

Arbeidsnotater

S T A T I S T I S K S E N T R A L B Y R Å

Dronningensgt. 16, Oslo-Dep., Oslo 1. Tlf. 41 38 20, 41 36 60

IO 71/21

13. november 1971

FORSØK MED MASKINELL OVERFØRING AV PERSONNUMMER TIL SOSIALHJELPSTATISTIKKEN

Av

Ole Førreisdahl og Jon Holmøy

I N N H O L D

	Side
1. Innledning	2
2. Formål og opplegg for forsøksoverføringen ..	2
A. Formål	2
B. Opplegg	3
3. Forsøksmaterialet	5
4. Forsøksoverføringen. Resultater	5
5. Forsøk med sikte på å ivareta flere person- nummeropplysninger	9
6. Hva oppnås ved maskinell overføring	12
7. Lister	16
8. Konklusjon	17
9. Diagrammer	18
10. Vedlegg I. Oversiktsdiagram	19

1. Innledning

I prosjektbeskrivelse av 18. september 1970 la 7. kontor fram en plan for innføring av fødselsnummer i sosialhjelpstatistikken. Planen gikk bl.a. ut på i én operasjon å overføre personnummer til årgangene 1966-1969. For å redusere prosjektets behov for manuell arbeidskraft ble det foreslått at arbeidet så langt som mulig burde utføres maskinelt. På prosjektmøtet for 7. kontor 18. november 1970 ble prosjektet vedtatt satt i verk.

I Byrået har vi lite erfaring med maskinell overføring av personnummer etter et slikt opplegg som 7. kontor har foreslått for denne statistikken.¹⁾

En fant således at før prosjektet ble satt ut i livet burde det gjennomføres forsøk, bl.a. for å finne fram til et hensiktsmessig kjennetegnsett (nøkkel) å bassere overføringen på. Slike forsøk ville trolig også gi svar på en rekke andre spørsmål av betydning for prosjektet.

2. Formål og opplegg for forsøksoverføringen

A. Formål

Hovedhensikten bak forslaget om maskinell overføring var først og fremst å redusere det arbeidskraftbehovet som ville oppstå dersom overføring av personnummer skulle skje manuelt. Hovedformålet med undersøkelsen blir

1) Med utgangspunkt i det materialet som ble innhentet i samband med jordbrukstillingen i 1969 opprettet Byrået et bruksregister. I stedet for å bruke de opplysningene jordbrukstillingen inneholdt om brukernes navn og adresse så overførte en disse opplysningene maskinelt fra personregisteret. Dette lot seg gjøre fordi en hadde opplysninger om brukernes fødselsnummer.

Fra Landbrukets Sentralforbunds produsentregister ble produsentnummer forsøkt overført til Byråets bruksregister. Her ble det testet bare på navn, men med krav om full overensstemmelse i alle posisjoner. Oppnådd resultat: Knappt 40 prosent av enhetene i bruksregisteret fikk tildelt produsentnummer.

således å finne ut hvor mye manuell arbeidstid vi kan spare på denne måten. Den manuelle arbeidsmengde vi kan spare er naturligvis avhengig av hvor stor del av materialet som tildeles personnummer maskinelt (overføringsprosent). Vi må imidlertid regne med at de records som ikke får tildelt personnummer maskinelt er de vanskeligste og mest arbeidskrevende. Den prosentvise besparelse i arbeidstid vil således alltid bli mindre enn den maskinelle overføringsprosent.

Det er likevel klart at vi må forsøke å finne fram til en nøkkel som gir en høy overføringsprosent. Når det gjelder tallet på feiloverføringer synes det umiddelbart rimelig å anta at det eksisterer en viss sammenheng mellom dette og overføringsprosenten, slik at en høy overføringsprosent medfører høy feilprosent. Nå er det vel også slik at overføringsprosent og feilprosent ikke bare avhenger av den nøkkel som brukes, men også i sterk grad av de øvrige betingelser som kreves oppfylt før overføring kan skje. Dersom ulike nøkler skal sammenlignes bør således de øvrige overføringsbetingelser såvidt mulig være like.

Siden det er aktuelt å overføre personnummer til flere årganger samtidig, er det av interesse å undersøke hvor mange gjengangere det finnes i materialet. Dette for å anslå hvor mye det koster ekstra å ta med 1966-1968-årgangene samtidig med 1969-årgangen.

Den manuelle etterkontroll og tildeling av personnummer må trolig skje ved hjelp av en eller flere lister. Under forsøket bør det derfor også legges vekt på å finne fram til en hensiktsmessig organisering av listen(e), slik at den etterfølgende manuelle kontroll og tildeling av personnummer kan skje så lett som mulig.

B. Opplegg

Av kjennetegn det er naturlig å bruke som nøkkel ved en maskinell overføring har vi i sosialhjelpmaterialet registrert fødselsdato, navn, kjønn og bosted. De samme opplysninger er også registrert i personregisteret. I sosialhjelpstatistikken er opplysningen om bosted noe usikker. Prosjektet omfatter dessuten flere årganger, slik at f.eks. flyttinger gjør bostedsopplysninger lite egnet som nøkkelelement. Nøkkelen må således baseres på kjennetegnene; fødselsdato, navn og kjønn. Fødselsdato og kjønn er nok de sikreste opplysningene i dette materialet, og nøkkelen bør derfor bygges opp omkring disse kjennetegnene. Selv om det må regnes med atskillige feil både i fødselsdato og kjønn synes det umiddelbart tvilsomt om overføringsprosenten kan økes vesentlig ved eksperimentering med disse data. Idet det nærmest må betraktes som en nødvendig betingelse for maskinell overføring under de forhold som her gjelder, at i hvert fall fødselsdato stemmer i alle posisjoner.

Det er således bare navn som gjenstår, og som gir muligheter for eksperimenter.

Rasjonaliseringsdirektoratet har foretatt noen forsøk med maskinell overføring av personnummer¹⁾. I den nøkkel som R-direktoratet har brukt er navn representert ved en kode på 6 karakterer som er konstruert slik:

Første bokstav i etternavn	1 kar.
Soundex-kode for etternavn	3 kar.
Første 2 bokstaver i fornavn	2 kar.

