

Datakvalitet i Ung i Oslo 2006

Kommentarer og vurderinger

Jon Ivar Elstad

NOTAT
NR 1/10

Datakvalitet i Ung i Oslo 2006

Kommentarer og vurderinger

JON IVAR ELSTAD

Norsk institutt for forskning om
oppvekst, velferd og aldring

NOVA Notat 1/2010

Norsk institutt for forskning om oppvekst, velferd og aldring (NOVA) ble opprettet i 1996 og er et statlig forvaltningsorgan med særskilte fullmakter. Instituttet er administrativt underlagt Kunnskapsdepartementet (KD).

Instituttet har som formål å drive forskning og utviklingsarbeid som kan bidra til økt kunnskap om sosiale forhold og endringsprosesser. Instituttet skal fokusere på problemstillinger om livsløp, levekår og livskvalitet, samt velferdssamfunnets tiltak og tjenester.

Instituttet har et særlig ansvar for å

- utføre forskning om sosiale problemer, offentlige tjenester og overføringsordninger
- ivareta og videreutvikle forskning om familie, barn og unge og deres oppvekstvilkår
- ivareta og videreutvikle forskning, forsøks- og utviklingsarbeid med særlig vekt på utsatte grupper og barnevernets temaer, målgrupper og organisering
- ivareta og videreutvikle gerontologisk forskning og forsøksvirksomhet, herunder også gerontologien som tverrfaglig vitenskap

Instituttet skal sammenholde innsikt fra ulike fagområder for å belyse problemene i et helhetlig og tverrfaglig perspektiv.

© Norsk institutt for forskning om oppvekst, velferd og aldring (NOVA) 2010

ISBN 978-82-78-94327-4

ISSN 1890-6435

Desktop: Torhild Sager
Trykk: Allkopi

Henvendelser vedrørende publikasjoner kan rettes til:

Norsk institutt for forskning om oppvekst, velferd og aldring
Munthesgt. 29 · Postboks 3223 Elisenberg · 0208 Oslo

Telefon: 22 54 12 00

Telefaks: 22 54 12 01

Nettadresse: <http://www.nova.no>

Forord

Høsten 2007 ble spørsmålet om hvordan en skulle vurdere datakvaliteten i Ung i Oslo 2006-materialet debattert blant forskere på NOVA. Bakgrunnen var blant annet en NOVA-rapport publisert denne høsten som omhandlet visse sider ved livssituasjonen og erfaringene til homofil, lesbisk og bifil Oslo-ungdom.

For å få et bredere grunnlag for å drøfte disse spørsmålene ba ledelsen ved NOVA undertegnede om å foreta en vurdering av dette datamaterialet.

Manuset var ferdig i februar 2008. Av ulike grunner ble offentliggjøringen utsatt, men høsten 2009 ble det besluttet at manuset skulle utgis som NOVA Notat. Bortsett fra en del mindre omredigeringer og språklige endringer, er notatet identisk med det manuset jeg utarbeidet i januar/februar 2008.

Oslo/NOVA, januar 2010

Jon Ivar Elstad

Innhold

Sammendrag	7
1 Problemstillingen.....	9
2 Problematiske sider ved Ung i Oslo 2006-datainnsamlingen	11
2.1 Introduksjon.....	11
2.2 Omfang, tidspress, muligheten for «feil» svar pga. utmattelse	13
2.3 Tidsangivelser som ikke blir fulgt	15
2.4 Skiftende tidsperioder og umenneskelige hukommelseskrav	17
2.5 Ordforståelse, leseforståelse, tallforståelse	19
2.6 Interesse, relevans, oppmerksomhet.....	20
2.7 Tvetydige spørsmål, spørsmål som er ulogiske, osv.	22
2.8 Oppsummering	23
3 Drøfting av kvaliteten på Ung i Oslo 2006-dataene	25
3.1 Argumenter for at datakvaliteten er høy.....	25
3.2 Frafallet er sosialt skjevt	27
3.3 Kriterievaliditet.....	30
3.4 Analyser av konsistens/reliabilitet	30
3.5 Sannsynlighetskriteriet.....	33
3.6 Oppsummering	35
4 Om bevisst feilinformasjon i Ung i Oslo 2006	36
4.1 Underrapportering av det sensitive, overrapportering av det sosialt verdsatte	36
4.2 Generelt om tullede feilinformasjon i Ung i Oslo-materialet.....	37
4.3 Et forsøk på å flagge tøyserne – hvordan kan vi finne deres signatur?	39
4.3.1 De åpne spørsmålene om fars/mors yrke	41
4.3.2 Umulige aldersangivelser.....	42
4.3.3 Kroppslige forhold – tøysesvar om funksjonshemninger?	43
4.3.4 Giganter og dverger, puslinger og fettberg.....	44
4.3.5 Utrolige sex-atleter	44
4.3.6 Dop-spørsmålene.....	45
4.3.7 Enorm kriminalitet	45
4.4 Er målingen av tøyserne valid?	46
5. Homo-tøysere?	49
5.1 Tøyse med homofili, lesbiske, bifile?	49
5.2 Hvordan tøsehomsene skiller seg fra homser uten tøys.....	51
5.3 Har tøsehomsene fordreidd resultatene i NOVA-Rapport 19/07?	53
5.4 Bør NOVA-Rapport 19/07 trekkes tilbake?.....	55
6 Anbefalinger.....	56
6.1 Generelle bemerkninger om datakvalitet i Ung i Oslo 2006.....	56
6.2 Tøysesvar og hva vi gjør med dem.....	56
6.3 Ny dokumentasjon	57
6.4 Lærdommer for nye datainnsamlinger	58
Litteratur	59

Sammendrag

I 2006 gjennomførte NOVA en stor datainnsamling blant Oslos elever i 14–17-årsalderen. Etter hvert som analyser ble foretatt, reiste det seg en debatt om datakvaliteten i Ung i Oslo 2006-materialet blant NOVAs forskere. En bakgrunn for denne debatten var blant annet NOVA Rapport 19/07 som ble publisert høsten 2007, og som ga til dels dramatiske tall om vanskelige sider ved livssituasjonen til homofil, lesbisk og bifil Oslo-ungdom.

For å få et bredere grunnlag for å drøfte disse kvalitetsspørsmålene ba ledelsen ved NOVA forfatteren om å foreta en nærmere empirisk vurdering av dette datamaterialet. Resultatene ble framlagt i et notat som var ferdigstilt i februar 2008 – og som publiseres her.

Notatet viser at en del sider ved organiseringen og opplegget for datainnsamlingen til Ung i Oslo 2006 kan ha ført til svekket datakvalitet i denne undersøkelsen. Undersøkelsen var usedvanlig stor (elevene skulle svare på over 500 spørsmål i løpet av knappe to skoletimer) og trøtthet og uoppmerksomhet på grunn av hvor krevende det var å gjennomføre undersøkelsen, kan ha svekket elevenes evne til å svare presist. I datamaterialet kan en se eksempler på at ikke bare en liten del av de 11.500 deltakende elevene, men flere tusen av dem, må ha misforstått spørsmål, feiltolket hva de skulle gjøre, og derfor avlevert svar som ikke er valide og som avviker mer eller mindre fra hva de ville ha svart gitt et bedre undersøkelsesopplegg. En del spørsmål stiller urimelige krav til hukommelse, en del ord kan ha vært vanskelige å forstå, og en del spørsmål har en uheldig og ulogisk utforming.

For noen deler av datamaterialet kan en derfor stille spørsmål om informasjonen som er blitt gitt, gir et adekvat uttrykk for ungdommenes holdninger, erfaringer og atferd. Tross dette må en likevel anta at mange deler av datamaterialet holder en rimelig god kvalitet, slik at analyser godt kan foretas. En må imidlertid legge merke til at den reelle svarprosenten ikke er så høy som dokumentasjonen gir inntrykk av, ettersom ganske mange ikke greide å fullføre undersøkelsen. Frafallet i løpet av undersøkelsen, det vil si at enkeltspørsmål mot slutten av undersøkelsen ikke blir besvart, var spesielt

stort blant elever med dårlige karakterer, blant innvandrerelever, og blant andre elevgrupper som hadde dårligere forutsetninger for å greie å gjennomføre en slik undersøkelse. Frafallet er derfor sosialt skjevt og det er derfor også et spørsmål om hvor representative svarene fra slutten av undersøkelsen er for hele ungdomsmassen i Oslo.

Notatet diskuterer videre hva slags framgangsmåter som kunne bli brukt for å undersøke datakvaliteten mer rigorøst. Videre blir spørsmålet om «tøysesvar» undersøkt nærmere, ettersom et tema blant NOVAs forskere har vært om elevene som ikke har besvart seriøst, men «tulla» og «kødda» i svarene sine, har vært så mange at enkelte spørsmål er vanskelige å analysere på en tilfredsstillende måte. Notatet foreslår en metode for å lokalisere slike tøysere, spesielt via spørsmål, særlig åpne spørsmål, der tøysere sannsynligvis vil markere seg. Via en slik framgangsmåte kommer notatet fram til at noe i overkant av 250 av de 11.500 deltakerne kan være elever som i hvert fall av og til synes å ha svart bevisst på en tøysete og usannferdig måte. Tøysere finnes, men likevel tyder disse resultatene på at de alt i alt ikke er mange. Tøys er altså en liten og underordnet kilde til svak datakvalitet i Ung i Oslo 2006 – dårlig datakvalitet skyldes i langt større utstrekning at elever, uten å ville det, har misforstått og misoppfattet og derfor uforvarende avgitt ikkevalide svar.

Notatet diskuterer konkret om det dramatiske bildet av opplevelser av trakassering og vold blant homofil, lesbisk og bifil ungdom, som tidligere analyser av Ung i Oslo 2006-materialet har tydet på, kan skyldes upålitelige svar fra tøysere. Denne gjennomgangen tyder på at forekomsten av vanskelige opplevelser blant homofil/lesbisk/bifil ungdom har blitt inflatert og overdrevet på grunn av useriøse svar. Men konklusjonen er likevel at også om tøysere tas ut av undersøkelsen, tyder resultatene blant de gjenværende på at homofil, lesbisk og bifil ungdom har klart høyere risiko for å oppleve slike vanskelige episoder og livssituasjoner enn den heterofile ungdommen.

Notatet avsluttes med forslag til tiltak om hvordan slike datainnstillinger kan forbedres slik at datakvaliteten blir høyere. Forslag om hvordan tøys-spørsmålet kan håndteres i administreringen av datamaterialet, blir også lagt fram.

1 Problemstillingen

Formålet med dette notatet er å drøfte noen sider ved datakvaliteten i NOVAs datainnsamling Ung i Oslo 2006. Dette var en skolebasert spørreskjemaundersøkelse. Forskerne fikk anledning til å benytte skoletimer og klasserom for å la elevene besvare spørsmålene i spørreskjemaet. Fordelene med slike skolebaserte spørreskjemaundersøkelser, sammenlignet med intervjuundersøkelser, er blant annet at innsamlingen av data er billig og frafallet er lavt siden datainnsamlingen foregår i skoletiden. Derfor kan svarutvalget bli mer representativt og i større grad utgjøre et «sant tverrsnitt» av dem vi ønsker å vite noe om, enn om frafallet hadde vært større. Metodelitteraturen antar dessuten at sensitive opplysninger kan bli bedre dekket gjennom spørreskjemaundersøkelser enn gjennom direkte personlige intervjuer, ettersom respondentene slipper å stå ansikt-til-ansikt overfor en annen person (intervjueren) med pinlige eller skamfulle opplysninger.

Men spørreskjemametoden – som enhver annen datainnsamlingsmetode – har også sine vansker og potensielle problemer. Informasjonen som registreres i datafilen kan avvike i større eller mindre grad fra den «riktige» informasjonen, dvs. – med metodelitteraturens språkbruk – målingenes *validitet* svikter, eller sagt annerledes: det vi har målt, overensstemmer ikke godt med det vi ønsket å måle.

Generelt antas det – med et mulig unntak for sensitive forhold – at spørreskjemaundersøkelser er *mer* utsatt for ikke-valide målinger enn personlige intervjuer. Spørreskjemaundersøkelser overlater arbeidet til respondenten. Det er en *selvadministrert* datainnsamling, og dette har mange viktige implikasjoner. I personlige intervjuer vil intervjueren løse respondenten gjennom spørsmålene og motivere underveis. Intervjueren vil langt på vei kontrollere at spørsmålene blir rimelig oppfattet, intervjueren kan veilede og oppklare misforståelser og har større trening i å krysse av/registrere opplysningene i samsvar med respondentens ytringer. Slik hjelp for respondenten mangler i spørreskjemaundersøkelser, som er utelukkende basert på hvordan respondentene reagerer på stimuli presentert i skrift og grafikk.

Likevel kan også spørreskjemaundersøkelser gi valide målinger, men det er mange fallgruber. Å lage *gode* spørreskjemaundersøkelser er vanskelig. Metoden har derfor vært tema for en omfangsrik metodelitteratur som drøfter, f.eks., kjennetegn ved respondentatferd i spørreskjemaundersøkelser, gode og uheldige spørsmålskonstruksjoner, hva slags ord kan brukes og kan ikke brukes, hvilken rekkefølge bør spørsmål ha, hvordan skal sensitive spørsmål formuleres, mv. (Gilbert 1993; Haraldsen 1999; Hjorth Andersen & Nygaard Christoffersen 1982; Mordal 1989; Olsen 1998).

Diskusjonen på NOVA om kvaliteten på data i Ung i Oslo 2006 har hatt en uheldig slagside. Det har vært hevdet at hovedproblemet er at dårlig datakvalitet, særlig i form av at en del opplysninger er upålitelige og ikke sannferdige, har oppstått fordi noen av respondentene (respondenter = de som har deltatt i undersøkelsen) har avgitt tulle svar og ikke vært seriøse og opptatt av å svare ordentlig. Derfor har det vært vanskelig å gjøre enkelte analyser og noen bestemte resultater «er ikke å stole på».

Spørsmålet om tulle svar skal jeg drøfte seinere, men hovedbudskapet i dette notatet er det følgende: *Hovedgrunnen til sviktende datakvalitet i spørreskjemaundersøkelser er at datainnsamlingsopplegget og spørreskjemaet har svak faglig kvalitet.* Dette mener jeg er den viktigste lærdommen fra studier av validitet i spørreskjemaundersøkelser. Om data som produseres i en datainnsamling ikke er valide, skyldes det i liten grad at det finnes tulle respondenter (de finnes alltid). Grunnen er først og fremst at undersøkelsesopplegget har sviktet. I del 2 vil jeg prøve å begrunne denne påstanden, ved å drøfte en del generelle aspekter av datainnsamlingsopplegget i Ung i Oslo 2006. Seinere i notatet vil jeg ta for meg andre temaer, blant annet spørsmålet om hvordan vi kan undersøke datakvalitet (del 3), om omfanget av tulle svar i Ung i Oslo 2006 (del 4), om konsekvensene av tulle svar for resultatene angående livssituasjonen for homser og lesber blant elevene (del 5). Tilslutt (del 6) sammenfatter jeg og trekker noen konklusjoner.

2 Problematiske sider ved Ung i Oslo 2006-datainnsamlingen

2.1 Introduksjon

Et første spørsmål om datakvaliteten i Ung i Oslo 2006 kan altså stilles slik: Hvor god er undersøkelsen med henblikk på å *produsere* valid informasjon? I hvilken grad kan egenskaper ved spørreskjemaet og ved datainnsamlingen ha bidratt til å generere ikke-valid informasjon?

Mange forskere – og enda flere utenforstående – undervurderer hvor avgjørende spørreskjemaet og datainnsamlingsopplegget er for datakvaliteten. Like mange overvurderer respondenters evne til å utføre den krevende oppgaven de pålegges, nemlig å administrere seg selv, uten hjelp fra intervjuere, til å avgi valide svar.

Her kommer et eksempel fra Ung i Oslo-dataene der vi kan se at respondentene – elevene – ikke oppfatter veiledningen de har fått. Tidlig i spørreskjemaet kommer spørsmål (om en oppfatter seg som norsk eller innvandrer, osv.) som bare skal besvares av de med «begge foreldre født utenfor Norge» (s. 4–6 i spørreskjemaet referert i Valset & Øia 2006)). Instruksjoner over fire linjer med fet skrift angir at spørsmålene som følger *ikke* skal besvares hvis far eller mor eller begge foreldre er født i Norge.

Om lag 11.500 elever alt i alt deltok i undersøkelsen og avleverte spørreskjema, og nær 7.800 blant dem hadde enten en far eller en mor (eller begge) som var født i Norge og derfor skulle de ikke besvare disse spørsmålene. Likevel har omlag 1400 av dem gjort det.

