

RUS OG VOLD:

Utdypende kritikk mot SIRUS-rapport

Sammendrag

Dette notatet oppsummerer og utdyper den kritikk jeg har levert mot SIRUS-rapporten "The impact of small changes in bar closing hours on violence. The Norwegian experience from 18 cities" av Ingeborg Rossow og Thor Nordström. SIRUS-rapporten skal publiseres som forskningsartikkel i tidsskriftet Addiction. Notatet sakser fra kritikk som er formidlet via Bergen Venstres hjemmesider den 30.9.2011, Dagbladet den 6.10, og Minervanett.no den 19.10. Sistnevnte var tilsvar til SIRUS sine svar på kritikken fra undertegnede og Menon Business Economics, publisert på SIRUS sine hjemmesider 12.10.2011.

Oppsummert går kritikken ut på at rapporten ikke redegjør for den statistikken som ligger til grunn for deres konklusjoner, og at rapporten er basert på feil metode (gitt den informasjonen som foreligger). Dette gjør at rapporten ikke er god nok som et beslutningsgrunnlag for oss politikere i vår bekjemping av utelivsvold.

Av Stian Skår Ludvigsen, ph.d.¹

¹ Forfatteren er cand.polit fra Institutt for sammenliknende politikk, UiB (2005), og ph.d. fra Institut for statskundskab, afdeling for offentlig politikk, Aarhus Universitet (2010). Avdelingens forskningsområder omhandler beslutningsgrunnlag for – og konsekvenser av – politikktutforming. Spesialiseringen var innen sekundæranalyse (meta-analyse) av forskningslitteratur, der de empiriske meta-analysene var av publiserte effekter av økonomi på britisk regjeringsoppfølging. Ludvigsen er nestleder i Bergen Venstre og er politisk engasjert i skjenketidsdebatten. Ludvigsen er ikke ansatt ved universitet eller forskningsinstitusjon, men jobber til daglig med statistikk og analyse. Kontaktinfo: stian.ludvigsen@venstre.no; tlf 41 32 27 34.

Innledning

Dette notatet oppsummerer og utdyper kritikken mot Rossow og Nordström (RN) sin rapport "The impact of small changes in bar closing hours on violence. The Norwegian experience from 18 cities",² hvor hovedkonklusjonen er at en endring i skjenketid på én time vil medføre en endring i vold på 17 prosent, noe som medfører fem færre anmeldelser av vold per time redusert skjenketid per 100 000 innbyggere per kvartal.

Det er en rimelig antakelse at beruselse medfører vold, men på grunn av feil bruk av metode og sammenblanding av ulike typer vold der ofrene har ulik motivasjon for anmeldelse, er det lite sannsynlig at effekten av skjenketidsinnskrenking er den RN har kommet frem til. Hvis det bare hadde vært et metodespørsmål, kunne effektene vel så gjerne vært større som mindre, men fordi det også mangler kontroll for andre faktorer som påvirker voldsnivået, er det lite sannsynlig at effektene er så sterke som RN vil ha det til.

Uten å vite hva de *faktiske* effektene av en skjenketidsinnskrenking er, er det vanskelig å bruke rapporten til SIRUS som grunnlag for politikkutforming. Rapporten er et særdeles svakt grunnlag for en politikk som tar inntekter og sysselsetting fra en stor og lovlig bransje, og som innskrenker en (for mange) viktig sosial arena.

Effektene av en skjenketidsinnskrenking bør diskuteres seriøst, men da må man ha et riktig utgangspunkt for diskusjonen. Det har man i dag ikke.

Jeg vil likevel avslutte dette notatet med en *hva-hvis-diskusjon*, altså: hva om effektene er reelle? Men før jeg kommer dit, vil jeg redegjøre for min kritikk av rapporten til Rossow og Nordström.

Notatet er ikke bare ment som kritikk av én rapport, men også som et innblikk i en del av forskningssosiologien som alt for sjeldent belyses: hvordan forskere publiserer sine resultat - og hvordan publiseringsprosessene kan lede til publisering av *feil* resultat.

Notatet er delt i fem deler i tillegg til innledningen og konklusjonen, der jeg først diskuterer dataene som er brukt i rapporten – og hvilke data som mangler, deretter diskuterer formidlingssvakheter i rapporten, deretter metodesvakheter, så noen betraktninger rundt publiseringsprosesser, og til sist den nevnte *hva-hvis-diskusjon*.

² http://www.sirus.no/filestore/Automatisk_opprettede_filer/RossowNorstrmvoldodskjenkingAddiction.pdf. Kritikken fra undertegnede er publisert på <http://www.venstre.no/hordaland/bergen/artikkel/39853/>, http://www.dagbladet.no/2011/10/06/nyheter/innenriks/vold/kriminalitet/skjenketider/18410426?utm_source=Dagbladet.no og <http://www.minervanett.no/2011/10/19/lettvinte-svar-fra-rusforskere/>.

Datasvakheter

Rapporten har ingen beskrivelse av voldsstatistikken som er lagt til grunn. RN skriver bare at statistikken er “the number of assaults reported to the police that occurred in the city centre at night-time (between 10 pm and 5 am) on week-ends (Friday-Saturday and Saturday-Sunday)” (s. 7-8). Videre bruker de “assaults” utenfor bykjernen som kontrollvariabel. Månedlige data fra Politidirektoratet ble aggregert opp til kvartalsvise data og omregnet til rater per 100 000 innbyggere.

De omtaler dataene som data om “violent crimes” (s. 8), men det fremgår ingen steder om hvilken voldelighetsgrad som måles. I Norge skiller vi mellom legemsfornærmelser og legemsbeskadigelser. Førstnevnte er av typen lette slag, spark, spytting, lugging som ikke krever legebehandling, mens sistnevnte er grovere vold som resulterer i legebehandling.

Skille mellom legemsfornærmelser og legemsbeskadigelser

Rick's har tidligere utarbeidet en rapport fra Bergen som viser en nedgang i anmeldelser av legemsfornærmelser, men ikke i de mer alvorlige legemsbeskadigelsene, ved endringen i skjenketid i 2008, noe de i tillegg hevder kan ha sammenheng med måten politiet mottok anmeldelser på.³ Det blir da svært problematisk at RN ikke skiller mellom forskjellige typer legemskrenkelser. Det er av stor betydning for publikum om det er grov vold som krever legebehandling som blir redusert ved en innskrenking av skjenketidene, eller om det “bare” er spytt og blåmerker. I Bergens tilfelle ville publikum, politi, legevakt, og media merket en endring på to tilfeller av grov utelivsvold per måned (ved en halvtimes utvidelse eller innskrenkelse som det står om) på en helt annen måte enn en endring på to tilfeller der offeret kan gå uskadet fra hendelsen. For oss politikere er det avgjørende at vi vet konsekvensene av vår politikk, og vår diskusjon om vi kan akseptere i snitt to tilfeller per måned, vil være en helt annen om det er snakk om invalidiserende vold enn om det er snakk om blåmerker.

RN svarer på dette på SIRUS sine hjemmesider:⁴

Det er til en viss grad tilfeldigheter som avgjør hvorvidt et voldssak ender opp som legemsfornærmelse eller legemsbeskadigelse, og derfor er det ikke alltid hensiktsmessig å skille mellom dem i analysen. Det kan også tenkes at praksisen i klassifiseringen varierer over tid og mellom byer; en sammenslåing av de to kategoriene øker da sammenlignbarheten i voldsdata over tid og rom. At tilfeller av legemsbeskadigelse er så få (ca 20% av all vold) betyr også at serien inneholder for mange tilfeldige svingninger (støy) til å være egnet for separat analyse. Vi må heller ikke bagatellisere legemsfornærmelse; det er ikke bare spytting, og lugging, det kan også være en voldsom slåsskamp, eller et kraftig slag i ansiktet. Bare det faktum at saken er meldt til politiet og har en strafferamme på bøter eller fengsel inntil seks måneder viser alvorret i slike saker.
(sic.)

Jeg vil selvsagt ikke bagatellisere legemsfornærmelser, de kan i aller høyeste grad være grove,⁵ men de er av en annen karakter enn legemsbeskadigelser, og motivasjonen for å

³ <http://www.regjeringen.no/pages/2243248/Sunkissed%20ved%200la%20Smith-Simonsen%20vedlegg%201.pdf>

⁴ <http://sir.us.no/SIRUS-forskere+besvarer+kritikk+av+artikkel+om+skjenketider+og+vold.d25-SMRjOZd.ips>

anmelde dem kan skille seg fra hverandre. Legemsfornærmelser kan lettere anmeldes i affekt, og det kan nok lettere forekomme at både offer og gjerningsmann anmelder hverandre, og det kan ofte være uklart hvem som er offer og hvem som er gjerningsmann. Ved legemsbeskadigelser vil gjerningsmannen derimot ha lite utbytte av å anmelde offeret.

