

NOTAT

Tromsøundersøkelsen anbefaler at det brukes anonyme data til studentoppgaver på masternivå

Ved bruk av anonyme datasett er det ikke nødvendig å søke Regional komite for medisinsk forskningsetikk (REK) og gjøre en personvernkonsekvensvurdering (Data Protection Impact Assessment, DPIA) for å få tilgang til data fra Tromsøundersøkelsen. Dette er prosesser som kan bli tidkrevende for studentprosjekter på masternivå.

Anonymisering innebærer at vi fjerner muligheten for å identifisere enkeltpersoner i et datasett. Anonymisering er et viktig virkemiddel for å hente ut verdifull innsikt ved dataanalyse, samtidig som risikoen reduseres for berørte personer. Når personopplysninger anonymiseres, regnes de ikke lenger som personopplysninger ([Veileder anonymisering av personopplysninger. Datatilsynet 2015](#)).

«Anonyme data» betyr at det ikke er mulig å finne tilbake til personen som opplysningene skriver seg fra (bakveisidentifisering), verken via en kodeliste, eller at opplysningene i seg selv avslører personen. Søknader om anonyme data sendes Tromsøundersøkelsen og behandles av Data- og Publiseringsutvalget (DPU). Før utlevering av anonymt datasett inngås det en avtale mellom Tromsøundersøkelsen, veileder og student som regulerer rammer for oppbevaring, bruk og sletting av datasettet.

DPU vurderer alle søknader om anonyme data. Rammeverk for anonyme datasett:

- Maksimalt antall variabler vil være omtrent 20, men antallet vurderes per prosjekt og kan være avhengig av hvilke variabler som etterspørres
- Det brukes 10-årige eller 5-årige aldersgrupper, og kun hvis nødvendig alder i år. Det lages en egen kategori 80 år og eldre for deltakere over 80 år
- Sannsynlighet for bakveisidentifisering vurderes. Dersom variabler inneholder ekstreme verdier blir de ekstreme verdiene enten satt til «missing», eller justert (trunkert) til passende nivå. Generelt vil sannsynlighet for bakveisidentifisering påvirkes av datasettets størrelse og type studie. Dersom et resultat kan knyttes til en liten gruppe (3-5 personer og/eller liten kildepopulasjon), og det foreligger offentlig tilgjengelige bakgrunnsopplysninger om populasjonen øker sannsynligheten for bakveisidentifisering. Dersom variablene omfatter lett gjenkjennelige ytre tegn, bosted eller yrke for den enkelte deltaker, øker muligheten for bakveisidentifisering
- Sensitivitet i aktuelle variabler vurderes. Ekstreme måleverdier, enkelte sykdommer, selvrapportert bruk av illegale rusmidler og spesielle typer av medisiner kan være særlig sensitive

Måleverdier for f.eks. blodtrykk, kolesterol osv. gjør det ikke mulig å bakveisidentifisere individuelle forskningsdeltakere. Høyde, vekt og kroppsmasseindeks kan i enkelte tilfeller være mulig å bakveisidentifisere dersom en forskningsdeltaker har ekstreme verdier.

For enkelte veiledere og studenter er det ønskelig å få utlevert aidentifiserte data og/ eller større datasett. Dette er særlig aktuelt dersom det er behov for mer omfattende analyser. Et slikt studentprosjekt krever søknad til REK og forskningsansvarlig institusjon må vurdere behovet for å gjennomføre personvernkonsekvensvurdering. Alle prosjekter som er forankret ved UiT Norges arktiske universitet og behandler personopplysninger skal meldes til NSD ([Norsk senter for forskningsdata](#)) i tråd med UiTs retningslinjer.