

# Når forskningsetikken nedprioriteres



Rune Flaaten Bjørk

Psykologisk institutt, Universitetet i Oslo

r.f.bjork@psykologi.uio.no

Det som er faglig og teknisk godt, er ikke moralsk godt selv om det ser fortreffelig ut på papiret.

Dette er en kronikk. Kronikken gir uttrykk for skribentens meninger. Omtalte forskerkolleger er informert om innholdet i kronikken og har fått anledning til samtidig imøtegåelse og tilsvare.



Rune Flaaten Bjørk

Foto: Tron Trondal

Som doktorgradsstipendiat jobbet jeg fra 2016 til 2021 på den norske Tuning in to Kids-studien (N-TIK) ved Universitetet i Oslo (UiO). TIK er et emosjonsfokustert foreldreveiledningsprogram hvor barna er 3 til 12 år. Målet er at foreldrene blir bedre til å møte barnas følelser ved å lære emosjonsveiledning, som på sikt kan bidra til å øke barnas emosjonelle kompetanse og redusere angst og atferdsvansker (Havighurst & Harley, 2007). Programmets metodeutvikler holdt til i Australia, og jeg var ansvarlig for gjennomføringen av studien i Norge. Jeg brukte fagnettverket mitt som kommunepsykolog og blogger for Psykologforeningen til å skape interesse for studien, og til å rekruttere gruppeledere og deltakere. Jeg utarbeidet forskningsprotokoll, satte opp et sikkert datainnsamlingssystem og veiledet mange hovedoppgaver på prosjektet.

Les tilsvaret til denne artikkelen: Response to “When research ethics is downplayed”

## Uredelig forskningspraksis



Nærheten til prosjektet gjør at jeg kjenner dataene svært godt. Av den grunn kjenner jeg på et særskilt ansvar når resultatene denne våren publiseres i *Journal of Applied Developmental Psychology* (preprint er tilgjengelig her: <https://osf.io/ebpsv/>). Jeg trakk meg fra N-TIK-studien i januar i år fordi jeg ikke kan stå inne for resultatene, og fordi prosjektet etter mitt syn bryter med grunnleggende forskningsetikk og -moral.

Norske foreldre bombarderes daglig med råd fra eksperter på foreldreveiledningsfeltet. Det er ekstremt viktig at rådene er vitenskapelig kvalitetssikret og edruelig formidlet. Målet med kronikken er *ikke* å henge ut forskerkollegene mine. Målet er først å opplyse klinikerkolleger som tilbyr programmet i Norge, om at de bør tilby et parallelt opplegg til barna dersom de ønsker å redusere angst og atferdsvansker ved hjelp av TIK. Dernest er målet å sette søkelys på uredelig forskningspraksis mer generelt og bidra til å redusere slik praksis. Kanskje har andre lignende historier å fortelle?

## Bekreftelsesfelle i TIK-forskningen

Det har tatt meg mange år å forstå hva jeg har vært med på. Da vi designet studien, trodde jeg at målet var å undersøke effekten av TIK i Norge. Da jeg flere år senere leste utkastet til effektivitetsartikkelen, stemte konklusjonene i manuskriptet etter mitt syn ikke overens med de faktiske funnene.

Atferdsvanskene hos barna i intervensjonsgruppa var riktignok redusert med 4,64 poeng, og studien kunne skilte med en statistisk signifikant reduksjon som følge av foreldreveiledningsprogrammet ( $d = -.14, p < .029$ ). Men når skalaen går fra 36 til 252 poeng, er reduksjonen i praksis ubetydelig for familiene. Jeg foreslo derfor å reformulere resultatet til at intervensjonen var effektiv i å forebygge, men ikke redusere atferdsvansker. Det ville etter mitt syn være en mer edruelig presentasjon av funnene, siden studien gjaldt en normalpopulasjon og målet var nettopp forebygging.

Forskerne fant en hovedeffekt av intervensjonen på atferdsvansker, men effekten var marginal. De fant ingen hovedeffekt på angst og lette videre inntil de fant en moderasjonseffekt på angst hos sjenerte barn som var statistisk signifikant, men vanskelig å tolke fordi forskerne bare oppgav den ustandardiserte regresjonskoeffisienten ( $b = -4.10, p = .024$ ). Med andre ord hadde forskerne nå funnet at intervensjonen fungerte på sjenerte barn med angst. Effektstørrelsen oppgav de ikke, men  $p$ -verdien var innafor .05-nivået.

