

Johs. Hjellbrekke, Helene Ingierd og Matthias Kaiser

DISKUTABEL FORSKNINGSPRAKSIS: holdninger og handlinger



UNIVERSITETET I BERGEN



De nasjonale
FORSKNINGSETISKE
KOMITEENE



Høgskulen
på Vestlandet

Johs. Hjellbrekke, Helene Ingierd og Matthias Kaiser

Diskutabel forskningspraksis: holdninger og handlinger

Andre delrapport fra arbeidsgruppen i forskningsprosjektet RINO
(Research Integrity in Norway)



UNIVERSITETET I BERGEN



De nasjonale
FORSKNINGSETISKE
KOMITEENE



Høgskulen
på Vestlandet

Oppsummering

Denne rapporten omhandler holdninger til, og egne erfaringer med diskutabile forskningspraksiser (QRP). En multivariat analyse av hhv. 8 holdningsvariabler og 6 praksisvariabler avdekker flere klare forskjeller mellom ulike grupperinger av norske forskere.

Vi finner seks undergrupper av forskerholdninger. Disse har vi døpt «De konsekvente», «De fleksible», «De tvisynte», «De splittede», «De selektive» og «De diskutabile». Et lite flertall av forskerne avviser alle former for diskutabile praksiser, men de som aksepterer en eller flere slike praksiser er altså ikke en homogen gruppe. Aksepten for spesifikke diskutabile praksiser er relativt hyppig forekommende, men en generell aksept er et marginalt fenomen.

De enkleste analysene viser at en relativ høy andel – drøyt 40% – av respondentene har begått en eller flere diskutabile praksiser i løpet av de siste tre årene. Den etterfølgende multivariate analysen av egenrapporterte diskutabile praksiser, som vi legger frem her, avdekker tre undergrupper. Den største grupperingen, «De rene og ranke» har, vurdert ut fra hva de selv oppgir, svært lave betingede sannsynligheter for å bryte med god forskningsskikk. «De gavmilde» har langt høyere betingede sannsynligheter for å ha gitt et gaveforfatterkap. Også i denne analysen avdekkes en gruppe av «Diskutabile», som har klart høyere betingede sannsynligheter for selv å ha utført en diskutabel forskningspraksis. Selv om gruppen er liten, med kun 5% av respondentene, utgjør den likevel en risikogruppe.

Resultatene kan ha flere forklaringer, men det er like fullt problematisk at et lite mindretall aktivt har brutt med god forskningsskikk. Vi håper funnene åpner opp for videre diskusjoner om forskningsetiske normers status.

Innledning: Om variasjon i forskernes atferd og holdning til forskningsetiske normer

I delrapport I fra forskningsprosjektet RINO (Hjellbrekke mfl., 2018), presenterte vi status for forskningsetikk blant forskere i Norge i 2018. Enkelte funn var overraskende: Gitt flere år med forskningsetisk arbeid på institusjonene, styrket med en lov om forskningsetikk, så vi at bare vel 60% av forskerne rapporterte å ha deltatt i forskningsetisk opplæring, og for halvparten av disse var omfanget kun en dag eller mindre. Vi fant også at Norge ligger betydelig lavere på selvrapporing av forfalskning, fabrikasjon og plagiering (heretter «FFP») sammenlignet med tall fra internasjonale meta-analyser (Fanelli, 2009). RINO-undersøkelsen avdekket en nærmest total normkonsensus i holdninger til de mest alvorlige bruddene på god forskningsetikk. Undersøkelsen avdekket også høy grad av normkonsensus om det man gjerne kaller diskutabile forskningspraksiser (heretter QRP [«Questionable Research Practices»]). Men i disse spørsmålene var det også en noe høyere prosentdel som oppfattet en eller flere QRP-praksiser som lite problematiske eller uproblematiske. En enkel, sammenlignende analyse avdekket noen, men ikke store forskjeller mellom de ulike vitenskapelige disiplinene. Gjennomgående er variasjonen internt i fagområdene klart større enn variasjonen mellom dem.

I denne delrapporten skal vi analysere holdningene til og de egenrapporterte diskutabile praksisene i større detalj. Ut fra en hypotese om at det finnes en sammenheng mellom holdninger til QRP, og viljen til å selv utøve en slik praksis, skal vi avdekke ulike undergrupperinger av QRP-holdninger hos et utvalg norske forskere. Å avdekke slike strukturer er i seg selv viktig. Forskningsetikk – og dermed også integritet i forskning – er ikke statisk, og krever både institusjonelle tiltak, fortløpende diskusjoner, og internasjonal koordinering. Også av den grunn er det viktig å fange opp variasjonene som karakteriserer norske forskere i vår undersøkelse. Dersom resultatene i den kvantitative delen av RINO prosjektet skal veilede våre forskningsinstitusjoner i deres arbeid rettet mot en styrking av forskningsetikk, er det avgjørende at vi har kunnskap om viktige forskjeller innad i forskersamfunnet, her uttrykt som klynger i det statistiske materialet. Slike klynger kan peke på hvor det er meningsfullt å sette inn tiltak, om det så er bare å etablere fagovergripende debatt og diskusjon om QRP.

