

Virker det nå?

Tomas Formo Langkaas

Forskningsinstituttet, Modum Bad Psykologisk institutt, Universitetet i Oslo

tomas.langkaas@gmail.com

Tilbakemeldingsverktøy er basert på et sviktende premiss om at det er mulig å følge med på om behandlingen faktisk virker.

Vi mangler gode metoder for å undersøke hvordan behandling virker i vanlig psykoterapeutisk praksis. Tilbakemeldingsverktøy og regelmessig utfallsmåling er ment å hjelpe klinikerne i å vurdere dette, men bidrar i stedet til at *endring i pasientens tilstand* forveksles med *den faktiske virkning av terapien*. Dermed kan tilbakemeldingsverktøy og utfallsmåling villedende behandle, forskere og pasienter i stedet for å rettlede dem. Dette er et problem vi bør ta på alvor.

Det er lett å forveksle endring og virkning

Et praktisk behov

Ingen behandling er garantert å være virksom i ethvert tilfelle. Å følge med på virkningen av tiltak er derfor en forutsetning for å justere behandlingen underveis. Observasjon av behandlingsevne danner grunnlaget for kliniske beslutninger om å videreføre, endre eller avslutte behandlingstiltak. En antakelse om at systematisk måling vil hjelpe klinikere i å følge med på behandlingsevne, har ført til utvikling av tilbakemeldingsverktøy (Lambert & Shimokawa, 2011) – i form av måleinstrumenter som brukes for å følge med på og gi tilbakemelding om enkeltpasienters behandlingsevne. Eksempler på veletablerte tilbakemeldingsverktøy er OQ-systemet (Lambert, 2015), *Clinical Outcomes in Routine Evaluation* (CORE; Barkham, Mellor-Clark & Stiles, 2015) og *Partners for Change Outcome Management System* (PCOMS; Duncan & Reese, 2015) – i Norge kjent som klient- og resultatstyrt praksis (KOR). Helsedirektoratet (2018) anbefaler bruk av tilbakemeldingsverktøy og regelmessig måling av behandlingsevne i pakkeforløp for psykisk helse og rus.

Å følge med på hvordan behandlingen virker, kan også brukes til å fange opp problematisk eller skadelig utvikling. Forskning på tilbakemeldingsverktøy antyder at tilbakemelding bidrar til å bedre behandlingsevner ved å fange opp problematisk utvikling tidlig (Lambert & Shimokawa, 2011). I litteraturen om negative effekter av behandling trekkes også bruk av tilbakemeldingsverktøy frem som anbefalt tiltak for å fange opp og forhindre skadevirkninger av behandling (Barlow, 2010; Rozental et al., 2018). En større forskningsoppsummering konkluderer imidlertid med at systematisk tilbakemelding ikke gjør noen forskjell (Kendrick et al., 2016).

Å vurdere behandlingsevne i vanlig praksis er fortsatt et uløst praktisk problem

Å kunne skille mellom når tiltak virker, og når de ikke virker, er avgjørende for å kunne justere og forbedre egne ferdigheter, med andre ord lære av erfaring. I ferdighetsforbedrende praksis (*deliberate practice*; Rousmaniere, Goodyear, Miller, & Wampold, 2017), der fagutøvere aktivt og målrettet arbeider for å bedre egne ferdigheter, er systematisk tilbakemelding en forutsetning for ferdighetsforbedring.

Kunnskap om behandlingsevne kan også brukes for å vurdere ekspertise. Behandlers resultater kan gi informasjon om den enkeltes ferdighet i å oppnå virksom behandling (Goodyear, Wampold, Tracey & Lichtenberg, 2017). Dette åpner for å både kunne studere gode behandlere og å vurdere kompetanseutvikling hos behandlere under opplæring.

