisar större akademisk framgång. Finns en sådan koppling kan vi alltså nyttja utfallet för i vilken utsträckning barnen äter middag med familjen för att uttala oss om barnens grad av akademisk förmåga, utan att behöva göra mätningar av det sistnämnda.

Här nedanför och på nästa sida finns ytterligare två exempel som illustrerar innebörden av prediktiv validitet.

Exempel: Prediktiv validitet för tester vid mönstring till militärtjänst

Vid inskrivningsprövningen (mönstringen) till militärtjänst genomförs en rad olika tester. Det ingår bland annat tester av fysisk arbetskapacitet, kognitiv förmåga och en medicinsk undersökning. Syftet är att kunna identifiera vilka personer som passar för en viss tjänst. Naturligtvis är det väldigt många egenskaper som behövs för att klara en lång utbildning, men tanken är att resultaten på vissa tester kan användas för att förutspå vem som kommer klara militärtjänstgöringen, och vem som löper risk att inte klara den.

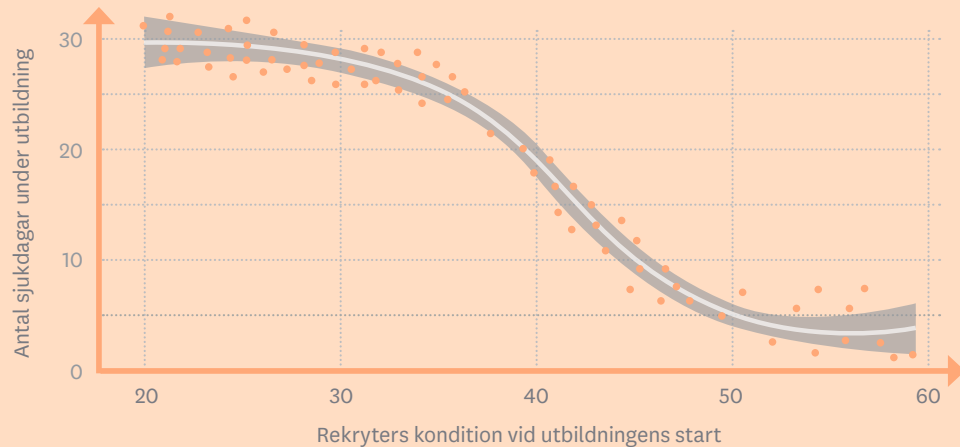
Om vi är intresserade av vilka individer som klarar en militärutbildning, är det möjligt att undersöka vilka av testerna vid inskrivningsprövningen som bäst kan förutse om en individ klarar sin utbildning. Och dessutom vilka värden på testerna som är tillräckligt bra för att med stor säkerhet klara utbildningen. En studie kring detta kan genomföras genom att sammanställa information från mönstringen, och jämföra den med information om vilka som klarat de olika utbildningarna.

Ofta är det dock kombinationer av vissa resultat som kan förutspå ett specifikt utfall (som vilka som klarar militärutbildningen), till exempel att en individ måste ha tillräckligt hög kondition och muskelstyrka samt vara fri från tidigare skador. Utifrån den här typen av studier har det exempelvis konstaterats att en individ som vill komma i fråga för en utbildning till jägarsoldat bland annat måste nå ett resultat på testet av muskelstyrka på 6, ett resultat på konditionstestet på 7 (båda på en niogradig skala), samt en godkänd läkarundersökning.

Det finns alltså en hög prediktiv validitet mellan denna kombination av testresultat och förmågan att klara utbildningen till jägarsoldat. Kroppslängd är en parameter som i stället har låg prediktiv validitet i sammanhanget, det vill säga en individs längd påverkar inte möjligheterna att bli jägarsoldat.

Exempel: Prediktiv validitet avseende rekryters kondition och antal sjukdagar

Till ett nyuppsatt förband i armén antogs 350 rekryter. Eftersom uppgifterna i förbandet mest var att lösa uppgifter via dator, och därmed sitta stilla, var gränsen för antagning när det gäller kondition mycket låg. Efter ett utbildningsår kunde det dock konstateras att en stor andel av soldaterna hade tvingats till sjukskrivning under utbildningen, främst beroende på stressrelaterade orsaker. En analys av sambandet mellan kondition och antal missade utbildningsdagar gav det resultat som presenteras i diagrammet nedan.