I det som følger beveger vi oss fra de univariate analysene som ble behandlet i delrapport I, over til en multivariat analyse av sammenhenger i holdninger og praksiser. Delrapport II er nødvendigvis litt mer teknisk orientert, men vi har prøvd å klargjøre forutsetninger og innhold i et språk som også er tilgjengelig for ikke-eksperter. Vi presenterer først to sett med problemstillinger, data og metode, før vi legger frem variasjonene vi har funnet. Til slutt gir vi vår avsluttende vurdering av datamaterialet.

1. Problemstillinger

Vi skal i to delanalyser ta opp to sett problemstillinger; et knyttet til holdninger til QRP og et knyttet til forskeres egenrapporterte praksiser:

- Hvor mange undergrupperinger av holdninger til QRP kan man finne blant norske forskere? Er det snakk om en, to eller flere grupperinger?
- Dersom det er flere undergrupperinger: Hvor store er de ulike gruppene? Er aksept for en eller flere former for QRP et relativt utbredt fenomen, eller er det snarere snakk om et marginalt fenomen som kjenner seg ut som et lite mindretall av forskere?
- Hvordan kan de ulike gruppene tolkes? Er det snakk om grupperinger av generell aksept, eller avvisning av QRP på tvers av flere variabler eller praksiser, eller bidrar enkeltvariabler eller praksiser til å definere en eller flere undergrupper i materialet?
- Hvor mange undergrupper av forskere finner vi når vi går over til å se på egenrapporterte diskutabile praksiser?
- Hva er de ulike grupperingenes «eksterne» eller strukturelle karakteristika? Hvilke sammenhenger finner vi mellom holdninger til QRP og variabler som alder, kjønn, stillingskategori og vitenskapelig disiplin?

2. Data

Data er hentet fra RINO-undersøkelsens spørreundersøkelse. Invitasjon til å delta ble sendt ut med e-post til 31 206 tilsatte ved norske universiteter, høyskoler og forskningsinstitutt. 7291 respondenter, eller 23,4%, fylte ut hele spørreskjemaet.¹

For å gripe bredden i materialet, har vi inkludert 8 variabler i analysen av holdninger til diskutabile praksiser:

- Å godta, bestemme eller tildele forfatterskap på annet enn vesentlig bidrag til et vitenskapelig arbeid («Gaveforfatterskap»)
- Å nekte eller å utelate forfatterskap til tross for vesentlig bidrag til et vitenskapelig arbeid («Nekte forfatterskap»)
- Å unnlate å informere kunnskapsbrukere og beslutningstakere om viktige begrensninger og/eller usikkerhet i datagrunnlag, analyse og/eller konklusjon («Unnlate å informere»)
- Å forandre en studies design, metode og/eller resultater etter press fra interessenter eller finansieringskilder («Endre design»)
- Å unnlate å varsle («Whistle Blowing») om alvorlige brudd på forskningsetiske retningslinjer («Unnlate å varsle»)
- Å inkludere irrelevante eller unødvendige kildehenvisninger i en publikasjon for å øke siteringsfrekvens til en kollega, et forskingsmiljø eller et tidsskrift («Irrelevant referanse»)
- Å bruke forskningsdata/materiale når eierskapet til dette er omstridt («Omstridte data»)
- Å gi inntrykk av å ha oppsøkt en kilde gjennom å kopiere andres kildehenvisninger («Kopiere kilder»)

¹ Selv om flere svarte på deler av skjemaet, har vi i de etterfølgende analysene avgrenset utvalget til kun å inkludere respondenter som fullførte undersøkelsen. Respondenter med manglende verdier på enkeltvariabler er også utelatt fra de multivariate analysene. I holdningsanalysen er i tillegg 20 respondenter med helt identiske svar på alle spørsmål ekskludert fra materialet, og i analysen av egenrapporterte diskutabile praksiser er alle respondenter som har svart «Ikke relevant for min type arbeid» utelatt. Det samme er 1 respondent som svarte at vedkommende hadde utøvd samtlige diskutabile praksiser. For ytterligere informasjon om datasettet viser vi til Delrapport 1 (Hjellbrekke mfl. 2018).

Opprinnelig er disse variablene målt på en 4-punktsskala, fra «Dette er svært problematisk» til «Dette er helt uproblematisk». I analysene som følger, er variablene gjort dikotome. «Svært problematisk» og «Ganske problematisk» er kodet sammen, da dette indikerer en avvisende holdning til den gitte praksisen. Likeledes er «Litt problematisk» og «Uproblematisk» kodet sammen, da dette indikerer en klart høyere eller en fullstendig aksept for praksisen. I de etterfølgende analysene skiller vi dermed mellom «Problematisk» og «Lite problematisk».

Til analysen av egenrapporterte diskutabile praksiser har vi kun tatt med seks av totalt ni variabler. At vi har utelatt flere praksisvariabler har både metodiske og substansielle årsaker. For det første avgrensers den statistiske metoden vi bruker, latent klasseanalyse (McCutcheon, 1987; Magidson & Vermunt, 2002), antallet variabler som kan inkluderes i analysen. Dette gjelder særlig når variablene er sterkt skjevfordelte, slik at en eller flere tabellceller blir tomme. For noen av variablene om egne diskutabile praksiser er dette tilfelle. For det andre har bakgrunnsanalyser avdekket et visst innslag av redundans mellom enkelte variabler. Der dette er tilfelle, har vi kun tatt med en av variablene. For det tredje har vi ønsket å få fram mulige forskjeller mellom vitenskapelige disipliner. Forskningen er organisert på ulike måter i ulike disipliner, og noen problemstillinger er mer aktuelle i enkelte disipliner enn i andre. Dersom dette også reflekteres i utførte QRP, er det viktig å avdekke.

