

Håvard Hungnes Lien

**Lønnsstatistikk for ansatte i helseforetakene
2010**

Dokumentasjon

© Statistisk sentralbyrå januar 2012 Ved bruk av materiale fra denne publikasjonen skal Statistisk sentralbyrå oppgis som kilde.	Standardtegn i tabeller	Symbol
ISBN 978-82-537-8269-0 (trykt)	Tall kan ikke forekomme	.
ISBN 978-82-537-8270-6 (elektronisk)	Oppgave mangler	...
ISSN 1891-5906	Oppgave mangler foreløpig	...
Emne: 06.90	Tall kan ikke offentliggjøres	:
Trykk: Statistisk sentralbyrå	Null	-
	Mindre enn 0,5 av den brukte enheten	0
	Mindre enn 0,05 av den brukte enheten	0,0
	Foreløpig tall	*
	Brudd i den loddrette serien	—
	Brudd i den vannrette serien	
	Desimaltegn	,

Forord

Notatet dokumenterer og beskriver prosessen med å klargjøre, kvalitetssikre, analysere og utarbeide lønnsstatistikken for ansatte i helseforetakene for 2010. Statistikken gir en oversikt over lønnsnivået og endringer i lønn per år for alle ansatte i næringen. Strukturstatistikken ble produsert første gang i 2001, mens kvartalsindeksen starter i tredje kvartal 2009.

Sammendrag

I lønnsstatistikk for helseforetakene benyttes et sett kontroll- og revisjonsrutiner som i utgangspunktet er standardiserte for alle lønnsstatistikkene.

Effekten av å kontrollere og revidere rådatamaterialet er i de fleste tilfeller betydelig, spesielt innen enkelte yrkesgrupper. En sammenligning mellom det ferdig reviderte datamaterialet og rådata fra oppgavegiverne viser klare forskjeller både med tanke på gjennomsnittsverdier og standardavvik i materialet. Av de innkomne observasjonene i 2010 var det mellom 89 og 95 prosent som ble godkjent til bruk i frigivning, med en synlig forbedring fra første til fjerde kvartal. Datafangsten i 2010 var god med tanke både på hastighet og samlet datainnngang. Dette har gjort det mulig å innlemme helseforetakene i den kvartalsvise lønnsindeksen.

Innhold

Forord	3
Sammendrag	4
1. Innledning	6
2. Populasjon og utvalg	6
3. Datafangsten	6
3.1. Datainnfang	6
3.2. Brutto inntekter	9
4. Behandling av data fra Oracle-basen til endelig produksjonsfil	9
4.1. Tilrettelegging av endelig produksjonsfil	10
4.1.1. Program 1 nedlesing fra Oracle-basen	10
4.1.2. Program 2 arbeidstidsforhold	10
4.1.3. Program 3 beregning av månedslønn	11
4.1.4. Program 4 yrkeskoding	13
4.1.5. 1 Program 5 godkjenning av observasjoner	14
4.1.5.2. Program 5b, etablering av heltidsekvivalenter	15
4.1.6. Frafall	15
5. Effekten av kontroll- og revisjonsrutinene	16
5.1. Arbeidstid	16
Figurregister	20
Tabellregister	20

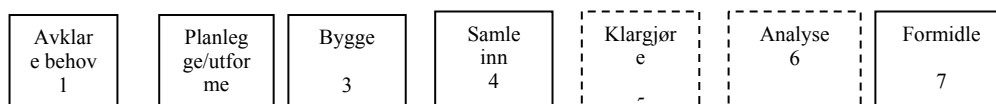
1. Innledning

Dette notatet dokumenterer arbeidsprosessene med å klargjøre, kvalitetssikre og analysere datagrunnlaget til lønnsstatistikken for ansatte i helseforetakene i 2010, både for den allerede eksisterende årlige strukturstatistikken og etableringen av den kvartalsvise lønnsindeksen.

