

Visuelle analyser av data—er det greit å ikke vite alt?

Erik Arntzen¹ og Jon Løkke²

¹Høgskolen i Oslo og Akershus og ²Høgskolen i Østfold

Artikkelen omhandler problemstillinger knyttet til bruken av visuelle analyser av data fra linje-grafer. Visuelle analyser eller prosedyrer dreier seg om slutninger om tiltak har ført til effekt. Videre er visuelt baserte slutninger mest benyttet i anvendt atferdsanalyse. Et supplement er slutninger støttet av statistiske prosedyrer, men statistikkmotstanden blant atferdsanalytikere har vært massiv—vi beskriver motstanden og årsakene. Et vesentlig moment er imidlertid at slutninger eller konklusjoner om effekt bør være uten for mye feil. Vi kommer inn på mulige feil ved visuelle analyser og ved slutninger om effekt. Generelt er det tre kilder til slutningsfeil; feil ved sansing, feil ved tenking og en kombinasjon. Visuelle analyser er sårbare for feil knyttet til synssansen, og et vesentlig spørsmål er om statistikk er et ønskelig hjelpemiddel for å forbedre visuelt baserte slutninger. Vi svarer *ja* på spørsmålet og anbefaler at statistiske teknikker inngår i atferdsanalytikernes repertoar.

Nøkkelord: Visuelle analyser, statistiske analyser, slutninger, feil i dataanalysen, grafiske framstillinger

Visuelle analyser eller inspeksjon er ofte kontrastert til statistiske analyser og indirekte knyttet til hvilken type design som bør benyttes. Ved bruk av *single case research design* (SCRD) er visuelle analyser den vanlige og mest benyttede prosedyren for å analysere og trekke slutninger i anvendt atferdsanalyse (Bourret & Pietras, 2013; Gast & Spriggs, 2014). Kazdin (2011) diskuterer ulike kriterier som kan brukes for å evaluere data i SCRD. Dersom forskjellene er store og umiddelbare mellom basislinjefasen og fasen hvor den uavhengige variabelen er implementert, er det åpenbart at vurderinger basert visuell inspeksjon er holdbare. For eksempel: Dersom atferden er hyppig forekommende, eller helt fraværende i

basislinjen, og det motsatte er tilfelle etter innføringen av den uavhengige variabelen, så er det åpenbart at visuell inspeksjon er tilstrekkelig eller holdbart. Det kan være snusing eller sigarett røyking som stopper fullstendig ved innføring av den uavhengige variabelen eller treningen.

Uansett hvilken type av design, gruppedesign eller SCRD, der det er betydelige forandringer kan det argumenteres for at visuell inspeksjon er holdbart. Det som er utfordringen er at i de fleste situasjoner så er det ikke så klare forskjeller som skissert ovenfor eller at databildet er sammensatt og komplekst. Et løsningsforslag har vært å utvikle kriterier eller hjelpemidler for å fortolke data visuelt. Kriteriene er inndelt i tre hovedgrupper: forhold som er relatert til (1) hvor stor endringen er, (2) endringsraten og (3) det totale mønstret (Kazdin, 2011).

Den første gruppen, størrelsen på endringen, dreier seg om vurderinger av

Takk til to anonyme fagfeller for konstruktive og nyttige kommentarer. Det er ingen konflikter med hensyn til dette manuskriptet og forfatterne er enige om at det skal sendes til fagfellevurdering i Norsk Tidsskrift for Atferdsanalyse. Spørsmål om artikkelen kan rettes til Erik Arntzen, erik.arntzen@equivalence.net

gjennomsnitt på tvers av faser og endring av nivået på tvers av faser. Den andre gruppen, endringsraten, dreier seg om forandringer i trenden av dataene eller gradienten på de kurvene i de ulike fasene, samt hvor raskt endringen kommer etter at den uavhengige variabelen er implementert (latens). Det betyr at jo raskere endringen skjer, jo større sannsynlighet er det for at tiltaket er årsaken til endringen og ikke andre variabler det ikke er kontrollert for (gjerne kalt for trusler mot indre validitet). Den tredje gruppen, det totale mønsteret, utgjør en samlet vurdering.