Vi har derfor valgt ut disse seks variablene til analysen av egenrapporterte diskutabile praksiser (variabelnavn i parentes):

- Å godta, bestemme eller tildele forfatterskap på annet enn vesentlig bidrag til et vitenskapelig arbeid («Gaveforfatterskap»)
- Å nekte eller å utelate forfatterskap til tross for vesentlig bidrag til et vitenskapelig arbeid («Nekte forfatterskap»)
- Å unnlate å informere kunnskapsbrukere og beslutningstakere om viktige begrensninger og/eller usikkerhet i datagrunnlag, analyse og/eller konklusjon («Unnlate å informere»)
- Å forandre en studies design, metode og/eller resultater etter press fra interessenter eller finansieringskilder («Endre design»)
- Å unnlate å varsle («Whistle Blowing») om alvorlige brudd på forskningsetiske retningslinjer («Unnlate å varsle»)
- Å inkludere irrelevante eller unødvendige kildehenvisninger i en publikasjon for å øke siteringsfrekvens til en kollega, et forskingsmiljø eller et tidsskrift («Irrelevant referanse»)

I spørreskjemaet kunne respondentene svare «Ikke relevant for min type arbeid», «Nei», «Ja, en gang», «Ja, noen ganger», «Ja, mange ganger». I analysene som følger er også dette kodet om til dikotome variabler, med verdiene «Nei» og «Ja».

3. Metode

Den statistiske analysen er gjort ved hjelp av latent klasseanalyse (McCutcheon, 1987). Dette er en teknikk som avdekker latente grupper i et sett kategoriske data, dvs. data som ikke har et naturlig nullpunkt, som f.eks. kjønn, fakultetstilknytning, stillingskategori etc. Metoden tar utgangspunkt i analyseenhetenes fordelinger på tvers av et sett variabler, slik dette kommer til uttrykk i den multiple krysstabellen som kan dannes på grunnlag av de samme variablene. Analysen blir gjort på frekvensene i de ulike tabellcellene, og ikke på selve individene. Målsetningen er å finne den enklest mulige modellen som best er i stand til å beskrive sammenhengen mellom variablene. Enkle modeller, eller modeller med få latente klasser eller grupper, har forrang framfor modeller med mange latente klasser.

I motsetning til mange andre statistiske teknikker, er målet i en latent klasseanalyse å finne en modell som får en p-verdi $>.05$. Dette indikerer at den gitte modellen gir en akseptabel beskrivelse av strukturene i datamaterialet. Ofte vil flere ulike modeller få en p-verdi $>.05$. Da må man sammenligne modelltilpasningen ut fra et sett kriterier, og til slutt velge den modellen som gir den enkleste, men samtidig beste beskrivelsen.

Det bayesianske informasjonskriteriet, BIC, er et slikt kriterium (Raftery, 1995). Jo lavere verdi for BIC, jo bedre passer modellen. Modeller med lave BIC-verdier har derfor forrang framfor modeller med høye BIC-verdier.

Modellens kvalitet må også vurderes opp mot prosentandelen analyseenheter den feilklassifiserer, målt ved det man kaller «Dissimilarity Index», eller D.I. Sagt litt annerledes, svarer indeksen på følgende: under en gitt modell, hvor mange av individene blir plassert i en annen latent klasse enn den de burde plasseres i? En verdi for D.I. på .07 forteller oss at 7%, eller 1 av 14 individer, blir feilklassifiserte av den gitte modellen. Er verdien for D.I. $<.01$, er derimot mindre enn 1%, eller 1 av 100 individer, feilklassifiserte.

Ofte vil disse kriteriene føre til samme resultat i valget av endelig modell. Men skulle en mer kompleks modell representere en statistisk signifikant forbedring fra en enklere, men likefullt egnet modell, har den komplekse modellen forrang, selv om begge kan gi en adekvat strukturbeskrivelse. Motsatt vil en kompleks modell forkastes, dersom den ikke representerer en statistisk signifikant forbedring. Dette testes ved det man kaller «bootstrapping».

Tolkingen av resultatene tar utgangspunkt i to enkle mål: Latent klasse-størrelse og betingede sannsynligheter.

Den latente klasse-størrelsen forteller oss hvor mange prosent av enhetene som plasseres i en gitt latent klasse eller gruppe. Verdien av den latente klassestørrelsen summerer seg til 1.0 eller 100 på tvers av alle de latente klassene.

De betingede sannsynlighetene forteller oss hvilke sannsynligheter respondentene i en gitt latent klasse har for å svare på en bestemt måte på et bestemt spørsmål i analysen. Verdiene summerer seg til 1.0 eller 100 for hver enkelt variabel som inngår i analysen.

Vi kan illustrere dette på en enkel måte. I en tenkt analyse finner vi tre latente klasser (heretter LK), med klassestørrelsene 0.80, 0.10 og 0.10. Av dette følger det at 80% av enhetene er plasserte i LK 1, 10% i LK 2 og 10% i LK 3. De betingede sannsynlighetene for variabelen «Forsker», med verdiene «Ja» og «Nei» er .70/.30 i LK 1, .30/.70 i LK 2 og .50/.50 i LK 3. En respondent i LK 1 har følgelig 70% sannsynlighet for å

være forsker, og 30% sannsynlighet for ikke å være det. I LK 2 er det akkurat motsatt, og i LK 3 er det like høy betinget sannsynlighet for å være forsker som å ikke være det. Slik avdekkes forskjeller og likheter mellom de ulike grupperingene i datasettet.