Disse praktiske fagområdene – å justere behandling, oppdage skadelig behandling, forbedre ferdigheter fra egen erfaring og vurdere ekspertise fra behandlingsresultater – er alle basert på et premiss om at det er mulig for den enkelte kliniker å fange opp hvordan behandling faktisk virker. Tilbakemeldingsverktøy – og utfallsmåling generelt – kan imidlertid kun gi informasjon om pasienter oppnår endring i løpet av behandling. Dette er ikke det samme som informasjon om virkning. Forveksling av endring og behandlingsrespons er et tema som Bruce Wampold, Asle Hoffart og jeg tidligere har utforsket i en engelskspråklig artikkel (se Langkaas, Wampold & Hoffart, 2018). Vi viste der hvordan tilbakemeldingsverktøy og rutinemessig utfallsmåling (*routine outcome monitoring*) tilrettelegger for feilinformerte beslutninger i klinisk praksis.

Denne problemstillingen er imidlertid ikke ny. At endring ikke nødvendigvis gir informasjon om virkning, er en kjent begrensning ved observasjonsstudier (se f.eks. Higgins et al., 2019). At behandlere, forskere og pasienter kan forklare gode utfall med behandlingen, men på feil grunnlag, er også påpekt før (se f.eks. Meehl, 1955). Det er imidlertid mer usikkert om behandlere, forskere og pasienter er klar over hvor misvisende det kan være å anse endring som tegn på behandlingsrespons – slik som bruk av tilbakemeldingsverktøy og rutinemessig utfallsmåling tilrettelegger for.

Med utgangspunkt i den nevnte artikkelen vår ser jeg her nærmere på de forutsetningene som må ligge til grunn når en kliniker vurderer virkning i vanlig praksis. Hvordan klinikere kan vurdere behandlingsrespons i vanlig praksis, er fortsatt et uløst praktisk problem. Jeg argumenterer her for at bruk av tilbakemeldingsverktøy og rutinemessig utfallsmåling inntil videre har begrenset nytteverdi i klinisk praksis.

Endring og virkning

Det er lett å forveksle endring og virkning. Om jeg for eksempel av en eller annen grunn tror at jeg kan få været til å skifte ved å gjennomføre et ritual jeg har funnet på selv, vil jeg etter hvert erfare at været faktisk skifter, bare jeg holder på lenge nok. Det er da lett å feilaktig tro at det var jeg som forårsaket endringen i været, siden det jeg prøvde på, faktisk skjedde. På samme måte er det lett å tro at endring som oppnås i løpet av behandling, er tegn på virkning av behandlingen.

I ytterste konsekvens kan behandlere i god tro ende opp med å bli bedre i å gjøre mer som ikke virker, eller som er skadelig

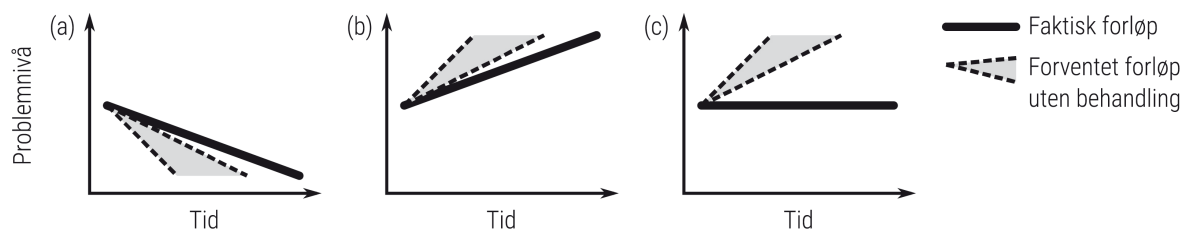
For å få vite om noe virker, hjelper det ikke å observere at endring inntreffer når vi gjør noe vi har tro på. Vi må sammenlikne med hva som ville skjedd om vi ikke hadde gjort noe: Gjorde det egentlig noen forskjell? Fordi mange ting er i stadig endring uansett – slik som været eller menneskers mentale tilstand – er det avgjørende spørsmålet om resultatet egentlig ville vært annerledes om vi ikke hadde grepet inn. Dersom vi kan sannsynliggjøre et annet resultat uten inngrep, har vi samtidig sannsynliggjort at det vi gjorde, hadde en virkning. Dette resonnementet danner generelt grunnlag for vurdering av virkning og omtales som *kontrafaktisk årsaksslutning* (Lewis, 1973).

