

# CONOR-databasen – et lite stykke Norge

Universitetene i Tromsø, Trondheim, Bergen, Oslo og Nasjonalt folkehelseinstitutt har gått sammen om å samle flere befolkningsbaserte helseundersøkelser i én database: Cohort of Norway (CONOR). Databasen inneholder informasjon om sykdom, helse, miljø og livsstil. I tillegg finnes nedfrost fullblod fra hver deltaker, og DNA er tilgjengelig for to tredeler av dem. CONOR er spesielt egnet for å undersøke sammenhenger mellom gener, eksponering og sykdom.

■ CONOR består i alt av ti helseundersøkelser gjennomført i tidsrommet 1994–2003 (tab 1). Data fra ytterligere to undersøkelser vil bli lagt til innen 2011 (Helseundersøkelsen i Nord-Trøndelag (HUNT) III og den sjette Tromsøundersøkelsen). CONOR ledes av en styringsgruppe med én representant fra hver av partene i samarbeidet (ramme 1). Databasen er koblet til Medisinsk fødselsregister, Dødsårsaksregisteret, Reseptregisteret, Kreftrregisteret, statistikk fra Forløpsdatabasen trygd (FD-Trygd) hos NAV og Statistisk sentralbyrås registre som inneholder sosiale og demografiske data. For å få flere sykdomsendepunkter vil også andre koblinger være aktuelle, f.eks. nasjonale og regionale registre med beintetthet, brudd og hjerteinfarkt og Norsk pasientregister, som ble personentydig i 2007. CONOR kan også kobles til registre med eksponeringsvariabler som Drikkevannsregisteret.

CONOR har sin egen hjemmeside (1) og er presentert også internasjonalt (2). Alle deltakerne er over 20 år, og aldersfordelingen varierer fra undersøkelse til undersøkelse. Det er noen flere kvinner enn menn som deltok.

## Datakilder

CONOR består av fire datakilder for hver deltaker: Den *første* kilden er svar på vel 50

CONOR-spørsmål innhentet fra spørreskjemaer: om livsstil (fysisk aktivitet, røyking, alkohol), utdanning, bomiljø, familie og venner, arbeid og yrke og bruk av medisiner. I tillegg spørres det om helse og sykdom (f.eks. tidligere underarms- og lårhalsbrudd, symptomer og sykdommer innen hjerte/kar, hørsnue og allergi, muskel- og skjelettsmerter, psykiske plager) samt sykdommer i familien. De fleste spørsmålene er identiske i alle undersøkelsene. Noen spørsmål, som i utgangspunktet var litt forskjellig formulert, er kodet om slik at de gir tilnærmet lik mening for alle helseundersøkelsene. Alle endringer, sammenslåinger og ulikheter er beskrevet i et eget dokument som følger utlevering av data fra databasen. Den *andre* kilden er målinger av blodtrykk, puls, høyde, vekt og liv- og hoftemål, mens den *trede* kilden er resultater fra analyser av glukose og lipider i serum tappet ved undersøkelsestidspunktet. Den *fjerde* kilden er EDTA-blod. Dette finnes nedfrost i CONORs biobank, lokalisert sammen med Helseundersøkelsen i Nord-Trøndelags biobank i Levanger. Det er ekstrahert DNA fra ca.  $\frac{2}{3}$  av deltakerne.

Det er per desember 2009 i alt 173 236 enkeltindivider i CONOR, der 7 309 personer ble fulgt opp i to undersøkelser. Etter at Helseundersøkelsen i Nord-Trøndelag III og den sjette Tromsøundersøkelsen blir

inkludert i CONOR, vil antall personer være om lag 200 000.

## Problemstillinger og design

CONOR er særlig nyttig for å studere om risiko for sykdom henger sammen med eksponering, enten ved å bruke hele databasen i kohort-, tverrsnitt- eller økologiske studier eller deler av databasen i kasus-kontrollstudier. For mer kompliserte problemstillinger der sammenhengen mellom gener, eksponering og sykdom skal studeres, kan nøstet kasus-kontrollstudier eller kasus-kohortstudier være aktuelle (3). Når vi skal velge mellom de to siste, er det flere hensyn som må veies mot hverandre, slik som tidsforskjellen mellom innsamlingstidspunkt og analysetidspunkt, tid fra rekruttering og lengde av oppfølging (4).

## Utvalgs- og informasjonsskjevheter

CONOR er en samling av flere utvalg og blir således en heterogen kohort, som er beheftet med seleksjonsskjevheter i forhold til den norske befolkning.

For studier av assosiasjon mellom eksponeringsvariabler og sykdom er det ikke så viktig om utvalget er representativt, siden det er den (mulige) sammenhengen som er av interesse og ikke det totale befolkningstallet. Ifølge Hjartåker og Lund (5), bør man utforme assosiasjonsstudier ut ifra

**Tabell 1** Oversikt over de ulike helseundersøkelsene som inngår i CONOR, forkortelser, innsamlingsår, alder på de inviterte og antallet