Selve analysen gjøres trinnvis. Først estimerer man de enkleste modellene med én latent klasse, to latente klasser osv. For hver analyse må utfallet vurderes opp mot den forrige. Når man har funnet en passende modell, må likevel også den etterfølgende (og mer komplekse) modellen estimeres. Dersom denne ikke utgjør en statistisk signifikant forbedring sammenlignet med den foregående, kan modellutprøvingen stanses.

4. Holdninger til diskutabile praksiser

Vi har først prøvd ut sju ulike modeller for å beskrive holdninger til QRP. Flere ulike modeller viser seg å gi akseptable beskrivelser av holdningsmønstrene, men vi har til slutt valgt å beholde modellen med seks klasser fordi den gir den beste beskrivelsen av holdningsbatteriet:

Tabell 1: Modelltilpasning, holdninger til diskutabile praksiser. 8 variabler. N=7054

Modell	L-kvadrat	Frihets- grader	P-verdi	Kji-kvadrat	P-verdi	D.I	BIC
1 latent klasse	5335.84	247	.00	10070833.71	.00	.27	3147.09
2 latente klasser	1293.11	238	.00	7146.12	.00	.107	-815.89
3 latente klasser	475.93	229	.00	662.43	.00	.053	-1553.32
4 latente klasser	315.53	220	.00	349.89	.00	.042	-1634.07
5 latente klasser	235.81	211	.12	267.58	.00	.027	-1633.94
6 latente klasser	207.02	202	.39	205.54	.42	.023	-1582.97
7 latente klasser	190.40	193	.54	180.80	.73	.023	-1519.84

Med p-verdier på henholdsvis .39 og .42, og med kun 2.3% feilklassifiserte respondenter, må modell 6 sies å gi en god beskrivelse av data².

I første omgang kan vi derfor konkludere med at holdningene til QRP varierer blant norske forskere. Selv om vi tidligere har avdekket en betydelig grad av normkonsensus rundt problemstillingene vi spør om, er variasjonen likevel for stor til at en enkel 1- eller 2-klasse modell kan gi en god beskrivelse av holdningsgrupperingene. Og selv om analysen avdekker to store latente klasser, som summerer opp 52% og 36% av respondentene, finner vi også fire mindre undergrupper med hhv. 6%, 3%, 1.6% og 1.4% av respondentene.

² Selv om modellen med 5 klasser får en lavere verdi for BIC, viser statistiske tester like fullt at modellen med 6 klasser i sammenligning utgjør en statistisk signifikant forbedring. Modellen med 7 klasser representerer ikke en slik forbedring. Selv om den får høyere p-verdier, feilklassifiserer den like mange respondenter. Verdien for BIC er også høyere.

Forskjellene mellom de ulike grupperingene går til dels klart fram av mønstrene i de betingede sannsynlighetene:

Tabell 2: Holdninger til diskutabile praksiser. N=7054.

	Latent klasse 1	Latent klasse 2	Latent klasse 3	Latent klasse 4	Latent klasse 5	Latent klasse 6
Størrelse, latent klasse	.52	.36	.06	.03	.016	.014
Gaveforfatterskap, Lite problematisk	.05	.16	.20	.58	.74	.87
Gaveforfatterskap, Problematisk	.95	.84	.80	.42	.26	.13
Nekte forfatterskap, Lite problematisk	.02	.02	.04	.28	.90	.90
Nekte forfatterskap Problematisk.	.98	.98	.96	.72	.10	.10
Unnlate å informere, Lite problematisk	.00	.03	.35	.35	.04	.99
Unnlate å informere, Problematisk.	1.00	.97	.65	.65	.96	.01
Unnlate å varsle, Lite problematisk	.01	.04	.21	.52	.01	.96
Unnlate å varsle, Problematisk.	.99	.96	.81	.48	.99	.04
Irrelevant kildehenvisning Lite problematisk	.03	.51	.81	.72	.12	.95
Irrelevant kildehenvisning Problematisk.	.97	.49	.19	.28	.88	.05
Kopiere kilde, Lite problematisk	.03	.35	.83	.52	.04	.91
Kopiere kilde, Problematisk	.97	.65	.17	.48	.96	.09
Endre design, Lite problematisk	.01	.06	.23	.38	.07	.94
Endre design, Problematisk.	.99	.94	.27	.62	.93	.06
Omstridt eierskap, Lite problematisk	.02	.16	.57	.53	.05	.95
Omstridt eierskap, Problematisk	.98	.84	.43	.47	.95	.05

Latent klasse 1, med 52% av respondentene, har gjennomgående svært høye betingete sannsynligheter for å oppfatte alle de 8 praksisene som klart problematiske. Vi kan kalle denne undergruppen for «De konsekvente». Blant disse respondentene er aksepten for diskutabile praksiser svært liten, om den er der i det hele tatt.