Skadenivå og tilfeldige svingninger

Når RN snakker om støy (tilfeldige svingninger), er det ikke utenkelig at det vil være den ikke-legebehandlede volden som er støykilden her. Det kan selvsagt forekomme at det er tilfeldigheter som skiller mellom en legemsfornærmelse og en legemsbeskadigelse, hvor det som var ment som en fornærmelse ender opp som en beskadigelse, men det vil nok være sjeldent at et voldelig angrep ment som en beskadigelse ender opp som en fornærmelse. I studier av effekter av krig og terrorisme, som jeg har befattet meg med tidligere, har det vært vanlig å telle den alvorligste grad av skade (drepte) og holde mindre alvorlig grad av skade utenom. Med andre ord: lavere skadenivå anses som mer støyfullt. Dette kan godt være tilfelle ved skjenkerelatert vold også.

En meget god måte å redusere støyen på vil være å utelate alle henlagte anmeldelser, og alle frifinnelser. Altså å sitte igjen med de voldstilfeller hvor det foreligger en straffereaksjon etter at påtalemyndighetene har prøvd sakene.

Ingen minimering av støy og ekstremavvik

RN beskriver ingen vasking av dataene for å minimere støy og ekstremavvik. De har (øyensynlig) ikke tatt bort gjengangere, de har ikke delt i to tilfeller der gjerningsmann og offer har anmeldt hverandre, og de har ikke kjørt separate analyser av legemsfornærmelser og -beskadigelser, eller behandlet dem som separate (om enn korrelerte) variabler. Jeg ville forvente at man gjorde seg bedre flid med dataene for å minimere risikoen for feilslutning, men det synes ikke å ha vært gjort av RN.

Ingen oversikt over sammenheng mellom beskadigelser og fornærmelser

Den siste parentesen i avsnittet over er viktig, fordi det i databeskrivelsen ville vært nyttig om RN redegjorde for forholdet mellom legemsbeskadigelser og legemsfornærmelser i datamaterialet sitt, og eventuelt rettfærdiggjorde sitt valg om å slå dataene sammen uten å kjøre separate analyser. Gitt rapporten fra Rick's, vil jeg ha grunn til å anta at det ikke er så perfekt samvariasjon mellom beskadigelser og fornærmelser som RN forutsetter ved å slå dataene sammen.

Det er derfor grunn til å spørre om rapporten til RN faktisk beskriver de sammenhengene de er ute etter å analysere.

Formidlingssvakheter

RN beskriver ikke hvilken variasjon det er i voldsstatistikken. Er det stor variasjon og mange mulige årsaksfaktorer, eller er bildet entydig? Her mangler typiske mål som gjennomsnitt, minimum, maksimum, standardavvik og avvikende observasjoner ("uteliggere"). Det mangler også en visuell presentasjon av data. Dess mer komplisert struktur dataene har, dess viktigere er det å gi en god beskrivelse av dataene.

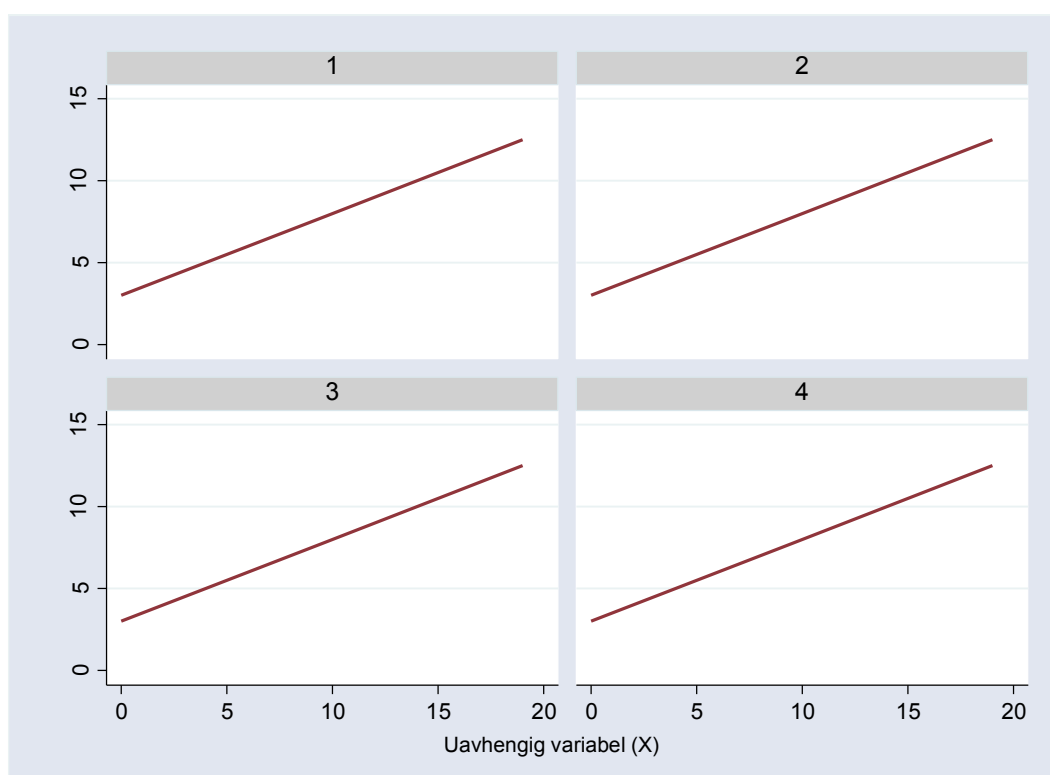
⁵ Det finnes også en kategori "legemsfornærmelse med skadefølge", men RN har (heller) ikke opplyst om dette.

RN gir oss ikke noe bilde av voldsstatistikken, og uten kunnskap om variasjonen kan vi ikke vite om funn som statistisk sett er signifikante faktisk er relevante. Det finnes mange måter å oppnå signifikans på, uten at funnet nødvendigvis er relevant. En visuell beskrivelse av dataene ville vært til stor hjelp.

Anscombes kvartett

Jeg vil illustrere dette med et eksempel av F.J. Anscombe: *Anscombes kvartett* er fire ulike datasett med (så nær som) det samme gjennomsnittet, samme gjennomsnittlige avvik fra gjennomsnittet, og samme samvariasjon mellom avhengig og uavhengig variabel. Bruker man ordinær lineær regresjonsanalyse (minste kvadrats metode, som er den vanligste) for å finne effekten den uavhengige variabelen har på den avhengige, får man den samme ligningen: $Y = 3 + 0,5X$.⁶

Grafisk får man følgende prediksjon i alle fire tilfeller:



Figur 1: Resultat av ordinær lineær regresjon på "Anscombes kvartett"

Alle fire datasett leder til det samme resultatet ved ordinær lineær regresjon, og forklaringskraften til regresjonene er høy: alle fire forklarer 67 % av variasjonen i den avhengige variabelen; og er høyst signifikante ($p = 0,002$). Man ledes her til å trekke nøyaktig den samme konklusjonen for alle fire sett: nemlig at en økning i X på én enhet leder til en økning i Y på en halv enhet.

