



Kilder og metoder for å identifisere byfolk på tidlig 1900-tall

Sources and methods for identifying urban people in the early 20th century

Gunnar Thorvaldsen

Professor emeritus, UiT Norges arktiske universitet

gunnar.thorvaldsen@uit.no

Sammendrag

Denne artikkelen er skrevet med tre hovedformål. For det første tar den for første gang i bruk den transkriberte utgaven med åpne individdata fra folketellinga 1920 i forskning. Arkivverket gjør denne kilden tilgjengelig i komplett versjon i 2023, noe som gir nye muligheter til å studere en rekke aspekter ved norsk befolknings- og sosialhistorie på tidlig 1900-tall, blant annet detaljert innsikt i tidlig mellomkrigstid ved hjelp av nominative folketellingsdata. For det andre illustrerer artikkelen metoder og problemer ved å lenke sammen individdata fra folketelling i et bysamfunn. Dette er mer komplisert enn for rurale strøk især på grunn av den mer omfattende migrasjonen. For det tredje presenterer artikkelen resultater om migrasjon og migranter inn og ut av et bysamfunn. Sandefjord er velegnet til å studere kompleksiteten i den urbane migrasjonen, en relativt liten by med stor spredning mht. både flyttestinasjoner og fødestedskommuner blant innbyggerne. Her var alle fylkene representert blant innflytternes fødested i 1920, og blant utflytternes destinasjoner manglet bare Troms fylke i 1910.

Nøkkelord

Lenking, historiske individdata, migrasjon, bysamfunn, tidlig 1900-tall

Abstract

This article is written with three main objectives. For the first time, it uses the transcribed edition of individual data from the 1920 census in research. The National Archives made a complete version of this source available in 2023, providing new opportunities to study multiple aspects of Norwegian population and social history in the early twentieth century, including detailed insights into the interwar period with nominative census data. Secondly, the article illustrates methods and problems when linking individual data from censuses in an urban community. This is more complicated than for rural areas, especially due to more extensive migration. Thirdly, it presents results on migration and migrants into and out of an urban society. Sandefjord – a relatively small town with a large variety in terms of both destinations and birthplace municipalities among the migrants – is well suited to studying the complexity of urban migration. All the Norwegian counties were represented among the immigrants' birthplaces in 1920, and only Troms county was missing among the emigrants' destinations in 1910.

Keywords

Record linkage, historical individual level data, migration, urban society, early 20th century.

Innledning

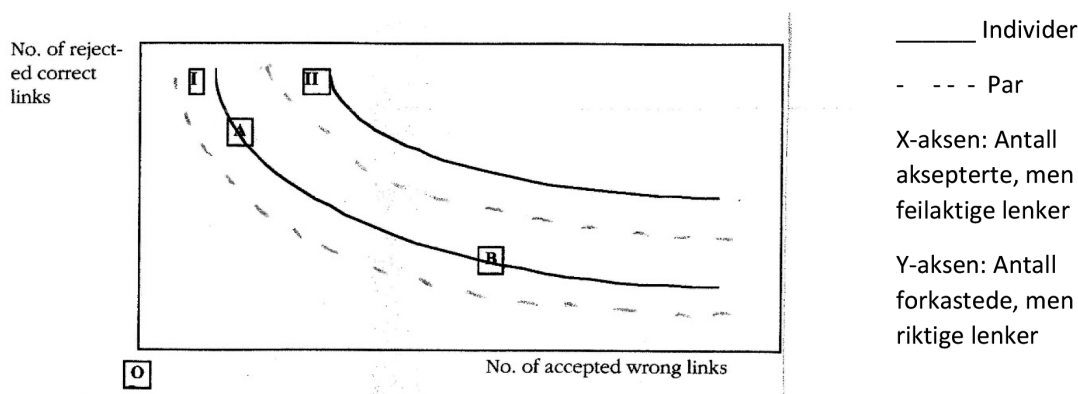
Historisk befolkningsregister (HBR) rekonstruerer befolkningens livsløp ved å lenke sammen individdata fra nominative kilder som folketellinger og kirkebøker. Dermed kan en rekke problemstillinger innen sosial- og befolkningshistorie analyseres empirisk, samtidig som sammenligning av kildene på individnivå avslører mange feil i innførslene. Før Det sentrale folkeregister (DSF) ble startet i 1964, hadde ikke Norge noe enhetlig system å kartlegge migrasjon mellom kommunene med. Mange kommuner manglet lokale folkeregistre inntil 1946, og flyttelistene i kirkebøkene er notorisk mangelfulle. For å bøte på migrasjonsproblemet må Historisk befolkningsregister følge flytterne for å kunne rekonstruere også disse innbyggernes livsløp tilbake til år 1800. Skal vi kunne inkludere migrantene i våre demografiske og sosialhistoriske analyser, må HBR dekke hele landet, ikke bare enkeltkommuner eller -fylker.

Denne artikkelen tar særlig for seg 1910- og 1920-folketellingene, det eneste paret av åpne folketellinger som inneholder fødselsdato på i prinsippet hele befolkninga og derfor gir unike muligheter til å identifisere og følge innbyggerne gjennom ett og etter hvert flere tiår. I den grad datoene er korrekte, gir da folketellingene kildekritiske muligheter til å analysere uoverensstemmelser mellom informasjonen i ulike persondatakilder. Selv om folketellingene 1900 og 1891 bare har fødselsdato på barn under to år, kan disse sammen med kirkebøkene brukes til å finne og rette opp uoverensstemmelser mellom opplysningene om familiene i ulike kilder. Dermed oppfyller vi det kildekritiske formålet som, sammen med migrasjon og mange andre empiriske studier, er hovedmotivene for å skape Historisk befolkningsregister.

Å følge individer og grupper fra kilde til kilde forutsetter lenkingsmetodikk, på engelsk *record linkage*, og store datamengder forutsetter *automatisk* lenking. Manuell lenking av enkeltpersoners livsløp gir riktignok fleksibilitet og gjør det mulig å ta hensyn til spesialtilfeller som bare er beskrevet i tradisjonen til hver enkelt familie, genealogier eller lokalhistorisk kunnskap. I tillegg til å være mer effektiv vil automatisk eller maskinell lenking ved hjelp av dataprogrammer sikre mer enhetlig håndtering av kildene. Slike rutiner eller algoritmer kan imidlertid introdusere feil på grunn av manglende konsistens og unikheter i kildene, for eksempel føre sammen navn som ingen genealog ville tolket som reelle varianter av samme navn. Erfaring med lenking av individdata i en rekke land viser at algoritmene kan gi varierende andel riktige lenker med tid og sted på grunn av innholdet i kildene, som grad av navnelikhet og migrasjonsgrad. Lenkingsalgoritmene i denne artikkelen er i utgangspunktet bare basert på informasjonen i to kildeinnførsler. Da må dataene være relativt unike og konsekvente mht. navn, fødselsdato, fødselsår og fødested, slik tilfellet skal være i folketellingene fra 1910 og 1920. Heldigvis kan vi kontrollere mot supplerende kilder og data om relaterte personer, for eksempel ektepar. Lenkingsteknikkene er et kompromiss mellom å lage så mange riktige lenker som mulig samtidig som man ikke introduserer feilaktige lenker, noe den danske historikeren Hans Chr. Johansen har illustrert i figur 1.¹ Kurvene beskriver forholdet mellom to variabler: X-aksen (horisontalt) representerer antallet falske positive lenker, dvs. aksepterte, men feilaktige lenker, og Y-aksen (vertikalt) representerer antallet rette negative, dvs. manglende lenker som burde vært laget. Hovedpoenget i figuren er at strengere lenkingskriterier (punkt A i figuren) vil motvirke introduksjonen av feilaktige lenker, men samtidig forårsake at flere rette lenker mangler i analysene. Slakkere regler, derimot, vil øke antallet feilaktige lenker, men samtidig sikre at flere rette lenker kommer med (punkt B i figuren). Johansen mener at det mer eksakte innholdet i kildema-

1. Johansen 2002

terialeet utover på 1800-tallet gjorde det lettere å finne et kompromiss mellom de to hensy- nene enn tilfellet er i tidligere kilder. Dette er illustrert med de to heltrukne kurvene, hvorav kurve I representerer nyere og bedre kilder og kurve II eldre og mer diffuse kilder. Intro- duksjonen av fødselsdatoer i folketellingene fra 1910 forbedret presisjonen ytterligere og er utgangspunkt for analysen av lenkingsproblemer i denne artikkelen. De stiplede kurvene i figur 1 illustrerer at relasjoner, for eksempel mellom ektefeller, kan utnyttes til mer presis lenking, slik vi skal se i eksemplene på lenkede livsløp sist i artikkelen. Vi kommer tilbake til lenkingsprinsipper og praktisk lenking nedenfor.



Figur 1. Lenking av individer og par. Kurve I: Senere perioder, kurve II: tidlige perioder. Punkt A: Strenge lenkingsregler, punkt B: Slakkere regler. O: Origo.