Notatet er i utgangspunktet todelt, der første del tar for seg datafangsten til lønnsstatistikken, mens del 2 tar for seg kontroll- og revisjonsopplegget og den effekten dette har. Årsaken til denne todelingen er at datafangst og kontroll- og revisjonsrutiner henger tett sammen, og at de til en viss grad er betinget av hverandre. Det vil si at kvaliteten på datafangsten virker inn på hvordan kontroll- og revisjonen skal legges opp, og motsatt.

I lønnsstatistikken oppfordres oppgavegiverne til å levere data elektronisk via Altinn eller cd/diskett, og det finnes også en backup-løsning med levering på papirskjema. Bearbeidingen av data frem til ferdig reviderte tall kan deles inn i fire trinn: Mottak, registrering, innlesning til Oracle-tabell, og etablering av produksjonsfil. Dataene blir kontrollert og revidert på hvert av disse trinnene, men dette notatet konsentrerer seg, i tillegg til beskrivelse av populasjonen, tempo og kvalitet i innkviktering og lagring av data, om den siste delen kontroll- og revisjonsopplegget. Den skjer etter nedlesing av data fra Oracle-tabellen ved etablering av endelig produksjonsfil (se figur vedlegg 1).

Statistisk sentralbyrås virksomhetsmodell for standardiserte arbeidsprosesser i statistikkproduksjonen er en modell med 7 statistikkprosesser som bygger på hverandre:



Dette notatet beskriver prosess 5 og 6, klargjøring og analysering av datagrunnlaget.

2. Populasjon og utvalg

I helseforetakene, som i resten av offentlig sektor, har vi fulltelling, og dermed ikke noe problematikk knyttet til utvalgsusikkerhet.

3. Datafangsten

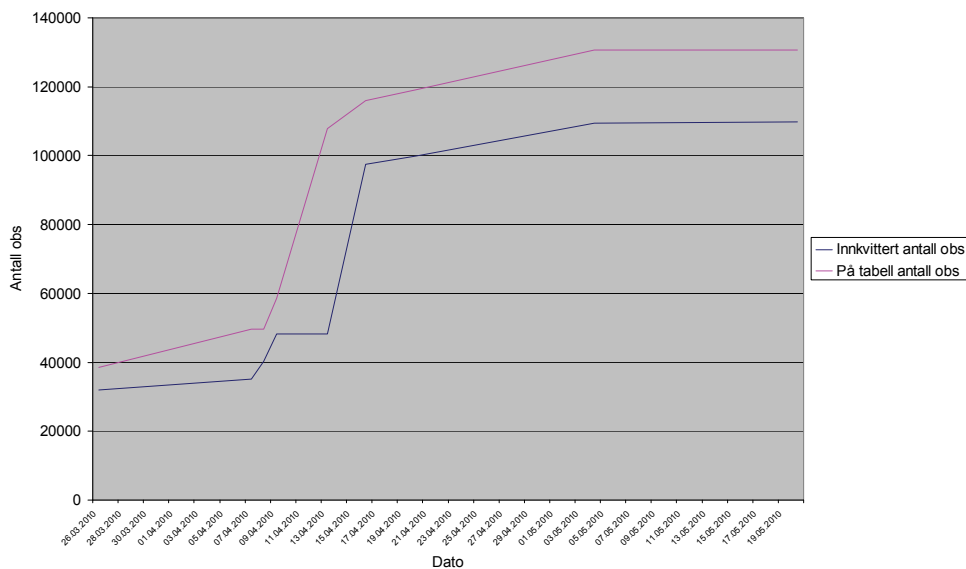
Datafangst for alle kvartal i 2010 resulterte i at 100% av foretakene leverte data, etter noe purring. I og med at indeksserien ikke ble publisert før i februar 2011 har noe sen datainngang i de foregående kvartalene ikke vært noe problem denne gangen, men dette bør følges tettere opp nå når indeksen er satt i vanlig drift.