⁶ F.J. Anscombe (1973): "Graphs in Statistical Analysis," *American Statistician* 27, s. 17-21. Data hentet fra http://en.wikipedia.org/wiki/Anscombe%27s_quartet

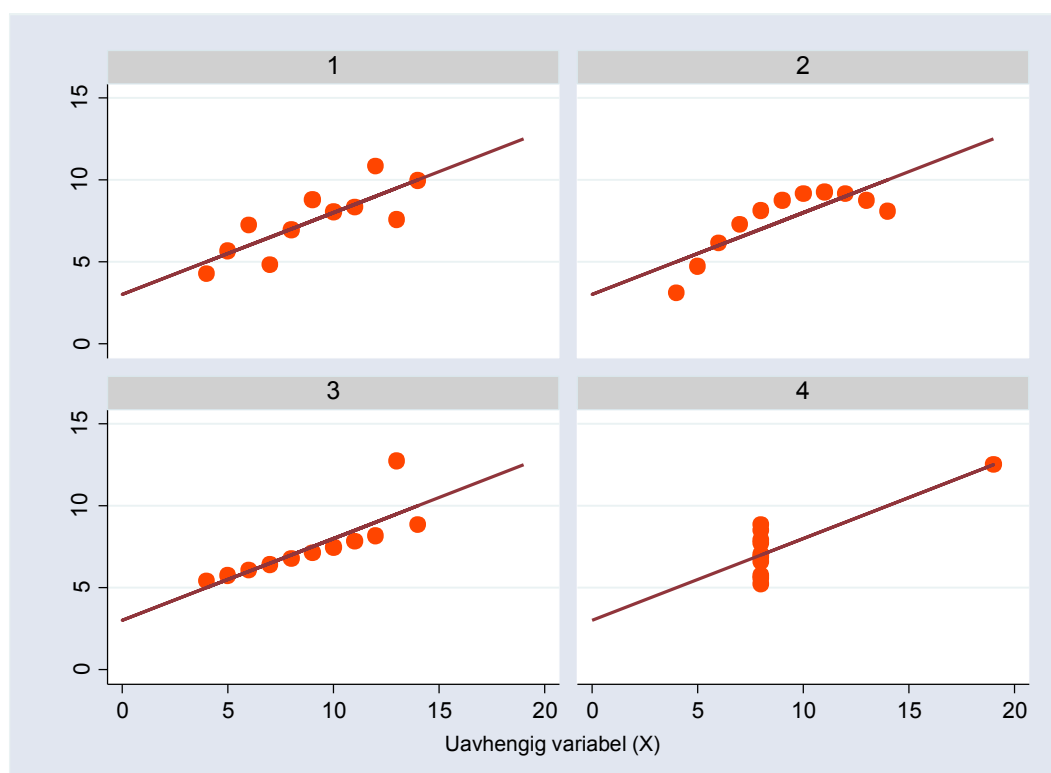
En slik effekt vil i de fleste tilfeller oppfattes som sterk (noe den også er, med en standardisert regresjonskoeffisient på 0,82; der 0 = ingen effekt og 1 = full effekt).⁷

Uansett hvilke bakgrunnsdata man har brukt i eksempelet over, kan man ledes til å konkludere med at man har et godt funn. Det er både sterkt og signifikant.

Men er det korrekt?

Nei. Det er det bare i det første tilfellet at det er korrekt å kjøre ordinær lineær regresjon på dataene. Tabben man her har gjort er at man har studert skogen uten å se trærne. Dette ligner på det Rossow og Nordström har presentert for oss: de har gitt oss regresjonsresultatene sine uten å vise oss hvordan resultatene passer til dataene.

Hvis man i Anscombes kvartett ser nærmere etter på dataene *bak* regresjonene, finner man at ordinære regresjonsmodeller slettes ikke passer så perfekt til dataene som man skulle tro (figur 2).

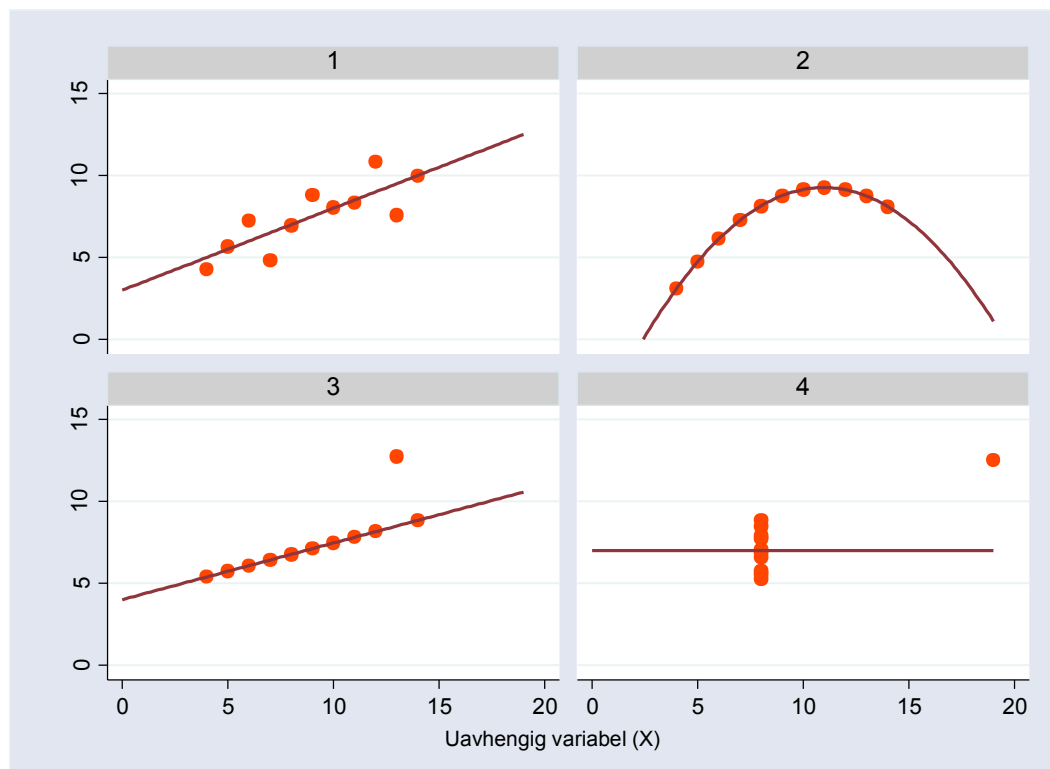


Figur 2: "Anscombes kvartett" med data. Ordinære lineære regresjoner

Vi ser her at regresjonslinjen bare passer til dataene i det første eksempelet. I det andre eksempelet skal det kjøres en ikke-lineær (kvadratisk) regresjon, og i de to siste tilfellene bør uteligger enten fjernes eller håndteres via robust regresjon (robust regresjon justerer for av-

⁷ De fire regresjonene bak figur 1 kan leses i tabell 1 i vedlegget.

vikende observasjoner – i dette tilfellet er robust regresjon i praksis det samme som å fjerne uteliggerne).⁸ Figur 3 viser dette.



Figur 3: "Anscombes kvartett" med data. Datatilpassede regresjonsmodeller

Ved kvadratisk regresjon i det andre eksempelet eller robust regresjon i det tredje eksempelet, øker man forklart varians til 100 % (den observerte variasjon i X forklarer *all* variasjon i Y), og ved robust regresjon i det siste tilfellet reduserer man forklart varians til 0 %. I det siste eksempelet har X den samme verdien (8) for alle verdier av Y, og kan derfor ikke forklare Y.

I figur 1 og 2 øker Y alltid med 0,5 for hver hele økning i X, og Y er alltid 13 hvis X er 20, men som man ser, endrer figur 3 dramatisk på prediksjonene som kan utledes av modellene.

Tabell 2 i vedlegget viser forskjellen på ordinær lineær regresjon og kvadratisk regresjon i eksempel 2, og tabell 3 og 4 viser forskjellen på ordinær lineær regresjon og robust regresjon i eksempel 3 og 4.

Anscombes kvartett og SIRUS

Disse eksemplene viser hvor viktig det er med en korrekt håndtering av data, og at man ikke mekanisk kjører en standard regresjonsanalyse og sier seg fornøyd med resultatene uten å ha forsikret seg om at man har benyttet korrekt metode på korrekt modell.

⁸ Det anbefales normalt ikke at man kjører robust regresjon på små datasett uten å korrigere standardfeilene (Andersen, 2008: *Modern Methods for Robust Regression*, Series: Quantitative Applications in the Social Sciences 152, Los Angeles: Sage Publications), men dette blir unødig komplisert for denne illustrasjonen. Standardfeilene er heller ikke av interesse her.

Rossow og Nordström har som nevnt gitt oss regresjonsresultatene uten å vise oss at modellene de har brukt faktisk passer til dataene som ligger til grunn. De bruker en lineær modell på data som de ikke har gitt oss noen som helst informasjon om. Vi er med andre ord ikke i stand til å vurdere om de har benyttet riktig modell. Vi kan ikke vite om RN har brukt en lineær modell der de skulle brukt en ikke-lineær modell (som i eksempel 2), og i verste tilfelle sitter vi med en rapport der signifikans og høy forklart varians er oppnådd som i det fjerde tilfellet i Anscombes kvartett: altså uten reell forklaringskraft.

I tillegg har de ikke opplyst om hvor mye av endringen i vold som kan forklares av endring i skjenketid. Det betyr lite om funnene er signifikante hvis endring i skjenketider bare forklarer noen få prosent av variasjonen, men dette får vi ikke vite. Uvanlig nok presenteres regresjonsanalysene uten noen form for teststatistikk som kunne fortalt oss dette.