I neste sekvens (spørreskjemaet, s.5–6) skal de som har begge foreldre født utenlands *ikke* svare. Likevel blir spørsmålene besvart av ca. en fjerdedel av elevene med «ikke-norsk» foreldrebakgrunn.

Dette er et instruktivt eksempel. Det opptrer tidlig i spørreskjemaet da konsentrasjonen og oppmerksomheten trolig fortsatt er høy. Det gjelder spørsmål der ingen vil tro at noen ville bry seg om å tulle. Likevel: Nesten én av fem av de «norske» og én av fire av de «utenlandske» elevene greide ikke å forholde seg adekvat til en svært tydelig og ikke spesielt vanskelig instruksjon.

I seg selv trenger ikke dette å bety at avgitte svar ikke er valide. Men eksemplet viser hvor vanskelig det er å styre respondentenes atferd presist i et selvadministrert spørreskjema. Tilsynelatende uten grunn viser det seg at en betydelig andel – kanskje hver sjettede, femte, eller hver fjerde – av elevene ikke oppfatter, overser, mistolker, eller ikke får med seg, noe som fortoner seg som enkle anvisninger. De gjør det de tror de skulle gjøre og forter seg deretter videre til neste spørsmål, uten å bry seg om å finlese mellomstekster og rettleidninger.

Når dette skjer, tidlig i besvarelsene da oppmerksomheten er relativ høy, må en selvsagt også anta at noe liknende skjer i enhver fase, for ethvert spørsmål, blant en ikke ubetydelig (men ofte ukjent) andel av respondentene – uten at det dreier seg om bevisste forsøk på å gi misvisende opplysninger.

Et annet eksempel fra Ung i Oslo 2006: På s.18 i spørreskjemaet (Valset & Øia 2006) gis følgende instruksjon: «Hvis du har en funksjonsnedsettelse, hvor ofte i løpet av det siste året har du ...» – og så kommer fire spørsmål. Her skal altså bare de som «har en funksjonsnedsettelse» svare. Det skulle vært omlag 500, ut fra svarene gitt på forrige spørsmål. Likevel svarer over 4000 av respondentene på spørsmål som bare skulle besvares av omlag 500 med funksjonsnedsettelse.

Er det 3500 tullere blant de 11.500 respondentene? Neppe. Det dreier seg om en kombinasjon av tidspress, overseelser og ubevisste feiltolkninger som fører til at – ikke bare noen titalls, men *flere tusen*, ikke har fått med seg – ikke oppfattet – hvem som skulle svare og hva de skulle svare på.

En kunne tro at slikt skjer fordi respondentene i Ung i Oslo er ukonsentrerte og utålmodige tenåringer, og at det er særegenhetene ved dette spesielle utvalget som er grunnen til slike misforståelser. Dette er nok dels riktig, og det må jo de som konstruerer undersøkelsen ta hensyn til.

Men det er bare riktig i noen grad, for liknende ting oppstår også når respondentene er voksne, seriøse og motiverte. Når respondenter blir spurt om sin alder, f.eks., vil noen skrive fødselsår i rubrikkene. De tenker ikke over det, eller de tenker at «så kan jo forskerne sjøl regne ut hvor gammel jeg er». I personlige intervjuer blir slike misforståelser fort oppklart, men ikke i spørreskjemaundersøkelser. Om spørreskjemaet avleses optisk vil svaret ofte se hinsides ut og respondentens svar kan bli forkastet.

Noe av det vanskeligste en kan gi seg inn på i spørreskjemaer, er instruksjoner av typen: «Dette spørsmålet skal du besvare, gitt en bestemt betingelse eller gitt en bestemt omstendighet». Det viser seg i praksis at mange – også voksne, seriøse – respondenter ikke greier å følge slike anvisninger. De hopper over betingelsen, legger ikke merke til den og greier i alle fall ikke å inkorporere den i sin atferd. De svarer altså usant uten at det dreier seg om bevisst feilinformasjon. Slikt skjer både når det gis anvisninger i egne tekstsekvenser og når rettleidninger gis i selve spørsmålsteksten.

Slike erfaringer er omfattende beskrevet i metodelitteraturen. I det følgende skal jeg drøfte noen flere sider ved Ung i Oslo 2006 som – potensielt – kan ha generert ikke-valide målinger, og jeg skal også vise eksempler på hvordan mange av elevene åpenbart ikke har forstått oppgavene.

2.2 Omfang, tidspress, muligheten for «feil» svar pga. utmattelse

Statistisk sentralbyrå har som hovedregel at intervjuer ikke bør belaste respondentene i mer enn omlag 45–60 minutter. Grunnen er at metodeundersøkelser viser at villigheten til å svare og muligheten for å holde oppmerksomheten oppe, synker markant om intervjuet varer for lenge.

Dette problemet er svært relevant for bedømmelsen av datakvaliteten i Ung i Oslo 2006, som er en uvanlig *stor* undersøkelse. Ungdommer i 14–17-årsalderen må forholde seg til omlag 550 *spørsmål* som skulle besvares i løpet av to skoletimer uten pause i mellom (Valset & Øia 2006). Ettersom noe tid går til oppstart, kan vi regne med at elevene stort sett har hatt anslagsvis 80 minutter til rådighet.

Størrelsen på Ung i Oslo 2006-undersøkelsen er spesiell. Mange skolebaserte spørreskjemaundersøkelser har mindre omfang. HEVAS-undersøkelsene i Bergen, for eksempel, som riktignok gjelder en litt annen aldersgruppe (11–15 år) har vanligvis stilt i underkant av 200 spørsmål (Wold et al. 1995). Skoleundersøkelsene til SIRUS (Pape et al. 2007; Rossow & Bø 2003) nøyer seg med spørreskjemaer som besvares i løpet av én skoletime.

Ung i Oslo-undersøkelsen forventer altså at elevene skal besvare, eller forholde seg til, på en valid måte, omlag 7 spørsmål i minuttet i gjennomsnitt, kontinuerlig gjennom om lag 1½ time. Hver rettleidning og hvert

spørsmål må leses såpass nøye at de får med seg hva det dreier seg om. Deretter skal de fundere over hvilket svar/svaralternativ som er riktig og krysse av på rett plass, eventuelt bestemme at de ikke kan eller ikke vil svare. I gjennomsnitt har de mindre enn ti sekunder til hvert spørsmål. De skal – for å si det mildt – håndtere denne store spørsmålmengden eksepsjonelt kjapt og presist for å besvare valid på alt, om de skal gjøre det forskerne forventer av dem.

Elevene har jo trening (ihvertfall en del av dem) fra vanskelige to-timers prøver, men spørsmålet er vel om ikke Ung i Oslo 2006 overgår det meste de tidligere har vært utsatt for?

Av betydning er også at spørsmålene i Ung i Oslo dreier seg om vidt forskjellige felt og temaer – fritid, skole, forhold til foreldre, rusmiddelbruk, osv. Respondentene må stadig skifte oppmerksomhet fra ett område til et annet.

Det er knapt mulig å trekke noen annen konklusjon enn at mange, kanskje til og med de fleste, ikke vil være i stand til å leve opp til disse forventningene. En kan tenke seg at mange starter med noenlunde gode intensjoner, og de første spørsmålene besvares ordentlig. Etterhvert gjør trøttheten seg gjeldende, de ser de får dårlig tid, de greier ikke å holde oppmerksomheten oppe, de kan ikke ha full konsentrasjon i så lang tid. Etterhvert vil kanskje svarene bli avgitt mer og mer tilfeldig, de har ikke tid til å tenke seg om, eller de begynner å hoppe over enkeltspørsmål eller hele spørsmålssekvenser. Noen begynner å plassere avkryssninger i hytt og vær bare for å få det unna, noen blir frustrerte og begynner å sette kryss uten å lese spørsmålet ordentlig, eller de setter kryss helt uten å vite hva spørsmålet var. Noen kan kanskje «kødde» i frustrasjon, noen gir kanskje helt opp.

Det virker sannsynlig at selv elever som prøver å være «flinke» respondenter, vil komme i en situasjon der avkryssningene etterhvert skjer ganske ubetenksomt. Antakeligvis vil en del skrive feil i farten – når en skal foreta 500 avkryssninger er det ikke annet å vente, og de har jo ikke tid eller mulighet til å viske ut det gale svaret for å plassere krysset riktig. Noen kommer til å krysse av noe de ikke hadde ment, eller oppgir et antall ganger de har gjort noe som mer eller mindre er tatt ut av luften.

Omfang og tidspress vil derfor generere lav validitet, men i hvilket omfang?

Spørsmålet er ikke lett å svare på. Det som i hvert fall framgår, er at andelen ubesvarte spørsmål øker utover i skjemaet og er relativt høy mot slutten. Vi kan undersøke dette for seks spørsmål som er lette å besvare: spørsmålene om klassetrinn (et av de første spørsmålene), om røyking (omtrent spørsmål nr. 120 i spørsmålsrekkefølgen), om en kjenner venners foreldre (ca. nr. 250), om en har mobiltelefon (ca. nr. 380), om en trener i et idrettslag (ca. nr. 470) og om en er/har vært med i en forening (ca. nr. 520). Andelen som ikke har avgitt svar (dvs. registrert med *sysmis* i datafilen) stiger systematisk med hvor seint i skjemaet spørsmålet kommer. Andelen ubesvart er 1,9 prosent helt i begynnelsen, deretter 2,4 prosent for spørsmål nr. 120, 3,9 prosent for spørsmål nr. 250, 5,5 prosent for spørsmål nr. 380, 12,9 prosent for spørsmål nr. 470 og endelig 15,7 prosent for spørsmål nr. 520.

Ofte framheves Ung i Oslo-materialet på grunn av sin høye svarprosent. For så vidt er ikke det feil, for mange har deltatt, men en realistisk vurdering er likevel at for de siste 20–30 prosentene av spørsmålene er den *reelle* svarprosenten ikke stort høyere enn hva personlige intervjuundersøkelser oppnår. Det problematiske med å stille *for mange* spørsmål ble allerede påpekt for den tilsvarende datainnsamlingen som foregikk ti år tidligere – Ung i Oslo 1996. Da var det ikke mindre enn 800 spørsmål som elevene skulle ta stilling til i løpet av to skoletimer (Bakken 1998). Ung i Oslo 2006 er noe mindre, men fortsatt antakeligvis for stor til at flertallet greier å få til særlig omtenkksomhet i hvordan de fyller ut skjemaet.

2.3 Tidsangivelser som ikke blir fulgt

Tidsangivelser – når foregikk noe, hvor mye har noe foregått i løpet av en viss tidsperiode – er alltid vanskelig å få til å fungere i spørreskjemaundersøkelser. Her er et eksempel fra Ung i Oslo 2006-materialet. Noe etter halvveis i skjemaet (altså når mange av elevene ikke lenger er så oppmerksomme som de var til å begynne med) kommer en serie på 22 spørsmål om aktiviteter en gjør i fritida, som besøke nabo, vært på kafé, vært i fritidsklubb, drevet med trening, hjulpet til hjemme, osv. Dette dreier seg altså om stort sett trivielle aktiviteter som mange ungdommer har vært borti.

I teksten øverst på siden angis det at elevene skal fortelle hvor mange ganger aktivitetene er gjort «*siste uke, siste 7 dager*». Et tall skal skrives inn. Aktivitetene er av en type som sjelden vil kunne ha foregått flere ganger i løpet av en dag. Med andre ord: Om en elev svarer ti ganger eller mer for en viss aktivitet, er dette med all sannsynlighet et ikke valid svar.

Men mange har gjort det. Omlag 2300 elever har minst én gang på de 22 aktivitetene skrevet inn et tall som er 10 eller høyere, altså noe som i nær sagt alle tilfelle er en ikke valid opplysning. (Dette er beregnet fra arbeidsfilen med de originale data for vel 11.440 respondenter. I den ferdig-redigerte datafilen med 11.500 respondenter som jeg ellers bruker her, er alle slike urimelig høye svar blitt omkodet til «6 ganger og mer». Denne kategorien kamuflerer hvor mange som har oppgitt et tall som ofte er langt større enn ukas sju dager.)

Det er neppe over 2000 av respondentene som juger om slike aktiviteter. Forklaringen på mønsteret er enkel og vel kjent i metodelitteraturen. I spørreskjemaundersøkelser vil mange ikke legge merke til eller oppfatte at svar skal gis *betinget* av et eller annet («for siste uke», «bare hvis du...», osv.). Her har øyensynlig mange elever oversett «siste sju dager» og bare lest teksten «Hvor mange ganger har du ... besøkt nabo?» osv. Så setter de ned et tall, ut fra litt uklare tidsforestillinger – noen for hele livet, noen for den siste tida, osv., og de som oppfattet rettleidingen har svart for siste sju dager.

Noen har argumentert for at urimelig høye tall på disse spørsmålene representerer tulle svar. Selvsagt har også noen tullet. Om en går inn i datafilen, vil en f.eks. se en og annen som har svart med 20 på første spørsmål, 30 på neste, deretter 40, fjerde spørsmålet får 50, osv. – og dette er nok bevisst tøys.

Men det er selvsagt ikke så mange som 2000 elever som tuller. En må ta i betraktning at spørsmålene er banale – de er ikke av den typen som pirrer til bevisst feilinformasjon. Derfor: De aller fleste av de som har svart ti eller mer, har svart *riktig* ut fra sin egen subjektive forståelse av spørsmålet.

Altså: I hvert fall 2000 av de om lag 11.500 elevene har forstått dette spørsmålet annerledes enn forskerne mente de skulle. Men naturligvis er det reelle antallet som har misforstått, høyere. Også mange av de som har brukt lave svaralternativer (som 0, 2 eller 5) og som derfor ikke påkaller seg forskernes oppmerksomhet, for en person kan jo lett ha vært 5 ganger på kafé

en uke, vil jo helt sikkert *også* ha misforstått hva slags tidsrom det skulle svares for, akkurat på samme måte som de som skrev 10 eller 20. I praksis er det umulig å beregne hvor mange som misforsto – var det 3000, 5000?

Slike feil i konstruksjonen av spørreskjemaer er ofte fatale, i hvert fall hvis en er ute etter eksakte anslag over hyppighet/forekomst av aktiviteter. Den sosialpolitiske interessen i disse spørsmålene knytter seg f.eks. til «å ha spilt på spilleautomat med pengepremie». Hvor ofte foregår slikt? Fra Ung i Oslo 2006-dataene er det umulig å beregne, ettersom det er umulig å vite hvor mange av respondentene som har svart for siste uke, hvor mange som har svart for «siste tida» og hvor mange som har svart for «i det hele tatt».

2.4 Skiftende tidsperioder og umenneskelige hukommelseskrav

Et tema som har opptatt metodelitteraturen er altså kartleggingen av omfang og hyppighet av bestemte former for atferd. Dette kan gjøres på mange forskjellige måter. En kan spørre om respondenten har gjort atferd A «noen gang i det hele tatt», eller om en vanligvis gjør A «aldri, sjelden, ofte, svært ofte», eller om en gjorde A «i går», eller om en har gjort A «aldri, noen ganger, flere ganger» i løpet av «siste uke» eller «siste måned» eller «siste år», eller om eksakt antall ganger en har gjort A, osv.

Ulike formuleringer kan gi svært ulike anslag. For eksempel: I Statistisk sentralbyrås intervjuundersøkelser (personlige intervjuer) blir voksne spurt om «antall ganger hos lege siste 14 dager» og deretter om «antall legebesøk sist år». Begge spørsmålene kan brukes for å beregne antall legebesøk i gjennomsnitt pr. person pr. år. Slike beregninger gir ofte et resultat på ca. 6–7 med den første spørsmålsformuleringen, men bare omlag 2–4 for den andre spørsmålsformuleringen.

Tidsangivelser kan fungere presist om en spør om noe ble gjort i går («Var du på kino i går?» «Spiste du middag hjemme i går?»), men en kort tidsangivelse fører alltid til så få positive svar at analyseproblemene blir store. Derfor velger forskere lengre tidsperioder, men da dukker problemet fort opp – folk husker ikke når noe foregikk, de greier ikke å plassere atferd A i tid – var det siste 14 dager, i siste måned eller ikke, ble A utført for 11 eller 14 måneder siden?

Når det gjelder antall ganger, er det lettere å besvare spørsmålet valid om atferd A *aldri* er gjort, eller bare gjort en eller to ganger – gitt, riktignok, at dette er noe folk husker. Metodeundersøkelser i USA der en sammenligner selvrapporterte opplysninger om operasjoner ved sykehus med sykehusets egne kartoteker, viser at overraskende mange blant de som fikk utført ikke bagatellmessige operasjoner siste år, synes å ha glemt det når de spørres om det.