Navn	Forkortelse	Innsamlingsår	Inviterte (alder i år) <sup>1</sup>	Antall inviterte	Antall deltakere
Den fjerde Tromsøundersøkelsen	Tromsø IV	1994–95	25+	37 558	26 925
Helseundersøkelsen i Nord-Trøndelag II	HUNT II	1995–97	20+	94 196	65 018
Helseundersøkelsen i Hordaland	HUSK	1997–99	40–44, 46–47, 70–72	38 587	25 530
Osloundersøkelsen II	Oslo II	2000	48–77	14 209	6 919
Helseundersøkelsen i Oslo	HUBRO	2000–01	30, 31, 40, 45, 46, 59/60, 75/76	58 660	22 015
Helseundersøkelsen i Oppland og Hedmark	OPPHED	2000–01	30, 40, 45, 60, 75	22 327	12 402
Den femte Tromsøundersøkelsen	Tromsø V	2001	30+	10 353	8 077
Helseundersøkelsen i Oslo 2002 – Innvandrere	Innvandrer-HUBRO	2002	20–60	12 088	3 683
Helseundersøkelsene i Troms og Finnmark	TROFINN	2002	30–77	16 229	9 327
Mosjon på Romsås II	MoRo II	2003	34–70	5 535	1 995

<sup>1</sup> Detaljert informasjon om de enkelte helseundersøkelsene er dokumentert på CONORs nettside (1)

muligheten av å kunne utføre valide sammenlikninger mellom grupper, mer enn at gruppene skal være representative. Det er derfor viktig å vurdere størrelse og sammensetning av disse gruppene. Hvorvidt dette kravet er oppfylt for CONOR, må vurderes i hvert enkelt tilfelle. Forfatterne har undersøkt hvordan ulike nøkkelvariabler som alder, kjønn, utdanning, røykevaner, alkoholbruk og fødeland samsvarer med befolkningen for øvrig. Disse undersøkelsene viste at CONOR svarer til Norges befolkning (ikke publiserte data).

Mange av opplysningene i CONOR er selvrapporterte gjennom spørreskjemaer. Vi må derfor anta at disse variablene er beheftet med målefeil (6). Vi må også anta at flere av CONOR-spørsmålene om f.eks. røyking og alkoholbruk er beheftet med underrapportering. Slike målefeil eller informasjonsskjevheter vil produsere underestimerer eller overestimerer avhengig av om målefeilen er avhengig (differensiell) eller uavhengig (ikke-differensiell) av sykdom. CONORs styringsgruppe prioriterer å undersøke validiteten av selvrapporterte helsemaal.

#### Tilgang til datamaterialet

Søknad om datautlevering sendes til Folkehelseinstituttet og behandles av CONORs styringsgruppe. Styringsgruppen legger vekt på at forskningsprosjektet ligger

innenfor formålet med CONOR, prosjektets faglige originalitet og kvalitet samt det faglige miljøet rundt søkeren. Det er utarbeidet retningslinjer for utlevering av data (1).

Avtale om utlevering skrives mellom Folkehelseinstituttet og den institusjonen som forskeren representerer og blir normalt gitt for en toårsperiode. Forskeren får tilgang til anonymiserte data. Dersom andre datakilder skal kobles til CONOR-data, må forskeren søke om egen konsesjon.

#### Geir Aamodt

geir.aamodt@fhi.no  
Divisjon for epidemiologi  
Nasjonalt folkehelseinstitutt  
Postboks 4404 Nydalen  
0403 OSLO

#### Anne Johanne Søgaard

Divisjon for epidemiologi  
Nasjonalt folkehelseinstitutt  
og  
Institutt for allmenn- og samfunnsmedisin  
Universitetet i Oslo

#### Øyvind Næss

Divisjon for epidemiologi  
Nasjonalt folkehelseinstitutt  
og  
Institutt for helseledelse og helseøkonomi  
Universitetet i Oslo

#### Anne Cathrine Beckstrøm

Divisjon for epidemiologi  
Nasjonalt folkehelseinstitutt

#### Sven Ove Samuelsen

Matematisk institutt

Universitetet i Oslo  
og  
Divisjon for epidemiologi  
Nasjonalt folkehelseinstitutt

*Vi takker CONORs styringsgruppe ved Grethe S. Tell, Inger Njølstad, Espen Bjertness og Lars Vatten for gjennomlesing og verdifulle kommentarer.*

*Oppgitte interessekonflikter: Ingen*

#### Litteratur

1. Nasjonalt folkehelseinstitutt. CONOR. [www.fhi.no/conor](http://www.fhi.no/conor) (15.12.2009).
2. Næss O, Søgaard AJ, Arnesen E et al. Cohort profile: Cohort of Norway (CONOR). *Int J Epidemiol* 2007; 37: 481–5.
3. Langholz B, Thomas DC. Nested case-control and case-cohort methods of sampling from a cohort: a critical comparison. *Am J Epidemiol* 1990; 131: 169–76.
4. Rundle AG, Vineis P, Ahsan H. Design options for molecular epidemiology research within cohort studies. *Cancer Epidemiol Biomarkers Prev* 2005; 14: 1899–907.
5. Hjartåker A, Lund E. Kohortstudier. I: Laake P, Hjartåker A, Thelle DS et al, red. *Epidemiologiske og kliniske forskningsmetoder*. Oslo: Gyldendal Akademisk, 2007: 185–209.
6. Laake P, Thoresen M, Veierød MB. Målefeilsproblemer. I: Laake P, Hjartåker A, Thelle DS et al, red. *Epidemiologiske og kliniske forskningsmetoder*. Oslo: Gyldendal Akademisk, 2007: 128–68.

*Manuskriptet ble mottatt 11.3. 2009 og godkjent 10.1. 2010. Medisinsk redaktør Trine B. Haugen.*