Etter dette opplegget må imidlertid både mottakerfilen og registerfilen først gjennom et eget program som genererer en nøkkel for hver enkelt record. Deretter må begge filene sorteres (etter nøkkelen), før overføringen kan gjennomføres. Når så store files som personregisteret inngår i overføringsprosessen, legger denne metode beslag på atskillig maskintid. Det blir således en forholdsvis dyr framgangsmåte.

Med utgangspunkt i disse betraktninger vil vi forsøke et enklere opplegg hvor det ikke genereres noen spesiell nøkkel for hver record og hvor bare mottakerfilen må sorteres. Dette siste betyr at registerfilen forutsettes brukt i en av de allerede foreliggende sorteringer.

Opplegget går i korthet ut på at nøkkelen i tillegg til fødselsdato og kjønn skal bestå av et visst antall posisjoner av recordens navnefelt regnet forfra. Tallet på navneposisjoner vil vi variere med sikte på å finne fram til det gunstigste antall. For at overføring skal skje vil vi stille tre betingelser. For det første at det er full overensstemmelse i nøkkelen mellom sosialhjelp- og registerrecord, dernest at det bare finnes én record i registeret som har den aktuelle nøkkelverdi, og for det tredje at denne record ikke har registertype 6-9 (utgått fødselsnummer). Alle records i sosialhjelpfilen som har lik nøkkel tildeles samme personnummer.

Hva kan vi så vente som resultat av forsøket? I sosialhjelpfilen er navn registrert på samme måte som i registerfilen, med rekkefølgen etternavn, fornavn, mellomnavn og med inntil 26 bokstaver. Når denne rammen på 26 bokstaver overskrides, er det i registeret fulgt bestemte regler for forkortinger. Tilsvarende regler er ikke gjort gjeldende for sosialhjelpmaterialet. Når det gjelder navn, vet vi også at det har lett for å oppstå forskjellige typer av skrive- og stavfeil. Dersom vi således går ut fra at det er en viss sannsynlighet for forskjellige typer av feil for hver posisjon i navnet, vil det bety at sannsynligheten for feil i nøkkelen øker når vi tar med flere navneposisjoner. Disse egenskapene ved navn har altså en

reduserende effekt på overføringsprosenten når tallet på navneposisjoner i nøkkelen øker. På den annen side vil sannsynligheten for nøkkeldubletter (mer enn en redord i registerfilen som har samme nøkkel) i registerfilen avta når antall navneposisjoner i nøkkelen øker, slik at overføringen ikke hindres av den grunn. Etter dette skulle vi altså vente å finne fram til et bestemt antall navneposisjoner i nøkkelen som gir en optimal overføringsprosent. (Det er dermed ikke sagt at det ikke kan finnes nøkler konstruert etter andre prinsipper som gir en høyere overføringsprosent.) Det er trolig at feilprosenten vil synke når tallet på navneposisjoner i nøkkelen øker. Informasjonsinnholdet i nøkkelen økes etter hvert og den får således større evne til å skjelne mellom ulike records.

3. Forsøksmaterialet

Som forsøksmateriale ble valgt alle stønadmottakere som er født i 1920, og det omfatter årgangene 1966-1969. Fødselsår 1920 er ikke valgt på annet grunnlag enn at dette er ett av de år som forekommer hyppigst i sosialhjelpstatistikken i disse årgangene. I alt omfatter materialet 2 793 records og utgjør således ganske nær en femtiendedel av de totalt ca. 140 000 records.

Tallet på gjengangere, både innenfor samme statistikkårgang og fra årgang til årgang, er trolig ikke like stort i alle aldersklasser. Dette kan ha en viss betydning bl.a. for anslag over merarbeid ved å ta flere årganger samtidig. Totalt sett mener vi imidlertid at undersøkelsen, tross de ovennevnte skjevheter og materialets utvalgsteoretiske skrøpeligheter ellers¹⁾, skal kunne gi et brukbart grunnlag for å vurdere hva som kan oppnås ved en maskinell overføring etter det opplegget vi har antydnet.

4. Forsøksoverføringen. Resultater

Forsøket gikk som tidligere nevnt bl.a. ut på å undersøke hvorledes overføringsprosenten endrer seg når antall navneposisjoner i nøkkelen varierer. Prøveoverføringer ble foretatt med nøkler som inneholdt de fra 4 til 12 første navneposisjoner og dessuten med hele navnefeltet. For hvert av disse 10 alternativene ble det kjørt ut en datosortert liste som inneholdt alle records i forsøksmaterialet samt alle de registerrecords som hadde fått sitt personnummer overført til records i forsøksmaterialet. På listene var registre-recorden plassert like foran den første sosialhjelprecord som hadde den

1) Dersom vi hadde bassert undersøkelsen på et tilfeldig utvalg trukket fra hele sosialhjelpfilen hadde vi vært nødt til å søke gjennom hele personregisteret ved hver prøveoverføring. Dette ville anslagsvis ha lagt beslag på maskinen i ca. 4 timer pr. prøveoverføring. Hver prøveoverføring tok nå ca. 6 min.

tilsvarende nøkkel. Samtidig med prøveoverføringene ble det kjørt ut enkle tabeller som for hver overføring gav noen hovedtall for overføringsresultatene. Disse tallene er stilt sammen i tabell 1.