3.1. Datainngang

Datafangsten kvartalsvis gjennom 2010

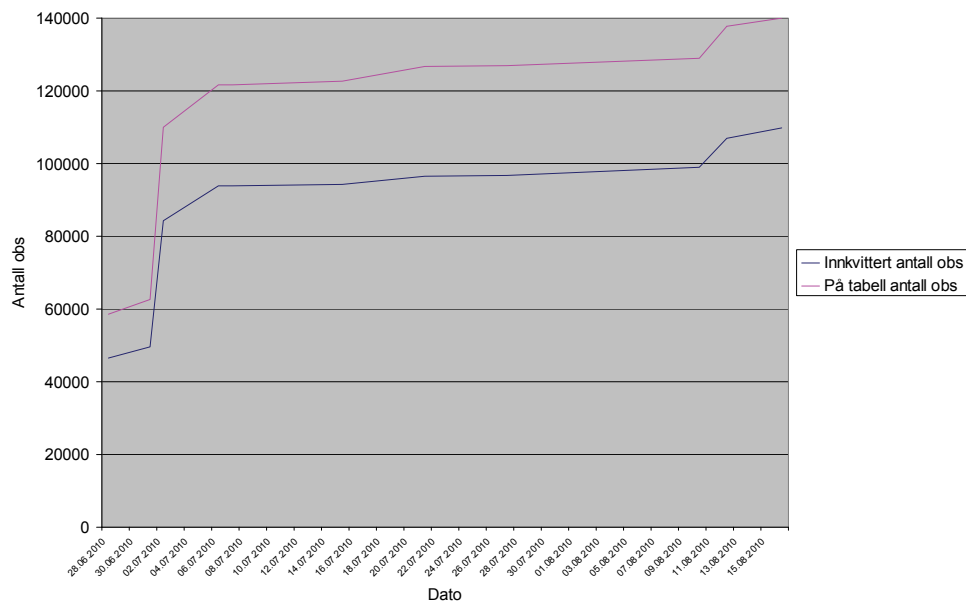
Dato	Hendelse	Antall foretak
19/3-2010	Utsendelse 1. kv	27
9/4-2010	Svarfrist	11
16/4-2010	Tvangsmulkt	24
23/4-2010	Svarfrist tvangsmulkt	5
28/5-2010	Publisering	27

Antall innkvitterte observasjoner og antall observasjoner på Oracle-tabellen pr dato, første kvartal 2010



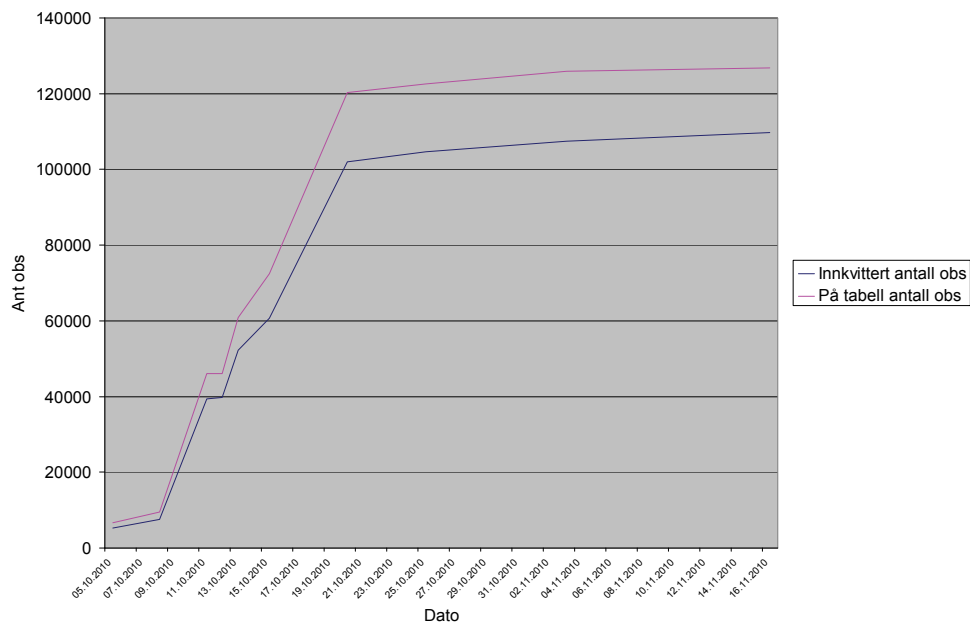
Dato	Hendelse	Antall foretak
18/6-2010	Utsendelse 2. kv	27
7/7-2010	Svarfrist	21
16/7-2010	Tvangsmulkt	22
4/8-2010	Svarfrist tvangsmulkt	24
27/8-2010	Publisering	27