Til dette har RN svart at deres hovedmodell forklarer 43,5 % av variansen i voldsdataene. Dette høres imponerende ut, men når modellen inkluderer 18 variabler for å kontrollere for variasjon mellom de ulike byene, vil det hovedsakelig være variasjonen mellom byene som fanges opp i de 43,5 prosentene. RN har ennå ikke redegjort for hvor mye av variasjonen i utelivsvold som skyldes endring i skjenketid.

Uten å vite passformen og forklaringskraften til modellene, er det vanskelig for oss politikere å vurdere endringer i skjenketid opp mot andre tiltak vi kan benytte for å bekjempe utelivsvold. Innskrenking av skjenketid er muligens det mest kostbare og mest inngripende tiltaket vi kan benytte oss av.

RN har unnlatt å svare på kritikken min om manglende beskrivelse av dataene som er brukt i studiet. I stedet svarer RN dette på SIRUS sine hjemmesider:

Det kan godt være at andre faktorer, som været eller lønnsutbetalinger, forklarer mer av variasjonen i vold, men disse er vanskeligere å endre ved politiske midler.

Dette blir rett ut flåsete. Det har blitt lagt ned mye arbeid i å redusere utelivsvolden utover å redusere skjenketider. Det leder oss til de neste mangler i RNs rapport.

Modellsvakheter

Det kan være en rekke andre faktorer som har betydning for endring i utelivsvolden. Jeg savner modeller som kan beskrive hvordan endring i skjenketid kan ses opp mot andre tiltak.

”Bivariat” analyse

RN ser bare på sammenhengen mellom skjenketider og vold, uten å kontrollere for andre tiltak som kan påvirke voldsnivået. Selv om RN inkluderer kontrollvariabler for andre byer, er analysen deres i prinsippet en såkalt *bivariat* analyse, der effektene er målt uten kontroll for andre variabler som kan ha innvirkning på resultatet. Det mangler altså en *multivariat* analyse der disse kontrollvariablene er inkludert. Dersom det finnes utelatte variabler som er relatert til vold eller skjenketider, er resultatet til RN vridd i retning av en sterkere effekt enn den virkelige effekten.

Andre tiltak

Fra Tønsberg lister politimester i Vestfold, Marie Benedicte Bjørnland, i en kronikk i Dagbladet 9. oktober opp: samarbeid mellom politiet, kommunen, skjenkekontrollen og restaurantnæringen; opplæring for å redusere overskjenking; ordensvaktkurs og -godkjennelse; målrettet skjenkekontroll; rådgivning; bedret nattbusstilbud; og politi til fots. I København er det etablert et prosjekt (*Tryg Den Af*) mellom København kommune, Diageo (en av verdens største alkoholprodusenter), og International Harm Reduction Association (en internasjonal NGO) som har kjørt målrettede kampanjer mot det unge nattelivspublikummet, sertifisering av serveringssteder, kursing, forum for aktørene, og et bølleregister som gjør det mulig å utestenge bråkmakere (det siste vil nok av personvern hensyn være vanskelig i Norge). I Stockholm driver Centrum for psykiatriforskning et prosjekt kalt STAD, med et utelivsprogram som er et strukturert, langsiktig samarbeid mellom bransjen og myndighetene, og inneholder både kursing og kontroll. STAD-samarbeidet er grundig evaluert via en rekke vitenskapelige arbeider, deriblant med bidrag fra Rossow og Nordström.⁹

RN sier seg "enige i at det ville ha vært interessant å kunne relatere betydningen av skjenketidsendringer til andre tiltak som kan begrense voldsomfanget", men de skriver videre at slike modeller vil være enormt omfattende og trolig ikke gjennomførbare. Videre skriver de at "i et fremtidig arbeid vil vi imidlertid utvikle diskusjonen om hvordan den effekten vi har funnet av endringer i skjenketid, er sammenlignet med effekt av andre tiltak". Med andre ord innrømmer RN at funnene deres ikke er entydige, og at modellen deres trenger forbedring. Dette alene er nok for meg til å konkludere at studiet ikke kan nyttes når vi skal vurdere tiltak mot utelivsvold.

Vi vet fremdeles ikke hvor mye av voldsreduksjonen som faktisk kan tilskrives skjenkereduksjon, mest av alt fordi studiet deres mangler disse omfattende kontrollvariablene.

Tidsserieanalysens problem

Dersom flere tiltak iverksettes samtidig, vil man i en tidsserieanalyse ikke kunne avdekke effekten av enkelttiltak, fordi man bare har informasjon om tidspunktet. Hvis det både kommer nattbusser, mer politi, flere utesteder og flere taxiholdeplasser samtidig som det kommer en endring i skjenketidene, vil man ikke kunne se på voldsstatistikken før og etter det gitte tidspunktet og peke på hvilke(n) faktor som førte til en endring i volden. Det eneste man kan si noe om er hvorvidt den samlede pakken av tiltak hadde en effekt.

Har man flere byer med ulike tiltakspakker, der den eneste fellesnevneren er endring i skjenketid, vil man likevel kunne argumentere for at endringen i skjenketid hadde en effekt, men man vil ikke få ut en isolert effekt av skjenketid uten at man også har kontrollert for de andre tiltakene.

For at det skal være mulig i tidsserieanalyse å isolere effektene, må de ulike tiltakene komme separat fra hverandre i tid. Dette er viktig å ha i mente dersom man planlegger evaluering av fremtidige tiltak.

Ingen kontroll for ikke-lineære sammenhenger

En annen modellsvakhet ved RNs analyse er at de bare sjekker om det er lineære sammenhenger i dataene. Dette blir spesielt underlig når de beskriver en mulig interaksjonseffekt

⁹ <http://www.stad.org/vetenskapliga-arbeten/>. Tryg Den Af har også blitt evaluert: <http://trygdenaf.dk/assets/files/Evalueringssrapport%20inkl.%20appendix%20%28skrivebeskyttet%29.pdf>

mellom alkoholkonsum og kundetetthet (s. 4). Interaksjoner kan ta mange former, men de er sjeldent lineære.¹⁰ Hvis man bruker en lineær modell på (den teoretiske) interaksjonen mellom konsum og kundetetthet, sier man i praksis at vold avtar gradvis med økende alkoholkonsum og med økende kundetetthet. Jeg ville trodd det var motsatt.

Ideelt sett skal en regresjonsmodell som inneholder en interaksjon også inneholde hver av de separate variablene som er interagert. Jeg har forståelse for at det ville vært vanskelig i dette studiet å kontrollere både for alkoholkonsum og kundetetthet, men hadde de gjort det ville de kunne brukt et faktisk mål på interaksjonen heller enn et teoretisk mål (skjenketider).

Proxy

Man kan nok bruke skjenketider som en proxy (erstatningsvariabel) for alkoholkonsum, men fordi skjenketider vil være sterkt korrelert med kundetetthet, burde det vært med en kontroll for kundetetthet i studiet. En proxy på dette kunne vært antall innbyggere per serveringssted.

Byenes størrelse

RN har inkludert antall innbyggere i sitt mål på vold, ved at de har regnet antall voldstilfeller per 100 000 innbyggere per kvartal. Her fremgår det av tabell 2 at det er innbyggertall i bykommunene som legges til grunn. Et av problemene med dette er at en by ikke er det samme som en kommune. Byene betjener omland, og er sentre i bo- og arbeidsmarkedsregioner. Skal man beregne antallet voldstilfeller per 100 000 innbyggere i Bergen, bør man vurdere å ta med den regionen det går nattbuss til (Askøy, Sund, Fjell, Øygarden, Os, Osterøy, Lindås, Meland, Radøy). Da går Bergen fra å regnes som en kommune med 260 000 innbyggere til å være et regionsenter med 370 000 innbyggere (2011-tall).

Et annet problem er at det fremgår som om byene med flere runder med endring i skjenketider bare er vektet mot sin innbyggerstørrelse én gang. For Bergens tilfelle er det oppgitte innbyggertallet 2009-tall. Bergen hadde en liberalisering av skjenketidene i 2004. På det tidspunktet var det 14 000 færre innbyggere i Bergen, og 24 000 færre innbyggere i "nattbussregionen".

Metodesvakheter

For simple metoder

Uten beskrivelse av dataene er det min oppfatning at RN har benyttet for simple modeller, og for simple metoder. Jeg er tilhenger av å bruke så simple metoder som mulig, men da må man rettfærdiggjøre bruken av dem. Det har RN ikke gjort. Det vi vet er at RN bruker data om anmeldte voldstilfeller i 18 byer, og at variasjon mellom byene kontrolleres for på to måter: først en modell der byenes ulike voldsnivå er kontrollert for via kontrollvariabler for hver by, deretter separate analyser av hver by som slås sammen i et uvektet gjennomsnitt.