Resultaten visade att risken för att missa utbildningsdagar var kopplad till kondition, och att de som hade ett konditionstestvärde på under cirka 40 hade avsevärt fler sjukdagar – och missade alltså mer av utbildningen. Trots att kondition mäter en förmåga att arbeta fysiskt, hade det måttet en god prediktiv validitet för att identifiera individer som hade högre risk för att missa utbildning (även om uppgifterna mest handlade om att sitta stilla). Det kan självklart spekuleras i varför sambandet finns. Kanske är det så att kondition är kopplad till risken att drabbas av stressrelaterad ohälsa, eller så finns det en annan förklaring. Den prediktiva validiteten finns där, vilket räcker. Vi behöver inte kunna förklara sambandet.

I dessa situationer är utfallen i sig alltså inte det viktiga, utan vad de samvarierar med. Ett sätt att förstå den prediktiva validiteten är därför att betrakta det som det huvudsakliga utfallet samvarierar med som ett ja/nej-mått, det vill säga om något har eller inte har inträffat.

Vi vet egentligen inte vad som påverkar vad, men vi kan tydligt se att de hänger ihop. I exemplet ovan förutser (predicerar) ett konditionstest volymen sjukdagar för individen. Om testet verkligen mäter

kondition eller ej är inte det viktiga; det är värdena från testet som predicerar utfallet.

När det avslutningsvis gäller mått på den prediktiva validiteten är dessa ofta uttryckta i så kallade risk-mått, till exempel relativ risk. Sådana beskriver sannolikheten för att något ska ske utifrån att något annat händer. Ett exempel på det skulle kunna vara att för varje 0,1 poäng högre på resultatskalan från högskoleprovet ökar sannolikheten att klara högskoleutbildning med X %.

Sammanfattning – Prediktiv validitet

- Prediktiv validitet handlar om i vilken utsträckning det finns ett samband mellan två olika egenskaper, utan att det för den skulle betyda att de påverkar varandra – när det ena förekommer så förekommer också det andra, antingen samtidigt eller längre fram i tiden.
- I dessa situationer är utfallen i sig alltså inte det viktiga, utan vad de samvarierar med.
- Mått på den prediktiva validiteten är ofta uttryckta i så kallade risk-mått, till exempel relativ risk som beskriver sannolikheten för att något ska ske utifrån att något annat händer.

Samtida validitet – motsvarande utfall via olika mätinstrument

Ibland uppstår situationer där vi identifierat en studie som fokuserar på effekten av en likartad insats som vi själva tillhandahåller, där vi skulle vilja göra en motsvarande mätning. Primärt i syfte att kunna jämföra relevansen av respektive insats. Låt säga att det handlar om program för att minska ett upplevt utanförskap. Den mätmetod som används i den aktuella studien verkar dock vara alldeles för omfattande för vår målgrupp. Efter lite sökande hittar vi i stället en annan metod som också mäter utanförskap, men är betydligt mindre i omfattning (och därmed mer passande för vår målgrupp). För båda metoderna saknas information om kriterievaliditeten, och det finns heller inga möjligheter att på egen hand göra tester mot en kriteriemetod.