Forberedelse av kildematerialet – standardisering og koding

Siden historiske individdata er relativt heterogene med omsyn til utforming, begrepsbruk og stavemåter, lønner det seg å standardisere eller kode innholdet før innførslene lenkes. I utgangspunktet var hovedformålet med kodingen å lage statistikk på grunnlag av informasjonen i folketellingene. Det finnes i hvert fall fire grunner til at sosialhistorikere og demografer ikke nøyer seg med tabellene Statistisk sentralbyrå publiserte etter hver folketelling fra 1920-tellinga i 13 bind.² For det første ble grensene mellom de administrative tellingsenhetene ofte endret i perioden mellom folketellingene. For det andre er sammenligninger over tid vanskelig fordi systemet for kategorisering av informasjonen i folketellingene ble revidert. Det gjelder for eksempel yrkesgrupperingene. Den tredje grunnen er at variablene i publisert statistikk kan bli kombinert på gruppenivå, ikke på individnivå. Dette øker faren for å begå såkalte økologiske feilslutninger, dvs. å trekke slutninger fra gruppers egenskaper til egenskaper ved individene. For eksempel kan statistikken vise at det stedvis var mange bønder, og at de eldre utgjorde en stor andel av befolkningen. Å konkludere med at de fleste bønder var eldre mennesker, behøver ikke være riktig av den grunn.³ En fjerde grunn er at de standardiserte kodene gjør det enklere å lenke personer fra kilde til kilde. For eksempel er kommunenummer 0724 lettere å følge i kildene enn skiftende kommunenavn som Sandeherred og Sandar.

Historikernes ønske om å studere samfunnsutviklingen over tid er altså i sum hovedgrunnen til at vi koder og lenker kildene sammen med ønsket om å produsere statistikk og å lenke kildene uten at små og uvesentlige ulikheter i begrepsbruk, stavemåter o.l. forstyrrer. Mens enkle felt, som kjønn, sivilstand og alder, krever lite standardiseringsarbeid og få regler, er

2. Statistisk sentralbyrå 1923

3. Langholm 1976

koding av yrke, familiestilling, fødested og etnisitet komplekse oppgaver (se tabell 1).⁴ Da datamaskintid og lagringsplass var kostbart, ble informasjonen gjort om til tallkoder manuelt, før den ble punchet og lest inn, noe som gjorde forskningsprosessen lite fleksibel. Endringer krevde ny bearbeiding av originalkildene. Derfor ble det besluttet å transkribere kildematerialet ordrett inn i datamaskinen. Sorteringsalgoritmer kan deretter brukes til å lage lister som gir oversikt over innholdet i folketellinga, og anvende dataprogram som langt på vei automatiserer kodingen, for eksempel et relasjonsdatabaseprogram. Dette kan komprimere innholdet i hver variabel, slik at for eksempel hvert yrke bare forekommer én gang sammen med en frekvens og yrkeskoden.

For hver kilde eller eventuelt distrikt der folketellingene er kodet, blir kodene overført til en databasetabell, og etter hvert gjør prekoding med denne kodesamlingen at kodearbeidet blir både mer konsistent og mer effektivt. Det vil imidlertid dukke opp mange nye variabelverdier (les: egenskaper) i hver nye kilde. I tillegg bør man passe på at koder ikke blir kopiert ukritisk fra et distrikt til det neste. For eksempel kan «tømmermann» bety en sjømann ved kysten, mens han i realiteten sikret gruveganger i et bergverkssamfunn. Etter prekoding er det rasjonelt å spesifisere at alle verdier som inneholder en viss bokstavsekvens, for eksempel «gårdbruker», bør oversettes til en spesifikk kode, såkalt automatisk koding. Endelig er det mulig å taste inn bestemte koder for de lavfrekvente forekomstene, som sogneprest eller lege, noe vi kaller manuell eller interaktiv koding. Til sist dekomprimeres kodefila, slik at hver personforekomst utstyres med relevante koder i henhold til innholdet i hver enkelt variabel. Dermed har vi laget en standardisert utgave av folketellinga som kan brukes i et statistikkprogram eller sammen med den bokstavrette tekstutgaven av kilden når man lenker.

Relasjoner mellom personer i samme hushold eller familie er mulig å analysere i den kodede versjonen fordi den er utstyrt med spesielt konstruerte variabler. For eksempel gir en relasjonsvariabel opplysning om hvem som var ektefeller, fordi den gjensidig inneholder ID-nummeret til ektefellene for gifte personer i hver husholdning. Ved å bruke disse ID-numrene, såkalte «pekere», blir det eksplisitt hvilken ektemann og kone som hørte sammen, selv om det var flere enn ett par i husstanden. Dermed blir det lettere å identifisere ektefellene i andre kilder og studere hvem som giftet seg med hvem. Tilsvarende variabler «peker» fra barna til hver av foreldrene, noe som blant annet er nyttig i studiet av fertilitet. Vær oppmerksom på at i store, kompliserte husholdninger kan dataprogrammet som skaper disse variablene, lage enkelte feil, for eksempel blande sammen barn fra to ektepar som bodde i samme husholdning.

Koding av fødested og standardisering av navn

En norsk folketelling vil vanligvis spesifisere en persons fødested på kommune- eller prestegjeldsnivå. Dette er en viktig variabel både fordi den skiller mellom personer med vanlige navn, og fordi den gir en oversikt over livsløpet til migranter. Som med yrke og familiestilling gjør varierende navn på samme kommune standardisering nødvendig, så alle personposter blir utstyrt med en firesifret kommunekode. Alle kommunenummer i samme fylke har de to første sifrene felles, noe som forenkler studiet av migrasjon på dette nivået. Når det tredje sifferet er null, betyr det at den aktuelle kommunen er en by. For eksempel var 0706 Sandefjord by i Vestfold fylke, 0724 var omkringliggende Sandar i samme fylke, mens 0823 var Eidanger landkommune i nabofylket Telemark. Antall kommuner har

4. Thorvaldsen 1995

endret seg over tid, fordi prestegjeld eller kommuner har blitt delt eller slått sammen, og arealer er overført fra den ene kommunen til den andre. Hvert område som noen gang har vært en egen kommune, har blitt tildelt en unik kode, og en spesiell tabell gir en oversikt over endringene.⁵ Databasen inneholder også de ulike navnene som har blitt assosiert med en gitt kommune gjennom de to siste århundrene. For å håndtere immigranter er systemet utvidet med landskoder for resten av verden.⁶

Selv om kodesystemet er dynamisk og godt dokumentert, oppstår det problemer ved praktisk bruk. Flere kommunenavn er tvetydige, siden noen kommuner deler samme navn. Det gjelder for eksempel Bø i Telemark og i Nordland, som har henholdsvis kommunenummer 0821 og 1867. Når det er stor avstand mellom alternativene, kan tvetydigheten løses ved å velge den mest sannsynlige koden, som er den for stedet nærmest der folketellinga ble holdt. Når en slik løsning er urimelig, må koden for ukjent brukes. Lignende problemer kan oppstå med navn på mindre områder, som gårdsnavn, siden samme navn kan finnes i flere prestegjeld (se nedenfor). Noen ganger angis navnet på et territorium som omfatter flere kommuner. Hvis hele området ligger innenfor samme fylke, kan fylkeskoden brukes, med to nuller for kommunedelen av koden. «Lofoten» får kode 1800, siden alle kommunene lå i Nordland fylke. Registreringsentralens kodete data fra folketellingene i 1865, 1875, 1900 og 1910 er tilgjengelige via internett fra vår partner ved University of Minnesota (se nappdata.org).

Tabell 1. Kodete tekststrenger og folketall i de nasjonale folketellingene 1865 til 1910 og i Sandefjord 1920. Hjemmehørende folkemengde i 1920 var 2 649 775. 1875-tall er avrundet.

Felt	1865	1875	1900	1910	1920 Sandefjord
Familie/hushold	19 188	22 000	25 172	20 050	1390
Yrke	75 734	78 000	368 598	334 079	2044
Fødested	41 813	55 000	73 893	58 227	1083
Antall individer	1 701 756	1 813 424	2 240 092	2 391 782	5764

Standardisering av navnevarianter er nyttig både ved kvantitative analyser av navnefrekvenser og ved lenking av dataposter hvor dette øker størrelsen og kvaliteten på lenkede datasett. Effekten av navnestandardisering på lenking av historiske data er testet for personposter fra folketellinger i USA og i Norge.⁷ Både for- og etternavn har blitt standardisert i samarbeid med professor i nordisk språk Gulbrand Alhaug.⁸ Det skjer via en modell med tre nivåer (se figur 2).