Tabell 1. Resultater ved prøveoverføringene etter tallet på navneposisjoner i nøkkelen. Records og nøkkelverdier¹⁾

Tallet på navneposisjoner i nøkkelen	Records			Nøkkelverdier ¹⁾					
	Av disse tildelt personnummer			Ikke matchet			Matchet		
	I alt	Records	Prosent	I alt	I alt	Av disse p.g.a. uoverensstemmelse i kjønn	I alt	Flere like i personregisteret	Overført
4	2 793	1 503	53,8	1 556	302	28	1 254	442	812
5	2 793	1 588	56,9	1 570	357	25	1 223	366	857
6	2 793	1 649	59,0	1 583	370	23	1 213	323	890
7	2 793	1 700	60,9	1 598	412	20	1 186	271	915
8	2 793	1 805	64,6	1 614	451	21	1 163	186	977
9	2 793	1 834	65,7	1 620	484	14	1 136	145	991
10	2 793	1 921	68,8	1 649	530	8	1 119	73	1 046
11	2 793	1 927	69,0	1 665	572	6	1 093	36	1 056
12	2 793	1 847	66,1	1 678	627	6	1 051	29	1 022
26	2 793	1 170	41,9	1 804	1 128	3	676	3	671

1. Tallet på ulike nøkler i sosialhjelpfilen
2. Inklusive nøkkelverdier som matchet, men som hadde registertype 6-9.

Vi ser at tallet på ulike nøkkelverdier øker når tallet på navneposisjoner i nøkkelen øker. Dette medfører visse vanskeligheter for tolkingen av de tallene som gjelder nøkkelverdier. Vi ser imidlertid at når vi øker tallet på navneposisjoner i nøkkelen, stiger overføringsprosenten, helt til nøkkelen har 10-11 navneposisjoner.

Av 1 665 nøkkelverdier fikk 1 056 eller 63,4 prosent tildelt personnummer med 11 navneposisjoner. Det samme prosentvise resultat ble oppnådd med 10 navneposisjoner. Når det gjelder records var det 1 927 records eller 69 prosent som fikk tildelt personnummer når nøkkelen hadde 11 navneposisjoner, mot 1 921 eller 68,8 prosent med 10 navneposisjoner. 11 posisjoner synes således å gi det gunstigste resultat. Men 10 posisjoner gir et resultat som er nesten like godt. Dersom vi øker til 12 navneposisjoner i nøkkelen,

synker overføringsprosenten betraktelig både for nøkkelverdier og redords.

Alle listene ble undersøkt manuelt med sikte på å finne feiloverføringer. Helt pålitelig blir naturlig nok ikke en slik gjennomgåelse, bl.a. fordi det er vanskelig å finne feilen i de tilfelle hele navnefeltet stemmer overens i alle posisjoner i sosialhjelp- og registerfile. Bostedsopplysningene gav i noen tilfelle indikasjon på slike feil, men på grunn av muligheten for flyttinger og dessuten noe tvil om godheten av bostedsopplysningene, var det i de fleste slike tilfelle vanskelig å avgjøre om overføringen var feil eller ikke. Resultatene av gjennomgåelsen er tabellert i tabell 2.

Tabell 2. Records med feil tildelt personnummer. (Tallene i parentes angir tallet på personer).

Tallet på navneposisjoner i nøkkelen	Feiloverføringer	Tallet på navneposisjoner som trengs i nøkkelen før overførtefeil rettes								
		5	6	7	8	9	10	11	12	13 eller flere
4	48(34)	12(7)	2(1)	6(6)	8(6)	4(4)	3(3)	6(5)	6(1)	1(1)
5	37(28)	.	2(1)	6(6)	9(7)	4(4)	3(3)	6(5)	6(1)	1(1)
6	38(30)	.	.	7(7)	10(8)	4(4)	3(3)	6(5)	7(2)	1(1)
7	33(25)	.	.	.	10(8)	5(5)	3(3)	7(6)	7(2)	1(1)
8	27(19)	6(6)	6(4)	7(6)	7(2)	1(1)
9	21(13)	6(4)	7(6)	7(2)	1(1)
10	19(12)	11(9)	7(2)	1(1)
11	8(3)	7(2)	1(1)
12	1(1)	1(1)

Når nøkkelen inneholdt 4 navneposisjoner fant vi 48 records med feil tildelt personnummer, av disse fikk 12 records rettet feilen når vi økte til 5 navneposisjoner i nøkkelen. D.v.s. at fra og med 5 navneposisjoner i nøkkelen fikk disse 12 records ikke tildelt personnummer. Fra og med 6 navneposisjoner fikk ytterligere 2 records feilen rettet osv. Bare 1 av de 48 records som fikk tildelt feil personnummer med 4 navneposisjoner trengte mer enn 12 navneposisjoner i nøkkelen før feilen ble unngått.

Med 5 navneposisjoner i nøkkelen fant vi 37 records med feil tildelt personnummer. Ovenfor så vi at av de 48 records som fikk feil tildelt personnummer med 4 navneposisjoner i nøkkelen, fikk 12 rettet feilen når nøkkelen hadde 5 navneposisjoner. Men $48 - 12 = 36$. Dette betyr at når

tallet på navneposisjoner i nøkkelen økte fra 4 til 5 var det 1 ny record som ble tildelt feil personnummer.

Tabell 3. Feilprosent etter tallet på navneposisjoner i nøkkelen

Beregningsgrunnlag	4	5	6	7	8	9	10	11	12
Records i alt	1,7	1,3	1,4	1,2	1,0	0,8	0,7	0,3	0,04
Records med tildelt p.nr.	3,2	2,3	2,3	1,9	1,5	1,1	1,0	0,4	0,05

Feilprosenten avtar med stigende antall navneposisjoner i nøkkelen. Med 11 navneposisjoner, som gav den høyeste overføringsprosent, var bare 0,4 prosent av overføringene feil, mot 3,2 prosent med 4 navneposisjoner.

For records som har fått tildelt feil personnummer ved den maskinelle overføringen må det riktige personnummer letes fram manuelt, slik som for records som ikke har fått tildelt personnummer maskinelt. I praksis må således den egentlige maskinelle overføringsprosent beregnes på grunnlag av størrelsen: Records med maskinelt tildelt personnummer minus feiltildelinger.

En kan vel neppe regne med å oppnå så liten feilprosent ved en maskinell overføring at det blir unødvendig å foreta en etterfølgende manuell kontroll. (Ved å teste på alle navneposisjoner vil vi nok unngå de fleste feiltildelinger vi kan ha håp om å rette ved en manuell gjennomgåelse, men overføringsprosenten blir da bare vel 40.)