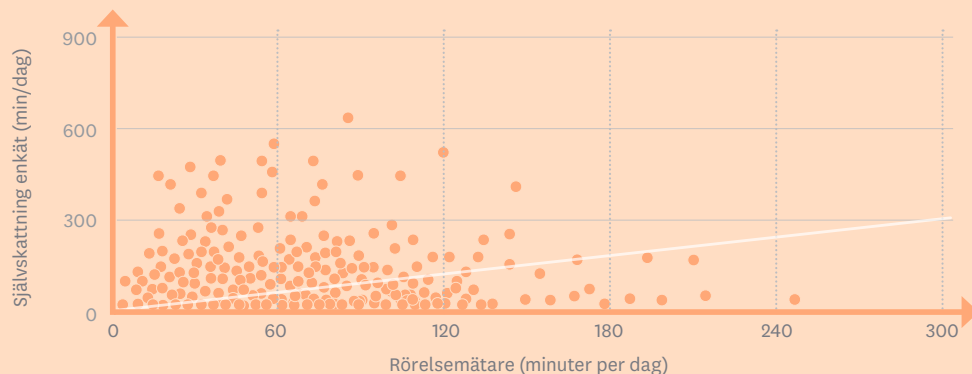
I en sådan situation kan det fylla en funktion att testa de två metoderna mot varandra, vilket synliggör dess *samtida validitet*. Det vill säga i vilken utsträckning de levererar motsvarande utfall gällande det vi vill mäta. Visar ett sådant test att den samtida validiteten är hög, innebär det att vi på god grund kan jämföra utfallet för utanförskap i våra egna mätningar med utfallet från den aktuella studien. Detta trots att olika mätmetoder har använts i respektive studie. Anledningen är att vi vet att de två metoderna fångar motsvarande grad av utanförskap i varje enskild situation.

Ett annat exempel gäller olika mått på alkoholkonsumtion. I vissa studier av detta används data från Systembolagets försäljning. Det är inte ett perfekt mått eftersom det finns andra sätta att skaffa sig tillgång till alkohol än att köpa via Systembolaget. Andra studier fokuserar sina mätningar på individers självskattade alkoholkonsumtion. Vanligen är data från Systembolaget ungefär dubbelt så hög som den från självrapporteringen, vilket indikerar att den samtida validiteten är låg. Det går alltså inte utan vidare att jämföra värden mellan studier som använt dessa två olika mått. Snarare är värden som erhålls i respektive studie beroende av vilken metod har använts.

Den samtida validiteten är alltså lite mer ”ödmjuk” än kriterievaliditeten, eftersom den senare kräver att den ena metoden är ”det perfekta måttet” (golden standard). Det innebär i sin tur att om vi tar del av en studie där två metoder har ställts mot varandra, och vi är osäkra på om den ena av dem är en kriteriemetod, är det bättre att prata om kopplingen mellan dessa två metoder som samtida validitet. Detta i stället för att chansa på att den ena är en kriteriemetod och därmed prata om resultaten som uttryck för kriterievaliditet. För att ge en bild av hur ett test av den samtida validiteten kan gå till, avslutar vi avsnittet med två olika exempel. Ett med utgångspunkt i fysisk aktivitet, och ett med fokus på stress.

Exempel: Samtida validering vid mätning av fysisk aktivitet

I en stor studie användes en självskattningsenkät samt en rörelsemätare (accelerometer) för att fånga i vilken utsträckning individer är i måttlig samt intensiv fysisk aktivitet. Diagrammet nedan visar att några av individerna rapporterade in flera hundra minuter per dag (se prickarna högt upp till vänster i figuren), medan deras rörelsemätare visade under 60 minuter per dag. En orsak till skillnaden kan vara att personerna vill framstå som mer aktiva än vad de egentligen är. En annan förklaring är att de utövar fysiska aktiviteter som en rörelsemätare inte klarar av att mäta, exempelvis cykling och styrketräning.



Andra individer rapporterade ingen fysisk aktivitet alls, medan mätaren visade att de hade rört sig över 60 minuter per dag (se prickarna som ligger närmast x-axeln). Orsaken till detta kan vara att individerna är väldigt aktiva i vardagen med sådant som de själva inte kopplar till måttlig eller intensiv fysisk aktivitet. För barn kan det handla om lek och spel, och för vuxna hobbies och olika former av hushållsarbete. Oavsett orsakerna går det att konstatera att utfallen från de två metoderna har svag överensstämmelse, och att den samtida validiteten därför är låg.

Exempel: Samtida validitet relaterad till ångest

En annan studie hade som syfte att mäta ångest hos ungdomar i högstadiet. I tidigare studier hade enkäten Spence Children's Anxiety Scale (SCAS-S) använts. Enkäten innehåller frågor om fobi, separationsångest och generell ångest, och vad viktigare är har enkäten svenska referensvärden. Det betyder att vi vet vilka värden som är normala i Sverige, och framför allt hur många som har alltför höga värden.