Burde brukt tonivåmodell

Når vi har variasjon i nivå mellom byer, og variasjon over tid *innad* i byer, vil det være bedre å bruke en tonivåmodell der man kan kontrollere for variasjonen på hver sitt nivå. En såkalt *fixed-effects modell* som RN har brukt tillater ikke at det kan være ulik effekt av endring i

¹⁰ Hvis produktet av interaksjonen faktisk har en lineær effekt, må effekten av hver av faktorene avta gradvis tilsvarende kvadratrotten av produktet.

skjenketid i ulike byer. Det eneste den kontrollerer for, er at byenes voldsnivå kan variere. Effekten av tiltaket blir derimot målt likt.

I stedet burde RN brukt en modell som tillater at *både* voldsnivået og effektene varierer. De har gjort noe som ligner, ved å kjøre separate analyser for hver by, og så regne ut en gjennomsnittseffekt. En av fordelene ved å legge alt inn i én modell fremfor å kjøre separate analyser for hver by, er at man lettere kan regne ut ikke-lineære effekter i førstnevnte. Ved hjelp av statistikkprogrammet STATA (som de har brukt til noen av analysene) har de redskapene til å gjøre dette.

Hvis vi godtar at sammenhengen er lineær, er RNs metode grei nok, men den er tungvinn, og den burde vært hovedanalysen, ikke en av kontrollanalysene.

Feil å bruke uvektet gjennomsnitt

Uvektet gjennomsnitt er feil metode til sammenstilling av sett med liknende regresjonsanalyser, fordi forskjeller mellom regresjonene ikke kommer frem ved uvektet gjennomsnitt alene.

Det finnes metoder for å samle regresjonsresultat på en slik måte at man kan få frem variasjon mellom regresjonene. Dette kalles metaregresjon (regresjon av regresjonsanalyser). Dette skulle vært gjort, men ikke uten også å presentere primærregresjonene (enkeltregresjonene for hver by). Man kan spørre seg hvorfor primærregresjonene *ikke* er presentert. Er det her enkeltresultat som forstyrrer det overordnede bildet forfatterne ønsker å presentere?

Jeg savner altså først en enkel tabell som oppsummerer funnene for hver by (er det for eksempel enkeltfunn her som forteller en annen historie enn hovedfunnet, er det enkeltfunn som "trekker" i gjennomsnittet?), deretter savner jeg en utregning av vektet gjennomsnitt, der vektingsgrunnlaget kunne vært byenes størrelse. Hvis vektet og uvektet gjennomsnitt er likt, er metoden grei nok, men hvis det er forskjell på det vektete og det uvektete gjennomsnittet vil det være variasjon mellom de separate analysene som ikke fanges opp i et gjennomsnitt. Der kommer metaregresjon, altså en sekundær regresjonsanalyse, inn for å fange opp denne variasjonen.

Dette er med andre ord tungvint. Anbefalingen er altså å kjøre tonivåmodellen innenfor samme analyse.

Meta-analyse hadde vært mulig

Hvis RN hadde oppgitt resultatene fra primærregresjonene i artikkelen, komplett med teststatistikk (metadata), ville jeg kunne kjørt metaregresjonsanalysen for dem. Men slik artikkelen nå foreligger gir de ikke informasjon som gjør dette mulig. I en meta-analyse måtte jeg altså forkastet artikkelen fra datagrunnlaget. Det mangler for mye informasjon rundt analysene som er gjort, og artikkelen er svært vanskelig å gjennomskue. Vi vet ikke hva som faktisk er målt, og hvor gode disse målene er. Et elementært krav til forskningsartikler er at de skal være reproduerbare av andre enn forfatterne, med den informasjonen som er gitt i artikkelen. Etter min mening er denne det ikke.

Publiseringsprosess

Rossow og Nordström skriver i sitt tilsvarende på SIRUS sine hjemmesider at de ikke har rapportert estimatene bare fordi de er statistisk signifikante; vi har rapportert estimatene vi fant ved

hjelp av tre ulike tilnærminger. Alle disse estimatene er signifikante, men vi hadde rapportert dem, selv om de ikke hadde vært signifikante". Videre spør de "om Menon Business Economics og Stian Ludvigsen i så fall hadde publisert sine kritiske kommentarer til vår artikkel".

Det siste spørsmålet har kommet flere ganger fra SIRUS-forskere. Jeg fikk høre det på et møte som helseministeren arrangerte i Bergen 24. oktober, og Thor Nordstrøm mente i Dagbladet 6. oktober at "det er påfallende at kritikerne har egeninteresser i skjenkebestemmelsene".

Dette er å gå etter mannen, heller enn ballen – noe som ikke kan oppfattes som noe annet enn et svakhetstegn fra SIRUS. Når de ikke kan forsvare seg mot kritikken, angriper de kritikeren. Min kritikk er fullt ut etterprøvable, og står på egne ben uavhengig av hvilke interesser jeg har i å kritisere SIRUS.

Jeg kan godt innrømme at jeg ikke hadde brukt tid på å gå rapporten deres i sømmene dersom de hadde kommet til et resultat som ikke var interessant. *Heri ligger viktig forskningssoziologisk innsikt.* For det første ville det blitt lite inspirert forskning om forskere arbeidet med temaer de ikke hadde noen interesse i. Vi må derfor akseptere at forskere har interesse i det de gjør. For det andre styrer hva som er interessant også hva som blir publisert.

Robusthetssjekk

Sett at man utfører et dårlig designet og gjennomført studie, og kommer til et uventet resultat. Da må man kunne forsvare resultatet sitt mot kritikere og andre forskere som er tilhenger av en rådende forventning om en helt annen sammenheng. Dette kan gjerne være forskeren selv. Man setter seg så ned, ser om man har gjort noe galt, finner feilene, retter på dem og eventuelt forkaster hele studiet før det i det hele tatt har blitt sendt til en redaktør. Sender man det til en tidsskriftsredaktør, vil redaktøren gjøre de samme vurderingene før han evt. forkaster det eller sender det til fagfeller for vurdering om studiet er egnet for publisering. Fagfellene gjør den samme kritiske vurderingen. Det dårlig designede og gjennomførte studiet som kommer til et uventet resultat har særdeles dårlige odds for å komme gjennom publiseringsprosessen.

Dels kommer det ikke gjennom fordi det er dårlig, dels kommer det ikke gjennom fordi det kommer til et uventet resultat, men mest av alt kan det stoppe opp fordi det uventede resultatet får alle som er involvert i publiseringen til å kikke nøye gjennom artikkelen for å finne eventuelle feil – feil som kunne blitt oversett hvis forskeren hadde kommet til et forventet resultat.

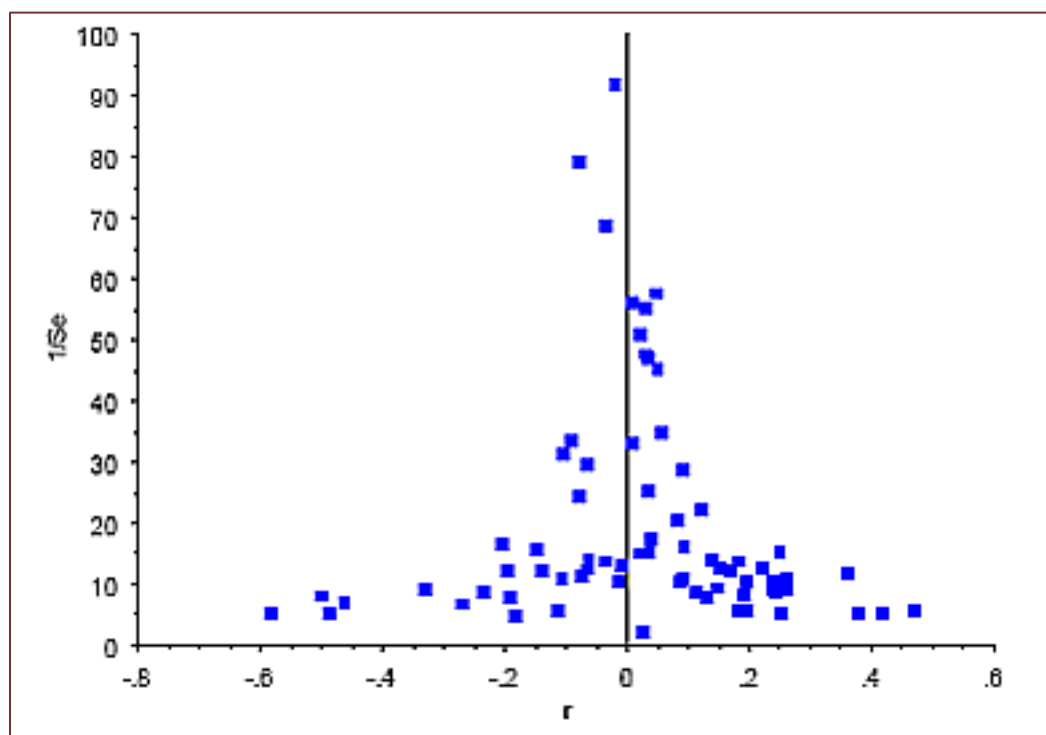
For de enkelte par av studier (ventet/uventet resultat) er sannsynligheten stor for at feil uansett blir fanget opp. Dette er delvis grunnen for at man har så omfattende publiseringsprosesser for forskning. Men for en samlet litteratur vil den ørlille sjansen det er for at det blir stoppet flere dårlige studier med uventet resultat enn dårlige studier med ventet resultat føre til at det blir publisert mer feil med "rett" resultat enn feil med "feil" resultat.