Latent klasse 2, med 36% av respondentene, er mer splittet særlig i synet på to av praksisene: Å gi irrelevante referanser og å kopiere kildehenvisninger. Denne gruppen kan vi kalle «De fleksible». Gjennomgående stiller de seg kritiske til diskutabile forskningspraksiser, men med et mulig unntak for strategisk sitering og for kopiering av kildehenvisninger.

Latent klasse 3, som rommer 6% av respondentene, framstår som en enda mer splittet gruppering. Noen diskutabile praksiser, som å nekte noen et medforfatterskap tross et vesentlig bidrag, er det liten aksept for. Men de betingede sannsynlighetene for irrelevante referanser og for kopiering av kildehenvisninger er så høye som .80. Vi kaller derfor denne gruppen «De tvisynte».

De tre neste klassene er alle klart mindre:

Latent klasse 4, 3% av respondentene, framstår som mer heterogen enn den gruppen vi har døpt «de tvisynte». For så å si samtlige variabler finner vi betingete sannsynligheter i intervallet .35-.65. Klasse 4 kalles derfor «De splittede».

Latent klasse 5, 1,6% av respondentene, blir særlig definert av to variable; Gaveforfatterskap og Nekte forfatterskap. Begge deler oppfattes gjennomgående som lite problematisk. For de øvrige praksisene er toleransen tvert imot svært lav. Vi har kalt gruppen «De selektive», da det kun er 2 av 8 diskutabile praksiser som får høy aksept.

Endelig utgjøres latent klasse 6, med 1,4% av respondentene, av en gruppe der aksepten for diskutabile praksiser gjennomgående er høy, og så høy at vi har valgt å kalle gruppen for «De diskutabile». De betingede sannsynlighetene for å oppfatte praksisene som lite problematiske, er svært høye; i intervallet .87-.99. Så langt har analysen avdekket en todeling. Et lite flertall av forskerne stiller seg konsekvent avvisende til alle former for diskutabile forskningspraksiser. De med større aksept for slike praksiser er imidlertid ikke en homogen gruppe. Noen typer diskutabile praksiser framstår som mer akseptable enn andre, og da særlig Irrelevante referanser og dels også Gaveforfatterskap. Men fremdeles er det slik at det kun er et lite mindretall som er splittet i synet på de ulike praksisene, og et enda mindre mindretall som kan omtales som «diskutabile» i sine holdninger.

Mens aksepten for *spesifikke*, diskutabile praksiser kan være mer utbredt, er en *generell* aksept for slik praksiser helt klart et mer marginalt fenomen. Men i hvilket omfang omsettes holdninger til handlinger? I neste del skal vi ta for oss denne problemstillingen gjennom en mer detaljert analyse av egenrapporterte diskutabile praksiser.

5. Egenrapporterte diskutable praksiser

Tidligere analyser har vist at jo mer alvorlig en praksis blir oppfattet å være, desto mindre blir den også rapportert å være utbredt. Dette gjelder både når man spør forskerne om de har observert slike praksiser, og når man spør om de selv har utøvd de gitte praksisene. Motsatt oppgir en høyere andel både å ha observert og selv også å ha utført diskutable praksiser som oppfattes som mindre alvorlige, f.eks. gaveforfatterskap og salamisering (Hjellbrekke mfl. 2018).

For å gi et samlet uttrykk for utbredelsen av slike praksiser, har vi i Tabell 3 summert respondentenes verdier på tvers av ni variabler:

Tabell 3: Rapporterte egne, utførte diskutable praksiser³. N=7223.

Antall utførte diskutable praksiser	Frekvens	Prosent	Kumulativ prosent
Ingen	4373	60,5	60,5
En	1766	24,4	84,9
To	741	10,3	95,2
Tre	231	3,2	98,5
Fire	77	1,1	99,6
Fem eller flere	35	0,4	100
Totalt	7223	100	

Resultatene bør påkalle oppmerksomhet. At nesten 4 av 10 forskere svarer at de selv har utført en diskutabel praksis i løpet av de siste tre årene, kan man trygt hevde indikerer en vidstrakt utbredelse. Samtidig tyder fordelingen på at en liten gruppe står for en relativt høy andel av de bruddene som er egenrapporterte. 4,8% av respondentene rapporterer selv å ha utført en diskutabel praksis mer enn tre ganger de siste tre årene. At nesten 1 av 20 forskere relativt hyppig har brutt med god forskningsskikk, kan kalles urovek-kende.

Tabellen forteller oss likevel ikke noe om hvilke typer kombinasjoner som er de mest vanlige, og om hvilke undergrupper av forskere vi kan finne i denne delen av materialet. For å finne svar på dette, har vi nok en gang valgt å gjennomføre en latent klasseanalyse.

Men som nevnt ovenfor, vil skjevfordelinger på enkelte variabler lett kunne skape stabilitetsproblemer for en latent klasseanalyse. Vi har derfor avgrenset oss til kun å ta med de 6 variablene som er presentert i avsnitt 2.0.Data, på side 4. Utvalget er dels gjort for å unngå redundans mellom to eller flere variable, men også for å unngå de mest ekstreme skjevfordelingene.

Framgangsmåten er den samme som i modelleringen av holdningsvariablene. Etter en samlet vurdering har vi konkludert med at Modell 3, modellen med 3 latente klasser, er den som gir den beste modelltilpasningen. Selv om Modell 4 oppnår bedre verdier for noen av kriteriene, utgjør den ikke en statistisk signifikant forbedring sammenlignet med Modell 3:

³ Respondenter med manglende verdier for mer enn fem variabler er utelatt fra analyse. For de resterende respondentene er manglende verdi kodet sammen med «Nei». Prosentdelen for kategorien «Ingen» kan derfor være marginalt lavere enn det tabellen oppgir.