I den nya studien användes av olika orsaker en annan enkät – The Revised Children's Anxiety and Depression Scale (RCADS). Den mäter symptom på oro, ångest och depressiva besvär, men det fanns inga svenska referensvärden för skalan. Det var därför oklart vilka värden som var vanliga i Sverige vid tidpunkten. Vad viktigare är, det var också oklart vilka värden på RCADS som ska anses vara alltför höga.

För att lösa detta beslutades det att göra en studie av den samtida validiteten mellan de två enkäterna. En stor grupp barn fyllde i båda enkäterna, och det sammantagna resultatet från de två jämfördes. Sambandet var ganska starkt, det vill säga de som fick höga poäng på den ena enkäten fick också höga poäng på den andra (och vice versa).

Sambandets styrka beräknades till en korrelationskoefficient på 0.85, vilket klassas som ett starkt samband. Av de som svarade på SCAS-S hade 21 % ett värde över gränsvärdet för mild till måttlig oro/ångest. Det starka sambandet mellan enkäterna innebär därmed att det motsvarande gränsvärde för den andra enkäten (RCADS) finns på den nivå som delar de 21 % med högst svarsvärden från övriga 79 %. Genom denna studie kunde därför gränsvärden även för den andra enkäten synliggöras.

Kopplat till det sistnämnda exemplet finns det flera viktiga utmaningar när det gäller mätning av just ångest. En sådana är att del befintliga enkäter för detta endast mäter just ångest, medan andra mäter både ångest och depression. En annan är att det i en del enkäter finns svenska normvärden, medan andra har normvärden från andra länder (eller inga alls). En tredje utmaning är att många av de

befintliga enkäterna har blivit validerade rätt långt tillbaka i tiden, då faktorer som korrelerade till ångest inte är riktigt desamma som dagens faktorer (som till exempel skärmanvändning och sociala medier). Det innebär en risk att dessa enkäter inte på ett tillräckligt relevant sätt fångar ångesten utifrån dagens samhällskontext.

Sammanfattning – Samtida validitet

- Samtida validiteten beskriver i vilken utsträckning olika metoder (eller mätinstrument) ger samma utfall, där ingen av dem anses vara kriteriemetod.
- En hög samtida validitet innebär därmed att utfallen från olika studier med olika metoder kan jämföras med varandra på ett relevant sätt.
- Tar vi del av en studie där två metoder har ställts mot varandra, och vi är osäkra på om den ena av dem är en kriteriemetod, är det bättre att prata om kopplingen mellan dessa två metoder som samtida validitet. Detta i stället för att chansa på att den ena är en kriteriemetod och därmed prata om resultaten som uttryck för kriterievaliditet.

Praktiska råd

I de föregående kapitlen har vi belyst innebörden av olika former av validitet. I det här avslutande kapitlet vill vi ge dig några råd på vägen i ditt fortsatta mätrelaterade arbete. Ett sådant viktigt övergripande råd är att alltid lägga fokus på det som är viktigast att ta reda på i undersökningen, det primära utfallet, och försöka optimera validiteten för just detta. Ofta är det många saker som vi vill mäta, och det kan då vara svårt att prioritera. Avsätta därför tid för att på djupet inventera vad som är av särskild vikt att fånga.

I grunden är det alltid en god idé att nyttja befintliga mätinstrument, vars validitet är testad på ett relevant sätt. Så när du vet vad du vill mäta, börja alltid att inventera vad det finns för relevanta mätinstrument för att fånga den aktuella företeelsen. Om det är en befintlig enkät du vill använda följer här några tips:

1. Om du inte har tillgång till den enkät du vill använda, är det en god idé att söka efter den på exempelvis Fbanken.se. Det är en databas med en stor mängd validerade enkäter. Ett annat sätt är att använda lämplig sökmotor på nätet för att försöka hitta den, eller leta i databaser för vetenskapliga studier (till exempel pubmed.gov).
2. Om du hittar enkäten, kontrollera om den finns på svenska, samt om det finns en kostnad förknippad med att använda den. Finns den inte på svenska är det vanligen en god idé att försöka översätta den. Ibland krävs det även att du har en viss profession för att få nyttja enkäten, så kolla även det.
3. Undersök om enkäten har använts tidigare i studier, och i så fall om dessa fokuserat på motsvarande målgrupp som du utgår ifrån. Hittar du sådana studier kan resultaten från dem vara intressanta för dig att jämföra med.
4. Undersök också om det finns studier som fångat enkätens validitet och reliabilitet, och vad de i så fall visar. Den informationen är av stor relevans för att du ska kunna uttala dig om just din undersöknings validitet (se mer om detta senare i kapitlet).
5. Kontrollera avslutningsvis om det finns resurser hos er för att distribuera enkäten, eventuellt finnas på plats vid ifyllande av enkät för att svara på frågor, skicka ut påminnelser om att fylla i enkäten, bearbeta insamlad data genom att lägga in den i excelfil eller statistikprogram, samt utifrån det genomföra en analys.

Det är vanligt att en sökning efter befintliga mätinstrument inte ger positiva resultat, utan att ett eget mätinstrument behöver skapas. Här följer därför några tips för situationer där

utgångspunkten är skapandet av en egen enkät (tipsen är dock giltiga också för andra typer av mätinstrument):

- Börja med att tydligt definiera syftet med din enkät. Vad är det du vill undersöka? Försök att vara så specifik du kan, för då blir det lättare att göra en relevant operationalisering.
- Identifiera den målgrupp som ska svara på enkäten. Vilka vill du undersöka? Kommer det vara lätt eller svårt att få tag i dem så att de kan besvara enkäten?
- Skapa enkäten så att den förstås av, och är anpassad till, målgruppen. Inkludera gärna målgruppen redan i den här fasen för att få ett så relevant innehåll som möjligt. Detta gäller särskilt om målgruppen är barn (för mer information om just det, se *Idéer för livets guide "Enkäter till barn och unga – viktigt att tänka på utifrån ett barnrättsperspektiv"*).
- Var noggrann med att formulera tydliga och neutrala frågor.
- Formulera tydliga svarsalternativ på frågorna så att de har liknande upplägg. Frågor kan till exempel gälla "gårdagen", "förra veckan", "förra månaden" eller "vanligtvis". Det är lättare att svara om svarsalternativ för olika frågor inte skiljer sig åt för mycket.
- Använd inte för många frågor så att deltagarna väljer bort att svara eller svarar "slarvigt".
- I skapandet av enkäten kan hela (eller delar av) validerade enkäter användas och sättas ihop i din enkät. Detta då tillsammans med dina egna frågor som inte är validerade. Det är då viktigt att utvärdera de validerade frågorna och dina egenformulerade frågor var för sig, med särskild tonvikt på de sistnämnda.
- Pilottesta enkäten på ett mindre antal individer i målgruppen genom att lämna ut en testversion av enkäten. Se steg 1 i tabellen på nästa sida för mer information om detta.
- Utifrån den återkoppling du får från pilottestet gör du nödvändiga revideringar av enkäten. Är dessa av större karaktär är det vanligen en god idé att göra ytterligare ett pilottest för att kontrollera att förändringar är relevanta.
- Innan enkäten lämnas till målgruppen är det ofta bra att ha skapat ett slags försättsblad där det förklaras vem som gör studien, varför den genomförs, och var resultaten från den kommer finnas att ta del av. Den här typen av information gör att de som svarar på frågorna lättare kan förstå vad de deltar i.

Inom akademisk forskning läggs mycket tid och möda på att synliggöra en studies validitet. Det kan innebära att särskilt undersöka de mätinstrument som används, eller säkerställa att den grupp som ska undersökas är lämplig. Detta kräver ofta tid, resurser och god vana att utföra studier. För den som vill göra en studie utanför den akademiska kontexten är det viktigt att känna

till att även mindre åtgärder gällande detta kan ha stor betydelse för undersökningens kvalitet. I tabellen nedan ger vi en bild av hur du kan arbeta med validitet även i en mindre studie, där tid och resurser för detta är begränsade. Steg 1 i tabellen bör alltid genomföras, medan du nyttjar de andra stegen utifrån möjligheter och ambitionsnivå.