Samme logikk danner også grunnlag for årsaksslutning i eksperimentell metode (Shadish, Cook & Campbell, 2002) – der poenget med kontrollbetingelser er å vise hva utfallet ville vært uten noe

inngrep. Dersom resultatene er forskjellige i eksperimentbetingelsen og kontrollbetingelsen, har vi sannsynliggjort virkningen av å gripe inn. Om eksperimentbetingelsen viser endring eller ikke, er derimot ikke relevant.

Tilstand, tid og riktig sammenlikningsgrunnlag

For å vise at endring ikke nødvendigvis sammenfaller med virkning, beskrev vi (Langkaas, Wampold & Hoffart, 2018) hvordan positiv terapeutisk endring kan sammenfalle med en faktisk negativ virkning av terapien – og omvendt – som illustrert i figur 1a og 1b. For tilstander som ofte bedres av seg selv uten behandling, slik som for eksempel belastningsreaksjoner eller depresjon, er forløpet i figur 1a en reell mulighet. Her har pasienten opplevd bedring i løpet av behandling, men bedringen ville sannsynligvis ha inntruffet tidligere uten behandling. Dette er et eksempel på at positiv endring kan sammenfalle med negativ virkning – behandling har sannsynligvis forhindret bedring i å inntreffe tidligere.



Figur 1. Det er ikke gitt at endring og virkning sammenfaller. Illustrasjon av (a) positiv endring med negativ virkning, (b) negativ endring med positiv virkning, og (c) null endring med positiv virkning.

Figur 1b beskriver en omvendt situasjon. Pasienten har her opplevd forverring i løpet av behandling, men forverringen ville sannsynligvis ha inntruffet tidligere og blitt større uten behandling. Dette er et eksempel på at negativ endring av tilstanden kan sammenfalle med positiv virkning av terapien – behandling har sannsynligvis bidratt til å bremse og utsette forverring.

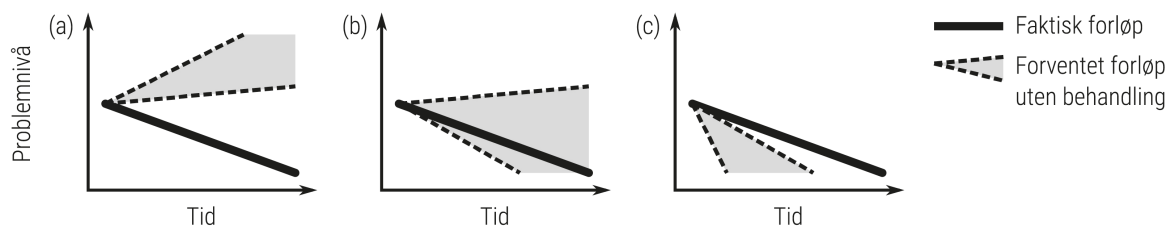
Figur 1c beskriver en liknende situasjon, men her sammenfaller null endring i tilstanden med positiv virkning av behandlingen. Dette er et eksempel på at endring i seg selv ikke er en forutsetning for virkning. Her har behandling sannsynligvis hindret forverring ved å holde tilstanden konstant – uten behandling ville tilstanden ha vært verre.

Disse tre eksemplene illustrerer at endring og virkning ikke nødvendigvis henger sammen, og at det ikke er mulig å entydig slutte seg til hva som er en virkning, kun fra å observere endring. De viser også at for å kunne fange opp om behandling virker, er vi avhengige av (a) å vite hva slags *tilstand* vi har med å gjøre, (b) å ha et kunnskapsbasert *sammenlikningsgrunnlag* som viser hvordan tilstanden sannsynligvis ville utviklet seg uten behandling, og (c) å anvende et sammenliknbart *tidsperspektiv*.