Da jeg stilte spørsmål ved konklusjonene deres, sa metodeutvikleren at jeg kunne trekke meg fra studien dersom jeg var uenig. Ingen av mine innspill ble tatt til følge.



Illustrasjon: hilde Thomsen

## Uriktig fremstilling



Foreldrene som deltok i programmet, ble bedre på å snakke med barna sine om følelser og opplevde barnas atferdsvansker som mer håndterlige etter kurset. Men de opplevde ingen særlig *reduksjon* i barnas angst eller atferdsvansker. Programmet fremstilles mer egnet til å avhjelpe angst og atferdsvansker enn dataene tilsier. Begrensningene burde ha kommet tydelig frem i artikkelsammendraget, men skjules av måten forskerne presenterer funnene på. Jeg har vært ekstremt kritisk til fremstillingen internt i gruppa. Da jeg imidlertid ble gjort oppmerksom på at én av medforfatterne opplevde kritikken min som ubehagelig, til og med trakasserende, sluttet jeg umiddelbart å mase om mine innvendinger.

I planleggingen av studien var jeg den eneste forskeren i N-TIK-gruppa som jobbet klinisk med familier i Norge. Jeg rekrutterte følgelig de fleste terapeutene til studien. Nettopp fordi jeg selv har deltatt aktivt i promoteringen av TIK, føler jeg et ansvar når resultatene formidles. Hadde jeg visst det jeg vet nå, ville jeg ikke ha promotert studien på samme måte. Det er derfor jeg går offentlig ut med mine erfaringer.

## Å pynte på sannheten

Den manglende viljen til kritisk selvrefleksjon rundt forskningsarbeidet har etter hvert fått meg til å tvile på hensikten med studien. Jeg trodde målet med slike studier var å identifisere styrker og svakheter ved en gitt intervensjon, for så å bruke kunnskapen til å gi bedre og mer treffsikker hjelp. Dessverre tilsier min erfaring med N-TIK-studien at formålet her har vært å vise at TIK virker, koste hva det koste vil.

Fra egen forskning på barns emosjonelle kompetanse vet jeg at det kan friste å pynte litt på sannheten i forskningssammenheng. Å selektivt fokusere på deler av datamaterialet, slik at resultatene passer bedre med forskningshypotesene, er nok mer vanlig enn vi liker å tro. Problemet oppstår når man ikke tar lærdom av andres korreksjon. Selektiv utvelgelse og tilpasning av data blir særlig problematisk når kommersielle interesser er drivende i prosessen og forskningsetikken settes til side for å oppnå faglig status eller kommersielle mål.

Min historie er neppe unik. Forskningslitteraturen er dessverre full av lignende historier. I *Håndbok om foreldreveiledning* skriver Øyvind Kvello følgende om metodeutviklerne på foreldreveiledningsfeltet (Bjørk, 2025):

Metodeutviklerne og de som skal drifte metodene, er gjerne «sterk troende» og opptrer nærmest som euforiske – og av og til karismatiske, emissærer. De skal sikre innsalg av metodene [s. 31]. [...] Et betydelig volum av evalueringene av foreldreveiledningsmetodene er gjennomført av forskere som har bindinger til metodene [...] [og] det er grunn til å ta høyde for at målingene som gjøres og forskningsmetodene som benyttes, primært styres av at resultatene skal bli så gode som mulig, slik at forskningsdesign velges med det for øyet og data «friseres» [s. 46].

## Mangelfull preregistrering

Det er et viktig prinsipp i forskning at hypoteser, metoder og analyser preregistreres, eller forhåndsregistreres, etter gjeldende regler. Det betyr at forskerne offentliggjør, for eksempel her: <https://www.clinicaltrials.gov/study/NCT04651465?term=b%C3%B8lsted&rank=2>, hva de har tenkt å forske på, *før* de analyserer innsamlet data. Slik unngår vi at forskere formulerer hypoteser etter å

ha fisket i data, såkalt HARKing («hypothesizing after the results are known»; Kerr, 1998). Jeg foreslo tidlig da vi designet N-TIK-studien, at studien måtte preregistreres, nettopp for å motvirke fising og frisering av data.