Kazdin bruker begrepet ”visuelle analyser” i 1982 mens han i 2011 (s. 286) bruker visuell inspeksjon som ”refers to reaching a judgment about the reliability or consistency of intervention effects by visually examining the graphed data.” Vi bruker disse begrepene synonymt i dette manuskriptet selv om det kan være noe forskjell i innhold; eksempelvis er inspeksjon en forutsetning for analyse.

Problemstillingen i dette ”Om begreper” manuskriptet dreier seg om et kritisk blikk på visuelle analyser og anvendelsen. Vi ønsker debatt og økt interesse for temaet. Visuelle analyser av data, og slutninger om effekt, bør inneholde minst mulig feil. Det er prinsipielt sett tre kilder til slutningsfeil; feil ved sansing, feil ved tenking og en kombinasjon. Visuelle analyser er sårbare for feil knyttet til synsansen og påfølgende tenking, og vi ønsker å problematisere følgende: Er statistiske analyser et godt og ønskelig supplement og hjelpemiddel for å forbedre slutninger med basis i data?

Fra starten av

Tittelen ”Visuelle analyser av data – er det greit å ikke vite alt?” henspiller på en artikkel av en av foregangsmennene for anvendt atferdsanalyse, Donald Baer, og hans, i hverfall noen kretser, berømte artikkel «*Perhaps it would be better not to know everything*» fra 1977. Artikkelen til Baer er skrevet med bakgrunn i at statistiske prosedyrer i design med en eller få personer begynte å bli aktuelt og diskutert. Eksempelvis har Kazdin

i den klassiske metodeboken fra 1982 et appendiks om statistiske analyser av SCRDs (tilsvarende i boken hans fra 2011). Baers (1977) anliggende var fra et atferdsanalytisk ståsted svært konservativt anlagt: Data fra såkalte N=1 design eller SCRD, altså planer for datainnsamling med få personer inkludert, var, og er, gullstandard i evalueringen av data i anvendt atferdsanalyse. Visuelle analyser av grafer med synssansen er analysemetoden – og den eneste aksepterte analysemetoden. Poenget med de visuelle analysene er å trekke slutninger om effekt. Store forskjeller mellom kontrafaktiske faser (basislinje og senere faser som skal fungere som sammenlikningsgrunnlag) og intervensjonsfaser utgjør mønster som en anvendt atferdsvitenskap kan bygges på i følge Baer. Når visuelle analyser og store forskjeller er gullstandard vil det sjeldent være slik at konklusjonene om effekt, eller slutningene om effekt, er feil. Det er det samme som å hevde at det i atferdsanalyse begås svært få Type 1 feil som innebærer å konkludere med effekt uten at en slik effekt er virkelig. Med andre ord innebærer Type I feil å konkludere med effekt når det ikke er forskjell på kontrollbetingelsen (eksempelvis basislinjen) og intervensjonsbetingelsen. Men, atferdsanalytikere kan komme til å overse små, men kanskje viktige effekter, og risikerer dermed å begå mange Type II feil. I følge Kazdin (2011) er sannsynligheten for å gjøre Type I feil ved visuell inspeksjon ukjent og vi vet derfor ikke hvor alvorlig problemet er.

Små forskjeller, som *kan* oppdages med statistiske analyser, er i ”Baer-tradisjonen” mindre vesentlige, og påstanden er at vi bør ikke kaste bort tiden på å lære å bruke statistiske analyser. Statistiske analyser “could divert us from the much needed further development of that technology we almost have in hand. If it did, we would cease to be distinctive, hopeful, or useful. The Greeks referred to that as Tragedy” (Baer, 1977 s. 172). Baers anbefaling i 1977 er å holde seg til de visuelle analysene i utviklingen av den ennå unge, anvendte atferdsanalysen.

Et tiår senere i en apologia, eller i et klassisk forsvarsskrift for sitt standpunkt, skriver Parsonson og Baer (1986 s. 184): “We believe that the fine-grained graphic analysis of data... provides researchers and research audiences with information and analytic power that cannot be matched by statistical analysis, or, indeed, in any other way”. Denne konservatismen når det gjelder analysemetode har til stor del holdt seg (sjekk eksempelvis *Journal of Applied Behavior Analysis*), og tragedien Baer (1977) antyder ser ut til å være unngått. Men, det finnes hederlige unntak. Shadish (2014) har over tid påpekt fordelene ved statistiske prosedyrer både ved analyser av enkelte dataserier og mulighetene for å foreta slutninger over flere studier med samme type atferdsanalytiske teknikker. Beregning av såkalt effektstørrelse er viktig når det gjelder å etablere evidensbaserte terapier. Denne sammenslåtte effektstørrelsen forutsetter statistiske prosedyrer.