Om atferd A derimot har forekommet hyppig, er det ikke mange respondenter som kan greie å huske antallet ganger noenlunde presist. De vil i hvert fall trenge en god del tid for å tenke seg om. Vi kan jo alle prøve sjøl: Hvor mange ganger var *du* på restaurant, på Vinmonopolet, kranglet med ektefellen, spiste pizza, kjøpte klær, var på kino, jobbet overtid – i løpet av siste år?

Vanskene med å angi antall ganger en aktivitet har foregått i et bestemt tidsrom, fører i praksis til at respondentene – ofte uten å være seg det helt bevisst – unnlater å reflektere spesifikt på tidsangivelsen. Mange oppfatter ikke at spørsmålet gjelder for et visst tidsrom, og metodelitteraturen har mange eksempler på hvordan slike tidsangivelser blir oversett. Ofte gir respondentene et slags anslag for hvor mye de føler at atferd A har forekommet, med varierende oppfatninger om tidsperiode – spør man om noe som forekommer sjeldent («har en bekjent begått selvmord?») vil man svare for hele livet, spør man om noe som foregår oftere («sett Dagsrevyen») vil man svare for de siste dagene.

Ung i Oslo 2006-spørreskjemaet har brukt ulike spørsmålsvarianter for å avdekke hvor mye visse typer atferd har forekommet. Mange ganger spørres det om eksakt antall ganger (f.eks. antall seksualpartnere, antall bøker en har lest), andre ganger om noe har foregått «ofte» eller «av og til», osv. I tillegg varierer tidsangivelsene, og det kan bidra til mer forvirring – noen ganger spørres det for hele livet, andre ganger om siste år, siste 12 måneder, siste måned, siste uke, osv. At respondentene hele tiden må skifte tidsreferanser er jo en egen kilde til at respondenter kan gå i surr.

Et karakteristisk trekk ved Ung i Oslo er at svært ofte er de vanskeligste variantene brukt, for eksempel i en serie spørsmål om regelbrudd. Her blir elevene spurt om eksakt antall ganger de har begått 23 typer regelbrudd i løpet

av siste år (som å ha sniffet, stjålet, vært beruset, innkalt til rektor, skulket, stjålet bil/motorsykkkel, kranglet med lærer, brukt hasj/marihuana, osv.).

Elever som ønsker å svare riktig, og som aldri har gjort noe galt eller kanskje bare en eller to ganger som de husker tydelig, vil nok kunne gi noenlunde valide svar på slike spørsmål.

Men for de som relativt ofte, over lengre tid, har «vært i slåsskamp» eller «sneket på trikken», er det i praksis umulig å gi et presist svar. De aner ikke, husker ikke, og har i hvert fall ikke tid til å tenke tilbake, måned for måned, hele siste året, om hvor mange ganger de har gjort ditt og datt.

Det blir en slags diabolsk logikk i dette, som umuliggjør valide svar fra hyppige regelbrytere. De som relativt ofte har vært borti slikt, og som prøver å svare ærlig, vil uvegerlig skyves inn i en opplevelse av å mislykkes, ettersom de ikke har den bokføringsaktige hukommelsen som er nødvendig for å svare sant. Om slike hyppige regelbrytere svarer, vil det aldri kunne skje på noen annen måte enn at de slenger ut det tallet de føler er passe – 11 ganger, 24 ganger, 50 ganger, osv. Dette tallet kan aldri være bokstavelig sant – i beste fall gir det et løst inntrykk av størrelsesordenen. Også om hyppige regelbrytere prøver å svare ærlig, vil de derfor ofte komme til å svare på en måte som gir et inntrykk at de er upålitelige respondenter, for eksempel om de svarer konsekvent med tallet 30 eller 50 på en hel rad regelbrudd-spørsmål etter hverandre.

2.5 Ordforståelse, leseforståelse, tallforståelse

Det er selvsagt at om alle elevene og ikke bare de «flinkeste» skal kunne svare riktig (valid) på spørsmålene, må ord og uttrykk som brukes bli forstått. Respondentene i Ung i Oslo er unge, ofte umodne, har begrenset erfaring og kunnskap, noen leser sjelden eller aldri aviser, de er skoleflinke i svært varierende grad, hos en del er ordtilfanget begrenset og lesekyndigheten svak. Rundt hver femte har innvandrerbakgrunn og en del av disse har ikke gode norskkunnskaper.

En feilkilde er at spørreskjemaer bruker ord som volder problemer hos respondenter. Et eksempel fra en spørreskjemaundersøkelse blant *voksne*: På spørsmålet «hvor mange er det i din husstand, inklusive deg selv?» svarte en ikke ubetydelig andel 0 (Axel West Pedersen, personlig kommunikasjon).

Problemet var åpenbart ordet «inklusive» som mange må ha trodd betyr «eksklusive». 0-svar kan selvsagt fortolkes som 1, men om 1 er et valid svar eller skal fortolkes som 2, osv., er umulig å svare på. En slik feil, generert av spørreskjemaet, kan altså oppstå om forskere tror at egen språkforståelse og ordkyndighet uten videre lar seg overføre til ordforståelsen blant folk flest.

De aller fleste spørsmålene i Ung i Oslo bruker ord og formuleringer som neppe er problematiske, men hva med ord som adoptert, folkegruppe, funksjonshemmet, bifil, samleie, spiritisme? For forskere selv er betydningen av «folkegruppe» en kinkig nøtt. Ordet «funksjonshemmet» er ikke mindre vanskelig. Skal en inkludere diabetikere som er avhengige av daglige sprøyter, men ellers lever normalt? Er det å ikke føle seg vel, sjenerthet, når en møter noen en ikke kjenner, en funksjonshemning? Når forskerne selv strides om et slikt begrep (en uenighet som fører til vidt sprikende anslag over andelen «funksjonshemmete» i befolkningen), hvor mye vil ikke oppfatningene sprike blant 15-åringer? En panelundersøkelse viste at ikke bare Bill Clinton, men også mange unge amerikanske kvinner, hadde problemer med ordet «samleie» («intercourse»). En rekke tidlige seksuelle erfaringer var vanskelig å definere definitivt som samleie eller ikke, og opplevelser som kvinnene først opplevde som samleie, ble seinere omdefinert. Forskerne knyttet dette til hvorfor en betydelig andel oppga vidt forskjellig alder for når de hadde hatt sitt «første samleie» i to spørsmålsrunder foretatt med noen års mellomrom (Upchurch et al. 2002).

2.6 Interesse, relevans, oppmerksomhet

Et velkjent fenomen i surveyundersøkelser er at forskere spør om noe som respondenten aldri har interessert seg for eller aldri har brydd seg om. Prototypen er spørsmålet norske meningsmålingsinstitutter ofte brukte i sin tid: Hvem ville du velge å spise middag sammen med – Gro Harlem Brundtland eller Kåre Willoch? Problemstillinger reises som respondenten ikke har kjennskap om, ingen mening om, eller ikke bryr seg om.

I metodelitteraturen omtales dette ofte som at temaer har ulik grad av «salience» (relevans, interessant for respondenten, framme i respondentens bevissthet, mv.). En typisk reaksjon vil være at respondenten hopper over spørsmålet. Typisk er det også å velge svaralternativer som «vet ikke», «har

ingen mening», «husker ikke», men om mange gjør det, kan en lett reise tvil om svarenes representativitet. Et vanlig mottiltak fra forskernes side er å utelate svaralternativer som «vet ikke/har ingen mening», slik at respondenten tvinges til å velge mellom to ubekvemme situasjoner: Enten blir respondenten en mislykket svarer som ikke fyller ut rubrikkene, eller han/hun tvinges til å markere en mening, en erfaring eller en posisjon som han/hun egentlig ikke har.

Med noen få unntak (som seksuell legning, interesse for politikk/samfunnsspørsmål, fars og mors utdanning) har Ung i Oslo-spørreskjemaet konsekvent utelatt «vet ikke» og liknende svarmuligheter, åpenbart for å gjøre det vanskelig for elevene å velge det uforpliktende «vet ikke/vil ikke svare»-alternativet. Dette er ikke i strid med anbefalinger fra metodeforskere. Mange mener at denne framgangsmåten alt i alt er å foretrekke. Men det fører til en overestimering av hva elevene har en mening om, ettersom «vet ikke»-gruppen fratras det svaralternativet som er valid *for dem*. Når respondenter tvinges til å svare på noe de er uinteressert i, har ingen mening om, ikke bryr seg om eller ikke har tenkt over, vil svarene ha lav validitet (de uttrykker ikke respondents mening/erfaring/posisjon). De vil dessuten ofte ha lav reliabilitet – med reliabilitet menes om det er konsistens mellom svarene, for eksempel mellom svar avgitt på forskjellige spørsmål i skjemaet eller samsvar mellom flere svar avgitt på forskjellige tidspunkter. Ettersom respondenten ikke føler sterkt for svaret sitt kan han/hun lett komme til å gi et annet svar ved en annen anledning. Om vi undersøker hvor pålitelige respondentene er, kan lav reliabilitet (manglende konsistens) på slike spørsmål i nesten omgang føre til at respondentenes svar på andre spørsmål urettmessig blir klassifisert som lite valide.

På en del spørsmål i Ung i Oslo 2006 har datainnsamlingsmåten konstruert et bilde av elevmassen som har problematisk validitet. En må ta i betraktning at respondentene er unge, ofte umodne, lite kunnskapsrike og kanskje ikke har særlig sterke eller klare oppfatninger om samfunnsspørsmål. Likevel stilles de 17 spørsmål om hvordan de forholder seg til sin folkegruppe, 20 spørsmål om politiske holdninger, 7 spørsmål om tillit til politikere, dommere og advokater, leger og sykehus osv., mv., – alt uten at de gis anledning til å markere (på annen måte enn å unnlate å svare) at de tilhører en kategori der slike temaer er langt unna det de bryr seg om.

2.7 Tvetydige spørsmål, spørsmål som er ulogiske, osv.

Det som er sagt over, gjelder generelle trekk ved Ung i Oslo-undersøkelsen. De peker alle på muligheter for at undersøkelsesopplegget vil generere ikke-valide data. En videreføring, som ikke vil bli gjort her, er å drøfte i detalj hvert enkelt spørsmål. Mange spørsmål er greit nok formulert og kan sikkert fungere bra, men en del enkeltspørsmål er problematiske. Et par eksempler:

I alkohol-sekvensen lyder første spørsmål slik: «Hvor gammel var du første gang du drakk alkohol, slik at du følte deg beruset?» (skjemaet, s.16 i Valset & Øia 2006).

Dette er et spørsmål av en type som ofte vil bli misforstått i spørreskjemaundersøkelser, ettersom det inneholder to ledd – et hovedledd når første gang drakk alkohol og et betingende sideledd slik at du følte deg tydelig beruset. Alle metodebøker advarer mot slike to-leddete budskap, for det viser seg alltid at mange respondenter ikke greier å forholde seg til de to budskapene samtidig.

Det har heller ikke forskerne som lagde skjemaet greid. Det første svaralternativet er å krysse av, eller la stå åpent, en rubrikk for «Jeg har aldri drukket alkohol». Hvor ble det av «slik at du følte deg beruset?». Tvetydigheten gjør at svarerne må bli usikre. Skal de tenke på hva den første teksten spør om («drukket, følt seg beruset») eller på hva avkrysningsalternativet spør etter («aldri drukket alkohol»- underforstått «drukket i det hele tatt, uansett grad av beruselse»)?

Hvis en ikke krysser av her, skal en skrive inn en aldersangivelse i den neste rubrikken. Om respondenten har skrevet en alder, er det umulig å vite om dette betyr «første gang de drakk alkohol i det hele tatt» eller «første gang de var beruset». De som har drukket alkohol, men ikke vært beruset, vil ha store problemer – hvor hører de egentlig hjemme i svaralternativene? Skal de ikke besvare dette spørsmålet i det hele tatt? Hva har spørsmålskonstruktørene egentlig tenkt?

Alkoholsekvens-spørsmålene har svak kvalitet. Selv respondenter med de beste intensjoner vil vikle seg inn i ulogiske kombinasjoner. Validiteten på svarene blir diskutabel på grunn av dårlige spørsmål, og dette gjenspeiles i at svarene på de ulike alkoholspørsmålene (det er flere av dem) ikke har udiskutabel innbyrdes sammenheng eller særlig høy konsistens.

Et annet eksempel, der i det minste en del respondenter får problemer (men dette har kanskje mindre konsekvenser): Skjemaet har en del spørsmål av typen «har familien bil», «har du eget rom», «hvilket yrke har far», osv. For de fleste er disse spørsmålene neppe spesielt kompliserte å svare på, men et visst problem oppstår for de som må forholde seg til foreldre, familie og hjem på en annen måte enn de som har en klar tilhørighet i en bestemt familie.

Omtrent 20 prosent i utvalget bor hos «mor og ny samboer» (eller «far + ny samboer»), «omtrent like mye hos mor og far», i «fosterhjem», hos «fosterforeldre», eller i «annet»? For en del vil hva som er «hjemme» være utvetydig, men noen vil forholde seg til mors nye samboer som til en far, noen forholder seg like mye til to farsfigurer, noen flytter ofte mellom far og mor, osv. I spørsmål som involverer hjem, familie, far osv. vil en del være usikre på hvilken far eller hvilken familie spørsmålene refererer til. Dette er ikke nødvendigvis et stort problem – kanskje de fleste har en umiddelbar selvsagt fortolkning – men spørsmålene tar aldri høyde for slike tvetydigheter. Dette gir seg utslag blant annet i at de med slike spesielle familie- og hjemrelasjoner oftere velger å avstå fra å besvare enkle spørsmål om familie, far, boforhold, osv. Det viser seg at spesielt de med fosterforeldre og fosterhjem har særlig høye andeler ubesvart på slike spørsmål.

2.8 Oppsummering

Denne diskusjonen peker uten tvil på at Oslo-datinnsamlingen inneholder en lang rekke *potensielle* kilder til lav datakvalitet. Kort oppsummert består disse kildene blant annet i:

- omfang og tidspres (mer enn 500 spørsmål, 10 sekunder pr. spørsmål)
- konsentrasjons- og oppmerksomhetsproblemer som en følge av omfanget
- mange respondenters tendens til å overse/misforstå det som står i skjemaet
- et betydelig antall respondenter som overser tidsangivelser
- ordbruk som ikke er tilpasset en del respondenters svake ord- og leseforståelse
- urimelige krav til hukommelse
- framprovosering av tvungne svar om temaer som ikke er relevant for respondentene
- spørsmål som er stilt ulogisk, med manglende/utydelige svaralternativer

Slikt kan ha generert ikke-valide målinger. Men det må også sies: Spørreskjemaer er vanskelige å konstruere. Mange av de diskutabile aspektene ved Ung i Oslo 2006-skjemaet vil en gjenfinne, i større eller mindre grad, i andre liknende undersøkelser. Det er mange grunner til at dette skjer, for eksempel press fra oppdragsgivere som vil ha svar på alt billigst mulig. Men også forskernes manglende kunnskap om metodestudier bidrar, samt gjerne at det ikke har foregått noen grundige tester og utprøvinger av spørsmålene, og at forskerne har for lite kunnskap og fantasi om hvor rart respondenter kan tolke ord og anvisninger.

Jeg ønsker selvsagt ikke å overdrive. Kritikken over kan balanseres ved å framheve at mange av spørsmålene faktisk er velprøvde og helt greit formulerte, forskerne har nedlagt mye arbeid i å informere og motivere lærerne som står for den praktiske gjennomføringen, osv. Antakeligvis har de fleste elevene bestrebet seg på å løse oppgaven de ble stilt overfor så bra de har vært i stand til.

Dermed er det *også* mange kilder til god datakvalitet. I neste hovedavsnitt vil jeg drøfte spørsmålet om datakvaliteten i Ung i Oslo 2006 via litt andre innfallsvinkler.