Prosjektet utarbeidet en liste med personnavn hvor variantstavinger og nærliggende lingvistiske varianter var standardiserte til samme oppslagsform. For eksempel var Kristian med K og med Ch kodet til oppslagsformen Kristian og Fredrek standardisert til Fredrik, mens forskjellen mellom Anne og Anna ble bevart, fordi begge navnene er vanlige. Sistnevnte eksempel illustrerer at standardiseringa er varsom. Derfor er det nødvendig å anvende et tilleggsprogram som beregner den fonetiske avstanden mellom navneformer ved hjelp av en spesiell algoritme.⁹ Dermed vil personposter med nærliggende navneformer

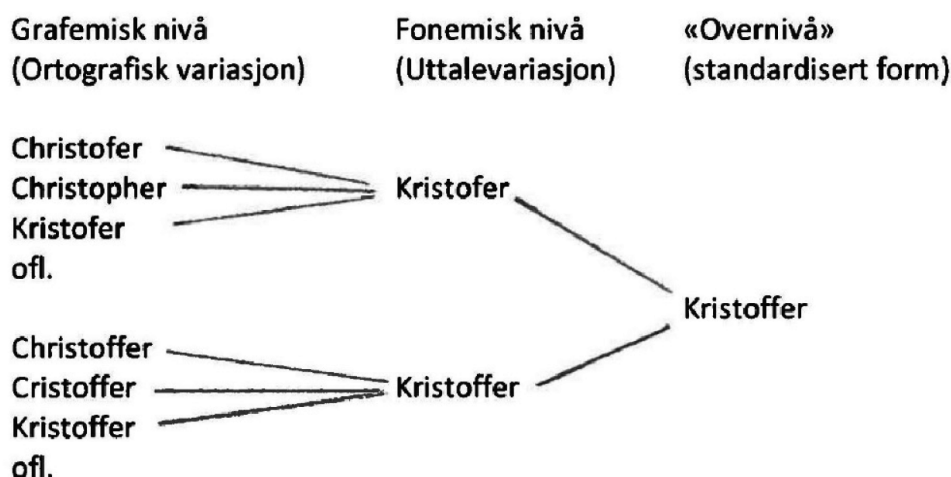
5. Brosveet 1979

6. Både norske og utenlandske historiske stedskoder ligger her: https://rhd.uit.no/koding/fs_koder.html

7. Vick & Huynh 2011

8. Alhaug 2005

9. Winkler 1990



Figur 2. Modell for standardisering av fornavn. Modellen er utarbeidet av G. Alhaug (2005)

bli lenket til tross for at den innledende standardiseringa var forsiktig. Sammenligninga av navnestandardisering i USAs og Norges folketellinger viste at antall lenker i de norske tellingene økte med 17 % på grunn av standardiseringa, og at det var arbeidet med etternavn som bidro mest.¹⁰ Et annet forhold er at navneskikkene endrer seg over tid.¹¹ Dette påvirker både frekvensen av ulike navn i ulike prestegjeld og hvilke navneformer som oppfattes som varianter av samme navn.

Den store jobben med å kode folketellinga 1920 startet i og med denne artikkelen og skal fullføres når en samlet utgave av denne tellinga foreligger i Digitalarkivet.

Ti regler og kort historiografi om lenking

Det er skrevet mange artikler og en del bøker om lenking av historiske individdata. Det er derfor en umulig oppgave å referere til alle regler og erfaringer her,¹² men vi kan oppsummere hovedpunktene slik:

1. Livsløpet lenkene beskriver, må være logisk. For eksempel kan ingen gifte seg etter sin død.
2. Historiske protokolldata inneholder for mange uoverensstemmelser til at man kan kreve full overensstemmelse mellom variablene man anvender til lenking. For eksempel vil det ofte være variantstavinger av navn eller fødesteder og feil i fødselsår fra kilde til kilde.
3. Det vil ofte være tvil om hvor store uoverensstemmelser mellom kildene man skal tillate eller ta hensyn til. Prester og folketellere bestemte ofte navns skrivemåte, så Niels og Nils er utvilsomt varianter av samme navn. Men er e-en en mer signifikant forskjell mellom Kristin og Kristine?
4. Man kan ikke lenke dubletter av identiteter, bare klassifisere dem som lenkingskandidater. Pga. større folketall gjør dette lenkingsproblemene større i mange byer. Et eksempel er seks blad Martha Olsen født i Kristiania i 1890 ifølge folketellinga 1900.

10. Vick & Huynh 2011

11. Fure 1990

12. Viktige referanser er Bloothoof et al. 2015; Christen 2012. Se vedlagte litteraturliste om lenking av historiske individdata.

5. Tvilstilfellene i regel nr. 3 og nr. 4 kan ofte avgjøres ved hjelp av komplementært kilde-materiale. For eksempel kan fødselsdato i folketellingene 1910 og 1920 gi grunnlag for lenking tilbake til dåpslister som ikke kunne lenkes til 1900-tellinga.
6. Også variabler som varierer gjennom livsløpet, kan brukes kritisk til lenking. De mest relevante er adresse og yrke. Vi kan altså lenke en snekkersvenn til en seinere innførsel om en snekkermester, selv om kildene inneholder konkurrerende dubletter med divergerende informasjon om yrke.
7. Det er lettere å lenke ektepar enn enslige personer. Det er vanlig å bruke relasjoner til andre personer – med kritisk sans – til lenking, selv om gruppeinformasjon ikke nødvendigvis er konstant over tid. Dette er illustrert med de stiplede kurvene for par i figur 1.
8. Lenkingen kan endres manuelt, som regel på grunnlag av informasjon i genealogiske oversikter. Dette må imidlertid flagges og dokumenteres, og det kan ha konsekvenser for andre lenker i databasen, jf. regel nr. 10.
9. Presisjonen i protokolldata ble allment forbedret over tid. Dermed kan man legge større vekt på små forskjeller i navn, fødselsår osv. når man lenker poster med lignende informasjon omkring år 1900 enn omkring år 1800, og allikevel unngå risikoen for å lenke dubletter.
10. Noen lenker vil bli oppløst fordi vi legger til nye kilder, forbedrer algoritmene eller sjekker mot genealogier. Det kan potensielt føre til mange endringer, som kjedereaksjon i databasen, men vil som oftest ha mer avgrensede konsekvenser. Allikevel må lenkene i databasen med jevne mellomrom kontrolleres med detaljerte datarutiner.

Noen norske og danske studier er relevante her fordi de lenker historiske individdata fra bysamfunn. Frode Ulvunds studie av innsatte i tvangsarbeids- og straffeanstalter i Bergen ca. 1850–75 behandler en ressurs svak gruppe, men samtidig personer som ble grundigere protokollført enn gjennomsnittsborgeren. Morten Bjørndal studerte bostedsflytting i Bergen ved hjelp av folketellingene 1865, 1875, 1891 og 1900 samt kirkebøkene og ligningsprotokollene for perioden med fokus på Nordnes. Kildene ble kodet og lenket i en relasjonsdatabase. I begge studier var navn hovedlenkingskriterium,¹³ og kombinasjon av manuell navnestandardisering og Soundex-algoritmer var banebrytende og effektiv, men Soundex er mer språkspesifikk enn Jaro-Winkler-algoritmen brukt i forbindelse med denne artikkelen.¹⁴ Inge Eikeland analyserte sjømannssamfunnet i Bergen fra 1875 til 1912 under overgangen fra seil til damp med henholdsvis den statlige og den kommunale folketellinga som ankerpunkter og med sjømannsrullene som utfyllende materiale. Han foretok ingen systematisk lenking av individdata, men illustrerte statistikken med utvalgte livsløp.¹⁵ Hans Chr. Johansen er den som mest kontant har diskutert metodiske utfordringer ved lenking av kilder fra by i forbindelse med sin studie av Odense på 1700-tallet, jf. figur 1 i innledningen:

På grunn av høye migrasjonsnivåer og behov for å kombinere informasjon fra flere kilder med svake identifikatorer, møter arbeid med urbane steder i perioder før det 19. århundre alvorlige lenkingsproblemer. (...) en metode bruker hus heller enn individer som den sentrale enhet for å kombinere individdata.¹⁶

For eksempel kunne han følge personer gjennom deres tilknytning til spesifikke bedrifter, typisk svenner som giftet seg med enker og fikk mesterstatus i en håndverksbedrift. Meto-

13. Ulvund 2002; Bjørndal 2002

14. Winkler 1990

15. Eikeland 2000

16. Johansen 1999: 115

den er imidlertid mer aktuell ved manuell lenking i mikrostudier av ca. 1000 hus i Odense enn når siktemålet er et omfattende befolkningsregister. Ingen av disse studiene fra tidsrommet før 1900-tallet benytter fødselsdatoer til lenking av individdata, slik det er vanlig å gjøre ved lenking av innførsler i svenske kirkebøker.¹⁷

Praktiske eksempler med lenking av folketellingene 1910 og 1920

Disse to folketellingene var som nevnt de første som spurte etter fødselsdato for hele befolkningen. I 1891 og 1900 ble datoen bare etterspurt for barn under to år. En bakgrunn for dette tillegget var at navneinformasjonen i folketellingene var blitt mer ustabil, siden kvinner flest nå skiftet etternavn når de giftet seg, og mange forlot skikken med patronymikon og tok andre typer slektsnavn, typisk gårdsnavnet der de bodde. Folketellinga 1910 ble transkribert i samarbeid mellom Registreringsentral for historiske data og Arkivverkets registreringsgrupper på grunnlag av originale folketellingsskjemaer. Kombinasjonen av profesjonelt transkriberingspersonale og kilder hvor lesbarheten ikke var svekket ved kopiering, skulle borge for høy kvalitet på arbeidet. Folketellinga 1920 er i hovedsak transkribert av frivillige som ble «betalt» med innsikt i denne kilden før klausuleringsperiodens utløp hundre år etter folketellingsdagen 3. desember 1920. De aktuelle skjemaene var skannet og lagt ut i Digitalarkivet med adgang for godkjente brukere. Fra desember 2020 er adgangen offentlig. Kvaliteten på skanningen er god, men fargeforskjellene er borte, noe som gjør det vanskelig å skille mellom hva som ble skrevet i felten, og hva som er lagt til av lokale administratorer eller i Statistisk sentralbyrå. Det er da grunn til å regne med noe lavere transkriberingskvalitet i den siste av de to tellingene. Det spiller også en rolle at 1920-tellinga består av personskjemaer, som gjør det vanskeligere å tyde skrift ved jamføring av opplysningene om ulike personer enn når vi har lister med persondata, slik som i 1910. Især kan dette gi utslag i bytelling, fordi disse ikke ble utfylt av lærere, slik de stort sett ble i landkommunene, men av huseierne selv. Vi skal imidlertid huske at en systematisk oversikt over uoverensstemmelser viser at de fleste er mellom originalkildene, slik at transkriberingsfeil skjer mest som unntak.¹⁸