Antall innkvitterte observasjoner og antall observasjoner på Oracle-tabellen pr dato, andre kvartal 2010



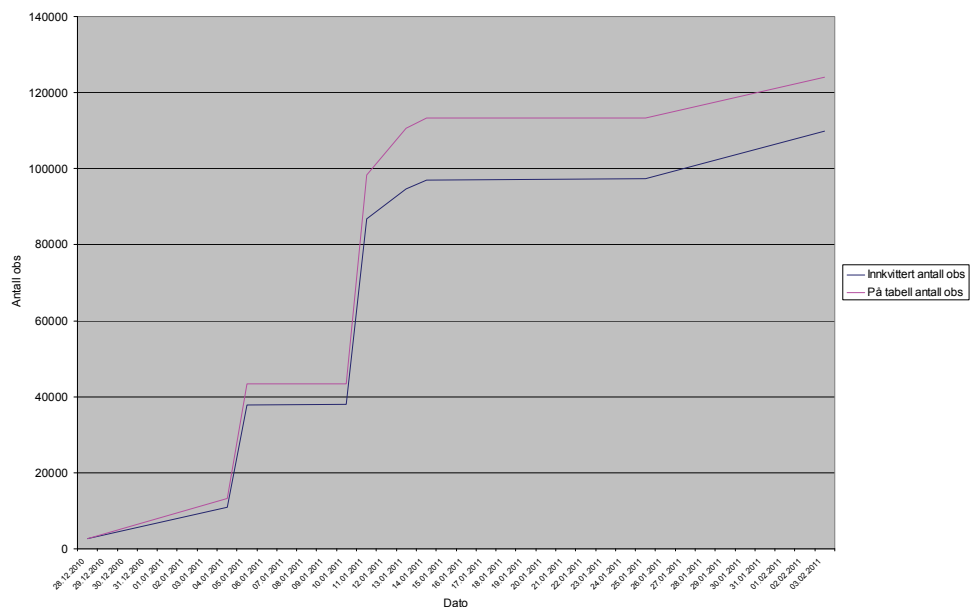
Dato	Hendelse	Antall foretak
17/9-2010	Utsendelse 3. kv	27
12/10-2010	Svarfrist	11
19/10-2010	Tvangsmulkt	16
29/10-2010	Svarfrist tvangsmulkt	25

Antall innkvitterte observasjoner og antall observasjoner på Oracle-tabellen pr dato, tredje kvartal 2010



Dato	Hendelse	Antall foretak
17/12-2010	Utsendelse 4. kv	27
7/1-2011	Svarfrist	10
14/1-2011	Tvangsmulkt	24
21/1-2011	Svarfrist tvangsmulkt	24
25/2-2011	Publisering	27

Antall innkvitterte observasjoner og antall observasjoner på Oracle-tabellen pr dato, fjerde kvartal 2010