Isolert sett er det en fordel at feil blir stoppet, men hvis det stoppes flere feil med "feil" resultat enn med "rett" resultat, vil den samlede litteraturen bli vridd i retning av det forventede.

Selektiv publisering

Faktisk er den omtalte "ørlille" sjansen ikke så rent liten. Mine tidligere kollegaer Tom Stanley og Chris Doucouliagos har gjennomgått meta-analyser fra 65 ulike forskningsfelt innen økonomi (rundt 2 000 studier) og funnet at rundt 2/3 av forskningsfeltene er preget av slik selektiv publisering.¹¹

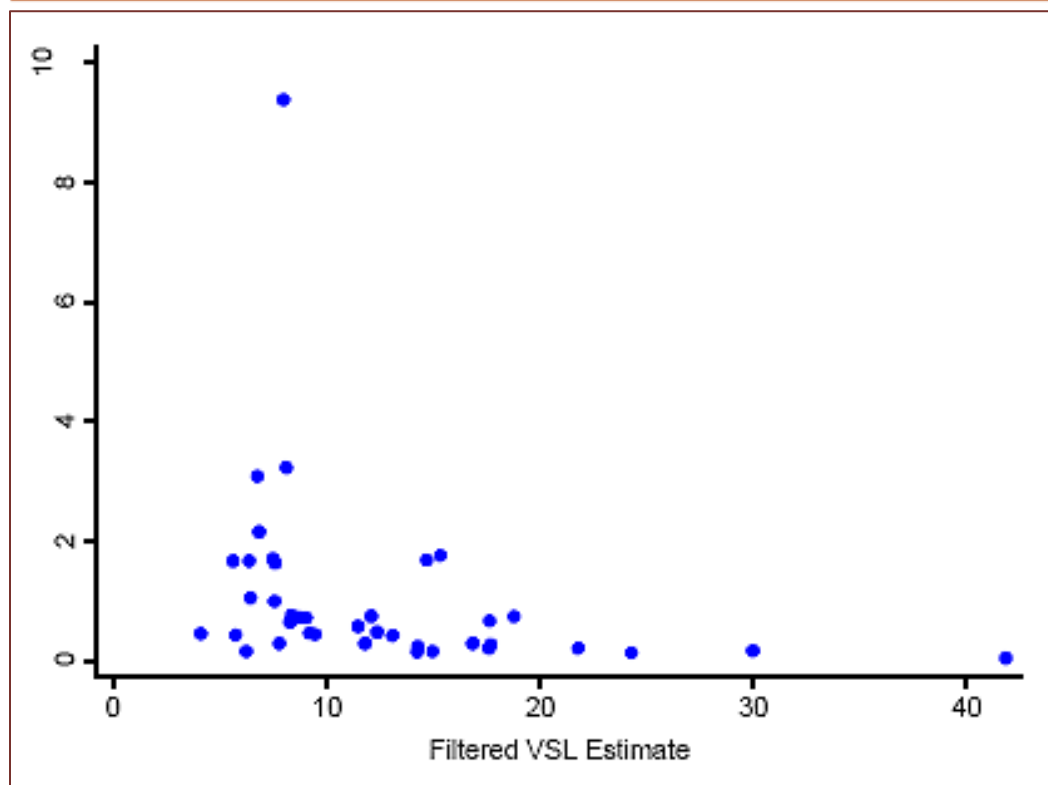
Selektiv publisering betyr rett og slett at noen resultater foretrekkes fremfor andre. Forfattere, redaktører og fagfeller kan foretrekke resultat som tydelig bekrefter rådende oppfatninger heller enn uklare eller motstridende resultat. Dette fører til en fordreining av virkeligheten, som reproduseres og forsterkes når resultatene blir samlet, referert, tatt inn i litteraturoversikter, osv. Den vitenskapelige produksjon reproduserer dermed sine feil, og dette kan få særdeles alvorlige konsekvenser, spesielt innen medisinsk forskning.



Figur 4: Presisjon og effekt i studier av fagforeningsproduktivitet¹²

¹¹ http://www.deakin.edu.au/buslaw/aef/workingpapers/papers/2008_06eco.pdf, se også Stanley, Jarrell & Doucouliagos 2010: "Could It Be Better to Discard 90% of the Data? A Statistical Paradox", *The American Statistician* 64(1): 70-77.

¹² Doucouliagos & Laroche (2003): "What Do Unions Do to Productivity? A Meta-Analysis", *Industrial Relations* 42 (4), 650-691.



Figur 5: Presisjon og effekt i studier av verdien av et statistisk liv¹³

Traktdiagram

Hvis man tenker seg et diagram der man plotter inn studiene etter resultat og presisjonsnivå, vil studier med høyt presisjonsnivå ha små feilmarginer, mens studier med lavt presisjonsnivå vil ha store feilmarginer. Hvis studiene publiseres uavhengig av hvilket resultat de kommer frem til, vil et slikt diagram tegne seg som en omvendt trakt, slik som figur 4.

Hvis man derimot har et forskningsfelt der man ikke publiserer resultat på "feil" side av streken, vil man bare få en halv trakt, som i figur 5.

Bare én mulig retning på resultatene?

Innen en del forskningsfelt er det helt urimelig å vente seg at resultatene kan gå i begge retninger. Sett for eksempel at man studerer økonomiske effekter av bruk av sykkelhjelm. *Alt annet like*, vil ingen foreslå at samfunnet sparer penger på økning i antallet hodeskader. Dermed vil alle studier som beskriver økonomiske effekter av sykkelhjelmbruk enten være ikke-signifikante eller positive, men sjelden negative.

Problemet blir da at man overdriver effekten av sykkelhjelmer dersom man refererer litteraturen på vanlig måte (studie A finner ..., studie B finner ..., ..., studie N finner ...; gjennomsnittet av studie A til N viser at ...). Siden negative resultater ikke kan forekomme, vil gjennomsnittet av resultatene være vridd i retning av de mest positive funnene, fordi det ikke finnes negative resultat som kan balansere alle de positive funnene. *Sykelhjelmer er ikke et tenkt eksempel*: Carpenter og Stehr har kommet fram til at nettoverdien av sykkelhjelmlover i USA

¹³ Doucouliagos, Stanley & Giles (kommer): "Are Estimates of the Value of a Statistical Life Exaggerated?", *Journal of Health Economics*. Se http://www.deakin.edu.au/buslaw/aef/workingpapers/papers/2011_2.pdf

er \$800 per syklist.¹⁴ Når Doucouliagos m.fl. justerer for denne manglende balansen, kommer de til at den estimerte verdien av sykkelhjelmer bare er \$200 per syklist.¹⁵ Man har altså fremdeles en økonomisk gevinst av sykkelhjelmlovene, men langt fra så sterk som den selektivt publiserte litteraturen hevder.

Andre eksempler vil være å finne innenfor studier som verdsetter verdien av et alminnelig menneskeliv, lønnsforskjeller mellom kvinner og menn, miljøskader, kriminalitet, og røyking (man kan vanskelig akseptere studier som sier at samfunnet tjener på at den jevne borger dør, at kvinner har høyere lønn enn menn, at luften blir renere av mer biltrafikk, at forsikringsselskapene sparer penger på forsikringssvindler, at røykere lever lengre) – og innen studier av alkoholkonsum og vold.

Alle disse forskningsfeltene kan man forvente vil fordele seg slik som figur 5. Enkelte studier innenfor disse feltene kan selvsagt ha kommet til korrekt resultat, men gjennomsnittet av studiene vil ikke ha gjort det. Selv om den enkelte forsker er redelig, vil litteraturen i sin helhet være vridd (ha bias).