Under gjennomgåelsen av listene la vi merke til at antall feiloverføringer ikke syntes å spille så stor rolle for den manuelle arbeidsmengde.

Feiloverføringene ble undersøkt for å klarlegge hvilke feil som lå bak feiloverføringen. Når tallet på navneposisjoner i nøkkelen var lite, skjedde de fleste feiloverføringer på grunn av at fødselsdato ikke var riktig. Når det ble brukt 9 eller flere bokstaver i navn, var feil i navnet den hyppigste årsak til feiloverføring. Dette er for så vidt ikke noe uventet resultat, idet det nok er rimelig at en sosialhjelprecord med feil i fødselsdato lettest finner makker i personregisteret når vi tester bare på noen få navneposisjoner.

Tabell 4. Feiloverføringer etter feiltype og tallet på posisjoner i navn. Records

Tallet på posisjoner i navn	Tallet på feiloverføringer	Feil i navn	Feil i dato	Feil både i navn og dato	Tilsvarende register-record utgått ¹⁾	Feiltype ikke fastslått
4	48(34)	8(3)	27(22)	1(1)	7(3)	5(5)
5	37(28)	8(3)	20(18)	-	4(2)	5(5)
6	38(30)	10(5)	20(18)	-	2(1)	6(6)
7	33(25)	9(4)	17(15)	-	2(1)	5(5)
8	27(19)	8(3)	9(9)	-	2(1)	8(6)
9	21(13)	8(3)	6(6)	-	2(1)	5(3)
10	19(12)	7(2)	6(6)	-	2(1)	4(3)
11	8 (3)	6(1)	1(1)	-	-	1(1)
12	1 (1)	-	1(1)	-	-	-

1) I disse tilfelle ble personnummer overført fra feil registerrecord fordi den riktige record var fjernet fra registeret. Når vi sammenliknet med tidligere utgaver av registeret fant vi den riktige record med full overensstemmelse i nøkkelen.

5. Forsøk med sikte på å ivareta flere personnummeropplysninger

Som vi tidligere har vært inne på har det å øke tallet på navneposisjoner i nøkkelen to hovedeffekter. For det første reduserer vi tallet på nøkkelverdier som forekommer mer enn en gang i registerfilen. Dette medfører at flere sosialhjelprecords får tildelt personnummer idet den riktige registerrecord etter hvert blir skilt ut. For det andre vil det etter hvert bli flere nøkkelverdier i sosialhjelpfilen som ikke har makker i personregisteret. Dette medfører at samtidig som vi oppnår å fjerne noen feiloverføringer kommer, på grunn av at vi etter hvert innfører noen flere navnefeil, noen av de records som fikk tildelt riktig personnummer med et lite antall navneposisjoner i nøkkelen, ikke lenger til å få tildelt personnummer.

For å anslå omfanget av dette har vi gjort noen enkle undersøkelser. Vi fant at av de records som fikk tildelt riktig personnummer med 4 navneposisjoner i nøkkelen var det 116 som ikke fikk tildelt personnummer med 11 posisjoner. Samtidig ble 41 av de 48 feiloverføringene vi fikk når nøkkelen inneholdt 4 navneposisjoner rettet. Ved å øke antall navneposisjoner fra 4 til 11 fikk således $1\ 927 \div (1\ 503 \div 41 \div 116) = 581$ nye records tildelt

personnummer. Vi har også undersøkt hva som skjer når vi øker tallet på navneposisjoner fra 10 til 11 og fra 11 til 12. Resultatene finnes i tabell 5, og de bekrefter at det er lite å vinne ved å ta i bruk mer enn 11 pos. i navn.

Tabell 5. Endringer i overføringene når vi endrer tallet på navneposisjoner i nøkkelen

Endring i tallet på navneposisjoner	Endring i tallet på records med tildelt personnummer		Records som ikke lenger får tildelt personnummer		Nye records med personnummer
	Fra	Til	Feil-overførte	Riktige overførte	
Fra 4 til 11	1 503	1 927	41	116	581
" 10 " 11	1 921	1 927	11	15	32
" 11 " 12	1 927	1 848	7	78	6

Tabell 1 viser at under overføringen med 4 navneposisjoner i nøkkelen var det 442 nøkkelverdier som ikke fikk tildelt personnummer fordi det fantes mer enn én record med samme nøkkel i registerfilen. Som nevnt overfor har vi ved å øke antall navneposisjoner i nøkkelen, for noen nøkkelverdier oppnådd å skille ut den riktige registerrecord og således fått overført personnummer. Men noen av nøkkelverdiene har ved denne økingen ikke matchet lenger, slik at det således har gått tapt opplysninger om personnummer for ikke så få records.

I et forsøk på å ivareta flest mulige av de tilgjengelige opplysninger om personnummer har vi gjort forsøk med en overføring etter litt endrede retningslinjer.

Vi baserte overføringen på en nøkkel som bestod av fødselsdato og de 4 første av recordens navneposisjoner. Alle sosialhjelprecords som hadde samme verdi for denne nøkkelen ble antatt å tilhøre samme person. Dette betyr at vi for hver nøkkelverdi som forekommer i sosialhjelpfilen bare trenger å teste på én record (den første). Dersom vi fant personnummer for denne ene ble således nummeret overført til alle andre sosialhjelprecords som hadde samme nøkkel.