Preregistrering kan motvirke uredelig forskningspraksis, men som Yamada (2018) er inne på, kan registrering i seg selv bli et problem når det gjøres i etterkant av databearbeiding. I N-TIK-studien unnlot forskerne å forholde seg til forskningsprotokollen, og registrerte utelukkende data de selv hadde håndplukket til effektivitetsstudien. De registrerte ikke at data var innsamlet fra begge foreldre, også den som ikke deltok på TIK-kurset, eller at data var samlet inn fra barnehagelærere. Man finner nemlig ingen effekt av intervensjonen dersom data fra barnehagelærere og den ikke-deltakende forelderen legges til grunn.

Innhenting av data fra uavhengige observatører bidrar til å styrke kvaliteten på forskningen (De Los Reyes et al., 2015). Hensikten er blant annet å redusere metodevarians (John & Benet-Martínez, 2000). Innhentes data med to forskningsinstrumenter fra én og samme forskningsperson, vil deler av variansen være forklart av varians i forskningspersonen selv. Dersom for eksempel mor går på kurs og opplever kurset positivt, vil hun kanskje være tilbøyelig til å se større endringer hos barna enn far, som ikke har gått på kurs. Dersom data hentes inn fra mor på opplevelsen av kurset og fra far på endringer hos barnet, unngår man slik metodevarians. Til tross for at det lå innbakt i forskningsprotokollen at studien innhentet data fra flere kilder, unnlot forskerne å registrere dataene, informere om dataene og analysere dem i effektivitetsstudien. Innsamlede data om gruppeledernes etterlevelse av programmet (fidelity data), brydde de seg heller ikke om å analysere. Alle innsamlede data skulle uansett ha vært registrert, uavhengig av om de ble brukt i effektivitetsstudien eller ikke. Poenget med preregistrering er nettopp å være åpen om innsamlede data som *ikke* blir brukt.

## P-hacking

Ifølge de regionaletiske forskningskomiteene går såkalt P-hacking, eller datafising, ut på å bruke ett sett av resultater til å teste mange ulike hypoteser, for så å bare rapportere tilfellene som har gitt treff (Stefan & Schönbrodt, 2023). P-hacking omtales som uredelig praksis, og er foreslått som én av årsakene til replikasjonskrisen i psykologien (Renkewitz & Heene, 2019). I N-TIK-studien ble data testet for alle tenkelige signifikante sammenhenger før forskerne gikk videre med analysene som undersøkte effekten av intervensjonen. Her hadde forskerne ingen analyseplan. Mitt inntrykk av at det ble fisket i data, ble forsterket av at de i egne preliminære t-tester noterte «feil vei» på t-testene som viste at barna hadde signifikant *mer* angst i intervensjonsgruppa enn i kontrollgruppa, etter intervensjonen. Deretter fulgte forskerne opp med et vanvittig antall ulike moderasjonsanalyser, inntil de til slutt fant signifikante effekter av intervensjonen på nettopp angst.

## Ingen voksne hjemme

Jeg meldte fra internt på instituttet om at mine forslag til endringer i artikkelen ikke ble tatt til følge, og at teksten slik den nå var formulert, opplevdes som uredelig forskningsformidling.

Instituttet nedsatte da en komité som skulle gjøre en uavhengig vurdering av forskningen. Komiteen anbefalte flere endringer i artikkelen, og var blant annet kritiske til at det ble gjort moderasjonsanalyser uten signifikante hovedeffekter på angst. I tillegg foreslo komitéen at kriteriet for signifikans (*p*-verdien) burde strammes inn for å kompensere for den omfattende testingen som ble

gjort. Komiteen påpekte videre betydelige mangler i preregistrering og rapportering av begrensninger, men gikk god for imputeringen av manglende data og rapporteringen av effektstørrelser.



Rådene fra komiteen ble ikke fulgt, men forfatterne gikk med på å legge til begrepet «exploratory» syv steder i teksten, uten å endre på konklusjonene sine. Komiteen aksepterte endringene som tilfredsstillende, og ledelsen gav tommel opp for publisering av artikkelen. Kritikken ble i all hovedsak lagt bort.