Eksempler på avvik i etternavn, fødested og fødselsdatoer i folketellingene 1920 og 1910 for Sandefjord

Vi skal i det følgende se på uoverensstemmelser i etternavn, fødesteder og fødselsdatoer mellom de to kildene, tre variabler som er særs relevante for lenking. Vi illustrerer metodene med eksempler fra Sandefjord, en by som i 1920 hadde vokst til 5760 innbyggere. Fødselsdatoene er til god hjelp når vi vil finne feil i andre opplysninger, men heller ikke denne opplysningen er alltid til å stole på. Av 1101 som var lenket ved hjelp av likt navn, fødestedskode og fødselsår, hadde 180 uoverensstemmelser i fødselsdatoen, hvorav 72 i de to første sifrene. For å få en «fasit» kan vi sjekke personer som var barn under to år i 1891 eller 1900 som de to folketellingene 1891 og 1900 skulle oppgi fødselsdato på. Trygve Gjertsen var født i år 1900 i Larvik. Hans fødselsdato var 1. april i 1920-tellinga, men 7. i 1910. Det siste stemmer med folketellinga 1900, hvor han som barn fikk notert fødselsdatoen. En kontroll mot 1920-tellinga i skannet versjon viser et transkriberingsavvik; med god vilje leser man 7/4. Datoene stemmer heller ikke overens for Rakel Hansen fra Sandar, men her er det overens-

17. Wisselgren et al. 2014

18. Fure 2000

stemmelse mellom 1900- og 1920-tellingene. Arkivverket har skannet 1910-tellinga i original til Digitalarkivet, og her viste det seg at Rakels fødselsdato ble ført som 13. i stedet for 14. september både i original og i transkribert versjon. Eivind Halvorsen, født år 1900 i Sandefjord, har flere mulige datoer: 6/5., 10/6. eller 6/6. i de tre folketellingene 1900–1920. I 1920 stemmer datoen med originalkilden, men problemet kan skyldes at Eivind var midlertidig fraværende i Amerika og ikke selv kunne informere folketelleren.¹⁹ For Emil Larsen, født 1899 i Tjølling, skyldes uoverensstemmelsen igjen et transkriberingsproblem. Som det framgår av figur 3, skrev folketelleren i 1920 åttetall slik at de ligner på femtall. Også i 1900- og 1910-tellingene var hans fødselsdato 28/5. 1899. Det gjør oss oppmerksomme på at Emil het Hartvigsen etter far Hartvig i 1900-tellinga, men i 1910 hadde hele familien endret etternavn til Larsen. Vi må altså være oppmerksomme på endrede etternavn og fødselsdatoer, for sistnevnte som regel bare ett element, enten dag, måned eller år.

1 desember 1920.

Skjema 1. Personseddel nr. 3

Sandefjord by. Tellingskrets nr. 8

Husliste nr. 238 Husholdningsliste nr. 2

1. Fullt navn: Emil Larsen

2. Mannkjønn¹. Kvinnekjønn¹.

3. Fødselsdag 28-5 i året 1899

4. Fødested: Tjølling Vestre

Opgi herred eller by i Norge eller fødeland utenfor Norge.

Figur 3. Eksempel på utydelige tall i folketellinga 1920 for Sandefjord.

I 1920 hadde endring av etternavn ved giftermålet blitt en norm for de fleste kvinner. Ida Mathilde, født i Sem i 1886, var et ekstremt tilfelle. Hun opplevde å ha fire ulike etternavn, ifølge folketellingene fra 1891 til 1920. Hun var først ført inn som Hansdatter i 1891, så som Hansen i 1900. Etter å ha giftet seg i 1910 ble etternavnet Mikkelsen, men dette var endret til Thorvaldsen i 1920-tellinga. Det siste skiftet skyldtes ikke gjengifte. Familietradisjonen forteller at mannens transportfirma konkurrerte med en annen transportør Mikkelsen, og han fant det derfor opportunt å skifte etternavn til et patronym laget med utgangspunkt i hans far Thorvald Mikkelsen. Skilsmisser var sjeldne, men fordi mange enker giftet seg på nytt, var Ida Mathilde langt fra unik. Vi kan saktens spørre hva det betød for kvinnenens identitet å skifte etternavn flere ganger gjennom livet. Ved å lenke kvinnene ved hjelp av fornavn, fødselsdato og fødestedskode kan vi se hvor uvanlig det var å beholde etternavnet. I Sandefjord i 1920 finner vi at 5 av 58 gifte kvinner hadde samme etternavn som de hadde i 1910, før ekteskapet ble inngått. Det var Ragnhild Jacobsen, Gunhilde Olsen, Hilda Marie Hansen, Aagot Hansen og Esther Sørensen, gift mellom 1911 og 1917. Samtlige var nok gift med menn som hadde samme etternavn i utgangspunktet. Det er tvilsomt at noen av dem hadde tatt hennes etternavn.

19. Ikke alle skannede lister fra 1910-tellinga er ennå ordnet sekvensielt, og det har derfor ikke lyktes oss å finne originalversjonen for alle de kontrollerte personene. Riksarkivet, Digitalarkivet. Folketellingene 1910 for Sandefjord kjøpstad. Skannet utgave. <https://media.digitalarkivet.no/ft/contents/36489>.

1. Fullt navn: *Ole Christian Christoffersen*

2. Mannkjønn¹. Kvinnekjøn¹.

3. Fødselsdag *15-7* i året *1848*

4. Fødested: *Westad Veøy* } Opgi herred eller by i Norge
eller fødeland utenfor Norge.

Figur 4. Selv om fargene er borte, avslører skriften bruk av kopiblyant hos SSB.

Fødestedene ble i denne perioden som regel oppgitt på kommunenivå; prestegjeldet var forlatt som administrativ enhet i folketellingene på 1800-tallet. De mange kommuneendringene skaper forvirring, men i første halvdel av 1900-tallet var oppsplitting hovedregelen, og mindre enheter gir større presisjon. Som vi så ovenfor, ble alle personposter utstyrt med en firesifret kommunekode.²⁰ Ofte er fylke føyet til for å unngå forveksling mellom tvetydige kommunenavn. At bare et lokalt navn er angitt, forekommer, og det skaper problemer. Ole Kristian Kristoffersen, født 1848 i Sandar, var det enkelt å lenke fra 1920- til 1910-tellinga ved hjelp av konsistent navn og fødselsdato, men fødestedet i førstnevnte kilde skapte tvil. Han var angivelig født i Veøy på Nordmøre. Dette var ingen transkriberingsfeil; derimot hadde Statistisk sentralbyrås standardisering av opplysningene tolket gårdsnavnet «Westad» dithen og ikke tatt hensyn til at samme gårdsnavn fins i Sandar (se figur 4). En mer naturlig feiltolkning fra SSBs side gjelder Kvelle sogn, som det er nærliggende å plassere i *Hedrum* kommune, mens tidligere folketellinger plasserte Augusta Larsen som født i 1857 i Kville i Sverige. I 1920 var hennes svenske ektemann død. Hun oppga ikke lenger svensk statsborgerskap til folketelleren; som «frihandlerske» var det kanskje greiere å framstå mer norsk?

Et komplisert tilfelle er Gustav Alfred Olsen, ifølge folketellinga 1920 født 19/10. 1871 i «V. Aker». Han var da fraværende som slakter ved en hvalstasjon på Syd-Georgia, og det er som nevnt vanskelig å få nøyaktige opplysninger om fraværende personer. Konsistente navn og fødselsdatoer rekker til å lenke ham til 1910-tellinga, men her ble fødestedet angitt som Fredrikshald. Det har ikke lyktes oss å kontrollere dette mot folketellinga 1900; han kan ha forsvunnet midlertidig fra landet. Derimot er det en interessant, men usikker lenkingskandidat i 1875-tellinga for *Aker*, hvor det fins en Gustav *Adolf* Olsen født 1864 i *Fredrikshald*. Ved søking i Digitalarkivet dukker en sikrere kandidat opp. Pøsemager Gustav Alfred Olsen født 1871 i «Vorter Aker» var barnefar i Sandefjord i 1920. Inspeksjon av den skannede utgaven av kirkeboka²¹ viser «Vestre Aker», men skriften er utydelig, og det må ha vært vanskelig for dem som transkriberte disse kirkebøkene for amerikanske genealogifirmaer i India, å tyde dette (se figur 5). Kirkebøkene blir ofte skannet fra mikrofilmkopier, noe som gir dårligere lesbarhet enn den direkte skanningen av de originale folketellingskjemaene i 1910-tellinga.