Ved overføringen testet vi altså først på fødselsdato og 4 navneposisjoner. Dersom det ikke fantes makker i registerfilen, ble sosialhjelprecorden(e) listet ut. Dersom det fantes en eller flere records i registerfilen som matchet, ble nøkkelen utvidet til 10 navneposisjoner. Dersom det fortsatt var records som matchet i registerfilen, ble personnummer overført hvis det bare var én record i registerfilen som hadde

nøkkelen. Var det mer enn én registerrecord som hadde nøkkelen, ble det forsøkt å skille dem fra hverandre ved å teste først på 11 og så på 12 navneposisjoner. Dersom dette ikke førte fram, ble personnummer ikke overført, men alle registerrecords som hadde nøkkelverdien ble listet ut. Dersom det ikke var makker i registerfilen når vi testet på 10 navneposisjoner, ble heller ikke personnummer overført, men også i dette tilfelle ble alle registerrecords som matchet med 4 navneposisjoner listet ut foran sosialhjelprecordene. De registerrecords som således kom med på listen uten at personnummer var overført, ble merket med x foran personnummer, mens registerrecords som hadde fått sitt personnummer overført ble merket med p foran personnummer. Se også nedenforstående beslutningstabell.

Beslutningstabell for overføring av personnummer.

Tallet på navneposisjoner i nøkkelen	4		10		11		12	
Ingen registerrecord som matcher	1		1		1		1	
En registerrecord som matcher		1		1		1		1
Mer enn en registerrecord som matcher			1		1		1	
List ut alle registerrecords som matcher med foregående nøkkel			X		X		X	X ¹⁾
List ut alle sosialhjelprecords med samme 4-navneposisjoners nøkkel	X		X		X		X	X
Overfør personnummer og list ut sosialhjelp- og registerrecords				X		X		X
Utvid nøkkelen med flere navneposisjoner		X X		X		X		

1) Alle registerrecords som matcher med 12 navneposisjoner i nøkkelen listes ut.

Etter dette opplegget ble 2 025 records eller 72,5 prosent av alle records tildelt personnummer. 53 records fikk tildelt feil personnummer. I realiteten har vi altså fått tildelt riktig personnummer til 70,6 prosent av materialet, og 821 records gjenstår. Under denne prøven ble det ikke testet på kjønn. 24 av feilene kunne vært unngått dersom det hadde vært gjort.

Ved hjelp av de registerrecords som var tatt med på listen lot det seg uten videre gjøre å tildele personnummer til 314 records til, slik at vi nå i alt har funnet riktig personnummer for $2\ 025 \div 53 + 314 = 2\ 286$ records eller 81,8 prosent av alle records.

Det ble så forsøkt å lete fram personnummer manuelt for de resterende 507 records. Det viste seg da at denne listen alene ikke var så godt egnet til formålet. Records som vi fant personnummer til i datolistene skapte for så vidt ingen problemer, men det var mange records vi ikke fant makker til ved å lete på fødselsdato. Vi måtte således forsøke å finne personnummer ved hjelp av andre kjennetegn. I første rekke navn. Som kjent er det lite rasjonelt å forsøke å finne personnummer i en registerliste som er sortert etter andre indikasjoner enn den listen personnummer skal overføres til. Ved gjennomgåelsen oppdaget vi også at det ikke var så sjelden at records som tilhørte samme person stod på forskjellige steder på listen fordi det var feil i fødselsdato. Dette førte oss da til den konklusjon at den manuelle overføring måtte foretas i to etapper:

1. Datolisten ble gjennomgått og personnummer ble tildelt alle records som det ble funnet personnummer til med en gang, enten ved hjelp av de registerrecords som stod på listen eller ved hjelp av en datosortert registerliste. Samtidig ble alle overføringer kontrollert og feiloverføringene rettet. Denne operasjonen tok i underkant av 4 timeverk. (For hele materialet tilsvarende det omtrent 200 timeverk.) Ved denne gjennomgåelsen ble personnummer funnet til i alt 434 records ($434 \div 314 = 120$ ble funnet i registerlisten), slik at det nå gjenstår $2\ 793 \div (2\ 025 \div 53 + 434) = 387$ records eller 13,9 prosent. (For hele materialet tilsvarende det omtrent 19 000 records.)
2. De personnummer som ble funnet ved den første gjennomgåelsen ble overført til forsøksmaterialet, som så ble sortert alfabetisk etter navn og listet ut på nytt. Ved hjelp av en tilsvarende sortert registerliste ble det så forsøkt å finne personnummer for resten. Denne operasjonen tok 14 timeverk. (For hele materialet tilsvarende det 700 timeverk.) Det ble funnet personnummer til i alt 270 av de 387 records som gjenstod. Det vil si at for 117 records (85 personer) eller 4,2 prosent av forsøksmaterialet ble det ikke funnet personnummer.

6. Hva oppnås ved maskinell overføring

De nedenforstående beregninger er bassert på de resultater som ble oppnådd ved overføringen etter det opplegget som er beskrevet i kap. 5.

Totalt sett tok de manuelle overføringsoperasjonene 18 timeverk, noe som tilsvarende ca. 900 timeverk for hele materialet på 140 000 records, dersom hele overføringen skulle skje manuelt og vi regner med at det kan

tildeles personnummer til 60 records pr. timeverk, ville det kreve ca. 2 350 timeverk. Det vil si at vi sparer fram imot 1500 timeverk ved å ta i bruk maskinell overføring.

Resultatene kan ikke direkte overføres til den enkelte statistikk-årgang fordi gjengangere fra år til år spiller en viss rolle for resultatet. Dersom vi tenker oss en tilfeldig valgt stønadsmottaker, kan en si at det er en viss sannsynlighet for å finne hans personnummer ved en maskinell overføring. Forutsetter vi at denne sannsynligheten er lik for alle stønadsmottakere og således uavhengig av om det er gitt hjelp i flere år, vil overføringsprosenten teoretisk bli den samme enten vi overfører til én årgang alene eller til flere årganger samtidig. Prosentvis vil det således stå tilbake like mange records som må tildeles personnummer manuelt om vi overfører til en enkeltårgang eller til flere årganger samtidig. Når vi overfører personnummer til flere årganger samtidig, inneholder de gjenstående records en rekke dubletter. (Fordi de tilhører stønadsmottakere som går igjen fra år til år.) Sammenliknet med en enkeltårgang blir det således forholdsvis færre personnummer som må letes fram manuelt.