3 Drøfting av kvaliteten på Ung i Oslo 2006-dataene

3.1 Argumenter for at datakvaliteten er høy

Argumentene *for* at datakvaliteten i NOVAs ungdomsundersøkelser er god, er oftest presentert i form av ett eller flere av de følgende fem punktene:

- (1) *Henvisninger til høy svarprosent.* Skolebaserte datainnsamlinger som Ung i Oslo 2006 gir alltid høy svarprosent, sammenlignet med for eksempel personlige intervjuer, der det kreves større frivillighet og en viss grad av selv-seleksjon for å delta, og i enda større grad sammenlignet med postale, hjemmebaserte, spørreskjemaundersøkelser.
- (2) *Inntrykk av at elevene arbeider seriøst med å svare.* Som en oppsummering av Ung i Oslo 1996-undersøkelsen sier: «En gjennomgang av skjemaene gir inntrykk av samvittighetsfull utfylling. Både tilbakemeldingen fra lærer og andre som var tilstede tyder på at ungdom flest tok utfyllingen svært seriøst.» (Bakken 1998). Samme sted noteres det at puncherne ble instruert om å fjerne skjemaer som «bar tydelig preg av tullede svar». Svært få skjemaer ble forkastet av denne grunn (25 av ca. 11 450). Det er ikke, så langt jeg kjenner til, foretatt en systematisk undersøkelse blant lærerne i etterhånd for å granske deres inntrykk på en mer rigorøs måte.
- (3) *Henvisning til andre metodestudier.* Noen studier basert på Ung i Oslo-data, se f.eks. Pedersen & Hegna (2003), henviser til utenlandske metodestudier som viser rimelig høy validitet i liknende undersøkelser. At stringente metodestudier av andre spørreskjemaundersøkelser har konkludert med at *de* undersøkelsene ga rimelig bra datakvalitet, kan selvsagt være en indikasjon på at også Ung i Oslo-dataene holder et rimelig bra nivå. Det er likevel et problem med slike henvisninger. Kvaliteten av data i spørreskjemaundersøkelser er svært sensitiv overfor små variasjoner i spørsmålsformuleringer, omfang, osv. Det er følgelig uklart om i hvor stor grad metodestudier av *andre* undersøkelser har relevans for Ung i Oslo-dataenes kvalitet.

- (4) *Henvisninger til rimelige univariate svarfordelinger.* En kilde til oppfatningen om at Ung i Oslos datakvalitet er høy, er de *univariate fordelingene* av hver enkelt variabel, for eksempel fordelingen av røykevaner, av familietyper, av fars/mors utdanning, av forekomsten av funksjonshemninger, andelen som sier de er homofile, lesbiske og/eller bifile, osv. Praktisk talt alltid viser fordelingene et rimelig mønster som i ganske stor grad er i samsvar med hva andre undersøkelser viser.

Problemet med dette argumentet er: De potensielle feilkildene som er skissert i del 1, vil stort sett (det finnes unntak) føre til at de ikke-valide svarene fordeler seg noenlunde jevnt rundt det valide svaret. En daglig røyker kan krysse av rubrikken for «røyker, men ikke daglig» i farten, eller fordi han gjerne vil slutte, mens en som ikke røyker daglig kan krysse av i rubrikken for «daglig røyking» om han har røykt de tre siste dagene, osv. Tilfeldige feil, gjort på grunn av misforståelser, manglende hukommelse, feilkrysninger, tidspress, osv, vil som regel ikke føre til at univariate fordelingene blir misvisende, ettersom feilene vil fordele seg tilfeldig rundt de sanne verdiene.

Dette betyr at de univariate fordelingene kan være valide for gruppen som kartlegges, selv om de enkelte svar for hver enkelt respondent ikke er valid. Andelen valide svar på individnivå kan både være svært høy (f.eks. 97 prosent) eller relativt lav (f.eks. 60 prosent), men i begge tilfelle kan den univariate fordelingen i hele utvalget tilsvare den sanne fordelingen. At de univariate fordelingene på utvalgsnivå er troverdige, gir ingen garanti for at validiteten på det individuelle nivået er høy.

- (5) *Teoretisk plausibilitet.* Et siste generelt argument kan være at de statistiske mønstrene – de statistiske *sammenhengene* – som framkommer i analyser av data i Ung i Oslo-undersøkelsen, overensstemmer grovt sett med eksisterende teori og med andres funn. Dette er også et relevant argument – og, så langt jeg kjenner til – finner en ofte statistiske mønstre i Ung i Oslo-dataene som er teoretisk plausible, som f.eks. sammenheng mellom sosial bakgrunn og karakterer. Dette kan bety at eventuelle problemer med validitet på individnivå ikke er så store at de skjuler de virkelige prosessene som foregår «der ute».

3.2 Frafallet er sosialt skjevt

Det er ofte vanskelig å vurdere argumentene skissert i avsnitt 3.1, ettersom det i praksis kan kreves spesialundersøkelser for å ha et solid grunnlag for å ta stillingen til dem. Men et punkt kan det være verdt å trekke fram her, nemlig spørsmålet om sosial skjevhet i frafallet. Som nevnt over (avsnitt 2.2) er ikke den reelle svarprosenten så høy som det ofte gis inntrykk av, ettersom undersøkelsens krevende omfang fører til at en del elever ikke greier å fullføre. Av datamaterialet framgår det at selv på de letteste spørsmålene har om lag hver sjettede elev latt være å avgi svar når det gjelder spørsmål som kommer seint, i den siste tredelen, av undersøkelsen.

Men selv om en del faller fra, har jo Ung i Oslo 2006 relativt høy svarprosent, men dette kan problematiseres enda mer om vi stiller spørsmålet om representativitet – spørsmålet om alle sosiale lag blant elevene er like godt representert. Opplegget i Ung i Oslo 2006 stiller jo store krav til de som skal svare, og for å gjennomføres på en adekvat måte trengs det stor leseforståelse og en høy evne til å arbeide kjapt og effektivt. De som ikke makter dette, vil komme til kort – de vil ikke greie å fullføre.

Ikke overraskende viser det seg at det særlig er de med svakere skolekarakterer som ikke får det til. Svært tidlig i spørreskjemaet blir elevene spurt om karakter i norsk skriftlig. Dette spørsmålet besvares av nesten alle (bare 3,9 prosent frafall). Blant dem oppga 562 en dårlig karakter (1 eller 2), mens 2907 oppga en god karakter (5 eller 6). Vi kan, slik det ble gjort i avsnitt 2.2, vise hvordan svarprosenten utvikler seg gjennom spørreskjemaet for de to gruppene. Fokus er på spørsmål som er enkle å svare på, spredt utover i spørreskjemaet: om klassetrinn (et av de første spørsmålene), om røyking (spm nr. 120), om en kjenner venners foreldre (ca. 250), om en har mobiltelefon (ca. 380), om en trener i et idrettslag (ca. 470) og om en er/har vært med i en forening (ca. 520).

Tabell 1. Andel (%) uten svar på enkle spørsmål, fra tidlig i spørreskjemaet til seint i spørreskjemaet. Elever som opplyser om dårlige (N=562) og gode norskkarakterer (N=2907).

Norsk-karakterer	Klasse-trinn	Røyking	Kjenner venners foreldre	Har/har ikke mobiltelefon	Trener i idrettslag	Er/har vært med i forening
Dårlige (1-2)	1,2	4,4	8,7	16,4	27,8	31,3
Gode (5-6)	1,0	1,5	2,1	2,6	7,6	8,8

Tabell 1 demonstrerer at alle, enten de sier de har dårlige eller gode norskkarakterer, starter med gode intensjoner. På det første spørsmålet er det knapt noen forskjell i svarprosent. Også etter omlag 120 spørsmål (når røykespørsmålet kommer) er det god svarprosent blant begge grupper, også blant dem med dårlige norskkarakterer. Så begynner de med dårligere karakterer å falle fra, og mot slutten av spørreskjemaet er det omlag 30 prosent som ikke er med lenger, selv på lette spørsmål. De med bra karakterer håndterer oppgaven bedre. Her begynner riktignok et frafall å gjøre seg gjeldende etter omlag 450 spørsmål, men helt til slutt har likevel fortsatt under 10 prosent av dem falt fra.

Mer detaljert blir slike mønstre framstilt i tabell 2. De høye svarprosentene for alle grupper i starten av spørreskjemaet betyr at det i utgangspunktet ikke er sosiale forskjeller i elevenes motivasjon for å delta i undersøkelsen. Men om intensjonene er like gode, er kapasiteten til å gjennomføre undersøkelsen forskjellig. Elever med fedre som er lavt utdannede og/eller uten arbeid, elever fra trolig mindre velstående familier (har ikke eget soverom), elever som selv går på yrkesfaglige linjer på VK I og ikke på allmennfaglig, og røykere, har i den siste halvparten av undersøkelsen markant høyere frafall enn de med mer privilegerte plasseringer.

Dette er ikke en uviktig kritikk av Ung i Oslo 2006-dataene. En viktig begrunnelse for å gjennomføre skolebaserte undersøkelser av denne typen, er at hensikten er å fange opp mange flere av dem som ellers, i f.eks. vanlige intervjuundersøkelser, ikke blir kartlagt. På denne måten skal en kunne gi mer realistiske og sannferdige anslag, for eksempel om sosiale problemer blant ungdommen som helhet.

Men opplegget i Ung i Oslo har en tendens til å støte ungdommene med mindre ressurser fra seg. Det er ikke et godt tegn når en så stor del av

ungdommen plassert i nedre deler av ulike sosiale hierarkier ikke greier å gjennomføre. De avviser jo ikke å delta i utgangspunktet, de svarer til å begynne med. Det stigende frafallet etterhvert dreier seg om at kravene de utsettes for, er tilpasset andre enn dem selv – litt spisst kunne vi si at kravene er tilpasset barn fra forskerfamilier, undersøkelsen ligner en skoleprøve der de «flinke» lykkes, mens de som ikke fikser skolekravene like bra har mer problemer.

Tabell 2. Andel (%) uten svar på enkle spørsmål, fra tidlig i spørreskjemaet til seint i spørreskjemaet, gruppert etter forskjellige bakgrunnsvariable*

	Klasse -trinn	Røyking	Kjenner venners foreldre	Har/har ikke mobil- telefon	Trener i idrettslag	Er/har vært med i forening
Fars utd.						
- lav	1,6	1,2	4,1	6,6	14,6	20,4
- høy	1,5	1,5	2,0	3,0	9,9	10,9
Fars arbeid						
- uten arbeid	1,8	1,8	5,1	7,6	17,5	21,9
- heltid	1,3	1,3	2,3	3,4	10,0	12,3
Eget soverom?						
- nei	1,0	1,4	4,3	8,1	18,7	22,0
- ja	1,5	1,3	2,5	3,6	10,4	13,0
Linje VKI						
- yrkesfagl	1,5	2,5	7,2	11,2	21,5	26,7
- allmennfagl	2,0	1,4	2,2	2,4	8,4	9,6
Røykevaner						
- røyker daglig	1,4	-	4,1	7,1	16,8	20,9
- aldri røykt	1,3	-	2,5	4,0	11,2	13,4

Fars utdanning lav=ungdomsskole, mindre enn ungdomsskole, vet ikke (N=2560). Fars utdanning høy = universitet eller høyskole (N=5160). Far uten arbeid=arbeidsløs, hjemmeværende, trygdet, død (N=1441), far heltid (N=8448). Linje videregående gjelder bare elever ved grunnkurs videregående (oftest 16–17 år), grovt sortert i yrkesfaglige linjer (N=825) versus vanlig allmennfaglig (N=2164). Eget soverom (N=9448), har ikke eget soverom (N=1757). Røyker daglig (N=899) og aldri røykt (N=7933).

Men mønsteret er velkjent. I Statistisk sentralbyrås undersøkelser viser studier av frafall jevnlig at det er sosialt skjevt – større frafall har de med lav utdanning, lav inntekt, osv. I elevundersøkelser er frafallet større på yrkesfaglige enn på allmennfaglige linjer (Hustoft & Vannebo 2000).

3.3 Kriterievaliditet

Diskusjonen om hvordan datakvaliteten i Ung i Oslo 2006 egentlig er, hemmes av at det ikke er gjort noen evaluering av svarene på individnivå i NOVAs ungdomsundersøkelser. Jeg skal nevne noen muligheter for framgangsmåter som kunne vært brukt, men ingen slike evalueringer er faktisk foretatt.

En framgangsmåte ville være å sammenligne respondentenes individuelle svar med andre kilder som vi *vet* holder høy validitet. Dette ville være å måle graden av såkalt *kriterievaliditet*, dvs. å finne et kriterium, målinger vi har all grunn til å anta er udiskutable og derfor valide, og sammenligne de faktiske observasjonene med dette kriteriet. For eksempel: Svarene om høyde og vekt kunne sammenlignes med helsesøstres målinger, svarene om karakterer kan sammenlignes med skolens protokoller, svarene om «kontakt med politiet» kan sammenlignes med politiets registreringer, osv.

Mange validitetsundersøkelser av skolebaserte ungdomsundersøkelser og andre undersøkelser av ungdomsgrupper har tatt i bruk slike metoder (for eksempel Kollisch & Oberwittler 2004; Maxfield et al. 2000; Wong et al. 2006). Ingen slike tester er foretatt når det gjelder NOVAs skolebaserte undersøkelser, så langt jeg kjenner til. Dette har selvsagt sine grunner (anonymitetshensyn, mangel på tid, mangel på penger, oppdragsgivere vil bare betale for resultater, men ikke for metodeutvikling, osv.) – men det medfører at sikre kunnskaper om validitet mangler.

En tilnærmet kriterievaliditets-test ville være å intervju, personlig og grundig, et antall respondenter som hadde (eventuelt med foreldrenes tillatelse) gått med på at intervjusvarene kan sammenlignes med hva de har krysset av i spørreskjemaene (men helst på etterhånd, så de ikke har det i tankene når de deltar i skoleundersøkelsen). En kunne også tenke seg at lærere ble intervjuet om sitt kjennskap til de enkelte elevene, men dette byr selvsagt på mange problemer, ikke minst hensynet til anonymitet – men å få til noen slike validitetstester vil nok være mulig om en prøver.

3.4 Analyser av konsistens/reliabilitet

En annen metode som ikke er avhengig av at opplysninger hentes eksternt, er å analysere i hvilken grad respondentene svarer konsistent. I metodelitteraturen ligger dette nær opp til vurderinger om svar er reliable eller ikke. I den

ferdige datafilen kan en sammenligne ett svar (eller et sett av svar) med andre svar gitt av samme elev, og tankegangen er altså at om eleven har svart på en bestemt måte et sted, bør han/hun svare på en bestemt måte på andre spørsmål.

Merk at dette kriteriet ikke forutsetter at svarene er valide – høy reliabilitet oppstår i seg selv bare på grunn av at svarene er konsistente med hverandre, men de trenger ikke nødvendigvis være sanne (morderen som benekter ugjerningen gjentatte ganger er jo konsistent, men hun svarer ikke valid).

Ung i Oslo-dataene gir noen muligheter til å analysere svarenes konsistens. Her er et eksempel: Ganske tidlig i skjemaet (s. 9, (Valset & Øia 2006)) kommer en rekke spørsmål om respondentens folkegruppe. Det første spørsmålet er: «Jeg har prøvd å finne ut mer om min folkegruppe, slik som historie, tradisjoner og skikker». Svaralternativene er «helt enig», «nokså enig», «nokså uenig» og «svært uenig». Et stykke lenger ned i spørsmålsrekka spørres det «Jeg har ikke brukt tid på å finne ut mer om min folkegruppes kultur og historie».¹

Selv om formuleringene ikke er eksakt overensstemmende, er de to spørsmålene innholdsmessig nesten identiske, men vendt ulik vei. Konsistente svar på disse spørsmålene vil altså generelt bestå i at de som svarer «helt enig» på det første svarer «helt uenig» på det andre, osv. Siden spørsmålene er vendt «ulik vei», vil ikke et slikt konsistent svarmønster bare oppstå om respondentene krysser av på alle svarene i venstre (eller høyre) rubrikker. Om svarene er konsistente, kan det tyde på at respondentene svarer bevisst, med en viss indre sammenheng (men i slike fordelinger vil en viss konsistens også alltid oppstå «spontant» på grunn av statistiske tilfeldigheter).

Disse spørsmålene kommer tidlig i skjemaet, og så mange som 95,3 prosent av hele utvalget på om lag 11.500 har besvart begge på en gyldig

¹ Variabelnavnene er ETNI21 og ETNI25. Jeg kommer ikke til å gå inn på slike detaljer her, men for dem som skal bruke datafilen: Merk at variabel-forklaringen for ETNI25 i datafilen er misvisende. Generelt er det en del slums i teksten i datafilen. Alle som bruker den bør sjekke nøye hva formuleringene *egentlig* var i selve spørreskjemaet. Et spørsmål om «antall personer» en har hatt samleie med, omskrives for eksempel i variabel-verdiene til «antall ganger» en har hatt samleie, osv. Slikt kan produsere svært misvisende feiltolkninger.

måte. En romslig definisjon av konsistens er at de som svarer «helt enig/nokså enig» på det første, må svare «helt uenig/nokså uenig» på det andre (og vice versa).

Det viser seg: 69,6 prosent svarer konsistent om vi regner på denne måten. En kunne kanskje si at om en tar i betraktning at dette er ungdommer som svarer på noe som ikke nødvendigvis har høy relevans for dem (folkegruppe), er en intern overensstemmelse på sju av ti ikke så dårlig. Men fra en mer krevende synsvinkel er ikke en slik konsistens særlig betryggende – betyr den, for eksempel, at om lag tre av ti ofte svarer på en måte som ikke har en meningsfull sammenheng med hva de egentlig mener eller hva de egentlig har opplevd?