Som vi ser har vi, ved første øyekast overraskende, til enhver tid flere observasjoner på Oracle-tabellen (produksjonsserveren, ikke tabellen som inneholder observasjoner som blir vraket med en gang) enn hva som er innkvittert. Dette skyldes at filene vi får fra helseforetakene inneholder datalinjer med personer som har sluttet, som har så lav rapportert lønn at vi antar at datalinjen er feilrapportert, dubletter og annet som blir fjernet i revisjonsprosessen. Kurvene har imidlertid samme form over tid, som er betryggende. I tillegg til hva som ligger på produksjonstabellen på Oracle er det også verdt å merke seg at vi eksempelvis for første kvartal alene kaster 27 485 observasjoner direkte, som ikke har noen form for lønn på noen variabler, verken hittil-i-år verdier eller for tellingstidspunktet, eller fordi personen den gjelder er markert som ikke ansatt på tellingstidspunktet. Noe senere i notatet vil vi se at vi for første kvartal etter alle revisjonsrutiner ender opp med 114 026 godkjente observasjoner, som er svært nært populasjonstallet på knappe 110 000 personer. Dette er et gjennomgående trekk for alle kvartalene, og viser at vi etter all sannsynlighet i hvert fall ikke får tilsendt for lite data. I tillegg viser det at revisjonsrutinene våre fungerer etter hensikten.

3.2. Brutto inkommet

Brutto inkommet antall foretak, bedrifter og antall ansatte er da 100% for samtlige fire kvartal.

Hva som ikke er så bra er at samtlige helseforetak leverer data pr kravspesifikasjon. Her er det et stor forbedringspotensiale i å fronte Altinn som leveringskanal.

4. Behandling av data fra Oracle-basen til endelig produksjonsfil

Å innhente lønnsinformasjon på individnivå fra helseforetakene har vist seg å medføre at alle opplysninger som kommer inn ikke nødvendigvis er korrekte. Det kan være ulike typer feil i materialet, fra relativt ubetydelige unøyaktigheter, til mer graverende tilfeller som medfører at observasjonen må forkastes. Det er derfor utarbeidet et system med kontroll- og feilrettingsrutiner for lønnsstatistikken, der hovedmålsettingen er å avdekke eventuell systematikk i typen feil og mangler i det som blir innrapportert. I tillegg er det en målsetting å minimere muligheten for at enkeltobservasjoner med åpenbare feil og mangler skal få uønsket innvirkning i statistikken. Systemet er bygget opp med enkle logiske mikrokontroller først (tabell 3 til tabell 9), før det kontrolleres mer på makronivå etterpå. Allerede ved innlesing av rådata, enten det måtte være fra Altinn-portalene, fra filer basert på kravspesifikasjonen for elektronisk innrapportering, eller fra optisk leste skjema blir det kontrollert for logiske feil og mangler. Dette skjer litt forenklet sagt ved at observasjoner merkes i Oracle-basen, gradert etter betydningen av feilen/mangelen. Ved betydelige feil og mangler blir oppgavegiveren kontaktet med det samme, og oppfordret om å levere nye data.

Denne dokumentasjonen tar for seg kvalitetssikringen av datamaterialet *etter* at den innledende mikro-kvalitetssikringen er gjennomført i databasen. Det er med andre ord allerede ”ryddet opp” i en del av de mest åpenbare feil og mangler. Tilrettelegging av data fra Oracle-basen fram til endelig produksjonsfil gjøres ved hjelp av fastlagte SAS-rutiner. Rutinene er lagt opp slik at de foregår i 5 trinn eller operasjoner, som i stor grad er standardisert og tilnærmet like fra næringshovedområde til næringshovedområde. Opplegget består av:

- Rutiner for *merking* av observasjoner med mulige feil
- Rutiner for *justering/korrigerings* av observasjoner med opplagte feil
- Rutiner for *å holde utenfor* observasjoner med opplagte feil

I korte trekk er rutinene bygget opp som følger:

Trinn 1: Nedlesing fra Oracle og tilrettelegging av SAS-fil. Observasjoner uten verdier på relevante variable tas ut.

Trinn 2: Grunnleggende kontroll og justering av bakgrunnsvariablene arbeidstid og avlønningsperiode.