Av tabell 6 ser vi at de 2 793 records i materialet representerer bare 1 473 forskjellige personer. Det vil si at ved en manuell overføring ville det vært tilstrekkelig å slå opp i registerlisten for 53 prosent av alle records. Under de betingelser som er nevnt ovenfor, betyr dette at de 821 records som måtte tildeles personnummer manuelt bare omfattet $821 \cdot 0,53 = 435$ forskjellige personer. Dette betyr igjen at vi bare fikk undersøkt 24 personer pr. timeverk¹⁾.

Ved overføring til en enkeltårgang vil det heller ikke være nødvendig å slå opp i registerlisten for alle records. I forsøksmaterialet blir summen av forskjellige personer i de 4 årgangene 2 562. Dette er 141 eller 5 prosent mindre enn recordtallet. Dette anslaget er trolig noe for lavt, idet en beregning på grunnlag av oppgaver i sosialhjelpstatistikkene for årene 1966-1968, gir som resultat at omkring 8 prosent av alle records er dubletter. For en enkeltårgang må det således manuelt finnes personnummer for $35\ 000 \cdot (1 \div 0,706) \cdot 0,92 = 9\ 470$ personer. Det er vel grunn til å anta at den manuelle operasjon vil være like arbeidskrevende - regnet pr.

- 1) Med utgangspunkt i dette kan vi regne ut timeverksbehovet.
Overføring til 4 årganger samtidig

$$\frac{140\ 000 \text{ records} \cdot (1 \div 0,706) \cdot 0,53}{24 \text{ records/timeverk}} = 909 \text{ timeverk}$$

Overføring til 1 årgang alene

$$\frac{35\ 000 \text{ records} \cdot (1 \div 0,706) \cdot 0,92}{24 \text{ records/timeverk}} = 395 \text{ timeverk}$$

Tabell 6. Stønadmottakere i de enkelte statistikkårganger fordelt etter hvilke år de er registrert som stønadsmottakere

År registrert som stønads- mottaker	I alt		Statistikkårgang				Sum "personer"
	Stønads- mottakere	Pro- sent	1966	1967	1968	1969	
66	248	16,8	248	.	.	.	248
67	163	11,1	.	163	.	.	163
68	185	12,6	.	.	185	.	185
69	174	11,8	.	.	.	174	174
66 og 67	102	6,9	102	102	.	.	204
66 og 68	38	2,6	38	.	38	.	76
66 og 69	33	2,2	33	.	.	33	66
67 og 68	76	5,2	.	76	76	.	152
67 og 69	13	0,9	.	13	.	13	26
68 og 69	96	6,5	.	.	96	96	192
66, 67 og 68	94	6,4	94	94	94	.	282
66, 67 og 69	28	1,9	28	28	.	28	84
66, 68 og 69	39	2,6	39	.	39	39	117
67, 68 og 69	53	3,6	.	53	53	53	159
66, 67 68 og 69 .	131	8,9	131	131	131	131	524
I alt	1 473	100,0	713	660	712	567	2 652

person - når en årgang behandles særskilt, som når de fire årgangene behandles samlet. Altså 24 personer pr. timeverk. Det betyr at for en enkelt årgang blir den manuelle arbeidsmengde nær 400 timeverk.

Ekstraarbeidet som er forbundet med å overføre personnummer til de tre årgangene 66, 67 og 68 samtidig med 69-årgangen kan således beregnes til omkring 500 timeverk. Med andre ord koster innføring av fødselsnummer i disse 3 årgangene på denne måten bare 100 timeverk mer enn hva den tilsvarende operasjon koster for en enkelt årgang, eller i underkant av 175 timeverk pr. årgang.

Hva sparer vi så av manuell arbeidskraft ved en slik maskinell overføring, når en enkelt årgang behandles særskilt? Under alternativet med 4 årganger ble det noe løslig anslått at personnummer manuelt kunne tildeles 60 records pr. timeverk. På grunn av et betydelig mindre antall "gjengangere" i en enkeltårgang må en i det tilfelle regne med vesentlig færre overføringer pr. timeverk. Vi er kommet til - fortsatt noe løst fundert - at omkring 40 records pr. timeverk er et noenlunde realistisk anslag her. En

ren manuell overføring (til liste) vil således kreve ca. 875 timeverk. Det vil si at en slik maskinell overføring som her er foreslått kan spare oss for opp imot 500 timeverk pr. år.

Under beregning av de forskjellige timeverkstall er det ikke tatt hensyn til at de forskjellige alternativene også har konsekvenser for andre avdelinger enn 7. kontor og Registerkontoret. I og med gjennomføringen av dette forsøket er det meste av system- og programmeringsarbeidet unnagjort, og noen videre belastning av betydning blir det ikke for Systemkontoret. På Driftskontoret berøres to avdelinger, Maskinsentralen og Dataoverføringsgruppen. Merarbeidet for Maskinsentralen ved en maskinell overføring vil vesentlig bestå i det merarbeidet matchingen av sosialhjelpfilen mot registerfilen med tilhørende overføring og utlisting medfører, sammenliknet med en ren listekjøring (datasortert). De etterfølgende operasjoner ved Maskinsentralen blir identiske med de operasjoner en manuell overføring til liste ville kreve (se vedlegg I). Antall overføringsrecords på de forskjellige trinn i prosessen blir imidlertid mindre enn ved en manuell listeoverføring, men dette spiller relativt liten rolle for belastningen på Maskinsentralen. Det reduserte antall overføringskort - ca. 70 prosent færre - betyr imidlertid en vesentlig reduksjon i den arbeidsmengde som Dataoverføringsgruppen ville bli pålagt ved en manuell overføring. Se for øvrig nedenforstående kostnadsoverslag, hvor de totale maskinrekvisita- og arbeidskraftkostnader er satt opp. Priser og lønninger pr. 1/7-71.