Anvendelse

Som kommentert på tidligere er visuelle analyser sentralt i SCRD (Kazdin, 2011). Det har derfor vært avgjørende å utvikle opplæringsprogrammer for visuelle analyser. Noen av programmene har vist lav reliabilitet blant de som har gjennomgått slik treningen (Lieberman, Yoder, Reichow, & Wolery, 2010), mens andre har vist høy grad av enighet (Kahng et al., 2010), noe vi kommer tilbake til senere. Det har nylig vært publisert opplæringsprogrammer med fokus på systematiske instruksjoner og med oppløftende resultater (Wolfe & Slocum, 2015). Forfatterne viste at to opplæringsmetoder, via et opplæringsprogram på en datamaskin og forelesning, hvor deltakerne ble trent i visuelle analyser av AB grafer var signifikant mer effektivt enn en betingelse uten disse to opplæringsmetodene.

Bruk av visuelle analyser har vært utbredt innen en rekke områder som for eksempel funksjonelle analyser (Chok, Shlesinger, Studer, & Bird, 2012; Hagopian et al., 1997). Innen dette område har visuelle analyser vært

brukt til å teste ut og iverksette effektive behandlingstiltak på en rekke felter og typer av atferd (e.g., Arntzen, 2014 og spesialnummer i NTA).

I mange sammenhenger er SCRD ansett som måten å sikre prediksjon og kontroll på innen atferdsanalyse (e.g., Bourret & Pietras, 2013; Iversen, 2013), men er samtidig også ledende innen rehabilitering når det gjelder å studere behandlingseffekter (e.g., Bobrovitz & Ottenbacher, 1998), selv om det for eksempel innen forskning på demens er forholdsvis liten bruk av slike design (Steingrimsdottir & Arntzen, 2015). SCRD er et alternativ til tradisjonelle design som baserer seg på statistiske sammenlikninger på tvers av gruppene. Analysene som gjøres i SCRD er ikke avhengig av statistisk hypotesetesting av data samlet inn fra et utvalg av forsøkspersoner. Strategien i atferdsanalyse kan sies å være induktivt anlagt.

Et typisk utslag av denne strategien innen atferdsanalyse har sprunget ut av en innstilling som er satt på spissen av Sidman (1960) ”I wonder what will happen if ” (s. 8). I andre vitenskapstradisjoner ville strategien også blitt beskrevet som undrende og eksplorerende med et fokus på enkeltorganismen. Det er mye å lære både om den enkelte organismen og om prinsipper ved SCRD. Men, vi vil understreke at det er viktig at problemstillingene innen atferdsanalyse også tar utgangspunkt i annen empiri (se for eksempel utviklingen av *naming hypothesis*; Horne & Lowe, 1996) eller inkludering av studier basert på gruppedata (e.g., Barnes-Holmes et al., 2005). Det er uansett om å gjøre å stille de riktige og gode spørsmålene. Atferdsanalyse er som andre naturvitenskaper basert på en stegvis og kumulativ utvikling med noen fakta som er godt kjent og en rekke nye problemstillinger i forskningsfronten (e.g., Catania, 2013).

Er visuelle analyser tilstrekkelige

Det er gjort flere undersøkelser hvor det er foretatt en sammenlikning av visuelle og statistiske analyser med inkonsistente funn