Konsistenskriteriet kan altså brukes for å klassifisere respondentene etter i hvilken grad de har besvart spørsmål på en måte som viser en innholdsmessig sammenheng og/eller overensstemmelse.

Hvorfor oppstår inkonsistente svar? I panelundersøkelser, der samme personer spørres på flere tidspunkter, er det ofte god anledning til å analysere konsistens. Her skal jeg gi et par eksempler på hvordan inkonsistens har blitt tolket. Den amerikanske undersøkelsen jeg nevnte over, fant betydelig inkonsistens i hva unge kvinner opplyste om sin alder for første samleie på to ulike tidspunkter. Hukommelsessvikt kunne åpenbart være en grunn, men forskerne mente også at grunnen var at tidlige seksuelle erfaringer kan ha en tvetydig tolkning og at oppfatningen om hva et samleie er, kan endre seg etterhvert som respondentene blir eldre/mer erfarne (Upchurch et al. 2002). En studie av inkonsistens når det gjaldt svar om debutalder for alkoholbruk knyttet dette på samme vis til hvordan alder og erfaring kunne føre til endringer i forståelse av hva det var å drikke (Rossow 2006).

Slike eksempler viser: Om en respondent avgir inkonsistente svar, trenger dette ikke å bety at respondenten (stort sett) gir invalide svar. Selvsagt kan inkonsistens oppstå på grunn av at svar er avgitt helt tilfeldig, uten noen tanke bak – eller som tull. Men inkonsistens oppstår også på grunn av at respondenter har uklare, sammensatte og kanskje motstridende forestillinger om sine erfaringer, som kan endres over tid. Noe i livssituasjonen trigger i visse fall hukommelsen, men undertrykker den i andre omstendigheter. Respondenters vansker med å svare konsistent kan ligge i at de er usikre på

meninger og holdninger og yter sin egen vakling rettferdighet ved å stemme litt på begge.

Med andre ord: Konsistenskriteriet kan – i visse fall – brukes for å vurdere datakvalitet, men med mange forbehold, ettersom inkonsistente svar kan vise seg å være valide svar i en viktig forstand.

3.5 Sannsynlighetskriteriet

Framgangsmåten for å vurdere datakvalitet på individnivå som ofte blir anvendt blant de som er interessert i Ung i Oslo 2006-dataene, har imidlertid anvendt det vi kan kalle *sannsynlighetskriteriet*. En oppdager svar – eller svarmønstre – som er såpass usannsynlige at «de bare ikke kan stemme». Derfor må disse svarene (og kanskje alle svar fra disse respondentene?) forkastes. Forklaringen på at slikt oppstår, blir ofte sagt å være at det finnes useriøse elever som bare tuller og «kødder» – en tolker altså usannsynlig svar som utslag av atferden til tøysete respondenter.

Her er to spørsmål involvert. Det første gjelder hva som er usannsynlig, dvs. opplysninger som må forkastes fordi de åpenbart ikke er valide. Ofte brukes ordet usannsynlig uten nærmere definisjon. Menes det umulig/fysisk umulig, altså noe som ikke kan forekomme? Eller menes det et svar/en svarkombinasjon som har veldig lav sannsynlighet, men som likevel kan forekomme hos, f.eks., noen få blant de 11.500 respondentene? (Dette ble opprinnelig skrevet i oppstarten av den amerikanske presidentvalgkampen 2007/2008, og da fantes det mange meninger om det definitivt usannsynlige, f.eks. om det komplett umulige i at en mann med svart far fra Kenya, oppvokst i Indonesia, som har innrømmet å ha brukt narkotika, kunne bli presiden i USA ...).

Det andre spørsmålet er hvorfor det dukker opp ikke-valide svar. Åpenbart kan tull være en grunn, men det er jo andre grunner, som diskutert over.

Et eksempel: På spørsmålet om kroppshøyde har 44 avgitt svar som kunne tyde på at de er under 120 cm høye. I et materiale av denne størrelsesordenen er det ikke umulig at det er én, eller kanskje et par–tre, kortvokste. Imidlertid er andelen blant kortvokste som er kortere enn

120 cm, svært liten. Med andre ord: Svarene som er registrert for disse 44 respondentene er ikke valide.

Men hvorfor har slike ikke-valide svar blitt registrert? Det er trolig få som ikke forstår spørsmålet, men det kommer på et tidspunkt da mange er trøtte, med lav konsentrasjon, lav oppmerksomhet – de har allerede forholdt seg til over 350 spørsmål. Noen kan skrive feil i farten (f.eks., de ville skrive 170, men det ble bare 70). Noen har kanskje skrevet riktig, men så utydelig at den optiske leseren leser feil. Og så er det de som bevisst har skrevet tøys – «jeg er 2 cm høy, hø hø». Når en ser på hvilke svar som konkret er gitt, virker det sannsynlig at ihvertfall en del har tøysa.

Et annet eksempel: Det er sagt at noen av de som sier de er adoptert, har opplyst om et opphavsland «som Norge ikke har adopsjonsavtale med» – ergo ikke-valide tøysesvar. Denne tolkingen kan jo være riktig, men hva om barnet ble adoptert før familien kom til Norge? Dessuten: I lys av det som er sagt tidligere, er det sannsynlig at en del ikke forstår eller ikke legger merke til ordet «adoptert» og svarer på hvilket land de kommer fra, ikke på hvilket land de var adoptert fra – dette er jo ofte fremmedspråklige elever som kan ha vansker med hva ordene betyr.

Et hyppig brukt argument har vært å hevde at en kan finne fram til respondenter som svarer usannsynlig høyt på mange typer atferd, samtidig. Det har blitt argumentert slik: Det er komplett usannsynlig at en og samme tenåring kan ha vært med på mange ulike slag kriminell aktivitet (både stjålet penger, gjort innbrudd, stjålet bil/motorsykkkel, etc.) mer enn 50 ganger av hver type i løpet av siste året. De som svarer dette, «kan ikke svare sant». En annen argumentasjon av samme type tar utgangspunkt i svarene om fritidsaktiviteter (se avsnitt 2.3). Argumentet er at de som sier de har gjort *alle* de 22 aktivitetene et stort antall ganger ikke kan svare sant.

Åpenbart er det aktivitetsmønsteret som slike respondenter opplyser om, bortimot fysisk umulig, om en tar spørsmålstekst og svar bokstavelig. Derfor er det god grunn til å være meget forsiktig med å bruke svarene på slike spørsmål som valid informasjon om hva det – bokstavelig talt – spørres om.

Men det er helt sikkert at mange urimelige svar – i hvert fall på fritidsaktivitetsspørsmålene – skyldes at respondentene ikke følger tidsanvisningene (se avsnitt 2.3). Respondentene har svart så godt de kunne, ut fra sin private

forståelse av spørsmålet. Svarene er altså ikke valide (i lys av spørsmålets bokstav), men svarene er ikke avgitt som bevisst feilinformasjon.

At et svarmønster framtrer som usannsynlig, vil altså kunne ha flere fortolkninger. Av og til er det usannsynlige det ekstremt uvanlige, og en bør være forsiktig med å avfeie opplysninger som kunne opptre hos en eller to helt spesielle respondenter. Videre kan en ikke sette likhetstegn mellom usannsynlig svar og bevisst tøysemaker. Som vist over er det mange kilder til usannsynlige og umulige svar – at spørsmålsteksten ikke blir forstått adekvat, hukommelsessvikt, svareren krysser av tilfeldig eller ukonsentrert for liksom å ha besvart prøven, osv. Men selvsagt: En mulighet er at det *bevisst* er gitt feil informasjon.

3.6 Oppsummering

Spørsmålet om datakvalitet – alt i alt – i Ung i Oslo 2006 og liknende undersøkelser, er altså ikke lett å besvare fullt ut, først og fremst fordi det mangler metodeundersøkelser, for eksempel kriterievaliditets-tester eller liknende typer undersøkelser. Slike er jo vanskelige å få til, men har det vært gjort noen forsøk?

Mye drar i retning av rimelig bra validitet for mange deler av materialet. En kan argumentere at for mange spørsmål – om de er lette å forstå, om de refererer til noe det er rimelig å forvente at elevene vet eller kjenner til, om spørsmålene er presist og forståelig formulert, om de refererer til relevante sider av elevenes livsverden, og om de knyttes til instruksjoner som lett blir misforstått – vil det være tilstrekkelig god validitet på individnivå for de fleste forskningsformål. Men svarene må tolkes fornuftig, og vi trenger dessuten bedre kunnskaper om hvordan spørsmål osv. faktisk fungerer, og til det trengs nok mer metodeundersøkelser og litt mindre synsing.

4 Om bevisst feilinformasjon i Ung i Oslo 2006

Til nå har jeg diskutert hva jeg mener er det viktigste problemet med datakvaliteten i Ung i Oslo 2006 – nemlig problemene med hvordan respondenter, undersøkelsesopplegg og spørreskjema har vært tilpasset hverandre. Her er det feilkilder, men mange av feilkildene ville føre til tilfeldige feil, og tross feilkilder kan det være at gjennomsnitt og univariate fordelinger på enkeltvariable i mange tilfelle blir noenlunde riktige, og om ikke validiteten er for lav, kan også statistiske mønstre, assosiasjoner og sammenhenger, bli noenlunde riktige.

En spesiell kilde til svak validitet er imidlertid bevisste feilsvar. Her er det altså ikke snakk om feilkryssninger i farten, misforståelser, avkryssninger i hytt og vær pga. manglende hukommelse og tidspress, osv. – men at elever svarer noe som de *vet* ikke er riktig.

4.1 Underrapportering av det sensitive, overrapportering av det sosialt verdsatte

Den mest typiske formen for slik bevisst feilinformasjon er at såkalte sensitive opplysninger ikke blir gitt. Det skjer en underrapportering av atferd og holdninger som på en eller annen måte kan framtre som stigmatiserende, pinlig, skamfylt eller ubehagelig. Typiske eksempler er stigmatisert seksualitet, lovbrudd, kriminalitet og annet som blir fordømt og sett ned på av omgivelsene.

Den viktigste begrunnelsen for selvadministrerte spørreskjema – i tillegg til kostnadsspørsmålet – er at den ekstra anonymiteten gjør det lettere for respondentene å fortelle om det han/hun ville kvie seg for å si rett ut til en intervjuer. Men på tross av anonymiteten i spørreskjemaundersøkelser, blir pinlig og stigmatiserende informasjon ofte ikke gitt, og metodeforskerne har spekulert mye på hvordan spørsmål skal formuleres slik at respondentenes motvilje og angst mot å rapportere slikt blir dempet.

En annen typisk form for bevisst feilinformasjon er at prestisjefylte eller sosialt verdsatte handlinger og egenskaper blir overrapportert. Således viser valgundersøkelser ofte at personer sier de deltok, selv om registerinformasjon viser at de ikke stemte. Noen undersøkelser viser en viss tendens til at menn overrapporterer sin kroppshøyde, riktignok ikke mye (blant kvinner/jenter er det heller en tendens til å underrapportere vekt). Et eksempel på overrapportering av sosialt ønskelig atferd ble funnet i en skolebasert undersøkelse der elever ved «high schools» hadde en tendens til å overrapportere sin treningsaktivitet, målt med funn fra medisinske tester (Wong et al. 2006).

Hvorfor skjer bevisst feilrapportering? Er den egentlig bevisst? Det kan jo være at respondenter også lurer seg sjøl, at de har illusjoner om seg selv, overdimensjonerer sine gode og underestimerer sine dårlige sider med subjektiv ærlighet. Røykere vet noen ganger ikke selv hvor mye de faktisk røyker, og de luller seg inn i forestillingen om at det er jo ikke «så mye». Leger har en tendens til å overdrive hvor mye de jobber uten at de er helt klar over det selv (Rutle 1981). Men det virker likevel som en hovedkonklusjon fra metodeforskningen er at slik feilinformasjon svært ofte er subjektivt bevisst, det er ikke et genuint selvbedrag, men en villet feilinformasjon (Tourangeau & Yan 2007).

4.2 Generelt om tullede feilinformasjon i Ung i Oslo-materialet

Å anslå forekomsten av bevisst feilinformasjon i Ung i Oslo-materialet i sin mest typiske form, nemlig underrapportering av ikke sosialt aktverdig og stigmatiserende atferd, er selvsagt vanskelig. At kriminell atferd, rusmiddelbruk, problematisk seksualitet som homofili, osv. ikke blir registrert i sannferdig omfang i datafilen, virker sannsynlig i lys av det vi kjenner til av metodeundersøkelser, men hvor mye dette gjelder for Ung i Oslo 2006 er vanskelig å besvare uten nærmere undersøkelser.

Jeg skal la spørsmålet om slik underrapportering ligge, og konsentrere meg om den type bevisst feilinformasjon som går under betegnelsen tull, tøys eller «kødding».

«Gutter er gutter» osv. Det er neppe tvilsomt at noen elever synes det er spennende, pirrende, gøyalt osv. å finne på noe for å sjokkere, eller for å ha noe å «briefe med» i skolegården etterpå. Hvorfor gjør de det? Jeg tror at en må forstå slik atferd som en egen sorts feilinformasjon som drives fram av spesielle, kanskje lokale, ungdomsnormer for hva som blir verdsatt. Å nekte å underlegge seg de strenge rammene som spørreskjemaundersøkelsen legger opp til, å leke med det sjokkerende eller spennende, groteske eller absurde, gøyale eller humoristiske, å fikle med identiteter (jfr. erfaringen mange kan ha med chatte-sider), å protestere i ergrelse mot at opplegget fortøner seg som en for vanskelig skoleprøve, osv. – slikt kan være grunner til å rapportere noe som de godt vet er usant. Å lure forskere, føre dem bak lyset, tøyse med deres gørr kjedelige gravalvor, kan jo gi et kick.

De kan altså si «jeg er mongoloid» eller faren min er månefarer eller «har pult 500 damer», i trygg forvisning om at anonymiteten sikrer dem mot personlige konsekvenser. På den annen side kan de dermed framstå som uredde, uvørne og villstyrige, som våger det andre ikke tør, som fantasifulle kjekkaser som underholder og gjør det litt gøyalt for venner og skolekamerater.

Hvilke temaer er det sannsynlig at feilinformasjonen vil bli gitt om? Sex, vold, kriminalitet og dop er sikre kandidater – kanskje også kroppslige lyter. Vil de feilinformere om boforhold, om ferier, om psykiske problemer, om hvor ofte de bruker mobiltelefon, om utdanningsplaner? Noen kan sikkert tenkes å gjøre det også. Antakeligvis er det noen som kan finne på å rapportere noe rart, merkelig, gøyalt eller vanvittig overdrevet om nesten hva som helst. Men det er vel rimelig å forvente at det nettopp er temaer som sex, vold, krim og dop det ofte vil dreie seg om.

Et problem i forsøkene på avdekke slik feilinformasjon, er igjen *mangelen* på metodestudier. Hvorfor blir ikke slike metodestudier gjort? Dette skyldes nok at slik feilinformasjon har vært betraktet som noe uønsket som skal «renskes ut» og «fjernes» fra datamaterialet. Det viktige forskningsformålet er jo å lage *valid informasjon* til departementer og sosialpolitikere osv. om virkeligheten når det gjelder ungdommens problemer og atferd. I dette perspektivet er jo ungdommer som sier de har vært 20 ganger eller mer i fritidsklubb siste 7 dager – dette gjelder 117 respondenter, ca. 1 prosent av

hele datamaterialet – ikke særlig til hjelp, selvsagt. Grunnen til slik ikke-valid informasjon ble diskutert over og er ikke tulle svar – for det kan vel ikke gi noe særlig status i venneflokken å skryte av at «jeg svarte kjempehøyt på antall ganger jeg hadde vært i fritidsklubb»?

Bevisste tulle svar (og ikke bare misforståtte spørsmålsformuleringer) kunne for øvrig gjerne vært tatt seriøst, som uttrykk for stemninger, atferd og kultur blant ungdom. Slik atferd burde vært et eget forskningsspørsmål, men har i stedet blitt betraktet som noe som må vaskes vekk, som utøy og skitt. Tøyserne er liksom blitt en slags fiender vi skal nedkjempe, i stedet for interessante – og valide – vitnemål om hva som foregår inne i tenåringshodene.

4.3 Et forsøk på å flagge tøyserne – hvordan kan vi finne deres signatur?

Blant de om lag 11.500 respondentene i Ung i Oslo 2006 er det helt sikkert at det er et visst antall som bevisst har levert feil informasjon av tøysekarakter. Lar det seg gjøre å finne ut hvem som har gjort det? Utvilsomt er det vanskelig, men her skal jeg argumentere for en mulig framgangsmåte.