Noen tilstander er kjent for å kunne gå over av seg selv, slik som depresjon og belastningsreaksjoner. Noen tilstander er ansett for å være mer stabile og varige ubehandlet, slik som personlighetsproblemer. Andre tilstander kan derimot føre til forverring uten behandling – slik som en begynnende psykose eller mani. Ved kun å følge med på endring uten å ta hensyn til forskjeller i tilstand er det ikke mulig å tolke eller sammenlikne behandlingsresultater på noen entydig måte. Det er derfor et poeng å skille mellom tilstander som utvikler seg ulikt uten behandling. Figur 2

viser hvordan samme endring kan være uttrykk for helt ulike behandlingsresultater som følge av at forskjellige tilstander utvikler seg ulikt uten behandling.

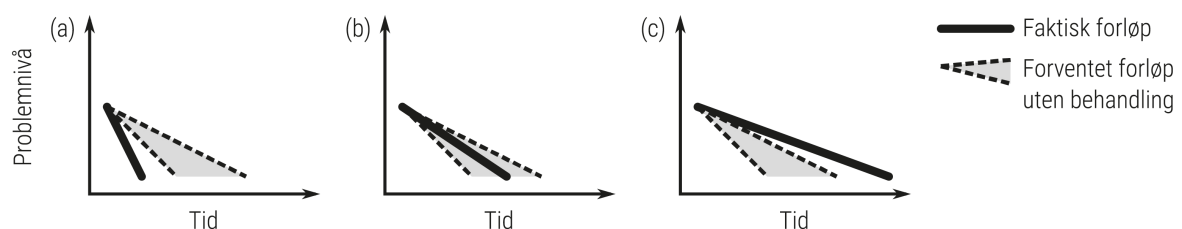
Eksisterende tilbakemeldingsverktøy som OQ, CORE og PCOMS/KOR skiller ikke mellom ulike tilstander – alle tilstander anses slik sett som om de skulle være like. Mens det ofte fremheves som en fordel at disse verktøyene er ment å brukes uavhengig av hvilken tilstand som behandles, illustrerer figur 2 hvordan det i stedet er en grunnleggende svakhet å ikke skille mellom ulike tilstander.



Figur 2. Samme endring kan være tegn på svært ulike behandlingsresultater ved ulike tilstander. Grafene illustrerer forskjellige tilstander med ulik utvikling uten behandling. Samme endring er her tegn på (a) virksom behandling, (b) uvirksom behandling og (c) skadelig behandling.

Å undersøke virkning av behandling i enkelttilfeller er en utfordrende metodisk problemstilling