Man kan helt sikkert argumentere, både statistisk og metodisk, for at N-TIK-studien er grei nok. Det er nettopp det som er problemet: Forskning baserer seg på at vi har tillit til forskerne. Så lenge man kan argumentere godt for den faglige og tekniske gjennomføringen, er studien god nok. Men ingen spør om forskningen er *moralsk* god nok. Spør du meg, er det svært problematisk å overselge forskning på bekostning av forskningsetikken.

Psykolog og filosof Ole Magnus Vik (2022) påpeker at vi psykologer utelukkende forholder oss til faglig og teknisk godhet i profesjonsetikken, samtidig som *moralsk godhet* ikke nevnes med ett ord i Etske prinsipper for nordiske psykologer. Hans poeng er at *moralsk godhet* er uavhengig av *faglig* og *teknisk* godhet. Man kan være teknisk og faglig god uten å være moralsk god, og vice versa. Det samme gjelder i forskningsetikken. Forskningen kan holde seg teknisk og faglig innenfor rammene for akseptabel forskningspraksis, men er ikke moralsk god nok av den grunn. En moralsk god studie kjennetegnes av en genuin søken etter sannhet.

## Nyttig idiot

Det viser seg at jeg har vært en nyttig idiot i N-TIK-studien. Det er jeg veldig lei meg for, og det beklager jeg. Det er lett å ende opp som nyttig idiot når man er uerfaren og ensom i akademien. De færreste tør å si ifra, og de som gjør det blir ikke nødvendigvis hørt.

Vi må likevel ikke gi oss. Vi må aldri unnlate å stille kritiske spørsmål, uavhengig av status og posisjon. Idiotti kan reduseres gjennom å tilby obligatorisk etikkundervisning på ph.d.-kursene og på ex.phil. Men som Vik (2022) skriver, må vi hele tiden reflektere over vår egen moralske praksis. Det er vårt selvstendige ansvar. Vi kan aldri tillate oss å glemme forskningsetikken. Det som er faglig og teknisk godt, er nemlig ikke moralsk godt selv om det ser helt fortreffelig ut på papiret.

## Referanser

- Bjork, R. F. (2025). Bokanmeldelse: Håndbok i foreldreveiledning. *Fokus på familien*, 53(1), 78–82. <https://doi.org/10.18261/fokus.53.1.6>
- De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E. & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, 141, 858–900. <https://doi.org/10.1037/a0038498>
- Havighurst, S. S. & Harley, A. E. (2007). *Tuning in to Kids: Emotionally intelligent parenting program manual*. University of Melbourne.
- Heck, R. H., Thomas, S. L. & Tabata, L. N. (2014). *Multilevel and longitudinal modeling with IBM SPSS*. (2. utg.). Routledge.
- John, O. P. & Benet-Martínez, V. (2000). Measurement: Reliability, construct validation, and scale construction. I H. T. Reis & C. Judd (Red.), *Psychology* (s. 339–369). Cambridge University Press.
- Kerr, N. L. (1998). HARKing: hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3), 196–217. [https://doi.org/10.1207/s15327957pspr0203\\_4](https://doi.org/10.1207/s15327957pspr0203_4)

- Renkewitz, F. & Heene, M. (2019). The replication crisis and open science in psychology: Methodological challenges and developments. *Zeitschrift für Psychologie*, 227(4), 233–236. <https://doi.org/10.1027/2151-2604/a000389>
- Stefan, A. M. & Schönbrodt, F. D. (2023). Big little lies: a compendium and simulation of <i>p</i>-hacking strategies. *Royal Society Open Science*, 10(2), 220346. <https://doi.org/doi:10.1098/rsos.220346>
- Vik, O. M. (2022). Psykoterapiens etiske dimensjon. *Tidsskrift for Norsk psykologforening*, 59(11), 1028–1034. <https://www-psykologtidsskriftet-no.ezproxy.uio.no/artikkel/2022as10ae-Psykoterapiens-etiske-dimensjon>
- Yamada, Y. (2018). How to Crack Pre-registration: Toward Transparent and Open Science [Opinion]. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.01831>