<i>Pøsemager Gustav Alfred Olsen</i>		1871	Vorter Aker.
<i>de Fredrikshald Olsen, Sandefjord</i>	<i>Født 19/10</i>	1871	Vorte

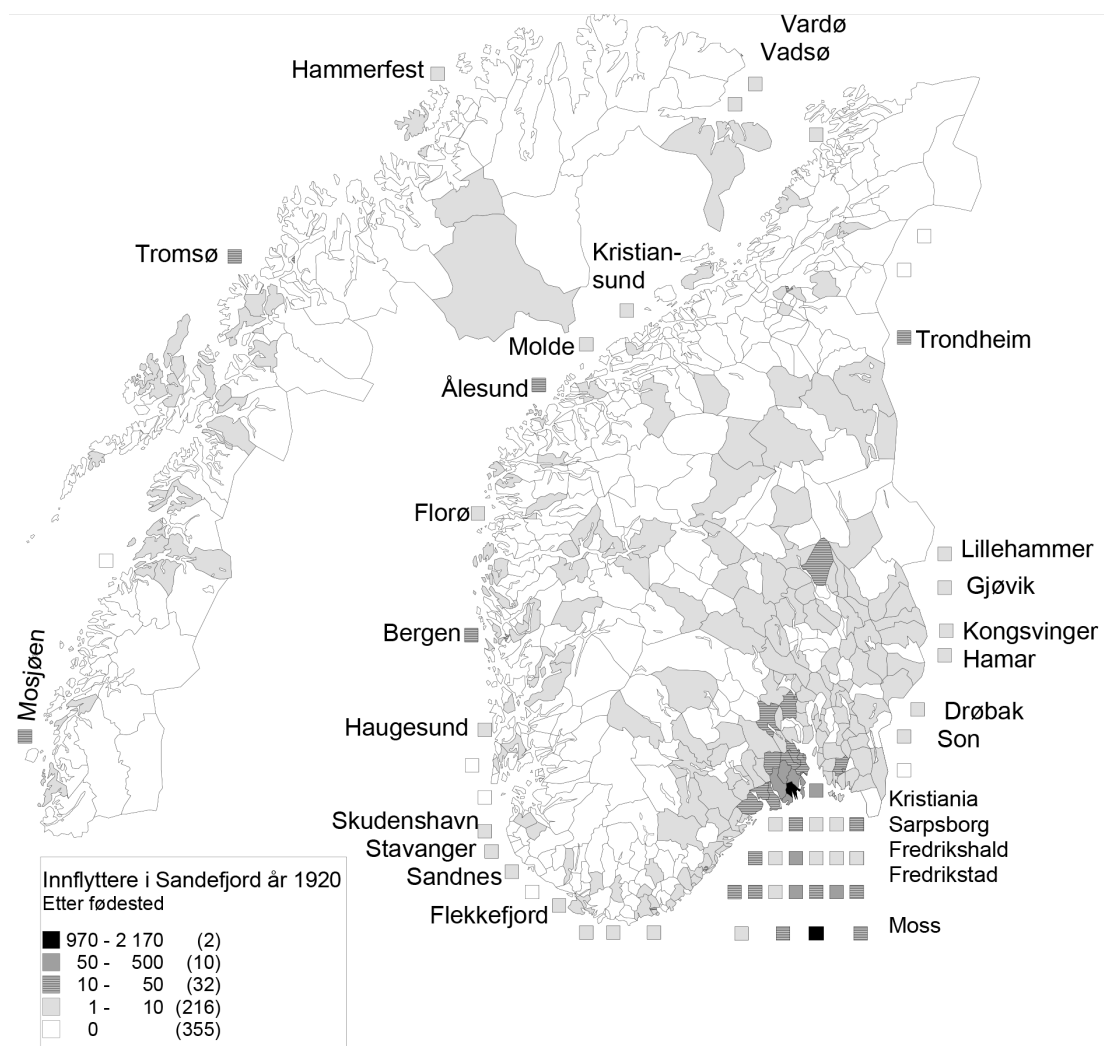
Figur 5. Fødestedet er Vestre Aker, ikke Vorter Aker.

20. Sandefjord er kodet som første kommune i 1920-tellinga, mens folketellingene 1865, 1875, 1900 og 1910 er kodet for hele landet ved Registreringsentral for historiske data, i hovedsak av Marianne Erikstad, se Thorvaldsen 1995.

21. Statsarkivet i Kongsberg, Fødselsregister for Sandefjord prestegjeld 1916-32 (0706Q). Født og døpte 1920, side-nummer 63, linjenummer 40

Innbyggernes fødested

Standardiseringa av fødested gir grunnlag for å måle andelen innflyttere i befolkninga på tidlig 1900-tall. De 1083 ulike betegnelse av fødested i Sandefjord ble standardisert til 294 ulike administrative enheter etter at ulikheter i skrivemåte og andre detaljer, som angivelse av gårdsnavn, var nøytralisert. I Norge er enheten kommune, mens den for fødte i utlandet er fødeland. Tabell 2 gir en grov oversikt på fylkesnivå og for regioner i utlandet. I samsvar med det heterogene flyttemønsteret ser vi at alle fylkene er representert. Tilbakevandringen fra Amerika sammen med barn født der var liten. Selv om nasjonale emigranttall på denne tida nærmet seg millionen, er dette ikke uventet, i og med at Vestfold hadde lave utvandringstall.²² Med omsyn til migrasjon var sandefjordingene altså ikke verdensborgere, slik det antydes i tittelen på bind I av byhistorien: *Strandsitter og verdensborger*.²³ I stedet dominerte immigrasjonen fra Sverige, med 111 menn og 78 kvinner født der, mens det bare var 45 med fødested i Danmark i byen.²⁴



Figur 6. Norske fødestedskommuner for innflytterne til Sandefjord ifølge folketellinga 1920.

Firkanter for byer på samme linje som henholdsvis Sarpsborg, Fredrikshald, Fredrikstad og Moss: Langesund, Horten, Åsgårdstrand, Honefoss, Sarpsborg. Skien, Stathelle, Tønsberg, Svelvik, Kongsberg, Fredrikshald. Risør, Porsgrunn, Brevik, Larvik, Holmestrand, Drammen, Fredrikstad. Tvedestrand, Kragerø, Sandefjord, Moss. De tre sørligste firkantene: Farsund, Mandal, Lillesand. Intervallene er fra og med, men ikke inklusive tallet til høyre.

22. Eide & Thorvaldsen 2011

23. Olstad 1995

24. Mye kan skrives for å plassere de norske folketellingene i en internasjonal sammenheng, se Thorvaldsen 2018.

Migrasjon fra steder i nærheten er et allment mønster (se især svensk forskning).²⁵ Innenlands vises flyttemønstret bedre ved å skravere kommunekart enn i tabellform. Kartet i figur 6 viser en klar konsentrasjon av opprinnelsessteder på Østlandet, især i nærområdene i Vestfold, men også fra områdene lenger vest, nord og øst. Flyttinga til Sandefjord fra sørlandskommunene langs kysten var imidlertid liten til tross for gode kommunikasjoner. I tillegg til nærflytting ser vi at det flyttemønstret fra innland til kyst som Sølvi Sogner fant allerede fra 1700-tallet,²⁶ stadig var aktuelt i mellomkrigstida. Byens tiltrekningskraft på innflytterne må forstås på bakgrunn av dens allsidige næringsliv.²⁷

Befolkningspyramiden basert på 1920-tellinga i figur 7 viser store variasjoner mellom aldersgruppene. Som forventet var overskuddet av gutter betydelig blant barn under fem år, men allerede i neste aldersgruppe var dette snudd til kvinneoverskudd, noe som vedvarte i alle eldre aldersgrupper bortsett fra i gruppa 10–14 år.²⁸ Fra konfirmasjons- til giftermålsalderen dominerte innflytterne befolkningsveksten især blant kvinner. Yrkesrubrikken viser at mange unge kvinner arbeidet som tjenestejente i byen i 1920. En mer detaljert framstilling av yrkes- og husholdsstruktur må av ressursgrunner vente til disse rubrikkene er standardisert og kodet i 1920-tellinga.

Tabell 2. Fødestedet til innbyggerne i Sandefjord ifølge folketellinga 1920 fordelt på fylker, regioner i utlandet og kjønn samt destinasjoner for utflytterne ifølge folketellinga 1910. I alle figurer og tabeller: K = kvinner, M = menn.