STEG 1: Gör ett pilottest

När du bestämt vad du vill undersöka och valt metod att samla in data, men saknar erfarenhet av forskningsarbete eller tid, kan du ändå göra mycket för din studies validitet. Ett viktigt första steg är att genomföra ett så kallat pilottest av ditt mätinstrument. Då låter du några få personer som inte ingår i själva studien, men som helst är från samma grupp, genomföra alla moment i själva informationsinsamlingen.

Därefter skapar du möjlighet för dem att berätta om sina upplevelser, exempelvis via intervjuer. Har du använt en enkät kan du ställa frågor om de förstod vad varje fråga innebar, om det fanns svåra uttryck eller ord, om frågorna kändes okej eller om de var påträngande, om enkäten var för lång, och så vidare. All sådan återkoppling gör att du får möjlighet att förbättra enkäten så att deltagare i den riktiga studien förstår frågorna och vill svara.

På det sättet kan du få syn på oklarheter i frågorna, eller andra saker i studiens tänkta genomförande, som inte riktigt fungerar som de ska. Utifrån det kan du göra nödvändiga justeringar. En eller två omgångar av sådana pilottester räcker ofta för att få studien att fungera bättre. På så sätt kan du ganska lätt få bort uppenbara hot mot din studies validitet.

STEG 2: Titta på hur andra har gjort

Har du möjlighet att lägga mer resurser på förberedelser kan du ta del av tidigare studier av liknande typ och använda snarlika eller identiska metoder. Så länge du refererar (det vill säga berättar varifrån du lånat informationen) är det alldeles korrekt att göra så. Även om metoderna i dessa andra studier kanske inte är de allra bästa, får du möjlighet att jämföra resultaten med tidigare studier. Något som kan vara av stort värde. Dessutom är det ofta så att du får nya idéer om genomförandet genom att ta del av detaljer i de tidigare studierna, både när det gäller metoder och tillvägagångssätt.

STEG 3: Välj ut den mest lämpliga metoden

Det finns ofta mer än en metod att välja mellan. Med än mer tid och resurser kan du därför läsa in dig på tidigare studier som använt de metoder som du tänker nyttja. De allra flesta enkäter, intervjumoduler eller testmetoder har någon gång validitetstestats, i de flesta fall flera gånger. Det görs ju för att ge viktig information till den som planerar att använda metoderna, för att på så sätt hjälpa till med valet av metod.

Det finns mycket att lära sig genom att ta del av tidigare studier. Sådana gör att du får stöd att välja mellan metoder, och ta den som passar bäst till den studie som just du planerar. Är metoderna testade på samma eller en liknande population som du fokuserar på? Hur stark är validiteten? Hur god är reliabiliteten? Finns det referensvärden? Är svaren relaterade även till andra egenskaper än de du primärt tänkt undersöka? Svar på dessa frågor hjälper dig mycket i arbetet med att skapa en studie där validiteten är tydlig.

STEG 4: Gör din egen studie

I akademisk forskning händer det inte sällan att forskarna själv validitetstestar de metoder som de ska använda. Ibland presenteras det i undersökningens metoddel, och ibland blir det en egen rapport. Har du mycket tid och erfarenhet av att göra studier, och dessutom planerar att använda en metod som inte tidigare testats, kan du själv göra en studie av validiteten. Inte sällan handlar det då om så kallad samtida validitet, där du jämför resultaten mellan en ny metod (kanske en du själv konstruerat) och en väletablerad metod. Detta för att undersöka hur lika de är.