Har Rossow og Nordström utøvd selektiv publisering?

Når RN spør meg om jeg ville kommet med min kritikk dersom de hadde kommet til et annet resultat, kan jeg spørre om det samme: ville *de* publisert sine resultat dersom de hadde kommet til et annet resultat?

Vi kan håpe at svaret er ja, men artikkelen til RN gir hint om det motsatte:

Some studies have demonstrated associations in the expected direction, i.e. an increase in violence rates following increased trading hours and vice versa, whereas other studies have reported no association or *even* a decrease in violence rates with increased trading hours. We will first suggest possible explanations of how changes in trading hours may affect violence rates and then address possible methodological explanations for the inconsistent findings.

(s. 4, min utheving)

Her har Rossow og Nordström en klar forventning til sammenhengen mellom skjenketider og vold, og de vil til og med søke etter forklaringer på de motstridende funnene (men ikke de forventede funnene). Dette er et signal om at de ville sett med helt andre øyne på sitt eget studie dersom de hadde kommet til et ikke-forventet resultat.

Det kunne vært interessant å se hvor RN sine resultat er plassert i forhold til resultatene innen sitt forskningsfelt, men dessverre har de ikke oppgitt nok informasjon til at det er mulig å standardisere resultatene på samme vis som i figur 4 og 5.

Hva hvis?

Det har liten hensikt å diskutere funnene i artikkelen til Rossow og Nordström før de kan dokumentere at funnene deres er korrekte, og at de er relevante for den utelivspolitikken de foreslår at vi fører. Likevel vil jeg kort kommentere funnet deres som om det var korrekt.

¹⁴ Carpenter & Stehr (2010): "Intended and Unintended Effects of Youth Bicycle Helmet Laws", *NBER Working Paper* 15658.

¹⁵ Jf. fotnote 13.

I sitt tilsvaer skriver RN at legemsfornærmelser utgjør 80 prosent av voldstilfellene som ligger til grunn for rapporten. I rapporten har de behandlet legemsfornærmelser og legemsbeskadigelser likt. Dette mener jeg man kan stille seg kritisk til, men hva nå om det er korrekt å behandle legemsfornærmelser likt med legemsbeskadigelser? I så fall vil én times reduksjon i skjenketiden bare medføre reduksjon på *ett* tilfelle av skade per 100 000 innbyggere per kvartal. De resterende fire tilfellene vil være legemsfornærmelser.

Det vil være lett for utelivsnæringen selv å manipulere statistikken på legemsfornærmelser, fordi dørvakter og serveringspersonale utsettes for det hver helg. Hvis bransjen gjør en ekstra innsats for å anmelde *alle* forhold, vil voldsstatistikken gå opp. Hva om bransjen gjorde en slik ekstrainsats etter at det kom en reduksjon i skjenketidene? Hvordan ville det blitt tolket av forskerne?

Det er noe annet med den volden som ender på legevakten eller med innleggelse.

Om det er én ekstra person eller fem ekstra som anmelder legemsbeskadigelse per 100 000 innbyggere per kvartal er av en viss betydning for publikum og myndigheter. Men skal vi tro Rossow og Nordström, har vi altså å gjøre med én ekstra person.

I Bergen var reduksjonen i skjenketid på en halv time, og en ytterligere innskrenkning vil da altså være på 30 minutter. Dersom RN har regnet korrekt vil Bergen, med 260 000 innbyggere, få fem – FEM – sparte legemsbeskadigelser PER ÅR! I ytterste konsekvens kan vi altså velge å påføre serveringsbransjen et stort inntektstap for å spare fem mennesker – som frivillig har gått på byen – fra et ufrivillig legevaktsbesøk.

Skal vi tro RN på at effektene er lineære og symmetriske, så vil en økning i skjenketidene til 03:00 for alle serveringsstedene i Bergen medføre fem flere legemsbeskadigelser per år.

På stående fot kjenner jeg ikke kostnadene per legemsbeskadigelse, men det vil være en rimelig smal sak å finne ut av. Noe av denne volden kan være invalidiserende, og således meget kostbar for både offer og samfunn, men likevel har Rossow og Nordström – hvis de har rett – gitt oss grunnlag for et regnestykke der det kan være dårlig samfunnsøkonomi i å redusere skjenketidene – og god samfunnsøkonomi i å utvide dem.

Hva så hvis man gjennomførte skjenketidsinnskrenkninger?

I debatter om norske skjenketider blir det ofte vist til at det er mange europeiske land med kortere skjenketider enn i Norge. Implisitt burde det jo da være mulig for norske serveringssteder å overleve på kortere skjenketid også?

Så enkelt er det ikke. Norske serveringssteder har høye lønnskostnader. Uten at jeg har tall på det, er det rimelig å anta at lønn utgjør en høyere del av driftskostnadene enn i de fleste andre land. Videre må norske serveringssteder forholde seg til et langt strengere lovverk enn de fleste andre land, og kravet til vakthold er meget strengt. Serveringsstedene må forholde seg til strenge forskrifter vedrørende matservering, alkoholservering, brannvern, arbeidstid, arbeidsmiljø, etc. Dermed påløper også strenge krav til internkontroll. Endringene i vaktloven gjorde at vaktholdet ble enda dyrere for mange norske utesteder (som ikke lengre kan drive egenvakthold). Norske utesteder må med andre ord ha nok inntekter til å kunne lønne sitt serveringspersonale, betale for innleide dørvakter og følge et vell av forskrifter.

Med unntak av endringene i vaktloven mener jeg alt dette er bra, men uten inntektene fra de tidspunktene med høyest omsetning vil mange steder berettiget kunne frykte at de ikke får

dekket inn sine driftskostnader. Risikoen blir da at "bare" de som opererer på akkord med forskriftene overlever. Det vil bety dårligere arbeidsmiljø, vakthold og internkontroll, og kan medføre det paradoksale i at utelivsvolden øker som en konsekvens av reduserte skjenketidder.

Konklusjon

Rapporten til SIRUS-forskerne Rossow og Nordström kan godt være god nok for publisering i *Addiction*, men den gir et dårlig grunnlag for politikktutforming. Rapporten er et særdeles svakt grunnlag for forslag om å redusere skjenketidene og ta inntekter og sysselsetting fra en stor og lovlig bransje, og for å innskrenke en for mange viktig sosial arena. Det er ikke gitt at metoder som er gode nok for publisering, er gode nok til å rettferdiggjøre en politikk som sender servitører og dørvakter på jakt etter nytt arbeid, og som reduserer allmennhetens kultur- og fritidstilbud. Beviskravet ved en politikk som potensielt har store menneskelige, kulturelle og samfunnsøkonomiske konsekvenser må være betydelig strengere enn beviskravet i en fagfelle vurdering der den eneste konsekvensen er hvorvidt forskerne får publiseringspoeng eller ikke.

Jeg har i dette notatet gitt *min* fagfelle vurdering av Rossow og Nordströms arbeid. Jeg er ikke rusforsker, men jeg er samfunnsviter med tung spesialisering i metode, og min fagfelle vurdering står på egne ben, uavhengig av min politiske interesse i temaet. Til forskjell fra RNs rapport, er min kritikk lett å etterprøve.

Temaet er interessant, og jeg skulle gjerne sett et godt utført studie rundt den problemstillingen som RN har forsket på. Det ligger et godt datamateriale i bunn, men det er dårlig utnyttet. På grunn av feil bruk av metode, tvilsom modellbruk, sammenblanding av ulike typer vold der ofrene har ulik motivasjon for anmeldelse, manglende flid med dataene for å minimere støy og ekstremavvik, ingen fjerning av gjengangere, ingen halvering av tilfeller der gjerningsmann og offer har anmeldt hverandre, manglende kontroll for kundetetthet, og manglende kontroll for andre faktorer som påvirker voldsnivået, er risikoen for feilslutning stor, og det er lite sannsynlig at effekten av skjenketidsinnskrenking er så sterk som SIRUS sine forskere har kommet frem til.

Rapporten mangler beskrivelse av voldsstatistikken som er lagt til grunn, og jeg har grunn til å anta at det ikke er så perfekt samvariasjon mellom legemsbeskadigelser og -fornærmelser som RN har forutsatt. Det er derfor grunn til å spørre om rapporten til RN faktisk beskriver de sammenhengene de er ute etter å analysere. I verste fall sitter vi med en rapport der signifikans og høy forklart varians er oppnådd som i det fjerde tilfellet i Anscombes kvartett: altså uten reell forklaringskraft. Min anbefaling til Rossow og Nordström er at de forsøker på nytt. De bør holde tilbake endelig publisering av rapporten inntil de har ryddet opp i de store svakheter som den har.