Kode	Fylke/Region	Fødested	K	M	Destinasjon	K	M
01	Østfold	136	74	62	70	33	37
02	Akershus	48	23	25	129	83	46
03	Oslo	195	90	105	550	354	196
04	Hedmark	58	31	27	15	11	4
05	Oppland	32	14	18	15	9	6
06	Buskerud	149	91	58	68	44	24
07	Vestfold	4353	2365	1988	951	495	456
08	Telemark	144	69	75	115	61	54
09	Aust-Agder	88	57	31	22	14	8
10	Vest-Agder	49	29	20	21	11	10
11	Rogaland	20	7	13	17	9	8
12	Hordaland	22	10	12	13	10	3
13	Bergen	20	9	11	14	8	6
14	Sogn og Fjordane	11	4	7	4	3	1
15	Møre og Romsdal	33	9	24	9	6	3
16	Sør-Trøndelag	25	16	9	20	11	9

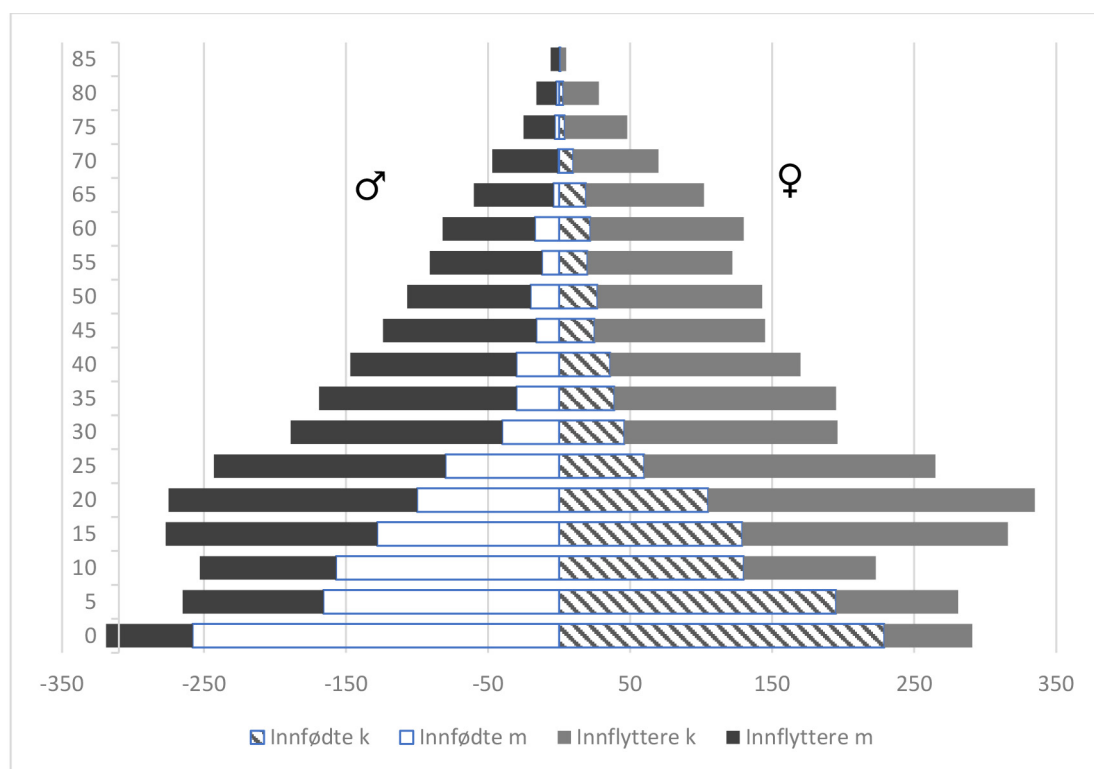
25. Hägerstrand 1947, 1953

26. Sogner 1979

27. Olstad 1995

28. Kommende forskning kan vise om dette skyldes et problem med mangler i 1920-tellinga, eller om det kan skyldes diskriminering av pikebarn i den vanskelige perioden under første verdenskrig og spanskesyken, Szoltysek et al. 2022

Kode	Fylke/Region	Fødested	K	M	Destinasjon	K	M
17	Nord-Trøndelag	5	1	4	8	4	4
18	Nordland	34	13	21	8	3	5
19	Troms	18	12	6	0	0	0
20	Finnmark	11	6	5	2	2	0
30	Amerika	36	18	18	2051	1171	880
40	Norden	241	99	142			
41–46	Resten av Europa	26	17	9			
52	Resten av verden	3	0	3			
99	Ukjent	5	3	2			
Sum		5762	3067	2695			

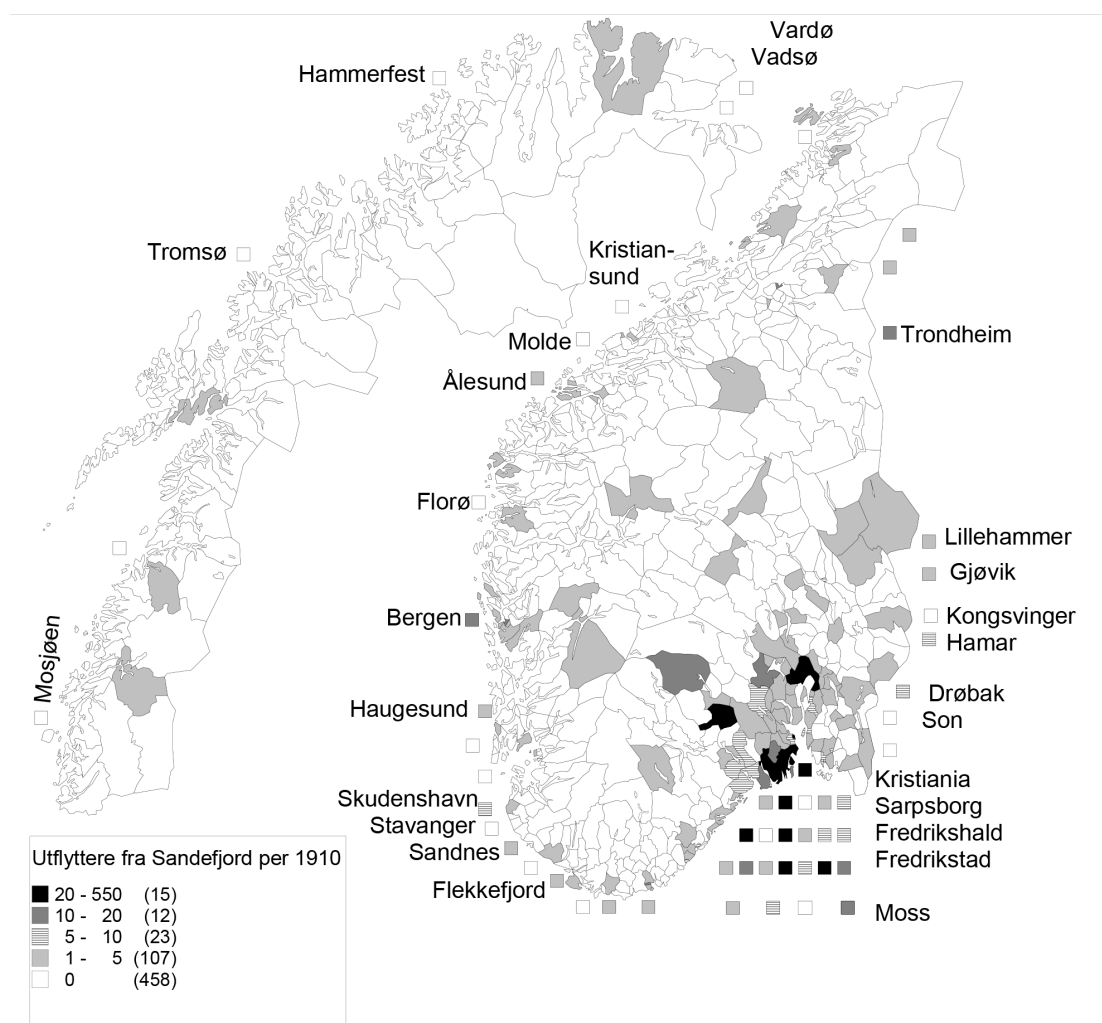


Figur 7. 5-årsaldersgrupper i Sandefjord i 1920 fordelt på kjønn og fødested ifølge folketellinga. Aldersgrupper fra 0–4 osv.

Migrasjon fra Sandefjord

Folketellinga 1920 er som nevnt ennå ikke standardisert med koder for fødested, yrke og stilling i husholdning. Tilgang til data fra 1920-tellinga for Aust-Agder fylke setter oss i stand til å slå fast at antall innbyggere her født i Sandefjord og Sandar kun var henholdsvis 56 og 6, altså enda lavere enn migrasjonen nordover til hvalfangerbyen. Den omfattende emigrasjonen fra Sørlandet antyder at mangel på levebrød der var grunnen. Detaljerte analyser av utflyttinga må heller bygge på den fullstendige folketellinga fra 1910, hvor kodearbeidet er fullført, og hvor destinasjoner er fordelt på fylker i tabell 2. Også utflytterne viser

stor variasjon i sine flyttemål. Alle fylker unntatt Troms er representert. Kvinnene domierte også utflyttinga og var i flertall i nesten alle fylkene. At en stor andel var nærflyttere til andre kommuner i Vestfold, er som forventet, men på de neste plassene kom Kristiania og Akershus og ikke nabofylkene. En spesiell flyttestrøm gikk fra Sandefjord og i retning Bergen (se kartet i figur 8). Vi kunne forventet at det dreide seg om sjøfolk som søkte hyre på Vestlandet. Det viser seg imidlertid at det blant de 20 sandefjordfødte innbyggerne i Bergen bare var to sjømenn, mens de andre arbeidet som lærer, på kontor, med servering eller som ingeniør. Selv om urbane destinasjoner dominerte, var utflyttinga ikke bare urban. Det var også sandefjordinger i kommuner på veien til Bergen gjennom Telemark og ved Hardangerfjorden, i Heddal, Tinn og Kinsarvik. I tillegg til familiemedlemmer var disse syssel-satt som rørlegger, anleggsarbeider, baker, fabrikkarbeider og til og med en «Modepynterske» – altså ikke typiske rurale yrker. Noe tilsvarende flyttemønster til nordenfjelske områder fins ikke. Her utgjorde Trysil, Åmot og Elverum den nordligste kommuneklyngen. Det blir spennende å følge utviklingen av byrker blant migrantene når hele 1920-tellinga er klar til analyse.