När du sedan genomfört din studie, och det är dags att presentera resultatet, är det lämpligt att skriva några rader om sådant som inte gick enligt plan. Det kan handla om att ni drabbades av avhopp från undersökningen, att vissa mätningar inte gick som planerat eller att ni tappat bort viss data. Allt detta påverkar undersökningens kvalitet. Samtidigt som det är viktigt att nämna dessa misstag eller avvikelser, är det av vikt att i ett andra steg kunna resonera kring vilka konsekvenser dessa kan ha för resultatet. Det kan ibland uppfattas som att avvikelser från idealet gör studien oduglig, men så är naturligtvis inte fallet. Det gäller att balansera mellan att å ena sidan redovisa avvikelser, och

å andra sidan kunna visa att det finns andra delar som stärker studien och dess resultats generaliserbarhet. Nedan finns ett exempel hur en sådan beskrivning skulle kunna se ut. Det är dessutom möjligt att sedan beskriva ytterligare hur du ser på studiens kvalitet genom att diskutera mer kring andra delar av datainsamlingen eller de metoder som använts. Det vanligaste är att inte på egen hand ha testat sin validitet, utan att information om de metoder som använts (inklusive tidigare studier som testat dess validitet) ligger till grund för detta. Ett lämpligt sätt är då att referera till dessa studier och ange mått på validitet från dessa.

Exempel: Resonemang om validitet i en studie om trygghet

Ett kort och koncist sätt att beskriva en studies validitet är följande: "I vår studie användes en enkät för att undersöka deltagarnas upplevelse av trygghet i bostadsområdet. Enkäten är inte testad för samtida validitet och vi har därför inte kunnat undersöka om fler i den undersökta gruppen upplever god trygghet, jämfört med andra studiers resultat. Vi hade en hög andel som ville vara med i studien, så vi bedömer att studiens externa validitet är god".

Med tipsen i det här avslutande kapitlet, och innehållet i handledningen som helhet, hoppas vi att du har fått en djupare insikt i vad validering handlar om, och om vad du själv kan göra för att stärka validiteten i de studier som du genomför – oavsett om de är av större eller mindre karaktär. Och kom ihåg, även små åtgärder kan få stor betydelse för validiteten. Så känn inte att validitet är något övermäktigt att hantera, utan börja gärna på en gång med att ta ett första steg i att stärka validiteten i din kommande studie. Därefter är du förmodligen redo att ta ytterligare ett steg.

Stort lycka till!



Vi vill skapa ett samhälle med bättre hälsa och ökad trygghet

Vi är många som vet att lyckade investeringar idag leder till god avkastning i framtiden. Men hos Skandias stiftelse Idéer för livet mäts resultaten inte bara i kronor, utan också i kraften och förmågan att förändra samhället till det bättre. För vi vet att gemenskap nu förebygger utanförskap sen och att trygghet nu förebygger kriminalitet sen.

Idéer för livet agerar kraftfullt med mjuka medel, och bygger starka vuxna genom att investera i dem som fortfarande växer. För oss är varje krona som spenderas på att få barn och unga att må bättre, känna mer delaktighet och tro på sig själva, en stor vinst för framtiden.

Vi verkar för bättre hälsa och ökad trygghet för barn och unga och finns med från idé till handling, i insatser som ger effekt och skapar framtidstro. Vi uppmuntrar kompetensöverföring, hjälper organisationer att växa och bidrar till att skapa fler skalbara insatser för en ökad social hållbarhet.

Stiftelsen initierar och stödjer forskning, organisationer och insatser, nationellt såväl som lokalt och är länken mellan många av dem. Vi har funnits i nästan 40 år och stöttat fler än 4600 projekt.

För att skapa en positiv samhällsutveckling krävs ett nära samarbete mellan civilsamhället, offentlig sektor, näringsliv och akademien. Genom vår roll som katalysator kan vi verka för nya strukturer för samarbeten, bidra till att utveckla effektiva förebyggande lösningar som agerar på utmanande samhällsutmaningar.

Vi vet värdet av att förutse risker och arbeta förebyggande och arbetar för utvecklingen av effektiva metoder för att mäta värdet av förebyggande sociala insatser. Att göra rätt insats vid rätt tid lönar sig – både för samhället i stort såväl som för alla barn och unga. Barns hälsa och trygghet är centralt för oss.