Tabell 4: Modelltilpasning, analyse av egenrapporterte diskutabile praksiser. N=6490

Modell	L-kvadrat	Frihetsgrader	P-verdi	Kji-kvadrat	P-verdi	D.I.	BIC
1 latent klasse	564.78	57	.00	12975.85	.00	.0759	64.44
2 latente klasser	99.44	50	.00	115.56	.00	.0124	-339.46
3 latente klasser	54.30	43	.12	59.03	.05	.007	-323.15
4 latente klasser	41.87	36	.23	36.76	.43	.005	-274.13
5 latente klasser	35.77	29	.18	34.60	.22	.005	-218.79

Både Modell 4 og 5 får også dårligere verdier for BIC enn Modell 3. Modell 2 får imidlertid de aller laveste verdiene for BIC, men Modell 3 representerer en statistisk signifikant forbedring i sammenligning med denne. Verdien for både L- og kji-kvadratet, samt de tilhørende p-verdiene, indikerer også at Modell 2 ikke gir en fullgod beskrivelse av strukturene i materialet.

De betingede sannsynlighetene får fram klare forskjeller mellom de tre grupperingene.

I latent klasse 1, 82% av respondentene, samles det store flertallet av forskere som rapporterer selv ikke å ha brutt med god forskningsskikk. Det betingede sannsynlighetene for å svare «Nei» på alle 6 spørsmål ligger i intervallet .92-.99. Vi kan kalle denne gruppen for «*De rene og ranke*».

Latent klasse 2, 13% av respondentene, har en klart høyere betinget sannsynlighet for å ha gitt et gaveforfatterskap. Men dette er også den eneste av de seks variablene som skiller seg ut med en relativt høy sannsynlighet for å ha brutt med god forskningsskikk. Denne gruppen kaller vi «*De gavmilde*».

Respondentene i Latent klasse 3, 5% av analyseenheterne, skiller seg ut med å ha relativt sett høye betingete sannsynligheter for å svare «ja» på 5 av 6 variabler. Kun i spørsmålet om å nekte medforfatterskap tross signifikant bidrag er de på linje med de andre respondentene. Selv om verdiene ligger i intervallet .21-.56, har denne gruppen av «*Diskutable*» følgelig en klart høyere sannsynlighet enn de to andre for å ha brutt med god forskningsskikk. Fra en slik synsvinkel utgjør gruppen også en risikogruppe.

Tabell 5: Egne diskutable praksiser. N=6490.

	Latent klasse 1	Latent klasse 2	Latent klasse 3
Størrelse, latent klasse	.82	.13	.05
Gaveforfatterskap, Nei.	.98	.41	.62
Gaveforfatterskap, Ja.	.02	.59	.38
Nekte forfatterskap Nei.	.99	.93	.91
Nekte forfatterskap Ja.	.01	.07	.09
Unnlate å informere Nei.	.99	.99	.65
Unnlate å informere Ja.	.01	.01	.35
Unnlate å varsle Nei.	.97	.89	.79
Unnlate å varsle Ja.	.03	.11	.21
Irrelevant kildehenvisning Nei.	.92	.75	.44
Irrelevant kildehenvisning Ja.	.08	.25	.56
Endre design Nei.	.97	.97	.69
Endre design Ja.	.03	.03	.31

Flere studier har avdekket klare disiplinforskjeller mellom måtene forskningen er organisert på, og også akademiske kulturer, sosialisering og kunnskapstyper utviser disiplinrelatert variasjon (se f.eks. Becher & Trowler, 2001: kap. 2 og 3). Innen mange naturvitenskapelige disipliner og innen medisin er det for eksempel vanlig at forskningen organiseres i store grupper. Innen samfunnsvitenskap og humaniora er det motsatt. Dette reflekteres også i publiseringen av forskningsresultatene. Eneforfatterskap er langt mer utbredt i humaniora enn i medisin, der antallet medforfattere kan være svært høyt. Det vil derfor ikke være uventet om vi finner disiplinforskjeller med hensyn til de tre latente klassenes sammenhenger med andre, «eksterne» og/eller strukturelle variabler. Klassenes viktigste kjennetegn er oppsummerte i tabell 6:

Tabell 6: Karakteristika ved latente klasser 1-3.

	LK1	LK2	LK3
Overrepresentasjon (>5% i fete typer)	Generell klynge	Offentlig orientert forskningsinstitutter Medisin/Helse MatNat 30-39 år	Privat-orienterte forskningsinstitutter Samfunnsvitenskap Postdok Forsker II Seniorforsker 30-39 år
Underrepresentasjon (>5% i fete typer)	Generell klynge	Humaniora Samfunnsvitenskap	Humaniora PhD-student 60-69 år

Som forventet er forskere innen medisin/helse og matematikk og naturfag overrepresenterte blant *De gavmilde*. I mange slike forskningsmiljøer er det en ikke uvanlig praksis at laboratorieleder tas med som medforfatter, selv om vedkommende ikke aktivt har bidratt til sluttproduktet. At humaniora og samfunnsvitenskap er underrepresentert, er også som forventet. Mer urovekkende er det at aldersgruppen 30-39 år også er overrepresentert. Mange forskere i denne aldersgruppen vil være juniorforskere eller stipendiater, og følgelig i en underordnet stilling. Dette vil også kunne gjøre dem mer utsatt for press fra forskere i ledende stillinger, for eksempel press til å inkludere meritterte forskere som medforfattere.