Trinn 3: Etablering av lønnsbegrepene (lønn per måned)

Trinn 4: Godkjenning av yrkeskoder

Trinn 5: Endelig godkjenning av observasjoner. Tar ut observasjoner med ekstremverdier i en eller flere lønnsvariable

Til slutt lages det heltidsekvivalenter, og kobles på utdanningsnivå. Etter disse automatiserte mikro-kontrollene kommer det en rekke av mer selektive makro-kontroller, basert på en top-down tilnærming, der vi ser nærmere på grupper av observasjoner ut i fra om de har stor innvirkning på resultatet, og om de viser store endringer i lønnsvariablene fra sist periode. De mest brukte grupperingene her er foretak, yrkeskode og strata.

Se vedlegg 1 for skjematisk oversikt over disse rutinene.

4.1. Tilrettelegging av endelig produksjonsfil

4.1.1. Program 1 nedlesing fra Oracle-basen

Tabell 1 viser gyldige og ugyldige fødselsnumre i helseforetakene gjennom hele 2010

Tabell 1. Kombinasjoner av fødselsnummer og d-nummer i helseforetakene 2010

	2010			
	1. kv	2. kv	3. kv	4. kv
I alt	128 259	132 704	126 887	123 681
Gyldig fødselsnummer og ikke gyldig d-nummer	126 343	130 352	124 988	122 033
Gyldig d-nummer og ikke gyldig fødselsnummer	1 914	2 347	1 898	1 647
Ugyldig fødselsnummer og ugyldig d-nummer	2	5	1	1

Vi ser at antall ikke gyldige observasjoner er svært lavt, og at andelen ansatte med d-nummer ligger på jevnt i nærheten av en og en halv prosent.

4.1.2. Program 2 arbeidstidsforhold

Program 2 kontrollerer og justerer lønnsperioder og arbeidstid. Kontrollene viser at enkelte observasjoner plasseres i feil lønnsperiode og må omplasseres slik at lønnsopplysningene får riktig håndtering videre i kontrollene.

Alle ansatte har en avlønningsperiode:

Lønnsperiode 1: Månedslønn 12 ganger i året

Lønnsperiode 2: 14-dagerslønn

Lønnsperiode 3: Lønn hver uke

Lønnsperiode 4: Månedslønn 12,5 ganger i året

Lønnsperiode 5: Månedslønn 13 ganger i året

Lønnsperiode 6: Timelønn

Lønnsperiode 7: Provisjons- eller prosentlønn

Tabell 2. Status etter kontroller og endringer

	2010			
	1. kv	2. kv	3. kv	4. kv
I alt	128 259	132 704	126 887	123 681
Disse må sjekkes. Fastlønte med timelønn i tillegg	9 833	11 001	7 899	7 388
Disse må sjekkes. Høy timefortjeneste og lav arbtid ..	227	279	323	306
Disse må sjekkes. Timelønte med fastlønn i tillegg	88	134	135	228
Månedslønte som mangler arbtid	4 473	6 195	1 937	1 283
Obs med lønnsperiode 6 som endres til 1	360	536	108	795
Observasjoner med annen lønnsperiode som gis lønnsperiode 6	57	68	178	58
Observasjoner som er uendret	106 812	107 283	108 140	106 948
Sjekkes. Heltid og fastlønn<10000 (kommafeil?)	834	1 356	519	293
Sjekkes. Timelønte med arbtid>45	24	22	60	28
Timelønte uten arbtid som gis arbtid=timeant/4.33	5 551	5 830	7 588	6 354

Vi ser her at innrapporteringskvaliteten varierer litt gjennom året, men ligger ganske jevnt på mellom 81 til 86 prosent, med best i fjerde kvartal og dårligst i andre kvartal. Det mest positive som kommer fram av denne tabellen er en markert bedring av rapportering av arbeidstid gjennom året.