Figur 8. Norske destinasjonskommuner for utflyttere fra Sandefjord ifølge folketellinga 1910.

Longitudinelle flytتهistorier

Lenkingen beskrevet ovenfor setter oss i stand til å beskrive mange av migrantenes bakgrunn, fordi de lettere kunne identifiseres i folketellingene 1910 og 1920. For å følge repre-

sentanter for migrantene til og fra Sandefjord over tid skal vi ta for oss noen individer som det har lyktes oss å følge ved hjelp av automatisk lenking i de tre tilgjengelige folketellingene på 1900-tallet. Edel Abrahamsen var født i 1863 i Kvelde sogn i Hedrum prestegjeld, ifølge 1865-tellinga som datter av Anders Jacobsen og Marte Mathisdatter. Hun kan representere de 2187 innflytterne til Sandefjord fra andre kommuner i Vestfold. 266 kom fra andre byer i fylket, mens 1921 kom fra landkommuner, hvorav 970 fra nærområdene i Sandar. I motsetning til et par av søsknene bodde hun imidlertid ikke sammen med faren i 1875, da han var blitt enkemann. Ifølge 1900-tellinga hadde hun flyttet til Sandar, men bodde ifølge 1910- og 1920-tellingene i Sandefjord, altså urban etappevandring via rural mellomstasjon. Hennes bosted i år 1900 var typisk nok nordvest for byen, dvs. på vei fra Hedrum. Hun var enke allerede i år 1900 og er derfor ikke så enkel å følge, men bodde sammen med en sønn født i 1894. I 1910 tok hun inn losjerende og var opptatt med «huslige gjøremål», mens hun i 1920 var badekone – hun bodde i Bjerggata, kort vei fra Sandefjord badeanstalt.²⁹ Hennes livsløp illustrerer behovet for lenking til data fra kirkebøkene, som må inneholde hennes vielse og barnefødsler. Så langt ser vi bare på grunnlag av sønnens patronym at hennes mann het Nils – nok et eksempel på nytten av parrelasjoner ved lenking.

Ragnvald Eugen Støkken kan representere de 184 innflytterne fra de to fylkene øst for Oslofjorden. Født i 1883 i Drøbak hadde han en broket etappeflyttingshistorie. Han hadde bosatt seg i Sandefjord før 1. desember 1900, men i 1910-tellinga var han i Sandar, for så å bli registrert i byen igjen i 1920. I år 1900 var han kontorist på et skipsrederkontor og hjemmевærende sønn i en Støkken-familie med syv barn i tillegg til mor og skipperfar. Fødestedene forteller om en ambulerende byfamilie. Foreldrene var fra Vestby og Ås i Akershus og hadde fått barn i Kristiania og Larvik i tillegg til i Drøbak. Bare tjenestejenta hadde rural bakgrunn, fra Kodal i Andebu kommune. I 1910 var Ragnvald Støkken nygift bankassistent uten barn, men i 1920 finner vi ham som «Aktiemægler» etter sigende med erfaring fra jobbetida under første verdenskrig. Han var siden 1910 gift med jevngamle Margaret, født i Fredrikstad, som hadde emigrert i 1905 og returnert i 1908. Hun kan sammen med Maren Sofie Olsen nedenfor representere de 41 kvinnene og de 73 mennene som i 1920 var hjemvendt norsk-amerikaner i Sandefjord.³⁰

Klinker på Framnes mekaniske verksted Ole Olsen var født i Amerika i 1889 og kan representere de 38 immigrantene derifra, hvorav 34 var fra USA. Han bodde i år 1900 i Sandar, men i Sandefjord i 1910 og i 1920. Da bodde han som enebarn sammen med sin mor, enke og «Flaskeskyllerske» Maren Sofie Olsen, som ifølge 1920-tellinga hadde vært i Amerika fra 1888 til 1896. I år 1900 drev hun jordbruk sammen med sin mann bygningssnekker Karl Olsen, som døde før 1910. Dette er eksempler på personer som lar seg lenke sikkert som par ut fra familierelasjoner, selv om de har vanlige navn, jf. de ulike kurvene i figur 1. Representanten for den større gruppa med 189 immigranter fra Sverige kan være Anders Fredrik Johannesen, født i 1853 i Krokstad sogn i Sverige, mellom Strömstad og Uddevalla. Han bodde i byen ifølge alle tre folketellingene først på 1900-tallet og var gift med Anna Marie med samme fødested, men ti år eldre. I år 1900 hadde de to hjemmевærende sønner og en datter. Guttene hadde forlatt redet i 1910, mens dattera fortsatt bodde hos foreldrene i 1920 og sikkert var til god hjelp for mora, som var svekket på grunn av alderdom. Da var hennes fødested oppgitt som identisk med foreldrenes, mens det i tidligere tellinger sto Horten. Enten en feilførsel eller et forsøk på å dekke over at hun var født utenfor ekteskap?

29. Badet førte til at mange badegjester oppholdt seg midlertidig i byen, men dette avspeiles i liten grad i folketellingene, som ble arrangert vinterstid.

30. Folketellingen 1920, fjerde hefte, tabell 14: Hjemvendte Norsk-Amerikanere fordelt etter bosteder 1 desember 1920, side 107.

Oppsummering

I og med Arkivverkets publisering av folketellinga 1920 i Digitalarkivet har perioden da vi kan studere Norges befolkning i detalj med åpne kilder, nådd mellomkrigstida. Mer spesielt tar denne artikkelen for seg muligheter, problemer og løsninger ved lenking av de nyeste allment tilgjengelige folketellingene på 1900-tallet især for å identifisere migranter. 1910- og 1920 tellingene er det eneste paret av åpne folketellinger som inneholder fødselsdato på alle. Dette gir nye muligheter for effektiv lenking av individdata. I tillegg til å være mer effektiv vil automatisk eller maskinell lenking ved hjelp av dataprogrammer sikre mer enhetlig håndtering av kildene, men dette kan introdusere feil på grunn av manglende konsistens i kildene og mangel på unike opplysninger om mange individer, især i byene. Heldigvis gjorde det mer eksakte innholdet i dansk og norsk kildemateriale fra annen halvdel av 1800-tallet det lettere å finne et kompromiss mellom hensynet til mer komplett lenking og ønsket om å unngå feillenker. Introduksjonen av fødselsdatoer i folketellingene fra og med 1910 er enda et skritt i denne retningen. Hovedpunktene i metodene ble oppsummert i ti regler om lenking av historiske individdata, mens detaljer kan studeres i den rikholdige litteraturen om lenking av historiske individdata.³¹

Standardisering og koding av opplysningene i folketellingene er nødvendig både for å produsere mer enhetlig statistikk og for mer rasjonell lenking på individnivå fra kilde til kilde. I den publiserte befolkningsstatistikken er sammenligning over tid ofte vanskelig, fordi kategoriene i folketellingene ble revidert, for eksempel yrkesgrupperinger og administrative grenser. Også for å unngå økologiske feilslutninger ønsker vi å legge til rette for å bygge analysene fra individnivå ved å standardisere og kode navn, fødested og andre variabler og lenke sammen informasjonen til kollektive biografier som gjør det mulig å foreta longitudinelle analyser. De nye mulighetene til å jamføre innførsler fordi både 1910- og 1920-tellingene har fødselsdato, forsterker også det kildekritiske potensialet. Kildene inneholder mange motstridende detaljer som påvirker lenkingen og analysene. Transkribering fra kilde til datafil, især i utlandet, hvor språkproblemer legger sten til byrden, fører i tillegg til uøyaktigheter. De kildekritiske mulighetene i et befolkningsregister som kombinerer data fra flere kilder, gir grunnlag for å estimere at fødselsdatoene har en nøyaktighet på ca. 90 %. Spesielt skaper opplysningene om kvinner som skiftet etternavn, og fraværende personer uoverensstemmelser. Også feiltolkning av kildene av statistikkpersonale uten lokal kunnskap kan virke inn.