Også i Latent klasse 3, *De diskutabile*, er denne aldersgruppen overrepresentert. Samlet understreker dette viktigheten av å styrke forskningsetikkutdanningen for stipendiater og postdoktorander, og også behovet for bevisstgjøring blant seniorforskere og veiledere. At Forsker II-kategorien og ansatte ved oppdrags- og privatorienterte forskningsinstitutter er overrepresenterte, kan dels skyldes det presset man kan bli utsatt for fra en oppdragsgiver, for eksempel til å endre design, metode etc. En høy andel ansatte ved slike institutter er også samfunnsforskere. Rammevilkårene som flere forskere i disse stillingene arbeider under, kan tenkes å øke sannsynligheten for at man gir etter for press, og i enkelte situasjoner går på akkord med de forskningsetiske normene vi undersøker her.

6. Diskusjon

I analysen av holdninger til QRP ender vi opp med en modell med seks latente klasser, der de to første samler et relativt betydelig antall forskere, mens de resterende fire er uttrykk for mer perifere syn. Etter vår oppfatning peker spenningen mellom *De konsekvente* og *De fleksible* – og til en viss grad også *De tvisynte* – på et diskusjons- og klargjøringsbehov som både forskersamfunnet og politiske myndigheter må ta alvorlig. Å slå seg til ro med at *De konsekvente* tilsynelatende deler oppfatningen som uttrykkes av forskningsetiske komiteer eller flertallet blant individuelle forskere i Norge, vil være korttenkt. De fleksible, og til dels *De tvisynte*, synes å sette spørsmålstegn ved deler av den den offisielle forskningsetiske kanon, uttrykt i første rekke gjennom nasjonale og internasjonale forskningsetiske retningslinjer. Selv om man fra «offisielt» hold har konkludert med at mange av de formene for QRP vi har inkludert her egentlig ikke er diskutabile, men heller er forskningsetisk uakseptable, så finnes det også et ikke ubetydelig antall forskere som mener at man i de konkrete tilfeller det her gjelder (strategisk sitering og kopiering av kildehenvisninger) ikke bør klassifiseres som QRP. Vi mener at dissens om disse praksisene ikke automatisk kan tolkes som mangel på refleksjon eller innsikt, og utdyper dette under.

Når det gjelder egenrapporterte praksisformer endte vi opp med tre latente klynger. Det klare flertallet, 82 %, rapporterte normkonformitet i egen praksis (*De rene og ranke*). De to andre klyngene, *De gavmilde* med 13 %, og *De diskutabile* på 5 %, har høyere betingete sannsynligheter for å ha utført en eller flere diskutabile forskningspraksiser. Også her kan man fristes til å trekke enkle konklusjoner om forekomsten av forskningsetisk sterke vs. forskningsetiske svake eller til og med kritikkverdige grupperinger. Ser vi mer nøye på over- og underrepresentasjon av undergrupper i de to siste klyngene, kan det imidlertid være grunn til å nyansere en slik tolkning. Eksempelvis vet vi fra før at urettmessig forfatterskap – inkludert gaveforfatterskap – i medisin og realfag, er et velkjent fenomen (Nylenna m.fl. 2014). Noe av grunnen til det ligger antakelig i at publikasjon av forskningsresultater i disse fagfeltene bygger i større grad på både et forskningsteam / en forskergruppe, og tilgjengelighet av krevende infrastruktur og ekstern finansiering. Det dannes således lett strukturer med segmentert ansvarsforhold og komplekse interaksjonsmønstre.

Vi kan ikke utelukke at respondenter i klynge 2 og særlig 3 er av den mening at veletablerte kriterier for rettmessig forfatterskap ikke tar høyde for bidrag som på mer indirekte måte er avgjørende for å nå bestemte vitenskapelige resultater. En laboratorieleder eller instituttleder vil ofte selv være forsker og som sådan avhengig av kontinuerlig merittering, samtidig som vedkommende ikke i samme grad kan bidra til alle prosjekter som samles under hans enhet. Her er det kanskje noen som ser gaveforfatterskap som en nøkkel til løsning. Videre vil også veiledere som er prosjektledere og har gjort et omfattende arbeid i å utvikle og drifte prosjektet kunne vurdere sitt bidrag som omfattende nok til medforfatterskap, selv om ikke alle kriterier for forfatterskap er oppfylt i henhold til internasjonale og nasjonale retningslinjer.

Det er viktig å diskutere mulige grunner til variasjonene vi har avdekket, for i neste omgang å kunne finne målrettede tiltak.