Å gi tøysesvar er en måte som ungdommer markerer seg på. Da må poenget med denne atferden være at den blir lagt merke til. Om den ikke blir lagt merke til, er det vel ingen vits? Naturligvis blir den lagt merke til når tenåringen med brautende stolthet doserer for måpende medelever om sine vanvittige avkryssninger. Tøysesvarene er et medium for å bli sentrum i flokken eller på festen.

Men det kan være sannsynlig at tøysesvarerne også ønsker å bli lagt merke til av forskerne. De vil kommunisere til forskerne at «dette gidder jeg ikke ta alvorlig», dette er for kjedelig, for tåpelig, dette «gir jeg ikke fem flate for». De kan kanskje gjøre det helt fra starten av besvarelsen. Men drivkraften kan også være trøtthet, at de føler seg overmannet av spørreskjemaets krav. Vi kan tenke oss en type karriere – en elev kan starte noenlunde seriøst og besvare skikkelig, men det blir for mye etterhvert, de begynner å tenke på å gi opp, de begynner å skrive inn ting bare for å fylle rubrikkene, bryr seg ikke som hva de svarer på, og deretter slutter de helt. Dette er altså en karriere fra

ordentlighet (rimelig høy validitet) til økende grad av villedende informasjon, for deretter å droppe det hele.

Om de ønsker å bli lagt merke til av forskerne, vil de legge igjen en signatur, omtrent som taggerne gjør. De vil gi en bevisst feilinformasjon som er slik at forskerne ikke kan la være å legge merke til den, om de undersøker nøye. Før den optiske lesingen av spørreskjemaene gjorde sitt inntog, den gangen hvert skjema ble gjennomgått individuelt og visuelt, ble slike tøysesvar plukket ut, på grunn av kluss, svarmønstre som dannet finurlige geometriske mønstre, tegninger av nakne damer i marginen, skjellsord rettet mot dusteforskere, osv. Men det har vist seg at et ytterst lite antall skjemaer ble forkastet av slike grunner – jf. at bare 25 av over 11.000 skjemaer ble tatt vekk fra Ung i Oslo 1996-materialet på denne måten (Bakken 1998).

Hvor kan tøysinformasjon vise seg, etter at skjemaene er optisk avlest og inkludert i datafilen? Den vil i mindre grad kunne vise seg der svaralternativene er bundet, altså der de må velge mellom et lite antall rubrikker som alle har en viss troverdighet. Den kan også vise seg her, riktignok, om et av de prekodete svaralternativene oppfattes som pirrende å velge.

Tøysinformasjon kan nok i større grad vise seg der respondentene har frihet på forholdsvis åpne spørsmål der de kan skrive inn ett eller annet uten å være bundet av prekodete alternativer. Her kan de jo utbre seg på mer villstyrkelige måter, hinsides enhver realisme.

Hva vil de tøysse med? Det virker umiddelbart sannsynlig at de særlig vil ta i bruk muligheter til å fortelle noe grotesk og absurd, sjokkerende eller bizart, og at de særlig vil søke å plassere signaturen sin på tabloide temaer som har å gjøre med sex, dop, krim, vold. Banale temaer som å gå på kafé, gå i fritidsklubb, hjelpe til hjemme, er ikke så sannsynlige kandidater. Noen har oppdaget enkelte usannsynlig høye tall for antall leste bøker siste måned og har foreslått at dette er en signatur, et vitnesbyrd om bevisst «kødding», men slikt er vel ikke særlig pirrende? Siden vi vet at respondentene ofte vil overse tidsangivelser, er det nok mer sannsynlig at slike fantastiske boklesere framtrer fordi de trodde de skulle svare på spørsmål om «antall leste bøker i alt», ikke bare for siste måned. Antall leste bøker siste måned er derfor heller ikke noen god indikator på bevisst feilinformasjon.

Her kommer forslag om hvordan en muligens kan avdekke tøysernes signatur på sju måter.

4.3.1 De åpne spørsmålene om fars/mors yrke

Spørreskjemaet har en del åpne spørsmål, og her kan elevene skrive hva de vil. Dessverre foreligger ofte ikke de originale svarene om for eksempel religionstilhørighet, men for fars og mors yrke, med tilhørende tilleggsinformasjon om hva de gjør, er respondentenes originale utsagn blitt scannet inn og finnes i datafilen. Jeg har gått igjennom 8920 svar der eleven har skrevet inn ett eller annet for fars yrke og 8657 svar for mors yrke. Opplysningene er lest samlet for far, mor, og hva foreldrene gjør i sitt yrke, for hver enkelt respondent.

Ikke sjelden er det ikke skrevet yrkesopplysninger i disse rubrikkene, det kan tvert om stå «død», «arbeidsløs», «hjemmeværende», «går på norskkurs», «vet ikke», og av og til kommer noe som henviser til arbeid uten at yrket kan identifiseres («noe på kontor», «IF»), men dette er jo ikke tøysesvar av den typen vi er ute etter. Men for 81 respondenter mener jeg det er god grunn til å anta at her er det et bevisst tøysesvar. Noen er nok helt definitive (far er «rompesleiker» og mor leker «gjemsel», far er «horekunde», «kriger» eller «leiemorder», far er «pimp» mens mor er «lol»(?), far er «modell» mens mor er «justisminister», far «temmer tigrer og bjørner» mens mor er «glamourmodell»).

Noen av svarene jeg har flagget som tøys kan imidlertid være riktige. Er alle de 6–7 svarene om at far er hasjdealer eller hallik usanne – hvor mange slike finnes faktisk som fedre til en elevmasse på over 10.000? Fire–fem har mødre som er «prostituert» eller «hore» eller «ludder» – sant eller usant? Det er altså tvilstilfeller. Kanskje noen av de 81 faktisk ikke tøysset. På en annen side kan svar som egentlig var markører på tøys være oversett. Tre–fire har skrevet kort og greit «narkoman» på far og et relativt vanlig yrke på mor (hudpleier, osv.), og de har jeg godtatt, for det står ikke noe annet rart i forklaringen, og blant respondentene må det jo være noen som faktisk har en narkoman far.

Altså: I datafilen er 81 respondenter blitt flagget på grunn av antatte tøysesvar om fars/mors yrke.

4.3.2 Umulige aldersangivelser

I spørreskjemaet skal respondentene oppgi sin alder da noe skjedde, og jeg vil trekke fram fire spørsmål: Alder da foreldrene ble skilt, alder da en kom til Norge, alder da en første gang drakk alkohol/ble beruset (hva de svarer på her er uklart pga. dårlige spørsmål), og alder ved første samleie – alder for eget første samleie kan jo også handle om å si noe sjokkerende om sex.

Et problem her er at svarene kan jo være tilfeldig feilskrift eller feil i den optiske lesingen. Dessverre ble elevens *egen* alder spurt om med prekodete alternativer – om dette var et åpent spørsmål ville vi kanskje hatt lettere for å finne tøysere?

Hvorfor en skulle tøys med alder om når foreldre var skilt eller når en kom til Norge, er jo uklart, men uansett – urimelige høye aldre (22+) er flagget som mulige tøysesvar (de er svært få).

Alder første gang en drakk alkohol: De som oppgir for høy alder (22 år +) må vel være ikke-valid, siden det knapt skal være så gamle med i undersøkelsen? Men er det tøys fra de 40 som oppgir alder 0–2 år som alder første gang de drakk alkohol? Mange unger har historier om at de drakk et glass øl da de var bitte små....

Alder ved første samleie er også problematisk. Igjen kan vi si at de med høy alder har ikke valide svar, uten at vi kan være sikre på at det er bevisst tøys, men hva med de 40 som oppgir 0–4 år? Når en ser nøye på svarene, tyder de på at en del av dem som har svart 0 på alder for første samleie, har misforstått og satt 0 for antall samleier – for disse betyr altså 0 egentlig «ikke hatt samleie».

Hva med de som oppgir særs lav alder, spesielt de som sier 0–3 år for «første samleie»? Noen kan ha vært utsatt for overgrep som småbarn, noen har «lekt doktor» så gjennomført at de synes de kan si det var et samleie. Jeg har valgt: Her må opplysningen vært gitt bevisst, i den forstand at de både har krysset av «ja» på «har hatt samleie» og opplyst om en alder. Jeg har valgt å flagge som svært tvilsomme, altså sannsynlig tøysere, de 18 som «har hatt samleie» første gang når de var 0–3 år.

Om vi teller sammen alle som har gitt noe som kan regnes som urimelige aldersangivelser blir dette 43 i alt. På grunn av muligheten for feil-avkryssninger har kanskje flere av disse 43 ikke tøysset bevisst.

4.3.3 Kroppslige forhold – tøysesvar om funksjonshemninger?

Noen synes lytehumor er kul. Siden de åpne svarene om konkret type helseproblem ikke er scannet inn i datafilen, kan en ikke spore de som f.eks. har sagt «oppstoppernese» eller «hjernedød» eller «syff» eller «angst for å spise gelé» eller annet gøyalt på det åpne spørsmålet. (På samme måte er de åpne svarene om «min folkegruppe» dessverre ikke tilgjengelig, her kunne jo tøysere blitt oppdaget blant eventuelle «hobbitter» og «apacheindianere».)

Noen har lagt merke til uventede svarmønstre blant de som er blitt klassifisert som «multifunksjonshemmet» og har foreslått at de multifunksjonshemmete kan «ha bølla» – det kunne jo være en variant av lytehumoren å kalle seg selv for multifunksjonshemmet. Men en bør være klar over at de 77 elevene som i datafilen framstår med verdien «multifunksjonshemmet» har fått denne verdien fordi de har krysset av minst to ganger på rubrikkene for bevegelseshemmet, synshemmet og hørselshemmet. Hvor mange av dem som bare har krysset av to ganger, f.eks. syns- og hørselshemmet, vet vi ikke (dette er jo ikke en usannsynlig kombinasjon). Det er altså viktig å være klar over at verdien «multifunksjonshemmet» slik den framtrer i datafilen, er konstruert av forskerne og ikke av elevene selv – elever som sier de er både synssvake og bevegelseshemmet har forskerne kalt for «multifunksjonshemmet», men elevene selv har ikke klassifisert seg slik.

Det er derfor vanskelig å avsløre noen som tøys i sjangeren lytehumor, men la oss likevel åpne for at noen av de som har krysset av minst to ganger på de tre rubrikkene for syns-, hørsels- og bevegelsesproblemer (og som forskerne dermed har kalt «multifunksjonshemmet») faktisk har tøysa. Her kan vi prøve et konsistenskriterium. Vi kan anta (men det er mye usikkert her) at om de har minst to av tre av disse hemningene, er dette en så dominerende del av deres selvbylde at de burde ha svart direkte og klart «funksjonshemmet» på spørsmålet stilt på forrige side (dvs. s. 17). Dette resonnementet er ikke vanntett pga. uklare ideer om «funksjonshemning» blant tenåringene. Men om vi bruker det, framstår 26 av de 77 forskerdefinerte «multifunksjonshemmete» som mulige tullevarere, og disse 26 har jeg altså flagget som mulige tøysere.

4.3.4 Giganter og dverger, puslinger og fettberg

En variant av spøk med bizarre kroppsformer kan oppstå når elevene skal opplyse om sin høyde og sin vekt – de kan være titaner på over 3 meter eller latterlige små, de kan veie et halvt tonn, osv.

Imidlertid er det jo også muligheter for at svar som åpenbart ikke er valide, kan ha oppstått på grunn av feilskrivning og feilaktig optisk lesning. De som tøyses vil nok ta noe i – ikke bare i grenseland for det mulige, men voldsomt inn i det absurde. Jeg har valgt en enkel framgangsmåte, som skiller ut ikke valide svar, men vi har egentlig små muligheter for å avgjøre om det er tøysesvar eller feilskrivninger. Jeg antar at de som skriver at de er under 120 cm eller over 215 cm, eller veier mindre enn 26 kg eller minst 190 kg, kan ha tøysa. Så mange som 87 ble flagget som mulig tøys her, men som sagt er det ikke lett å vite hvem av disse som er egentlige tullere og hvem som mer tilfeldig har fått registrert ikke valid informasjon.

4.3.5 Utrolige sex-atleter

Å tøyses med seksuell legning og samleier kan helt sikkert være en måte å legge igjen en signatur på. En kan sjokkere med å være «soper» eller «bifil» – hadde spørsmålet vært åpent, ville vi nok registrert en del andre varianter. Opplysningene om seksuell legning i seg selv er ikke bemerkelsesverdig – faktisk er antallet som har rapportert at de er homofil, lesbisk eller bifil i den størrelsesorden vi ville forvente.

Men alle spørres om antall seksualpartnere, og her kunne fiktive seksualatleter vise seg. Et problem er riktignok at pga. alt vi vet om hvordan spørsmål misforstås, kan det godt hende at en del som prøver å svare ærlig, sier et tall som er ment å dekke antall samleier og ikke antallet seksualpartnere. Her kan vi bare gjette – noen har kanskje flesket til for å ha noe å skryte av, men at en håndfull faktisk har hatt svært mange seksualpartnere er heller ikke utenkelig. For enkelhets skyld antar vi at de som sier «50 eller mer» på antall partnere (som også kan inneholde noen som mener de svarer om antall samleier) neppe kan være seriøse – dette er 28 av respondentene. Trolig har vi da også flagget en håndfull som ikke skulle være klassifisert som tøysere.

4.3.6 Dop-spørsmålene

Det blir spurt om «hvor mange ganger siste 12 måneder» en har brukt 11 typer rusmidler og kraftige narkotiske stoffer som LSD, kokain, heroin, morfin, ecstasy, «piller» og «andre stoffer» («piller» og «andre stoffer» kan være hva som helst), her spørres også om lett eksotiske stoffer som fleinsopp og khat. Vi bør ikke legge vekt på tidsangivelsen «siste 12 måneder», ettersom slike betingelser *alltid* blir oversett, men det spiller neppe noen stor rolle her. Dette vil opplagt være et felt som tøysere kan boltre seg i, men at noen elever har vært mye borti dop er jo også sannsynlig. Noen usannsynlig høye tall for dopbruk kan være gitt uten at det egentlig var for å tøyse, men for å si fra at disse elevene faktisk er «in deep shit»? I den redigerte filen med 11 500 respondenter er alle disse svarene omkodet slik at den mest ekstreme svarvarianten er «50 ganger eller mer». I den foreløpige arbeidsfilen som jeg også har hatt tilgang til, framgår hva respondentene egentlig svarte. Åpenbart er det merkelige svar her, som når en respondent svarer f.eks. brukt amfetamin 666 ganger. Noen elever har svart «365» på flere eller nesten alle stoffene – om disse kan vi i det minste si at de har oppfattet at svaret skulle avgis for de siste 12 månedene ...

Det er opplagt bevisste tøysesvar i denne svarmassen, men trolig oppblandet med noen som faktisk er i et dophelvete (men ville de som *virkelig* er på kjøret, greie å delta i undersøkelsen?). Jeg foreslår følgende: Tøysesvar på disse spørsmålene antas særlig å finnes blant dem som har svart mer enn 100 (101+) på minst to av følgende stoffer: LSD, kokain, amfetamin, opium, morfin, heroin, fleinsopp og khat. Disse 29 respondentene har jeg flagget på grunnlag av de opprinnelige svarene respondentene har avgitt og som er registrert i arbeidsfilen.

4.3.7 Enorm kriminalitet

Å sjokkere troskyldige forskere med uhyrlige mengder kriminalitet kan også være en karakteristisk tøyse-mulighet. Elevene er spurt om 23 typer regelbrudd hvor de har skrevet inn et eksakt antall ganger disse typene regelbrudd skal ha foregått «siste år». En del av regelbruddene kan opplagt ha forekommet ofte, som «voldsom krangel med lærer», nasket, «klort eller lugget», sneket på trikken, vært beruset, kanskje sniffet, osv.

I den redigerte filen med om lag 11.500 respondenter er allerede noen som rapporterte mengder av kriminalitet blitt vasket vekk, men vi kan også se på de gjenværende. Jeg foreslår at metoden for å flagge mulige tøysere er å lokalisere dem som har – i sine originale besvarelser – oppgitt skremmende høye tall for mer tung kriminalitet. Jeg fokuserer på følgende fem spørsmål: stjålet noe verdt «mer enn tusen kroner», stjålet penger fra en venn (selv hardbarka skurker gjør vel ikke det ofte?), gjort innbrudd for å stjele, stjålet bil/motorsykkel, vært i kontakt med politiet for å ha gjort noe galt. Det er knapt mulig å gjøre annet enn et vilkårlig valg her, og jeg har flagget de som har rapportert å ha gjort minst *to* av disse tingene *20 ganger eller mer* – i alt 36 respondenter.