(Deprospero & Cohen, 1979). Forfatterne gjennomførte en undersøkelse hvor 36 ABAB figurer ble sendt ut til 250 fagfeller i JABA og *Journal of the Experimental Analysis of Behavior* (JEAB). Fagfellenes oppgaver var å evaluere disse figurene på en ordinalskala fra 1–100 med hensyn til ulike kriterier som: (1) Mønstre i de ulike fasene. Altså, ideelt mønstre hvor det er en klar forskjell mellom A og B fasene og hvor B fasene er innbyrdes like, ikke-konsistent mønstre hvor det ikke var noen forandring i de tre første fasene og en økning av avhengig variabel i den siste fasen, og irreversibelt mønster hvor det er det ikke var en gjenskaping av mønstre i den andre A eller B fasen. (2) Graden av den gjennomsnittlige forandringen på tvers av betingelser eller faser. (3) Variasjonen innen betingelsene eller fasene. (4) Trender i data. Hovedfunnet i undersøkelsen var en mellomobservatørenighet på 0.61, og forfatterne konkluderer med at visuelle analyser ikke er spesielt reliabelt. Kahng et al. (2010) replikerte og utviklet prosedyrene fra Desprospero og Cohen (1979). Studien brukte de samme 36 figurene som i studien ovenfor ble vurdert av 45 fagfeller i JABA. Kahng et al. (2010) fant en mye høyere konsistens i vurderingene blant forsøkspersonene som skulle vurdere grafene enn tidligere studier. Dette kan skyldes flere forhold. Det poengteres for eksempel av forfatterne at atferdsanalyse som fagfelt har utviklet seg og vokst kraftig på 30 år siden de tidligere undersøkelsene ble gjort, samt den kurssekvensen som kreves av *Behavior Analyst Certification Board* (BACB) som tilsammen innebærer at fagfellene er bedre trent i å gjøre visuelle analyser. Videre at det er forskjeller i avhengige variabler. Kahng et al. hadde i tillegg til den ordinale skalaen 1–100 en dikotom skala, *ja* eller *nei*. Forfatterne argumenterer med at når eksperimentell kontroll skal vurderes er det vanlig å skåre det som enten eller og ikke i forhold til en gradering. Men det er uklart om den dikotome skalaen kan ha påvirket skåringen på den ordinale skalaen. Forskjellen kan

også skyldes forhold som instruksjonen som ble gitt, hvor Kahng et al. eksplisitt ga en definisjon av eksperimentell kontroll. De nevner også at de har selv et mer homogent utvalg, fagfeller i JABA som er mye mer vant til å gjøre visuelle analyser, mens i Desprospero og Cohen (1979) var det både fagfeller fra JABA og JEAB.

Visuelle analyser og statistiske analyser

Kazdin (1982) argumenterer for bruk av statistiske analyser i tre situasjoner, (1) dersom det er problematisk å få en stabil basislinje, (2) ved evaluering av en ny behandlingsmetode eller dersom behandlingsresultatene ikke er uten videre er tolkbare, og (3) for å vise eksperimentell kontroll i studier i naturlige settinger.

Bobrovitz og Ottenbacher (1998) undersøkte enigheten blant 32 ansatte som jobbet innen rehabilitering. Forfatterne sammenlignet visuelle analyser og statistiske analyser i førtito grafer av hypotetiske SCRD. Dataanalysene dreide seg hovedsakelig om hvor stort samsvar var det mellom de visuelle vurderingene og resultater fra de statistiske testene. Resultatene viste at det var en enighet på 86 % mellom visuelle analyser og statistiske signifikans. Videre viste funnene fra denne studien at både visuelle og statistiske prosedyrer samsvarte når det gjaldt vurderinger av middels to store behandlingseffekter i alle grafene.

Det er også verdt å merke seg at undersøkelser viser at fortolkningene av resultatene i SCRD er direkte påvirket av hvilke metoder som er brukt i dataanalysene (Nourbakhsh & Ottenbacher, 1984). Forfatterne sammenliknet resultatene fra tre statistiske metoder for 42 forskjellige grafer. De beregnet samsvaret mellom mellom de statistiske metodene ved dividere alle tilfellene hvor metodene ga samme resultat med det totale antall grafer multiplisert med 100. (1) *C-statistics* er en tidsserieanalyse som kan brukes på små datasett. Denne analysen brukes i utgangspunktet på basislinjedataene. Dersom man får bekreftet

at basislinjedataene ikke inneholder en signifikant trend, slås disse dataene sammen med dataene for intervensjonen og det gjøres analyser hvor det testes ut om det er signifikante forskjeller mellom basislinjedataene og intervensjonsdataene; (2) *the two-standard deviation band method*. Dette er en metode som brukes blant annet for å undersøke forandringer i variabilitet på tvers av faser; (3) *the split-middle method of trend estimation*, som er en metode for å analysere om trender er akselererende, avtagende eller stasjonære. Hovedfunnene var 38% samsvar mellom disse tre metodene som er et forholdsvis lavt samsvar. Dessuten at samsvaret mellom metode (1) og (2) var høyest med 71%. I tillegg var det større grad av enighet mellom de tre statistiske testene når det var signifikante forskjeller i motsetning til når det var ikke-signifikante forskjeller på tvers av betingelsene i de hver av de 42 grafene.