Kostnadsoverslag. 4 årganger samtidig

	Manuell overføring			Maskinell overføring		
	Timer	Pris/lønn	Kostnader	Timer	Pris/lønn	Kostnader
Arbeidskraft:						
Assistent ¹⁾	2910	18,-	52 380,-	1065	18,-	19 170,-
Operatører	40	26,-	1 040,-	30	26,-	780,-
Programmerere	10	33,-	330,-	70	33,-	2 310,-
I alt	3010		55 750,-	1165		22 260,-
Maskiner:						
Punchmaskiner	560	10,-	5 600,-	165	10,-	1 650,-
Sorterer	20	15,-	300,-	7	15,-	105,-
360/40	9	550,-	4 950,-	10	550,-	5 500,-
I alt			10 850,-			7 255,-
Rekvisita:						
Datalister		19,-	228,-		19,-	456,-
Hullkort		14,-	1 960,-		14,-	574,-
I alt			2 188,-			1 030,-
I alt			66 788,-			30 545,-

1) Omfatter også puncharbeidet som er beregnet til 560 timeverk ved manuell overføring og til 165 timeverk ved maskinell overføring.

7. Lister

Når det gjelder de listene som skal brukes under overføringen, var det ikke så mange forsøk som ble gjort. For hver record inneholdt de listene vi brukt opplysningene fødselsdato, navn, kjønn, bostedskommune og personnummer for de records som hadde fått tildelt det. I enkelte tilfelle hadde det vært en fordel om det lot seg gjøre å gå tilbake til skjema for om mulig å finne tilleggsopplysninger. Det bør således tas med opplysninger som åpner for en slik mulighet. Det vil si opplysninger om stønads-kommune og løpenr. På listen må det også avsettes plass til de data som trengs i samband med overføringen samt til skriving av personnummer og/eller eventuelle korreksjoner. Det vil si felter for aksjon, recordnr., feltnr. og ny verdi. Når det gjelder disse siste feltene gjorde vi forsøk med to forskjellige plasseringer, lengst til venstre og lengst til høyre på listen. Som en kunne vente viste det seg mest praktisk at disse feltene befant seg til høyre på listen. Bortsett fra dette spiller det for så vidt liten rolle hvorledes listens "hode" redigeres, men vi fant at nedenforstående rekkefølge gav god oversikt:

1. Fødselsdato
2. personnummer
3. Navn
4. Bostedskommune
5. Kjønn
6. Stønadskommune
7. Løpenr.
8. Aksjon
9. Recordnr.
10. Feltnr.
11. Ny verdi

Det anvendes to lister. Den ene datosortert og den andre alf-sortert. Begge lister med den ovenfor skisserte redigering i hode. På den datosorterte listen tas det også med noen registerrecords. Disse listes med følgende kjennetegn: Fødselsdato, personnummer, navn, bosted og kjønn. På listen plasseres disse opplysningene slik som tilsvarende opplysninger for sosialhjelprecordene, bortsett fra kjønn som er inkludert i fødselsnummeret.

Under den manuelle behandling av listen skal det bare foretas rettinger i fødselsnr. (tildeling av personnummer betraktes her som en retting av feltet). Vi fant det praktisk å dele fødselsnr.-feltet i to, fødselsdato og personnummer. Dette fordi det oftest bare er behov for å rette i ett av disse feltene. På denne måten blir således skrivearbeidet redusert samtidig som mulighetene for feil blir noe mindre (det er lettere å holde rede på et felt med bare noen få posisjoner). På listen skal det avsettes to linjer for hver record, en linje for hvert av de ovennevnte felt. Selv om denne løsning er noe plasskrevende, fant vi det likevel praktisk på denne

måten å gjøre plass til samtidig korreksjon av feltene.

8. Konklusjon

Sett på bakgrunn av at sosialhjelpmaterialet med tanke på en maskinell overføring av personnummer vel må karakteriseres som vanskelig, med mange unøyaktigheter og feil, synes vi resultatene som ble oppnådd ved forsøket er meget gode. Det kan således ikke være tvil om at bruk av maskinell overføring i dette tilfelle medfører betydelige besparelser.

Vi har konsentrert oppmerksomheten om de muligheter som foreligger for sosialhjelpmaterialet, men det er rimelig å anta at det kan oppnås gode resultater også for andre datamasser som nå får overført personnummer (fødselsnummer) manuelt, eller som det planlegges en slik overføring for. For å oppnå besparelse synes det imidlertid å eksistere et visst minstekrav til datamassens størrelse. Grunnen til dette er at den tiden det tar å søke gjennom personregisteret er relativt lite avhengig av tallet på records som skal tildeles personnummer. Det lar seg imidlertid ikke gjøre uten videre å fastsette en nedre grense for antall records, idet den nedre grense vil variere fra det ene materialet til det andre, og vil være avhengig blant annet av hvor høy overføringsprosent som oppnås, og hvilke nøkkelgenereringer og sorteringer som blir nødvendige. Materialet kan dessuten være begrenset på en slik måte at det ikke er nødvendig å søke gjennom hele personregisteret, og samtidig slik at det er billig å skille ut den aktuelle registermasse. Materialet som skal tilføres personnummer kan f.eks. bare omfatte personer født innenfor en bestemt periode eller personer bosatt innenfor et bestemt område. For et lite materiale kan det også være aktuelt å gjennomføre overføringen sammen med ett eller flere andre materialer. Denne teknikken stiller imidlertid visse krav som i vesentlig grad kan begrense mulighetene. Spesielt nevnes at nøkkelen i et slikt tilfelle bare kan bygges opp av kjennetegn som forekommer i alle materialene som tas med i prosessen. Selv om de materialene som tas med i en slik prosess har de samme nøkkelkjennetegn mister vi likevel muligheten til å "skreddersy" nøkkelen til de enkelte materialene.

For mindre materialer må altså mulighetene for å oppnå besparelser ved maskinell overføring nøye vurderes i hvert enkelt tilfelle. For større materialer synes det imidlertid ikke å være noen tvil om at maskinell overføring lønner seg.