4.4 Er målingen av tøyserne valid?

Slik har jeg altså flagget mulige tøysesvar på sju felt – informasjonen om fars/mors yrke, om aldersopplysninger, om mulig tøys om kroppslige lyter/funksjonshemming, om høyde/vekt, om antall sexpartnere, om tung dop og om urimelig kriminalitet. På samme måte som ellers, kan vi prøve å evaluere validiteten av denne klassifiseringen: Har vi funnet de virkelige, bevisste, tøyserne?

At tøysesvar skulle være en guttegreie, virker rimelig. Tabell 3 viser en *påtakelig* overvekt av gutter på seks av de sju tøyse-indikatorene. Når det gjelder svarene om mulige tøysesvar blant de som forskerne kalte «multi-funksjonshemmet» er imidlertid flertallet jenter. Dette kan tyde på at *denne* indikatoren for tøysesvar ikke er god. Vi velger likevel å beholde den.

Tabell 3. Gutteandel (%) på de sju flaggingene av mulige tøysesvar.

Fars/mors yrke	Aldersopplysninger	Multi-funksjonshemmet	Høyde/vekt	Sexpartnere	Tung dopbruk	Uhyrlig kriminalitet
75	74	44	74	93	78	89
(N=81)	(N=43)	(N=26)	(N=87)	(N=28)	(N=29)	(N=36)

N=det faktiske tallet av flaggete tøysere, mens prosentandelen gutter er beregnet fra dem av de mulige tøyserne som har svart på spørsmålet om kjønn (noen få mangler svar her).

Om vi summerer tøyserne, viser det seg at til sammen 259 blir flagget på grunn av mulige tøysesvar, men bare 53 av dem ble registrert som tøysere på

to eller flere av de sju kriteriene. Disse 53 har utvilsomt et svarmønster som tyder på gjentatt tøysing. Men det er altså 206 som bare ble flagget på én av de sju indikatorene. Dette kan tyde på at begrepet «konsekvente tøysere» ikke har så stor gyldighet. Noen tøysere ett sted, noen andre steder, og så lar de det kanskje bli med det?

En kan også undersøke: Fullfører de mulige tøyserne hele undersøkelsen? Om de er med helt til slutt, kan tyde på en større grad av motivasjon enn om de har droppet ut tidligere. I tabell 4 er dette blitt forsøkt målt med andelen av mulige tøysere som ikke har besvart et lett spørsmål helt på tampen av spørreskjemaet (om en er/har vært med i forening, klubb, lag, el. liknende). De mulige tøyserne har alle lavere svarprosent på slutten enn gjennomsnittlig i utvalget, men forskjellen er ikke stor for f.eks. tullevarere på fars/mors yrke- og høyde/vekt-spørsmålene.

Tabell 4. Andel (%) av de mulige tøyserne som *ikke* svarer på det siste lette spørsmålet i undersøkelsen om medlemskap i forening/klubb/lag. (Tilsvarende andel som *ikke* har svart på dette spørsmålet for hele utvalget på 11.500 er 16 prosent.)

Fars/mors yrke	Alders-opplysninger	Multi-funksjons-hemmet	Høyde/vekt	Sex-partnere	Tung dopbruk	Uhyrlig kriminalitet
21 (N=81)	42 (N=43)	35 (N=26)	20 (N=87)	29 (N=28)	38 (N=29)	31 (N=36)

N=det faktiske tallet av flaggete tøysere, mens prosentandelen gutter er beregnet fra dem av de mulige tøyserne som har svart på spørsmålet om kjønn (noen få mangler svar her).

Vi kan endelig bruke kriteriet om en viss konsistens mellom to svar avgitt om et folkegruppe-spørsmål. Konsistens i svarene her (avsnitt 3.4) for hele utvalget (dvs. for de som hadde besvart begge disse spørsmålene) ble anslått til å være nær 70 prosent.

Tabell 5. Andel (%) av de mulige tøyserne som *ikke* svarer relativt konsistent på to spørsmål om folkegruppe (se avsnitt 3.3). (Tilsvarende andel som svarte konsistent i hele utvalget på 11 500 var 69,6 prosent.) (Avgitte svar i parentes.)

Fars/mors yrke	Alders-opplysninger	Multi-funksjons-hemmet	Høyde/vekt	Sex-partnere	Tung dopbruk	Uhyrlig kriminalitet
51 (N=74)	51 (N=35)	58 (N=24)	49 (N=83)	52 (N=27)	40 (N=25)	58 (N=33)

Tabell 5 viser at de mulige tøysesvarerne er mindre konsistente enn gjennomsnittet, men forskjellen er faktisk ikke veldig stor (40–58 prosent konsistens blant mulige tøysere, mot 70 prosent blant alle).

En konklusjon kan være: Denne flaggingen av mistenkte tøysesvarere tyder på at de finnes, at de ikke er helt ubetydelige i antall (259 = 2,2 prosent av hele utvalget), at de stort sett er gutter, at de i mindre grad fullfører undersøkelsen, og at de svarer mindre konsistent. Men fra en annen synsvinkel kan en også si at om disse ca. 260 faktisk er elever som ofte feilinformerer bevisst, er dette ikke et alarmerende høyt antall. Mange datainnsamlinger, også blant seriøse respondenter, må nok regne med at svarenes validitet lett kan være sviktende for et par–tre prosent av svarutvalget.

Det er dessuten viktig å være klar over at denne klassifiseringen er svært usikker. Det er langt fra sikkert at de tøysesvarerne jeg har flagget er konsekvente tøysesvarere – noen kan svare bevisst gøyalt på noen steder i skjemaet, men de kan likevel være noenlunde seriøse andre steder.

5 Homo-tøysere?

5.1 Tøyse med homofili, lesbiske, bifile?

Debatten ved NOVA om datakvaliteten i Ung i Oslo 2006 har også beskjeftiget seg med hvordan denne datainnsamlingen bidrar til et bilde av vansker i homofile og lesbiske elevers livssituasjon. Et argument har vært at så mange har tøyset med opplysninger om seksuell legning og relaterte spørsmål at en ukritisk bruk av dette datamaterialet ikke vil gi et realistisk bilde av unge homofile, lesbiske og bifiles situasjon. På den annen side har forskere brukt materialet uten betydelig forbehold. I NOVA Rapport 19/07 antas det at temaer som «mobbing, utsatthet for vold, bruk av narkotiske stoffer og det å ha en lesbisk og homofil orientering» *ikke* «inviterer til overrapportering». Dette er temaer som kan være knyttet til skam og stigma, og «Norske tenåringsgutter og tenåringsjenter skryter ikke på seg eller tuller med homoseksuell i en anonym spørreundersøkelse» (Moseng 2007).

Denne holdningen har stort sett vært den framherskende i NOVAs ungdomsforskermiljø – nemlig at sex-relaterte forhold ikke er særlig problematisk å kartlegge via NOVAs spørreskjemaundersøkelser blant ungdom. NOVA-forskere har derfor presentert analyser av homofili, å selge sex, seksuell debutalder, samband mellom seksuell aktivitet og rusbruk, mv. (Hegna 2007; Pedersen & Hegna 2003; Pedersen et al. 2003) på basis av datamateriale som er samlet inn på om lag samme vis som dataene i Ung i Oslo 2006.

Men det er neppe riktig å anta at å tulle med homoseksuell er noe tabu for de som har lyst til å tøyse. Nettopp det pirrende, sjokkerende og tabubelagte med temaet, og homofiles *stigmatiserte* situasjon, kan invitere til tøyasing. I forrige del (del 4) lokaliserte jeg 259 respondenter som vi her vil kalle (mulige) tøysere. Om det å tulle med homoseksuell var noe som faktisk foregår, ville vi tro at blant disse 259 vil det være forholdsvis mange som kalte seg for homofile, lesbiske eller bifile.

I tabell 6 vises antall som har kalt seg selv homofil, lesbisk eller bifil – i øverste linje for hele utvalget (N=11.500) og i den neste linjen for utvalget *etter* at de 259 tøyserne er tatt ut (da blir altså totalutvalget på 11.241).

Tabell 6. Antall som har oppgitt at de er homofile/bifile gutter og lesbiske/bifile jenter i Ung i Oslo, om en har med og om en tar ut de mulige tøysesvarerne.

Tøysesvar inkludert eller ekskludert	Heterofile jenter og gutter	Homofile/bifile gutter	Lesbiske/bifile jenter	Annet (usikre, vet ikke, ikke svart, mv)	Antall i alt
Inkludert	8887	187	167	2259	11 500
Ekskludert	8757	136	154	2194	11 241

Tabell 6 tyder klart på at det er noen som tuller med sin seksuelle legning. Dette gjelder særlig gutter. Femten av de 187 guttene som i totalutvalget har kalt seg homofil eller bifil, var blant de 259 tøysene. Dette utgjør en relativt stor andel av den opprinnelige gruppen av homofile/bifile gutter (27 prosent). Blant de 167 jentene i det opprinnelige utvalget som sa om seg selv at de var lesbiske/bifile, er det imidlertid bare 13 som var blant de 259 mulige tøysene.

Betyr dette at en *fjerdedel* av de opprinnelige homofile/bifile guttene er tøysere? Målingen av tøysene har feilkilder. Et argument kunne være: Tøysene ble blant annet lokalisert ut fra ekstreme kombinasjoner av dopbruk, seksualpartnere og kriminalitet. Her kan det også ha blitt flagget unge gutte-homofile som er i en ekstremt marginalisert posisjon med hyppig atferd av denne typen, noe vi vet finnes i visse miljøer (Moseng 2007). Men merk at *de fleste* av disse 51 ble *ikke* lokalisert på dop- eller krimspørsmål, men på tøysesvar om fars/mors yrke og høyde/vekt (tabell 7).

Tabell 7. Oversikt over antall tøysesvarere blant de 51 mulige tøysesvarerne blant homofile/bifile gutter for de sju kriteriene for mulige tøysesvar.

Fars/mors yrke	Aldersopplysninger	Multi-funksjonshemmet	Høyde/vekt	Sex-partnere	Tung dopbruk	Uhyrlig kriminalitet
16	7	1	21	5	7	10

En troverdig konklusjon er derfor at en del av guttene som har sagt de er homofile eller bifile, har tøysa. De har kanskje lekt med identiteter for skøy, eller de har antatt den *motsatte* posisjonen av den de selv vil være. Å spøke med en stigmatisert identitet for å harsellere med dem er jo ingen ukjent festmoro – nazister som lager jødeparodier, rasister som lager parodier av svarte, avholdsfolk som spiller full mann, osv. At en del tøysere vil kalle seg homo er forståelig når en tar utgangspunkt nettopp i hva tøysere vil tiltrekkes av – det sjokkerende, stigmatiserte og seksuelt relaterte.

5.2 Hvordan tøysehomsene skiller seg fra homser uten tøys

Mens det ser ut som at tøysere som tar på seg en lesbisk/bifil merkelapp i liten grad forekommer blant jentene, er dette tilsynelatende mer utbredt blant guttene. Vi har altså – forsøksvis – delt de 187 guttene som har svart de er homofile eller bifile i to grupper – homser uten tøys (N=136) og homsetøysere (N=51). Å teste validiteten av denne inndelingen er vanskelig. Mye tyder på at blant de 51 er det en del tøys, men gjelder det alle? Og hvis det er en tilbøyelighet til å tøyse med homofil, kan det vel også skje at det er noen tøysere blant de 136 som ikke ble flagget? Og blant de heterofile guttene kan det være noen som ikke vil tre ut av skapet i denne undersøkelsen.

Til støtte for inndelingen i homser uten tøys og tøysehomsere kan en muligens bruke en del svarmønstre på spørsmål som en ikke uten videre ville anta er relatert til seksuell legning.

Jeg skal vise noen eksempler på at tøysehomsene har et gjennomgående svarmønster som tyder på svært lite valide svar, og at de avviker sterkt i så måte fra de jeg kaller homser uten tøys som ofte ligger nærmere de heterofile guttene på mange spørsmål som en i utgangspunktet ikke skulle tro er sterkt relatert til seksuell legning.

Tabell 8. Svarmønstre i Ung i Oslo 2006 blant gutter, tre kategorier: Heterofile gutter, homsegutter uten tøys (N=136), og tøysehomsere (N=51).

	Heterofile gutter (N=4452)	Homser uten tøys (N=136)	Tøysehomsere (N=51)
(1) Andel (%) som konsekvent bruker det mest negativt svaralternativet for 12 spørsmål om foreldre	0,1	0,7	19,6
(2) Andel (%) som er overbevist om at det finnes gjenferd og spøkelser (blant de som svarte)	11,6	20,3	55,6
(3) Andel (%) som har svart konsekvent «Ikke tillit i det hele tatt» på sju spørsmål om tillit til politiet, skolen, leger, osv.	2,4	13,2	43,1

Tidlig i spørreskjemaet (før spørsmålene om kriminalitet, dop, seksuell legning) kommer en spørsmålsrekke om foreldrene. En skal krysse av om diverse utsagn «stemmer helt», «stemmer ganske godt», eller «stemmer ikke i det hele tatt». Jeg har plukket ut 12 av disse spørsmålene som alle er konstruert slik at svaralternativet «Stemmer ikke i det hele tatt» markerer den mest negative holdningen til foreldre. Dette svaralternativet er plassert lengst

til høyre. En som ikke gidder å lese selve spørsmålsteksten, kan gjøre det lett for seg ved å konsekvent krysse av den ytterste høyre rubrikken, og derved framkommer et bilde av ekstremt negative svar om sine foreldre. Som det framgår av tabell 8, har bare noen promille blant heteroguttene og homser uten tøys krysset av det mest negative for alle disse 12 spørsmålene, men dette *ekstreme* svarmønsteret er brukt av 20 prosent av tøysheidsene.

Litt seinere i skjemaet (men fortsatt før spørsmålet om seksuell legning) kommer et spørsmål om en tror på spådommer, tankeoverføringer osv, deriblant et spørsmål om en er overbevist om eksistensen av «gjenferd og spøkelses». Blant de heterofile guttene svarer 12 prosent at de er overbevist om dette, mot 20 prosent blant homsene uten tøys, og hele 56 prosent blant tøysheidsene.

Mot slutten av skjemaet blir elevene spurt om de har tillit til sju institusjoner (politiet, politikere, skolen, barnevernet, dommere/advokater, leger/sykehus, Stortinget). Et ekstremt negativt svar ville være å krysse av rubrikken «Ikke tillit i det hele tatt». Nær halvparten av tøysheidsene har gjort det (43 prosent), mot 13 prosent blant homsene uten tøys og bare to prosent blant de heterofile guttene. Dette spørsmålet trenger ikke å være uten relasjon til seksuell legning – det er kjent at homser kan ha anstrengte relasjoner til helsepersonell – men likevel ser en her at homoer uten tøys er forholdsvis nær de heterofile guttene, mens det ekstreme avviket finnes hos tøysheidsene.

Dette er altså en del eksempler som viser at de jeg har kalt tøysheidsomer i denne sammenhengen har helt spesielle svarmønstre. Ofte er det ganske konsekvente rekker av de mest ekstreme eller negative svaralternativene, der det ikke ser ut til å ha noen betydning hva det egentlig er spurt om. Slike mønstre *styrker* en hypotese om at vi her står overfor svar med ekstrem lav grad av validitet og at dette kan være gutter som har krysset av svaralternativet homofil/bifil bare på tøys.

Men kan vi vite at de som vi har klassifisert som homser uten tøys – faktisk er ærlige svarere? Kan det være noen som tøys med homsesex også blant dem? Det kan det selvsagt være. Men vi ser i tabell 8 at på en rekke typer spørsmål som kunne tenkes å være relativt lite relatert til seksuell legning, er svarene til homser uten tøys ikke så veldig forskjellige fra svarene til heterofile gutter. Dette styrker en antakelse om at konsekvent tøys neppe er så utbredt blant de 136 klassifisert som homser uten tøys. Dette gjør at vi kan bygge videre på antakelsen om at blant de 187 guttene som har svart

homofil/bifil i hele utvalget, er det rimelig å anta at de 51 vi har kalt tøysehomser utgjør en brukbar tilnærming til å avgrense svarere som har kalt seg homofil eller bifil på tøys.

5.3 Har tøysehomsene fordreid resultatene i NOVA Rapport 19/07?

I denne forbindelse er det spurt om NOVA-Rapporten 19/07 «Vold mot lesbiske og homofile tenåringer» (Moseng 2007) har *overestimert* problematiske livssituasjoner blant homofil, lesbisk og bifil Oslo-ungdom. I lys av diskusjonen i avsnittene 5.1 og 5.2 kan en stille spørsmålet om tøysesvar kan ha påvirket resultatene i denne NOVA-rapporten. Der framstår Oslo-ungdommer som er homofile, lesbiske og bifile med en ekstremt problematisk og vanskelig situasjon på mange felt, sammenlignet med heterofil ungdom, og overhyppigheten av problematiske livssituasjoner og livserfaringer ser særlig ut til å gjelde for de homofile/bifile guttene.