4.1.3. Program 3 beregning av månedslønn

Dette programmet foretar først og fremst en beregning av månedslønn på grunnlag av innrapporterte verdier av fastlønn, bonus og uregelmessige tillegg. Det foretas også korreksjon og justering av overtidstimer og overtidsgodtgjørelse. Månedslønnsbegrepet består av en summering av avtalt lønn, uregelmessige tillegg og bonus. For å beregne månedslønn må det derfor først gjøres en beregning av de øvrige lønnslementer som månedslønna bygger på:

Avtalt lønn

Avtalt lønn er beregnet lønn siste måned for alle, både fastlønte, timelønte og provisjonslønte. Dette for å få sammenlignbare størrelser uavhengig av lønnsperiode. Innrapportert lønn siste avlønningsperiode knyttet til variabel for avlønning benyttes i utregningen av avtalt lønn:

Avlønnsperiode 1 = Oppgitt fastlønn siste periode (månedslønn)

Avlønnsperiode 2 = Oppgitt fastlønn per 14-dager multiplisert med 2.165 (antall 14-dagers-perioder i en måned).

- Avlønnsperiode 3 = Oppgitt fastlønn per uke multiplisert med 4.33 (antall uker i en måned).
- Avlønnsperiode 4 = Oppgitt fastlønn per uke multiplisert med (12,5/12) .
- Avlønnsperiode 5 = Oppgitt fastlønn per uke multiplisert med (13/12) .
- Avlønnsperiode 6 = Hovedregel er timesats (enten beregnet eller direkte oppgitt) multipliser med antall timer i en måned (avtalt arbeidstid per uke multiplisert med antall uker i en måned).
- Avlønnsperiode 7 = Hovedregel: (bonus hittil i år + bonus 4. kvartal i fjor)/12 + ev. oppgitt fastlønn siste periode. Bonusfeltene tømmes etter beregning av avtalt lønn.

Uregelmessige tillegg

Innhentes som aggregert størrelse fra 1. januar til tellingstidspunktet. Gjennomsnittsberegning beregnes per måned for denne perioden for strukturstatistikken, mens kvartalsindeksen derimot beregner uregelmessig tillegg kun for gjeldende kvartal.

Bonus

Innhentes både for 4. kvartal i fjor, samt som aggregert størrelse fra 1. januar til tellingstidspunktet. Beregningene gjøres ellers på samme måte som for uregelmessige tillegg, altså gjennomsnitt pr måned basert på hele perioden for strukturstatistikken, mens kvartalsindeksen beregner utvikling kun for gjeldende kvartal.

Månedslønn

Månedslønna framkommer ved en summering av de innrapporterte verdiene for avtalt lønn, uregelmessige tillegg og bonus. Dette betyr at *beregning av månedslønn er ulik for strukturstatistikken og kvartalsindeksen, fordi to av delverdiene er beregnet forskjellig*. Omregningsformelen fra strukturmetodens månedslønn til kvartalsmetodens månedslønn er som følger:

$$\begin{aligned} & ((\text{månedslønn strukturstatistikk} - \text{avtalt månedslønn}) * 3 * \text{kvartalsnummer}) - \\ & ((\text{månedslønn strukturstatistikk forrige kvartal} - \text{avtalt månedslønn forrige kvartal}) \\ & * 3 * \text{forrige kvartalsnummer}) \\ & / 3 \\ & + \text{avtalt månedslønn} \\ & = \text{månedslønn kvartalsstatistikk} \end{aligned}$$

Kronenivåene for kvartal 1, 2 og 4 regnes videre om til en indeks etter vanlig måte, der tredje kvartal 2009 = 100.

For ordens skyld tar vi også med formelen for omregning fra indekstall til prosentvis endring, der vi i dette eksempelet ønsker å finne endringen fra fjerde kvartal 2010 til første kvartal 2011:

$$\begin{aligned} & (\text{indekstallet for første kvartal 2011} - \text{indekstallet for fjerde kvartal 2010}) * 100 \\ & / \text{indekstallet for fjerde kvartal 2010} \\ & = \text{prosentvis endring} \end{aligned}$$

Overtidsgodtgjørelse

Inngår ikke i månedslønnsbegrepet. Innhentes som aggregert størrelse fra 1. januar til tellingstidspunktet. Det blir foretatt gjennomsnittsberegning per måned for denne perioden. For overtid gjøres en spesiell kontroll av forholdet mellom innrapporterte overtidstimer i forhold til overtidsgodtgjørelsen.