Et siste poeng her er endringen innen atferdsanalyse i synet på bruk av andre typer av design enn SCRD og statistikk. For eksempel *Journal of the Experimental Analysis of Behavior (JEAB)* som ansees å være flaggskipet innen eksperimentell atferdsanalyse har annonsert følgende på innsiden av omslagssiden: "primarily for the original publication of experiments relevant to the behavior of individual organisms." På tross av dette er det en endring i hva slags type artikler som blir publisert i JEAB med hensyn til eksperimentelle design og statistiske analyser. Over det siste tiåret er det sterk økning av gruppedesign og tilhørende statistiske analyser. Enkelte ganger er slike valg av gruppedesign egnet for å redusere mulighetene for effekter av rekkefølge, for eksempel innen forskning på emergente relasjoner (Arntzen, Nartey, & Fields, 2015). Eksemplet er brukt for å vise at det ikke nødvendigvis er et enten eller forhold, men at valg av design og tilhørende analyser er avhengig forskningsspørsmålet(ene) som til enhver tid stilles.

Hva kan visuelle analyser være?

Ved bruk av SCRD brukes ofte visuelle analyser. Forskerne eller terapeutene er da opptatt av trender, retningen og stigningen eller fallet på kurvene, nivåforandringer og variabilitet i basislinjedataene (Scruggs, Mastropieri, & Regan, 2006). Noe av utfordringene med visuell inspeksjon har vært lav interraterreliabilitet (e.g., Ottenbacher, 1984).

Ved visuelle analyser er det viktig å skille mellom analyser innen en fase eller betingelse, mellom tiliggende faser eller betingelser, og mellom liknende faser eller betingelser. Antall datapunkter for hver fase eller betingelse, antall variabler som er endret i tiliggende faser eller betingelser, stabilitetsnivå, stabilitet innen og mellom faser eller betingelser, trender, trend innen og mellom faser eller betingelse, og overlapp av datapunkter på tvers av faser eller betingelser (Gast & Spriggs, 2014).

Det er ingen standardiserte måter for valg av datapunkter som skal være med i en grafisk framstilling, og som poengtert av Johnston og Pennypacker (1993): "paradox of summarization: It can simultaneously obscure and reveal important features of the data" (s. 303). Derimot argumenterer Fahmie og Hanley (2008) for at det er noen generelle regler for hvordan man skal velge ut data som skal presenteres visuelt. Dataframstillingen bør sørge for at den øker forståelsen av sentrale sammenhenger og da heller ikke tilsløre viktige forhold. Dessuten bør framstillingen forenkles mest mulig ved kombinere deler som ellers ville være overflødig. Gast og Spriggs (2014) gir en detaljert gjennomgang av variabler som er viktige ved bruken av visuelle analyser innen og mellom betingelser. Forfatterne trekker spesielt fram rekkefølgen og varigheten av betingelsene, nivået og forandringer, samt trender innen en betingelse. Mellom betingelsene diskuteres antall variabler som blir forandret, forandring i trender og nivå, samt overlapp av datapunkter.

Det finnes en rekke forskjellige måter å framstille data på, som kumulative kurver, semi-logaritmiske diagrammer, *scatter plot*, stolpediagrammer og linjediagrammer. Forskere bruker en A–B–C notasjon for å referer til ulike betingelser. Dersom en forandring er gjort innen en fase brukes en apostrof, B'. Ved to forandringer brukes to apostrofer B'' osv. I studier hvor det er flere betingelser benyttes flere bokstaver i alfabetet. I tilfeller hvor to variabler (B og C) anvendes innen en betingelse brukes betegnelsen BC (e.g., Lane & Gast, 2014).

Datapunkter, uten de eksakte verdiene, er fordelt over tid langs den horisontale X-aksen og med verdier langs Y-aksen eller den vertikale aksene. Typisk ulike betingelser med data fordelt over faser. Fasene kan være kontrafaktiske eller faktiske; de kontrafaktiske fasene viser hva som skjer uten tiltak eller kontra noe som faktisk skjer, de faktiske fasene hva som skjer med tiltak og differansen mellom kontrafaktiske faser og faktiske faser er det samme som effekt. Poenget er å oppdage denne differansen ved å se på grafene. Det sees etter: Nivåforskjell mellom faser, latenstid eller tid fra det siste punktet i en fase til det første i den neste fasen, overlapp mellom kontrafaktiske og faktiske faser, variasjon innen faser og stabile forskjeller mellom ulike faser.