Tabell 9. Problematiske livssituasjoner (%) blant gutter, heterofile og homofile/bifile. A: Tøysesvar inkludert, B: Tøysesvar ekskludert. 95% konfidensintervall i parentes.

	Heterofile gutter	Homofile/bifile gutter	
		Tøysehomser tatt med	Tøysehomser ikke tatt med
Andel (%) debutalder opp til 11 år	1	24* (17,9-30,1)	13
Mobbet daglig	2	23 (17,0-29,0)	14
Utsatt for trusler om vold, en gang eller mer	36	51 (43,8-58,2)	47
- for vold som krevde legebehandling en gang eller mer	13	39 (32,0-46,0)	32
Utsatt for vold fra foreldre	2	16 (10,7-21,3)	9
Har røykt hasj eller marihuana	10	27 (20,6-33,4)	20
Tatt LSD	1	16 (10,7-21,3)	7
Brukt heroin	1	18 (12,5-23,5)	9
Brukt kokain	2	20 (14,3-25,7)	10
Selvordstanker siste uke	5	45 (37,9-52,1)	36
(N) (Samtlige svar, inkludert ubesvart)	(4452)	(187)	(136)

* Merk at estimatene for heterofile gutter er gjort med tøysesvar inkludert. Prosentanslagene blir bare marginelt endret om tøysesvarene tas ut, ettersom andelen tøyser blant de heterofile er bare 2,2 prosent.

Rapporten kommer fram til at generelt har homofile/lesbiske/bifile hatt mer erfaringer med sex, og særlig de homofile/bifile guttene har ofte svært tidlig debutalder. Homofile, lesbiske og bifile ungdommer, særlig guttene, er langt mer utsatt for mobbing, trusler og vold som førte til at de trengte legebehandling. Uvanlig ofte utsettes de for vold fra foreldre, de har oftere erfaring med narkotika, deriblant tunge stoffer, og de er mer plaget av selvmordstanker enn heterofil ungdom.

Hvordan framtrer dette bildet om vi tar vekk de svarerne vi har karakterisert som mulige tøysesvar? Dette framgår av tabell 9 for guttene. Tøysehomoene blant guttene har ført til en klar oppjustering av hyppighet og forekomst av problematiske livssituasjoner, noen ganger med dramatiske utslag. Andelen som har brukt kokain, f.eks., var 20 prosent blant homofile/bifile gutter når tøysesvarene var tatt med, men sank til 10 prosent når tøysehomoene ble tatt ut av materialet.

Men likevel: Også når tøysesvarene er utelatt, framstår de homofile/bifile guttene med klart mer problematiske livssituasjoner i gjennomsnittet enn blant heterofile gutter.

En enkel statistisk test er også foretatt i tabell 9. Om vi antar at estimatene der tøysehomoer ikke er tatt med, er de riktige, ser vi at for åtte av de ti indikatorene i tabell 9 er de riktige estimatene *utenfor* et 95 % konfidensintervall rundt tallene presentert i NOVA Rapport 19/07. Gitt slike forutsetninger ville altså konfidensintervall *stort sett ikke* ha fanget opp de riktige estimatene, men vi ser likefullt at nedre grense for konfidensintervallet stort sett er nær de reviderte estimatene basert på data der tøysehomoene ikke er med.

For analysen av situasjonen blant jenter som er lesbiske/bifile, blir utfallet av en slik analyse annerledes, fordi *så få* av jentene som sa de var lesbisk/bifil ble lokalisert som tøysere (bare 13 av i alt 167). Derfor gjør det små utslag om tøysere er med eller ikke blant de lesbiske/bifile jentene. En kan likevel si at tøysere blant jentene også har bidratt til å forsterke inntrykket av problematiske livssituasjoner blant lesbisk/bifile jenter (tabell ikke vist).

5.4 Bør NOVA Rapport 19/07 trekkes tilbake?

Min tolkning er: NOVA Rapport 19/07 presenterer anslag over omfanget av problematiske livssituasjoner (seksuell debutalder, utsatthet for mobbing, vold, dopbruk, selvmordstanker) blant homofile/bifile *gutter* som er klart overdrevet.

Et viktig moment er likevel: Også når tøysesvar blir tatt ut av materialet, er *hovedbildet* det samme – en markant større hyppighet blant homofile/bifile gutter når det gjelder slike problematiske livssituasjoner og erfaringer. Betrachtingene gjort her tilsier at det er sannsynlig at NOVA Rapport 19/07 overdramatiserer situasjonen for homofile/bifile gutter, men når det gjelder sosialpolitiske implikasjoner, blir de i praksis akkurat de samme om en foretrekker de reviderte (nedjusterte) estimatene presentert over. Også disse gir inntrykk av en betydelig overhyppighet av vanskelige livssituasjoner og opplevelser blant de homofile/bifile guttene.

Når det gjelder analysene av lesbiske/bifile jenter gir de reviderte anslagene bare noe lavere estimater for problematiske livssituasjoner, etter som ganske få jenter ble klassifisert som tøysere med å være lesbisk/bifil.

Alt i alt tilsier dette at det ikke er noen tvingende grunn til å trekke NOVA Rapport 19/07 tilbake, selv om en del av estimatene kanskje er overdrevne. Generelt er det neppe mulig å unngå at enkelte forskningsrapporter gir et noe upresist virkelighetsbilde, på grunn av svakheter i data eller analysearbeid. Det som er mest tungtveiende, er om bildet som presenteres, er misvisende eller feilaktig på en politisk betydningsfull måte. Hovedbildet som NOVA Rapport 19/07 presenterer, kan ikke sies å være misvisende i den forstand at den leder til feilaktige sosialpolitiske konklusjoner, og i sosialpolitisk forstand kan en neppe hevde at NOVA Rapport 19/07 er sterkt villedende.

6 anbefalinger

6.1 Generelle bemerkninger om datakvalitet i Ung i Oslo 2006

I dette notatet har jeg prøvd å vise at det er problemer med datakvaliteten i materialet samlet inn av Ung i Oslo 2006. En bør imidlertid ikke dramatisere dette i den forstand at en trekker en konklusjon om at materialet ikke kan eller bør brukes. Materialet kan brukes, men må brukes med forstand og godt forskerhåndverk. For mange typer spørsmål vil materialet gi muligheter for bra analyser av mange sider ved Oslo-ungdommenes livssituasjon. Data-materialet er en viktig kilde til kunnskap om denne ungdomsbefolkningen – men en må være klar over forbehold og feilkilder.

6.2. Tøysesvar og hva vi gjør med dem

Når over 11 000 Oslo-skoleungdommer besvarer en slik datainnsamling, er det knapt tenkbart at ikke noen vil bruke skjemautfyllingen som en anledning for å tulle og tøyse. Å lokalisere tøyserne byr imidlertid på mange vansker. Her har jeg brukt sju kriterier som alle synes som sannsynlige kandidater for tøyning, og via disse kriteriene ble 259 svarere flagget som tøyser. At omlag 1 av 50 er en tøysesvarer virker ikke urimelig, tatt i betraktning dagens ungdomskulturer.

Jeg foreslår: I framtidig bruk av datamaterialet i Ung i Oslo 2006 blir disse 259 tatt ut av datamaterialet, slik at analysene blir foretatt for de gjenværende 11 241 respondentene. Dette betyr f.eks. at om materialet skal benyttes av personer utenfor NOVA, skal det normalt være en fil med de 11 241 respondentene som blir distribuert til dem.

Selvsagt kan dette være en feilkilde. Det er ikke sikkert at alle de 259 er tøyser. Det er fullt mulig at en del av dem har tøyset på noen spørsmål, men ikke på andre. Det kan være at noen av tøyserne også representerer ekstremt vanskelige livssituasjoner. Men sjansen for å gjøre store feil ved denne framgangsmåten, er svært liten. Om respondenter som urettmessig er blitt klassifisert som tøyser blir tatt ut av datafilen, vil dette ikke påvirke

analyseresultater for helheten hvis de feilaktig klassifiserte tøyserne er noenlunde tilfeldig spredt i datamaterialet.

Men tøysesvarere skal ikke bli eliminert fra den originale datafilen – deres svar er en valid kilde til studiet av et ungdomsfenomen: Tilbøyeligheten til å avgi tøysesvar.

6.3 Ny dokumentasjon

Den foreliggende dokumentasjonen (Valset & Øia 2006) bør suppleres med følgende tillegg:

- (1) En realistisk vurdering av den *generelle* datakvaliteten i materialet. Her må det generelt vises til at validitet i spørreskjemadata er et komplisert spørsmål, og slike data har mange feilkilder. Dette må alle som analyserer materialet ta hensyn til. Det må videre framgå at stor spørsmålsmengde, samt vansker med å forstå spørsmålene blant noen, har gitt mulighet for en god del tilfeldige feil. Det må også påpekes at svarprosenten er synkende utover i undersøkelsen, og at særlig ungdom med mindre privilegerte sosiale plasseringer har synkende svarprosent i den siste halvparten av undersøkelsen.
- (2) Spesielle forhold knyttet til tidsangivelser: Det må påpekes at respondenter ofte *overser* hvilket tidsrom svar skal avgis for, og hukommelsesproblemer kan gjøre seg gjeldende. Forskere bør være svært forsiktige med å overvurdere respondentenes evne til å forstå spørsmålstekstene bokstavelig slik de var ment. Beregninger om hvor ofte visse typer atferd foregår bør foretas med forsiktighet og bør presenteres med romslige konfidensintervall og refleksjon over om respondenten har svart for angitt tidsrom. Materialet er bedre egnet til å analysere faktorer som samvarierer enn til å prøve å lage anslag av hvor mange blant ungdommene i Oslo som gjør et eller annet hver uke.
- (3) Dokumentasjonen bør også gjøre rede for spørsmålet om tøysesvar, at det er sannsynlig at det oppstår tøysesvar i en ungdomsundersøkelse av denne art, og hvordan NOVA har behandlet dette fenomenet. Generelt kan en også si at selv om tøysesvar forekommer, er dette en relativt liten kilde til svak datakvalitet i materialet som helhet. Kildene til vansker med datakvaliteten i det store og hele, er mye mer knyttet til at respondenter, uten å ha noe bevisst ønske om å gi feil informasjon, misforstår

spørsmål, overser anvisninger, ikke rekker å tenke igjennom svarene ordentlig, osv.

- (4) Dokumentasjon bør ikke gi en generell advarsel mot å analysere det som noen kaller «lavprevalente» fenomener. Om antatte tøysere er tatt vekk, vil forhåpentligvis en del feilkilder i analysen av slike fenomener være borte. Datamaterialet kan sikkert brukes til å analysere lavprevalente grupper, f.eks., de 447 jentene som er med i kristne foreninger eller de 242 guttene som sier de har 6 i matematikk. Det gjelder for analysearbeidet, uansett om det er små eller store kategorier som undersøkes, å kjenne materialet nøye og å være klar over feilkilder.

6.4 Lærdommer for nye datainnsamlinger

Diskusjonen som har oppstått rundt Ung i Oslo 2006 gir en del lærdommer, som jeg mener kan oppsummeres slik:

- (1) En eventuell framtidig Ung i Oslo 2016 bør planlegges ut fra en større bevissthet om hva som generelt kan svekke datakvaliteten i slike undersøkelser. Det betyr at undersøkelsen bør ha et mindre omfang (550 spørsmål gir for stor uttrøtting av respondentene). Spørsmålene må drøftes/testes ut mer grundig (Ung i Oslo 2006 har en del dårlige/ ikke-fungerende spørsmålskonstruksjoner). En må ta sikte på at alle grupper blant Oslo-ungdommen, ikke bare de som har stor leseforståelse og er høyeffektive, vil greie å delta *og fullføre* undersøkelsen.
- (2) Tøysesvarere er trolig et fenomen som vil oppstå igjen og igjen. Ved bruk av en del strategisk valgte åpne spørsmål kan en gi mye bedre muligheter for å lokalisere (og studere!) tøysesvarere enn det var mulig i Ung i Oslo 2006. De *åpne* svarene om bl.a. folkegruppe og helseproblemer ble denne gangen overlatt til firmaet som gjør den optiske lesningen. Så vidt vites er de originale besvarelsene nå blitt makulert. Dette var svært ubetenksomt.
- (3) De *originale* besvarte spørreskjemaene bør gjennomgås nøye, av to forskere, *før* de sendes til optisk lesning. Dette vil gi lærdommer og inntrykk som er nyttig for å bedømme kvaliteten.

Litteratur

Bakken, Anders (1998), *Ungdomstid i storbyen*. NOVA Rapport 7/98. Oslo: Norsk institutt for forskning om oppvekst velferd og aldring.

Gilbert, Nigel (1993), *Researching social life*. London: Sage.

Haraldsen, Gustav (1999), *Spørreskjemametodikk – etter kokebokmetoden*. Oslo: Ad Notam Gyldendal.

Hegna, Kristinn (2007), *Homo? Betydningen av seksuell erfaring, tiltrekning og identitet for selvmordsforsøk og rusmiddelbruk blant ungdom*. NOVA Rapport 1/07. Oslo: NOVA – Norsk institutt for forskning om oppvekst, velferd og aldring.

Hjorth Andersen, Bjarne & Mogens Nygaard Christoffersen (1982), *Om spørreskemaer. Problemer ved spørsmålformulering i interviewundersøgelser*. København: Teknisk Forlag.

Hustoft, Anne Gro & Berit Vannebo (2000), *En undersøkelse av frafallet i utvalgsundersøkelser i perioden 1997–2000*. Oslo: Statistisk sentralbyrå.

Kollisch, T. & D. Oberwittler (2002), «How honestly do male adolescents report their delinquent behaviour? Results of an external validation of self-reported delinquency». *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 56:708–735.

Maxfield, M. G., B. L. Weiler & C. S. Widom (2000), «Comparing self-reports and official records of arrests». *Journal of Quantitative Criminology*, 16:87–110.

Mordal, Tove L. (1989), *Som man spør, får man svar. Arbeid med survey-opplegg*. Oslo: Tano.

Moseng, Bera Ulstein (2007), *Vold mot lesbiske og homofile tenåringer*. NOVA Rapport 19/07. Oslo: Norsk institutt for forskning om oppvekst velferd og aldring.

Olsen, Henning (1998), *Tallenes talende tavshed. Måleproblemer i surveyundersøgelser*. København: Akademisk Forlag.

Pape, H., I. Rossow & E.E. Storvoll (2007), *Report of Study Methodology for the School Surveys 2004, 2005 and 2006*. www.sirus.no. Oslo: Statens institutt for rusmiddelforskning.

Pedersen, W. & K. Hegna (2003), «Children and adolescents who sell sex: a community study». *Social Science & Medicine*, 56:135–147.

- Pedersen, W., S. O. Samuelsen & L. Wichstrom (2003), «Intercourse debut age: Poor resources, problem behavior, or romantic appeal? A population-based longitudinal study». *Journal of Sex Research*, 40:333–345.
- Rossow, I. (2006), «Inferences of associations and implications for prevention: the case of early drinking onset». I: J. Elster, O. Gjelsvok, A. Hylland & K. Moene, red., *Understanding choice and explaining behaviour. Festskrift til Ole-Jørgen Skog*. Oslo: Universitetsforlaget.
- Rossow, I. & Anne Kristin Bø (2003), *Metoderapport for datainnsamlingen til Ung i Norge 2002. NOVA's intranettsider*. Oslo: Norsk institutt for forskning om oppvekst, velferd og aldring.
- Rutle, Olav (1981), *Primærlegen ut av skyggen. Analyse av situasjonen i dag, tanker om morgendagen*. Oslo: SIFF – Gruppe for helsetjenesteforskning.
- Tourangeau, R. & T. Yan (2007), «Sensitive questions in surveys». *Psychological Bulletin*, 133:859–883.
- Upchurch, D. M., L. A. Lillard, C. S. Aneshensel & N. F. Li (2002), «Inconsistencies in reporting the occurrence and timing of first intercourse among adolescents». *Journal of Sex Research*, 39:197–206.
- Valset, Kirsti & Tormod Øia (2006), *Dokumentasjon «Ung i Oslo 2006»*. Oslo: Norsk institutt for forskning om oppvekst velferd og aldring.
- Wold, B., H. Aasen, L.E. Aarø & O. Samdal (1995), *Helse og livsstil blant barn og unge i Norge. Resultater fra en landsomfattende spørreskjema-undersøkelse tilknyttet prosjektet «Helsevaner blant skoleelever – en WHO-undersøkelse i flere land (HEVAS)»*. Bergen: Universitetet i Bergen, HEMIL-senteret.
- Wong, S. L., S. T. Leatherdale & S. R. Manske (2006), «Reliability and validity of a school-based physical activity questionnaire». *Medicine and Science in Sports and Exercise*, 38:1593–1600.