Kritikken min er konstruktivt ment. I dette notatet har jeg kommet med en rekke forslag til hvordan SIRUS kan forbedre sin analyse av en eventuell sammenheng mellom skjenketidsendringer og vold. I tillegg til å påpeke åpenbare svakheter, har jeg foreslått at man bruker flernivåmodeller for å kontrollere for variasjon både mellom byer og innad i byer, som alternativ har jeg foreslått regresjon av regresjonsanalyser (metaregresjon), jeg har gitt forslag til hvordan man kan minimere støy rundt resultatene, og jeg har foreslått at man i fremtiden

iverksetter separate tiltak på separate tidspunkt dersom man ønsker å kunne evaluere effekten av tiltakene i ettertid.

Jeg har også brukt noen illustrasjoner for å begrunne min kritikk – og gi noen advarsler. Anscombes kvartett er fire ulike datasett hvor man kommer til samme konklusjon dersom man rutinemessig kjører ut statistikk på dem uten først å studere detaljene i datasettene. Dette illustrerer hvor viktig det er å gjøre en grundig vurdering av hvorvidt man har håndtert dataene riktig og benyttet korrekt metode.

Dette er viktig av andre årsaker også: Forskningslitteratur der det er "umulig" å tenke seg resultat som motstrider teori, vil naturlig falle for selektiv publisering der det blir publisert mer feil med "rett" resultat enn feil med "feil" resultat. Dette skjer ikke bare til tross for at man har redaktører og fagfeller som vurderer artiklene før publisasjon, men ofte også på grunn av at man har redaktører og fagfeller som lar seg styre av rådende oppfatninger.¹⁶ Da blir det ekstra viktig å påse at man ikke gjør feil.

Det er mulig Rossow og Nordström har gjort alt rett, men de har ikke gjort rede for hvordan, og i mine øyne brutt et elementært krav til forskningsartikler om at de skal være reproduerbare av andre enn forfatterne.

Skulle RN likevel ha rett, står vi med den – for mange provoserende – påstand at det kan være god samfunnsøkonomi i å *utvide* skjenketidene: I Bergen var reduksjonen i skjenketid på en halv time, og en utvidelse vil da være tilsvarende. I følge resultatene til RN gir det fem flere legemsbeskadigelser per år. Kostnadene ved dette kan være mindre enn inntektene, noe som kan gjøre *regnestykket* verd å vurdere. Når man legger til sannsynligheten for at RN har regnet feil på effektene, kan kostnadene bli enda mindre. RNs rapport er i seg selv like lite egnet som grunnlag for en utvidelse av skjenketidene som en innskrenkning. Det trengs mer forskning, men de har lagt grunnlaget for å undersøke om det kan ha noe for seg å utvide heller enn innskrenke skjenketidene.

Hvis vi likevel følger rådet om å innskrenke med ytterligere en halv time, kan regnestykket for serveringsbransjen medføre at det blir de som opererer på akkord med forskriftene som har best sjanse for å overleve. Det vil bety dårligere arbeidsmiljø og sikkerhet, og potensielt *mer* utelivsvold som konsekvens av reduserte skjenketider. Uten kunnskap kunne man sagt at det hadde vært en uintendert konsekvens av politikken, men når det i stedet er *feil* kunnskapsgrunnlag vil det være en grov forsømmelse.

¹⁶ Se f.eks. fotnote 11 samt Begg (1994): "Publication Bias" i Cooper & Hedges (red.): *The Handbook of Research Synthesis*, New York: Russel Sage Foundation, og Hunter & Schmidt (2004): *Methods of Meta-Analysis – Correcting Error and Bias in Research Synthesis*, Thousand Oaks: Sage Publications.

Vedlegg

Tabell 1: Regresjoner bak figur 1 og 2

. regress y1 x1, beta

Source	SS	df	MS			
Model	27.5100011	1	27.5100011	Number of obs =	11	
Residual	13.7626904	9	1.52918783	F(1, 9) =	17.99	
Total	41.2726916	10	4.12726916	Prob > F =	0.0022	
				R-squared =	0.6665	
				Adj R-squared =	0.6295	
				Root MSE =	1.2366	

	Coef.	Std. Err.	t	P> t	Beta
y1					
x1	.5000909	.1179055	4.24	0.002	.8164205
_cons	3.000091	1.124747	2.67	0.026	.

. regress y2 x2, beta noheader

	Coef.	Std. Err.	t	P> t	Beta
y2					
x2	.5	.1179638	4.24	0.002	.8162365
_cons	3.000909	1.125303	2.67	0.026	.

. regress y3 x3, beta noheader

	Coef.	Std. Err.	t	P> t	Beta
y3					
x3	.4997273	.1178777	4.24	0.002	.8162867
_cons	3.002455	1.124481	2.67	0.026	.

. regress y4 x4, beta noheader

	Coef.	Std. Err.	t	P> t	Beta
y4					
x4	.4999091	.1178189	4.24	0.002	.8165214
_cons	3.001727	1.123921	2.67	0.026	.

Tabell 2: Lineær og kvadratisk regresjon i eksempel 2

. regress y2 x2

Source	SS	df	MS			
Model	27.5000024	1	27.5000024	Number of obs =	11	
Residual	13.776294	9	1.53069933	F(1, 9) =	17.97	
Total	41.2762964	10	4.12762964	Prob > F =	0.0022	
				R-squared =	0.6662	
				Adj R-squared =	0.6292	
				Root MSE =	1.2372	

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
y2						
x2	.5	.1179638	4.24	0.002	.2331475	.7668526
_cons	3.000909	1.125303	2.67	0.026	.4552978	5.54652

. regress y2 x2 x2sq

Source	SS	df	MS			
Model	41.276274	2	20.638137	Number of obs =	11	
Residual	.000022378	8	2.7973e-06	F(2, 8) =	.	
Total	41.2762964	10	4.12762964	Prob > F =	0.0000	
				R-squared =	1.0000	
				Adj R-squared =	1.0000	
				Root MSE =	.00167	

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
y2						
x2	2.780839	.0010401	2673.72	0.000	2.778441	2.783238
x2sq	-.1267133	.0000571	-2219.22	0.000	-.126845	-.1265816
_cons	-5.995735	.00433	-1384.70	0.000	-6.00572	-5.98575

Tabell 3: Lineær og robust regresjon i eksempel 3

```
. regress y3 x3
```

Source	SS	df	MS			
Model	27.4700075	1	27.4700075	Number of obs =	11	
Residual	13.7561905	9	1.52846561	F(1, 9) =	17.97	
Total	41.2261979	10	4.12261979	Prob > F =	0.0022	
				R-squared =	0.6663	
				Adj R-squared =	0.6292	
				Root MSE =	1.2363	

y3	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x3	.4997273	.1178777	4.24	0.002	.2330695	.7663851
_cons	3.002455	1.124481	2.67	0.026	.4587014	5.546208

```
. rreg y3 x3
```

```
Huber iteration 1: maximum difference in weights = .08975626
Huber iteration 2: maximum difference in weights = .0285571
Biweight iteration 3: maximum difference in weights = .13303087
Biweight iteration 4: maximum difference in weights = .01007869
Biweight iteration 5: maximum difference in weights = .00288686
```

```
Robust regression
```

```
Number of obs = 10
F( 1, 8) = 9.3e+05
Prob > F = 0.0000
```

y3	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x3	.3453611	.0003578	965.29	0.000	.3445361	.3461862
_cons	4.005864	.0032635	1227.49	0.000	3.998339	4.01339

Tabell 4: Lineær og robust regresjon i eksempel 4

```
. regress y4 x4
```

Source	SS	df	MS			
Model	27.4900007	1	27.4900007	Number of obs =	11	
Residual	13.7424908	9	1.52694342	F(1, 9) =	18.00	
Total	41.2324915	10	4.12324915	Prob > F =	0.0022	
				R-squared =	0.6667	
				Adj R-squared =	0.6297	
				Root MSE =	1.2357	

y4	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x4	.4999091	.1178189	4.24	0.002	.2333841	.7664341
_cons	3.001727	1.123921	2.67	0.026	.4592411	5.544213

```
. rreg y4 x4
```

```
Huber iteration 1: maximum difference in weights = 0
Biweight iteration 2: maximum difference in weights = .11763749
Biweight iteration 3: maximum difference in weights = .00005683
```

```
Robust regression
```

```
Number of obs = 10
F( 0, 9) = 0.00
Prob > F = .
```

y4	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x4	(omitted)					
_cons	7.000496	.4244469	16.49	0.000	6.040331	7.960662