Hvor godt grunnlag for slutninger gir visuelle analyser?

Ettersom anvendt atferdsanalyse i all hovedsak har dreiet seg om undersøkelser av enkeltindivider med relativt avgrensede målatferder som kvantifiseres, så har det vært hevdet at atferdsanalytikere bør bruke visuell inspeksjon som sitt førstevalg når det gjelder dataanalyse (Bourret & Pietras, 2013). Det er viktig å understreke at det er framsatt en del kritiske kommentarer (Fisch, 2001) som argumenterer for at visuell inspeksjon er mangelfullt på flere måter og da spesielt i forhold til misforståelse i analyse av trender.

Baer (1977) og Parsonson og Baer (1986) har i denne konservative tradisjonen foreslått at atferdsendring fra de kontrafaktiske betingelsene (eksempelvis basislinjen) til de faktiske betingelsene (tiltaksfasene) bør være mulig oppdage med det blotte øyet. Effekter som er akseptable både for klient og kliniker bør være store og åpenbare. Det er ukontroversielt å hevde at anvendt atferdsanalyse er effektivt. Når vi sier at noe er effektivt, så har vi foretatt en slutning—siden det er snakk om tiltak er det en kausale slutning eller årsaksslutning. Slutningen er basert på sansedata og anvendt atferdsanalyse er en empirisk vitenskap der kunnskapsbasen bygges systematisk, sten på sten eller på kumulativt vis og med data som drivkraft. Data inneholder grovt sett tre komponenter; det vi kan kalle for ”sanne” data eller det vi ville funnet dersom data kunne vært samlet inn under liknende forhold mange ganger. I tillegg inneholder data alltid to typer feil og en kombinasjon av de to feilene som reduserer andelen sanne data. Det er feil knyttet til sanseapparatet—at vi eksempelvis ser feil—for lite eller for mye. Data blir ikke til å stole på eller er ikke reliable. Reliabilitet har vært sentralt i atferdsanalyse, i alle fall et grovt mål som enighet mellom to observatører som følger den samme protokollen eller regler for hva som skal registreres av det som er observert. Dersom det er mye feil, kan feilene rettes ved å justere protokollen eller lage bedre regler og øve opp sansene. Disse feilene karakteriseres ved at de er tilfeldig fordelte—det er ikke nødvendigvis noe system i feilene—noen ganger ser vi for lite og andre ganger for mye. Slik sett vil slike usystematisk fordelte feil fordele seg på hver side av den ”sanne” verdien over tid. Med små datasett kan feilene legge seg på en bestemt måte og dermed forkuldre fortolkningene.

Feil knyttet til tenking er faktisk mer problematisk. Dersom vi med utgangspunkt i data tenker feil får det konsekvenser det for alle datapunkter. Et eksempel kan være at slutningene ikke tar hensyn til at det kan

være en alternativ forklaring på effekten eller en forklaring i tillegg til tiltaket og denne uforklarte effekten gjelder alle datapunkter.

Konklusjon

Oppsummert er det slik at de ikke fins noen klare og entydige standardkriterier for visuelle analyser. I prinsippet er det snakk om kvalitative analyser preget av subjektive vurderinger. Det innebærer at slutninger basert på visuelle analyser kan medføre at vi foretar slutninger om effekter som ikke er helt holdbare. Vi anbefaler, i tråd med Kazdin (2011), derfor en kombinasjon av visuelle analyser og kvantitative analyser for å minimere sannsynligheten for feil. Sannsynligheten for å konkludere om tiltakseffekter som ikke er reelle (Type 1 feil) er lav når data er klart forskjellige i de ulike fasene. I slike tilfeller er visuelle analyser stort sett tilstrekkelig. Kvantitative, statistiske analyser kan hjelpe oss til å oppdage effekter som er små, men viktige eller skjult i et uryddig databilde. Dermed øker sannsynligheten for å oppdage slike små effekter og der gjennom redusere Type II feil. En kombinasjon av visuelle analyser og statistiske analyser minsker den totale andelen feil. Vi får i ”både pose og sekk” dersom bruken av statistiske analyser ikke fører til en drift vekk fra interessen for variabler som er vesentlige for atferdsanalyse som anvendt vitenskap. Har vi rett i våre antakelser, bør utdanningen og opplæringen av atferdsanalytikere også inkludere mer enn grunnleggende kjennskap til statistiske teknikker.