9. Diagram

Diagram 1. Overføringsprosent etter tallet på navneposisjoner i nøkkelen

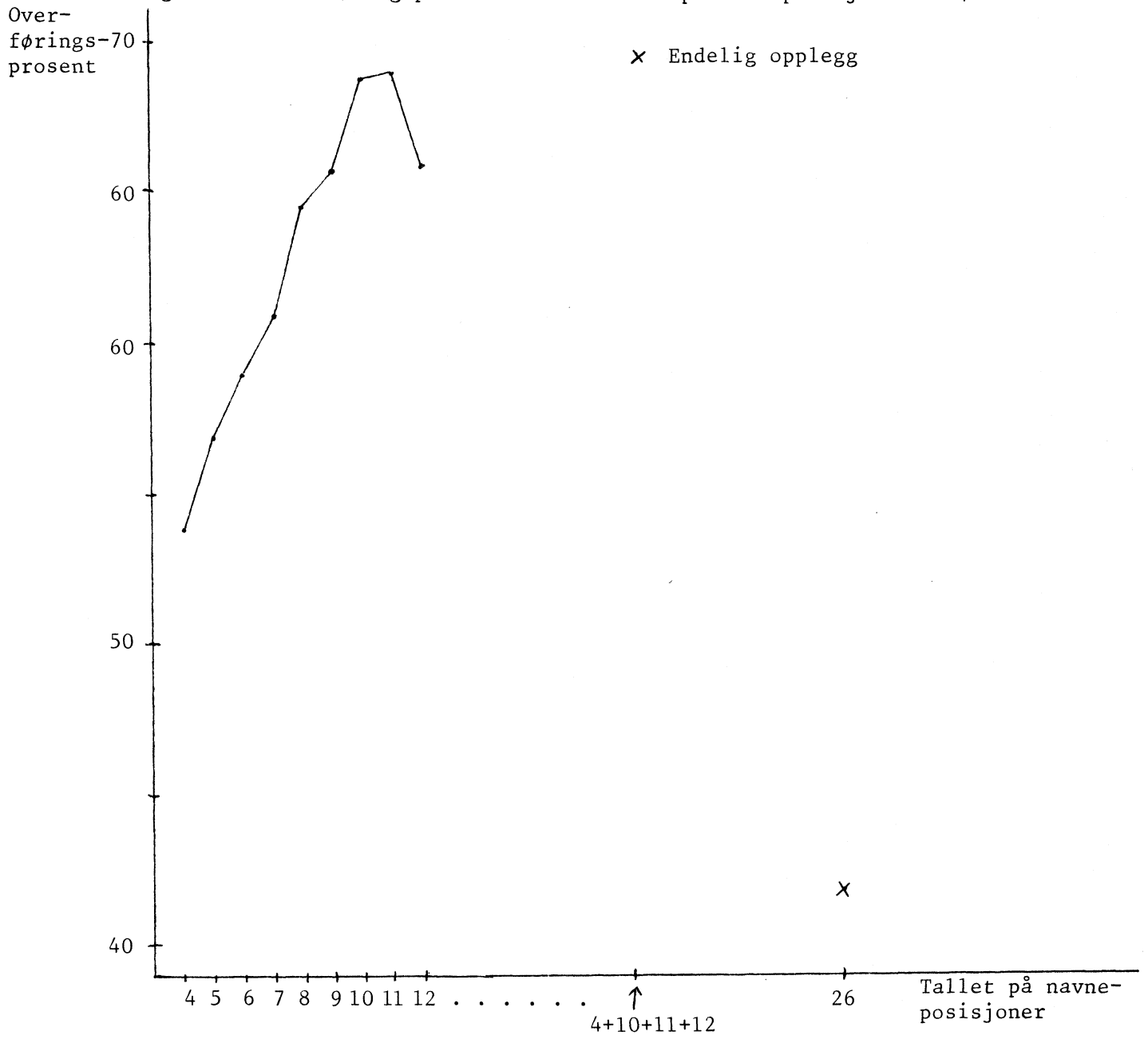
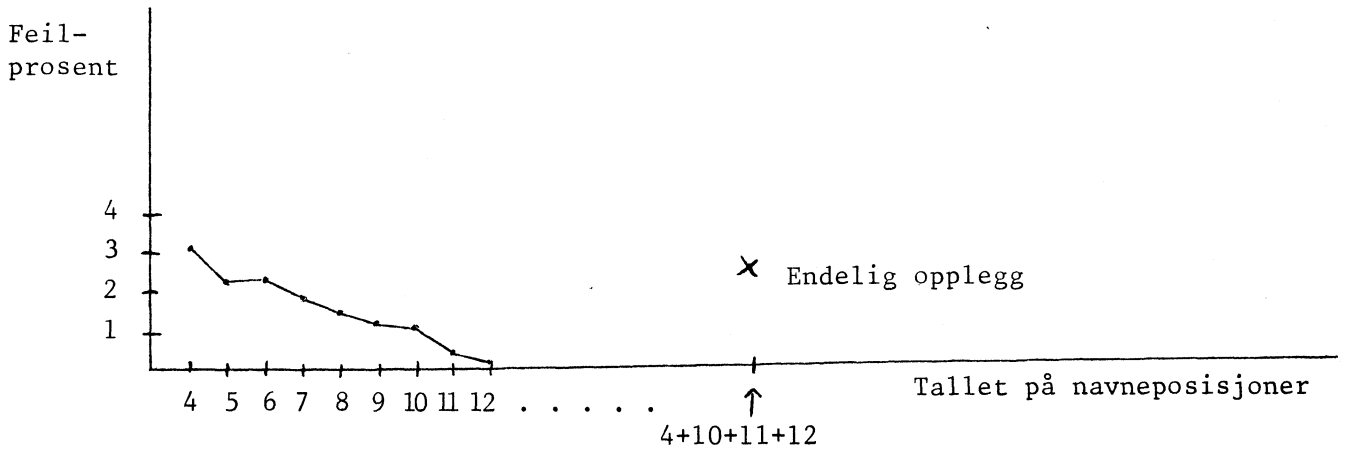


Diagram 2. Feilprosent etter tallet på navneposisjoner i nøkkelen



4131 SOSIALHJELP

OVERSIKTSDIAGRAM FOR INNFØRING AV P.NR.