For det første, det at vi finner noe variasjon både i holdninger til enkelte praksisformer og selvrapportert praksis i datamaterialet er i seg selv ikke nødvendigvis grunn til bekymring, men kan være et uttrykk for ulike tolkninger av normer innenfor ulike fagområder og kulturer. Forskningsetiske vurderinger forutsetter alltid et element av skjønn, og innebærer dermed noe mer enn mekanisk etterfølgelse av regler. Med nye teknologiske muligheter for bildemanipulasjon kan for eksempel til og med såpass bastante feiltrinn som fabrikkasjon kreve skjønnsmessige vurderinger, som må foretas med bakgrunn i god innsikt i de normer og metoder som gjør seg gjeldende innenfor fagområdet. Selv FFP kan altså være vanskelig å stadfeste, og når vi beveger oss mot QRP er de skjønnsmessige vurderingene ofte enda vanskeligere.

For det andre har arbeidsgruppen i RINO hatt som utgangspunkt at forskningsetisk aksept også er en funksjon av hvordan man oppfatter sosio-politiske og sosio-økonomiske forhold rundt forskningen. Er man av den oppfatning at slike forhold påvirker merittering og vurdering av vitenskapelig kvalitet i relativt stor grad, kan noen av disse praksisformer muligens oppfattes som responser på individuell avmakt i forskersamfunnet. Med andre ord, en svak fordømming av noen former for QRP (f.eks. strategisk sitering og kopiering av kildehenvisninger) behøver ikke å tyde på dårlig etikk, men kan også oppfattes som en individuell respons på strukturelle forhold i forskningen som man oppfatter som uetiske, men utenfor ens kontroll.

For de tredje er det grunn til å stille spørsmål ved enkelte av normene og om de må nyanseres eller omformuleres i takt med endrede forutsetninger. Det er viktig å diskutere dette, særlig i Norge, der vi nå har en ny forskningsetikklov (Lov om organisering av forskningsetisk arbeid) som bl.a. lovfester forskningsinstitusjoners ansvar for å håndtere mistanker om vitenskapelig uredelighet. Loven nevner FFP spesielt, men definisjonen av vitenskapelig uredelighet kan også inkludere tilfeller av QRP. Dersom forsker X blir anklaget for brudd på forskningsetiske normer fordi han/hun har gitt forsker Y et gaveforfatterskap, kan dette være en aktuell sak for et redelighetsutvalg ved institusjonen.

Forskere i våre latente klynger 2 og 3 kan ut fra dette risikere å bli beskyldt i en slik sak. Har man innsikt i komitéarbeid er det lett å forstå hvorfor man ofte vil kvie seg for å komme med klare fordømminger. En av grunnene er igjen at skjønn spiller en rolle: hva er for eksempel den rette balansen i komplekse forskningsstrukturer mellom «acknowledgement» (takksigelse) og forfatterskap? Her må man også ta i betraktning at forskersamfunnet lenge har ført en intern diskusjon om hvorvidt denne todelingen mellom forfatterskap og takksigelse fortsatt er funksjonell. Trenger vi andre og mer spesifikke opplistinger av ulike type bidrag? Det er kjent at faglige artikler fra CERN gjerne har forfatterlister som er lengre enn selve bidraget. Er våre standard forfatterkriterier da fortsatt realistiske og rimelige?

Avslutning

Resultatene fra denne analysen, og klyngene *De gavmilde* og *De diskutabile* spesielt, tyder på en spenning mellom norm og virkelighet som vi må ta alvorlig. Det er i det lange løp uholdbart at en gruppe forskere tilsynelatende aktivt engasjerer seg i praksisformer som et flertall av forskere anser som uakseptable.

Spenninger mellom norm og virkelighet kan skyldes moralsk svikt eller ulik utøvelse av forskningsetisk skjønn, men kan også indikere urealistiske og urimelige forventninger eller for stor rigiditet i det normative grunnlaget.

Utfordringen er å definere en felles forskningsetisk plattform basert på mest mulig realisme, faglig innsikt, og vitenskapelig kvalitet.

Vi håper funnene vi har presentert her åpner opp for diskusjoner om forskningsetiske normers status i et stadig mer komplekst forskersamfunn og en kunnskapsproduksjon med mange fasetter og målsettinger.

* * *

Referanser

- Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines* (2nd ed.). Philadelphia: Open University Press.
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLOS ONE*, 4(5). <https://doi.org/10.1371/journal.pone.0005738>
- Hjellbrekke, J., Drivdal, L., Ingierd, H., Rekdal, O. B., Skramstad, H., Torp, I. S., & Kaiser, M. (2018). *Etikk og integritet i forskning – resultater fra en landsomfattende undersøkelse. Første delrapport fra arbeidsgruppen i forskningsprosjektet RINO (Research Integrity in Norway)*. Oslo: Universitetet i Bergen, De nasjonale forskningsetiske komiteene, Høgskulen på Vestlandet. (https://www.uib.no/sites/w3.uib.no/files/attachments/rino_delrapport_1_2018_2.pdf)
- Magidson, J. & Vermunt, J. K. (2002). Latent class analysis. A nontechnical introduction to latent class models. (<https://www.statisticalinnovations.com/wp-content/uploads/Magidson2002.pdf>, lest den 15.11.2018)
- McCutcheon, A. L. (1987). *Latent class analysis*. Beverly Hills: Sage.
- Nylenna M, Fagerbakk F, Kierulf P (2014). Authorship: attitudes and practice among Norwegian researchers. *BMC Medical Ethics* 2014;15:53. <https://doi.org/10.1186/1472-6939-15-53>
- Raftery, A.E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111-163.



UNIVERSITETET I BERGEN



De nasjonale
FORSKNINGSETISKE
KOMITEENE



Høgskulen
på Vestlandet