Referanser

Arntzen, E. (2014). Funksjonelle analyser: Status, utfordringer og veien videre. *Norsk Tidsskrift for Atferdsanalyse*, 41, 59–73. Hentet fra <http://www.nta.atferd.no/>

Arntzen, E., Nartey, R. K., & Fields, L. (2015). Enhancing responding in accordance with stimulus equivalence

by the delayed and relational properties of meaningful stimuli. *Journal of the Experimental Analysis of Behavior*, 103, 524–541. doi: 10.1002/jeab.152

- Baer, D. M. (1977). “Perhaps it would be better not to know everything.”. *Journal of Applied Behavior Analysis*, 10, 167–172. doi:<http://dx.doi.org/10.1901/jaba.1977.10-167>
- Barnes-Holmes, D., Staunton, C., Whelan, R., Barnes-Holmes, Y., Commins, S., Walsh, D., . . . Dymond, S. (2005). Derived stimulus relations, semantic priming, and event-related potentials: Testing a behavioral theory of semantic networks. *Journal of the Experimental Analysis of Behavior*, 84(3), 417–433. doi:10.1901/jeab.2005.78-04
- Bobrovitz, C. D., & Ottenbacher, K. J. (1998). Comparison of visual inspection and statistical analysis of single-subject data in rehabilitation research. *American Journal of Physical Medicine & Rehabilitation*, 77, 94–102. Hentet fra http://journals.lww.com/ajpmr/Fulltext/1998/03000/comparison_of_visual_inspection_and_statistical.2.aspx.
- Bourret, J. C., & Pietras, C. J. (2013). Visual analysis in single case research. In G. J. Madden, W. V. Dube, T. D. Hackenberg, G. P. Hanley & K. A. Lattal (Eds.), *APA handbook of behavior analysis, vol. 1: Methods and principles* (pp. 199–217). Washington, DC, US: American Psychological Association.
- Catania, A. C. (2013). A natural science of behavior. *Review of General Psychology*, 17(2), 133–139. doi:10.1037/a0033026
- Chok, J. T., Shlesinger, A., Studer, L., & Bird, F. L. (2012). Description of a practitioner training program on functional analysis and treatment development. *Behavior Analysis in Practice*, 5, 25–36. Hentet fra <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3592487/pdf/i1998-1929-5-2-25.pdf>.
- Deprospero, A., & Cohen, S. (1979). Inconsistent visual analyses of intrasubject data.