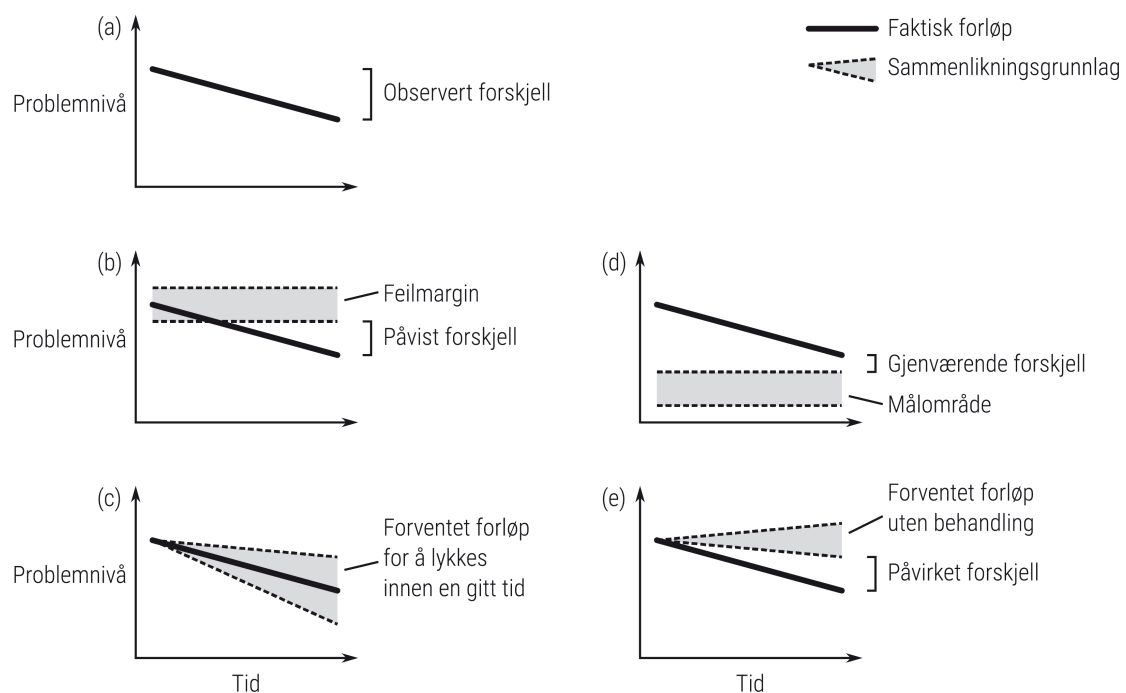
Men selv ved samme type tilstand kan samme endring være uttrykk for helt forskjellige behandlingsresultater dersom tidsperspektivet varierer, noe som er illustrert i figur 3. Dette er relevant i alle sammenhenger der behandlingstid kan variere, slik det ofte gjør i vanlig praksis. Det er ofte naturlig å forsøke behandling inntil ønsket endring opptrer, noe som da gjerne tas til inntekt for at behandlingen var virksom, og at det i enkelte tilfeller er nødvendig med langvarig behandling for å oppnå resultater. Selv om dette under noen omstendigheter kan være riktig, kan det i mange tilfeller være en helt feil konklusjon. Som figur 3 illustrerer, kan det for enkelte typer tilstander være slik at ønsket endring kan være tegn på virksom behandling innen et gitt tidsperspektiv, uvirksom behandling i et lengre tidsperspektiv og skadelig behandling i et enda lengre tidsperspektiv.



Figur 3. Samme endring kan være tegn på helt ulike behandlingsresultater ved ulik lengde på behandling. Illustrasjon av hvordan samme endring ved samme tilstand kan være tegn på både (a) virksom behandling, (b) uvirksom behandling og (c) skadelig behandling ved varierende behandlingstidsperspektiv.

For å vurdere virkningen av en behandling har vi sett at det er relevant å sammenlikne det observerte behandlingsforløpet med forventet forløp uten behandling. Dersom det ikke foreligger god kunnskap om hva som er et forventet forløp uten behandling, vil behandlere eller andre som tolker behandlingsresultatene, stå fritt til å tolke resultater i hvilken som helst retning de skulle ønske. Det er alltid mulig å forestille seg at det kunne gått annerledes uten behandling – både gode og dårlige utfall kan slik sett alltid feiltolkes som virksom behandling når det ikke foreligger et relevant kunnskapsbasert sammenlikningsgrunnlag.

I artikkelen vår beskrev vi fem typer sammenlikningsgrunnlag for å vurdere behandlingsutvikling, som vist i figur 4. Kun den siste av disse – *påvirket forskjell* (figur 4e) – kan gi informasjon om virkning av behandling. Dette er samtidig den eneste av disse fem sammenlikningsgrunnlagene som ikke er tatt i bruk for tilbakemeldingsverktøy eller rutinemessig utfallsmåling. Statistiske tilnærminger som i dag anvendes for tilbakemeldingsverktøy, samsvarer med de øvrige sammenlikningsgrunnlagene. Ingen av disse kan gi informasjon om virkning.



Figur 4. Fem former for forskjell å følge i praksis, med tilhørende sammenlikningsgrunnlag. Illustrasjonen viser (a) observert forskjell (observed difference), (b) påvist forskjell (detected difference), (c) forventet forskjell (predicted difference), (d) gjenværende forskjell (attainment difference) og (e) påvirket forskjell (induced difference). Kun den siste av disse kan gi informasjon om virkning. Basert på illustrasjon fra Langkaas, Wampold og Hoffart (2018). Opphavsrett: 2018 American Psychological Association (APA). Tilpasset og gjengitt med tillatelse. APA er ikke ansvarlig for kvaliteten på oversettelsen.