Folketellingene 1920 og 1910 ble brukt til en analyse av migrasjon til og fra småbyen Sandefjord. Som ventet dominerte kvinnene i de fleste aldersklassene. Lokalt var flytting fra og til rurale steder viktigst, men over lengre avstander var flyttinga mer urban, og alle fylkene var representert som fødesteder, mens bare Troms fylke manglet som destinasjon. Spesielt interessant er et urbant flyttemønster i retning Bergen. Imidlertid var migrasjonen fra Sandefjord bemerkelsesverdig lite internasjonal med omsyn til befolkningens store engasjement i global sjøfart og fangst. Noen utvalgte migranthistorier illustrerer kompleksiteten i oppbyggingen av et longitudinelt befolkningsregister som avspeiler både hendelser i enkeltindividenes livsløp og dynamiske relasjoner mellom personene.

Takk til Britt Gabrielsen for transkribering av det meste av folketellinga 1920 for Sandefjord som frivillig, til seniorrådgiver Lars Nygaard i Riksarkivet for tilgang på folketellinga 1920 for Agder og Vestfold og til to anonyme fagfeller og andre som har lest og kommentert. For kodingen av 1920-tellinga takker jeg meg selv.

31. En lengre historiografisk framstilling av metoder utviklet for behandling av norske historiske individdata er under publisering, se Thorvaldsen & Holden 2023.

Litteratur

- Riksarkivet, Digitalarkivet. Folketellingen 1910 for Sandefjord kjøpstad. Skannet utgave. <https://media.digitalarkivet.no/ft/contents/36489>
- Riksarkivet, Digitalarkivet. Folketellingen 1920 for Sandefjord kjøpstad. Skannet utgave. <https://media.digitalarkivet.no/ft/contents/73803?censuses%5B0%5D=18&municipalities%5B0%5D=0706&text=>
- Statistisk sentralbyrå (1923). Norges offisielle statistikk VII. 81. Folketellingen i Norge 1 desember 1920. Fjerde hefte. <https://www.ssb.no/historisk-statistikk/folketellinger/folketellingen-1920>
- Statsarkivet i Kongsberg, Fødselsregister for Sandefjord prestegjeld 1916-32 (0706Q). <https://media.digitalarkivet.no/kb/contents/4795R>

Sekundærlitteratur

- Alhaug, G. (2005). Støtte frå Noregs forskingsråd til standardisering av namn i folketeljningane. *Nytt om namn*, 41(2005), 47–49.
- Bjørndal, M. (2002). *Bostedsflytting i Bergen 1865–1900: Kartlegging av det interne flyttemønsteret* [Hovedfagsoppgave]. Universitetet i Bergen.
- Brosveet, J. (1979). *Kommuneendringer 1838–1978*. Norsk samfunnsvitenskapelig datatjeneste.
- Eide, O. J. & Thorvaldsen, G. (2011). Historisk befolkningsregister: Oversikt over emigrasjonen. *Heimen*, 48(3), 207–228.
- Eikeland, I. (2000). *Fra seil til damp: Sjømannssamfunnet i endring, Bergen 1875–1912* [Hovedfagsoppgave]. Universitetet i Bergen.
- Fure, E. (1990). Personnavn og tidsånd. *Namn og Nemne*, 7, 35–55.
- Fure, E. (2000). Interactive Record Linkage. The Cumulative Construction of Life Courses. *Demographic Research*, 3. 10.4054/DemRes.2000.3.11
- Hägerstrand, T. (1947). En landsbygdsbefolknings flyttningsrørelser. Studier över migrationen på grundval av Asby sockens flyttningslängder 1840–1944. I *Svensk geografisk årbok*, 114–142.
- Hägerstrand, T. (1953). *Innovationsförloppet ur korologisk synpunkt*. Gleerupska universitetsbokhandeln.
- Johansen, H. C. (1999). Urban social and demographic reconstitution. The case of eighteenth century Odense. *History and Computing*, 11(1–2), 115–128. <https://doi.org/10.3366/hac.1999.11.1-2.115>
- Langholm, S. (1976). On the scope of micro-history. *Scandinavian journal of history*, 1(1) 1–24. <https://doi.org/10.1080/03468757608578894>
- Olstad, F. (1995). *Sandefjords historie: Strandsitter og verdensborger* (Bd. 1). Sandefjord kommune.
- Sogner, S. (1979). *Folkevekst og flytting. En historisk demografisk studie i 1700-årenes Øst-Norge*. Universitetsforlaget.
- Szołtysek, M., Ogórek, B., Gruber, S. & Beltrán Tapia, F. J. (2022). Inferring «missing girls» from child sex ratios in historical census data. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 55(2), 98–121. <https://doi.org/10.1080/01615440.2021.2014377>
- Thorvaldsen, G. (2018). *Censuses and Census Takers: A Global History*. Routledge. <https://doi.org/10.4324/9781315148502>
- Thorvaldsen, G. & Holden, L. (2023). The development of microhistorical databases in Norway. A historiography. *Historical Life Course Studies* 13(5):127–147. <https://doi.org/10.51964/hlcs14315>
- Ulvund, F. (2002). «Byens udskud»? Ein studie av karrierar som innsett i tvangsarbeids- og straffanstalt i Bergen ca. 1850–75 [Doktorgradsavhandling]. Universitetet i Bergen.
- Vick, R. & Huynh, L. (2011). The Effects of Standardizing Names for Record Linkage: Evidence from the United States and Norway. *Historical Methods*, 44(1), 15–24. <https://doi.org/10.1080/01615440.2010.514849>
- Winkler, W. E. (1990). String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage 354–359. Konferanseinnlegg fra *Proceedings of the Section on Survey Research Methods*. American Statistical Association.

Litteratur om lenking av historiske individdata

- Bloothoof, G., Christen, P., Mandemakers, K. & Schraagen, M. (Red.). (2015). *Population Reconstruction*. Springer. <https://doi.org/10.1007/978-3-319-19884-2>
- Christen, P. (2012). *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Springer. https://doi.org/10.1007/978-3-642-31164-2_2
- Gutmann, M. P. (Red.). (1992). *Historical Methods: Theme issue on record linkage*. 55 (2). <https://www.tandfonline.com/toc/vhim20/53/2>
- Helgertz, J., Price, J., Wellington, J., Thompson, K. J., Ruggles, S. & Fitch, C. A. (2022). A new strategy for linking U.S. historical censuses: A case study for the IPUMS multigenerational longitudinal panel. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 55(1), 12–29. <https://doi.org/10.1080/01615440.2021.1985027>
- Holden, L., Boudko, S. & Thorvaldsen, G. (2020). Lenking og kobling i Historisk befolkningsregister. *Heimen*, 57(3), 216–229. <https://doi.org/10.18261/issn1894-3195-2020-03-04>
- Johansen, H. C. (2002). Identifying people in the Danish past. I H. Sandvik, K. Telste & G. Thorvaldsen (Red.), *Pathways of the past: Essays in honour of Sølvi Sogner on her 70th anniversary 15 March 2002* (s. 103–110). Universitetet i Oslo.
- Katz, M. & Tiller, J. (1972). Record-Linkage for Everyman: A Semi-Automated Process. *Historical Methods Newsletter*, 5, 144–150.
- Miller, R. & Thorvaldsen, G. (1997). Beyond Record linkage: Longitudinal Analysis of Turn-of-the-century Inter-urban Swedish Migrants. *History and Computing*, 9, 106–121. <https://doi.org/10.3366/hac.1997.9.1-3.106>
- Nygaard, L. (1992). Name standardization in record linkage: an improved algorithmic strategy. *History and Computing*, 4(2), 63–74.
- Pouyez, C., Roy, R. & Martin, F. (1981). The Linkage of Census Name Data: Problems and Procedures. *Journal of Interdisciplinary History*, XIV(1), 129–152.
- Ruggles, S., Fitch, C. & Roberts, E. (2018). Historical Census Record Linkage. *Annual Review of Sociology*, 44, 19–37. <https://doi.org/10.1146/annurev-soc-073117-041447>
- Thorvaldsen, G. (1995). The Encoding of Highly Structured Historical Sources. *Computers and the Humanities*, 28(4/5), 301–305.
- Thorvaldsen, G. (2008). Fra folketellinger og kirkebøker til norsk befolkningsregister. *Heimen*, 45(4), 341–359.
- Thorvaldsen, G. (2011). Using NAPP census data to construct the Historical Population Register for Norway. *Historical Methods, A Journal of Quantitative and Interdisciplinary History*, 44(1), 37–47. <https://doi.org/10.1080/01615440.2010.517470>
- Thorvaldsen, G., Andersen, T. & Sommersteth, H. L. (2015). Record Linkage in the Historical Population Register for Norway. I G. Bloothoof, P. Christen, K. Mandemakers & M. Schraagen (Red.), *Population Reconstruction* (s. 155–170). Springer.
- Wisselgren, M. J., Edvinsson, S., Berggren, M. & Larsson, M. (2014). Testing Methods of Record Linkage on Swedish Censuses. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 47(3), 138–151. <https://doi.org/10.1080/01615440.2014.913967>
- Wrigley, E. A. (1973). (Red.). *Identifying people in the past*. Edward Arnold.
- Wrigley, E. A. & Scofield, R. S. (1973). Nominal record linkage by computer and the logic of family reconstitution. I E. A. Wrigley (Red.), *Identifying People in the Past*. Edward Arnold, 64–101.