- Journal of Applied Behavior Analysis*, 12, 573–579. doi: 10.1901/jaba.1979.12-573
- Gast, D. L., & Spriggs, A. D. (2014). Visual analysis of graphic data. In D. L. Gast & J. R. Ledford (Eds.), *Single case research methodology. Application in education and behavioral sciences* (pp. 176–210). New York: Taylor & Francis Group.
- Fahmie, T. A., & Hanley, G. P. (2008). Progressing toward data intimacy: A review of within-session data analysis. *Journal of Applied Behavior Analysis*, 41, 319–331. doi: 10.1901/jaba.2008.41-319
- Fisch, G. S. (2001). Evaluating data from behavioral analysis: Visual inspection or statistical models? *Behavioural Processes*, 54, 137–154. doi: 10.1016/S0376-6357(01)00155-3
- Fisher, W. W., Kelley, M. E., & Lomas, J. E. (2003). Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs. *Journal of Applied Behavior Analysis*, 36, 387–406. doi: 10.1901/jaba.2003.36-387
- Hagopian, L. P., Fisher, W. W., Thompson, R. H., OwenDeSchryver, J., Iwata, B. A., & Wacker, D. P. (1997). Toward the development of structured criteria for interpretation of functional analysis data. *Journal of Applied Behavior Analysis*, 30, 313–326. doi: 10.1901/jaba.1997.30-313
- Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior*, 65, 181–241. doi: 10.1901/jeab.1996.65-185
- Iversen, I. H. (2013). Single case research methods: An overview. In G. J. Madden, W. V. Dube, T. D. Hackenberg, G. P. Hanley, & K. A. Lattal (Eds.), *APA handbook of behavior analysis, Vol. 1: Methods and principles* (pp. 3–32). Washington, DC, US: American Psychological Association.
- Johnston, J. M., & Pennypacker, H. S. (1993). *Strategies and tactics of behavioral research*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Kahng, S., Chung, K. M., Gutshall, K., Pitts, S. C., Kao, J., & Girolami, K. (2010). Consistent visual analyses of intrasubject data. *Journal of Applied Behavior Analysis*, 43, 35–45. doi: 10.1901/jaba.2010.43-35
- Kazdin, A. E. (1982). *Singe-Case Research Designs*. New York: Oxford University Press.
- Kazdin, A. E. (2011). *Singe-case research designs* (2nd ed.). New York: Oxford University Press.
- Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: brief review and guidelines. *Neuropsychol Rehabil*, 24, 445–463. doi: 10.1080/09602011.2013.815636
- Lieberman, R. G., Yoder, P. J., Reichow, B., & Wolery, M. (2010). Visual analysis of multiple baseline across participants graphs when change is delayed. *School Psychology Quarterly*, 25, 28–44. doi: 10.1037/a0018600
- Nourbakhsh, M. R., & Ottenbacher, K. J. (1984). The statistical analysis of single-subject data: a comparative examination. *Physical Therapy*, 74, 80–88. Retrieved from <http://ptjournal.apta.org/content/74/8/768.full.pdf+html?sid=197c4468-2968-4898-8632-a3bb0435f50a>
- Ottenbacher, K. J. (1993). Interrater agreement of visual analysis in single case subject decisions: Quantative review and analysis. *American Journal on Mental Retardation*, 92, 135–142. Retrieved from <http://aaidjournals.org/loi/ajmr.1>
- Parsonson & Baer (1986). Parsonson, B. S., & Baer, D. M. (1986). The graphic analysis of data. In A. Poling & R. W. Fuqua (Eds.), *Research methods in applied behavior analysis. Issues and advances* (pp. 157–186). New York: Plenum Press.
- Scruggs, T. E., Mastropieri, M. A., & Regan, K. S. (2006). Statistical analysis for single case research design. In T. E. Scruggs &

- M. A. Mastropieri (Eds.), *Applications of research methodology: Advances in learning and behavioral disabilities* (Vol. 19, pp. 315–335). Oxford, UK: Elsevier.
- Shadish, W. R. (2014). Statistical analyses of single-case designs: The shape of things to come. *Current Directions in Psychological Science*, *23*, 139–146. doi:10.1177/0963721414524773
- Sidman, M. (1960). *Tactics of scientific research. Evaluating experimental data in psychology*. New York: Basic Books.
- Stanovitch, K. E. (2014). How to Think Straight About Psychology. Essex: Pearson Education Limited.
- Steingrimsdottir, S. H., & Arntzen, E. (2015). On the utility of within-participant research design when working with patients with neurocognitive disorders. *Clinical Interventions in Aging*, *10*, 1189–1200. doi: 10.2147/CIA.S81868
- Wolfe, K., & Slocum, T. A. (2015). A comparison of two approaches to training visual analysis of ab graphs. *Journal of Applied Behavior Analysis*, *48*, 472–477. doi: 10.1002/jaba.212
-

Visual Analysis of Data: Is it okay Not to Know Everything?

Erik Arntzen¹ and Jon Løkke²

¹Oslo and Akershus University College and ²Østfold University College

The article discusses issues related to how visual analyses are employed. Visual analyses or procedures are concerned with inferences about effective treatments. Moreover, visually based inferences are the most used procedure within applied behavior analysis. A supplement is inferences supported by statistical procedures, but there has been some resistance towards the use of statistics—we describe the resistance and causes. An essential point is that inferences or conclusions about the effect should be without too many mistakes. We are discussing possible errors in visual analysis and conclusions about efficacy. In general, there are three sources of inference errors; error sensing, incorrect thinking, and a combination. Visual analysis is vulnerable to errors related to the sense of sight, and an essential question is whether the statistical procedures are desirable tools for improving visual-based conclusions. We answer *yes* to the question and recommends that statistical techniques contained in behavior analysts' repertoire.

Keywords: visual analysis, statistical analysis, interpretations, errors, graphic formats