Praktiske konsekvenser av å forveksle endring med virkning

Når klinikere anser endringer i pasientens tilstand som tegn på virkning av behandlingen, risikerer de å fatte kliniske beslutninger på feil grunnlag.

For å justere behandlingen på riktig måte trenger klinikerer informasjon om virkning – ikke endring. Figur 1 illustrerer konsekvensene av å forveksle endring med virkning, slik tilbakemeldingsverktøy tilrettelegger for. Figur 1a viser en tilsynelatende god behandlingsutvikling dersom vi kun har informasjon om endring, noe som vil gi behandler og pasient tilbakemelding om å videreføre det som allerede ser ut til å fungere bra. Men med informasjon om sannsynlig virkning vil tilbakemeldingen være helt motsatt – dette ser nå ut til å være en problematisk og potensielt skadelig behandlingsutvikling som bør avbrytes og revurderes. På samme måte viser figur 1b og 1c tilsynelatende problematisk behandlingsutvikling dersom vi kun har informasjon om endring, noe som vil gi behandler og pasient tilbakemelding om å avbryte og revurdere tilnærming. Men i realiteten er det motsatt – med informasjon om sannsynlig virkning ser behandling faktisk ut til å virke bra, noe som tyder på at den heller bør videreføres.

Observasjon av endring kan heller ikke legges til grunn for å fange opp problematisk eller potensielt skadelig behandlingsutvikling. Dersom vi kun har informasjon om endring, vil vi i tilfeller som illustrert i figur 1a, feiltolke skadelig behandling som virksom, og i tilfeller som illustrert i figur 1b, feiltolke virksom behandling som skadelig.

For å lære av erfaring er vi avhengige av informasjon om når iverksatte tiltak virker, og når de ikke virker. Uten mulighet for å avgjøre om endring forveksles med virkning, risikerer vi å trekke helt feil konklusjoner fra erfaring om når tiltak virker, og når de ikke virker. I ytterste konsekvens kan behandlere i god tro ende opp med å bli bedre i å gjøre mer som ikke virker, eller som er skadelig, uvitende om at dette er tilfelle. Forveksling av endring og virkning kan på samme måte forklare hvordan behandlere, forskere og pasienter kan bli lurt av egen erfaring til å ha tro på uvirksom eller skadelig behandling.

Vi kan også risikere å feilvurdere grad av ekspertise dersom vi forveksler endring og virkning. Om vi tenker oss en svært kompetent behandler som i hovedsak arbeider med vanskelige tilstander, der effektiv behandling handler om å hindre forverring, vil denne behandleren feilaktig bli vurdert som en av de med dårligst ekspertise dersom vi kun ser på oppnådd endring. Som kontrast kan vi tenke oss en inkompetent behandler som i hovedsak arbeider med enkle tilstander som ofte bedres av seg selv. Denne behandleren kan gjennomføre både uvirksom og skadelig behandling uten at noen fanger det opp, fordi ved å følge pasientene over lang nok tid bedres tilstandene uansett, som illustrert i figur 3b og 3c. Dersom vi kun ser på oppnådd endring, vil denne behandleren feilaktig bli vurdert som en av de med best ekspertise. Uten riktig sammenlikningsgrunnlag vil begge behandlere også kunne feilvurdere egen ekspertise, dersom de også selv vurderer oppnådd endring som tegn på egen ekspertise.

Mulig å undersøke virkning?

Å undersøke virkning av behandling i enkelttilfeller er en utfordrende metodisk problemstilling, som ofte blir undervurdert og ignorert i klinisk forskning. Mens det tradisjonelle randomiserte eksperimentet er en metode som kommer nær i å avklare virkning på gruppenivå, kreves eksperimentelle metoder basert på gjentatt veksling av betingelser (*repeated period cross-over*) for å anslå virkning av behandling hos enkeltpersoner (Senn, 2016). I vanlig klinisk praksis er det derimot sjelden mulig å anvende slike metoder.

Vi bør derfor utvikle metoder for å følge med på det vi har betegnet som *påvirket forskjell* (Langkaas, Wampold & Hoffart, 2018). Måling av påvirket forskjell forutsetter at behandlingsforløp sammenliknes med (1) hva som ville vært forventet forløp uten behandling (2) for tilsvarende tilstand

(3) i tilsvarende tidsperspektiv. Usikkerhet rundt hvilke andre og samtidige tiltak eller endringer i pasientens liv en eventuell virkning kan tilskrives, vil være en uunngåelig begrensning ved all ikke-eksperimentell metode. Selv om denne begrensningen også vil gjelde observasjon av påvirket forskjell, vil en slik tilnærming fortsatt gi et betydelig bedre grunnlag for å vurdere virkning i vanlig praksis enn det dagens tilnærminger kan gi.

Referanser

- Barkham, M., Mellor-Clark, J., & Stiles, W.B. (2015). A CORE approach to progress monitoring and feedback: Enhancing evidence and improving practice. *Psychotherapy*, 52(4), 402–411. <https://doi.org/10.1037/pst0000030>
- Barlow, D.H. (2010). Negative effects from psychological treatments: A perspective. *American Psychologist*, 65(1), 13–20. <https://doi.org/10.1037/a0015643>
- Duncan, B.L., & Reese, R.J. (2015). The Partners for Change Outcome Management System (PCOMS) revisiting the client's frame of reference. *Psychotherapy*, 52(4), 391–401. <https://doi.org/10.1037/pst0000026>
- Goodyear, R.K., Wampold, B.E., Tracey, T.J.G., & Lichtenberg, J.W. (2017). Psychotherapy Expertise Should Mean Superior Outcomes and Demonstrable Improvement Over Time. *The Counseling Psychologist*, 45(1), 54–65. <https://doi.org/10.1177/0011000016652691>
- Helsedirektoratet. (2018). *Nasjonal plan for implementering av pakkeforløp for psykisk helse og rus 2018–2020 (IS-2734)*. Helsedirektoratet.
- Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., & Welch, V.A. (2019). *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley. <https://doi.org/10.1002/9781119536604>
- Kendrick, T., El-Gohary, M., Stuart, B., Gilbody, S., Churchill, R., Aiken, L., ... Moore, M. (2016). Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. *Cochrane Database of Systematic Reviews*. <https://doi.org/10.1002/14651858.CD011119.pub2>
- Lambert, M. J. (2015). Progress feedback and the OQ-system: The past and the future. *Psychotherapy*, 52(4), 381–390. <https://doi.org/10.1037/pst0000027>
- Lambert, M.J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy*, 48(1), 72–79. <https://doi.org/10.1037/a0022238>
- Langkaas, T.F., Wampold, B.E., & Hoffart, A. (2018). Five types of clinical difference to monitor in practice. *Psychotherapy*, 55(3), 241–254. <https://doi.org/10.1037/pst0000194>
- Lewis, D. (1973). *Counterfactuals*. Blackwell.
- Meehl, P. E. (1955). Psychotherapy. *Annual Review of Psychology*, 6(1), 357–378. <https://doi.org/10.1146/annurev.ps.06.020155.002041>
- Rousmaniere, T., Goodyear, R.K., Miller, S.D., & Wampold, B.E. (2017). *The cycle of excellence: Using deliberate practice to improve supervision and training*. John Wiley & Sons. <https://doi.org/10.1002/9781119165590>
- Rozental, A., Castonguay, L., Dimidjian, S., Lambert, M., Shafran, R., Andersson, G., & Carlbring, P. (2018). Negative effects in psychotherapy: Commentary and recommendations for future research and clinical practice. *BJPsych Open*, 4(4), 307–312. <https://doi.org/10.1192/bjo.2018.42>
- Senn, S. (2016). Mastering variation: Variance components and personalised medicine. *Statistics in Medicine*, 35(7), 966–977. <https://doi.org/10.1002/sim.6739>

Shadish, W., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.