Tabell 3. Oversikt over overtidstimer og overtidsbetaling

Status observasjoner	2010							
	1. kvartal		2. kvartal		3. kvartal		4. kvartal	
	Frekv.	Pros.	Frekv.	Pros.	Frekv.	Pros.	Frekv.	Pros.
I alt	128 259	100	132 704	100	126 887	100	123 681	100
Observasjoner uten overtidstimer eller betaling	71 005		67 500		59 851		74 224	
Observasjoner med flere overtidstimer enn betaling	40	55,36	33	0,02	36	0,03	67	0,05
Observasjoner med verdier på betaling	54 263		62 289		64 316		47 080	
og av disse mangler det timer på	2 951	42,31	2 882	2,17	2 684	2,12	2 310	1,87

På basis av denne kontrollen blir overtidsgodtgjørelsen justert på observasjoner som åpenbart bare har fått innrapportert *overtidstillegg*, og ikke hele godtgjørelsen, for utført overtidsarbeid. Dette gjøres ved at det beregnes en lønn per time innen normal arbeidstid for den aktuelle observasjonen, som så *legges til* overtidsgodtgjørelsen per time. Vi ser også her at det er god kvalitet på dataene som rapporteres for overtidstimer og overtidsbetaling.

4.1.4. Program 4 yrkeskoding

I program 4 gjennomføres en logisk kontroll av gyldighet av innrapporterte yrkeskoder. Foretakene som leverer data til statistikken skal rapportere yrkeskoder i henhold til Standard for yrkesklassifisering, C521 (STYRK). I kvalitetssikringsarbeidet før frigivning brukes relativt mye tid på å kvalitetssikre yrkeskodingen. Blant annet sjekkes forholdet mellom yrkeskode og yrkesbetegnelse, og ved feil og mangler kontaktes oppgavegiverne i så stor grad som mulig.

Første siffer i yrkeskoden betegnes som yrkesfelt:

1. Lederyrker
2. Akademiske yrker
3. Høyskoleyrker
4. Kontor- og kundeserviceyrker
5. Salgs- og serviceyrker
6. Yrker innen jordbruk, skogbruk og fiske (ikke i bruk)
7. Håndverkere o.l.
8. Operatører og sjåførere
9. Yrker uten krav til utdanning

Tabell 4. Fordeling av observasjoner på 1-siffer yrkeskode gjennom 2010

Yrkeskode	2010			
	1. kv	2. kv	3. kv	4. kv
I alt	128 259	132 704	126 887	123 681
1	4 250	4 297	4 007	4 388
2	38 810	39 961	38 751	37 733
3	42 908	43 903	42 733	40 717
4	6 891	6 950	6 717	6 034
5	26 850	28 492	26 698	23 715
6	77	71	73	23
7	872	902	895	828
8	279	292	264	232
9	4 554	4 627	4 775	4 158
X	2 768	3 209	1 974	5 853

Verdien X er samle verdien for ugyldige innrapporterte yrkeskoder. Det er kun for strukturstatistikken, altså kvartal 3, at yrkeskoding betyr noe for oss, i og med at kvartalsindeksen ikke er fordelt pr yrke. Dette er grunnen til at vi legger mer arbeid i å revidere yrkeskoder akkurat dette kvartalet.

Det kobles på 7-sifret yrkeskode fra AA-registeret (Arbeidstaker-/arbeidsgiverregisteret) med oppdatering nærmest mulig tellingstidspunktet. Informasjonen fra AA-registeret brukes til kontroll av innrapportert yrkeskode til lønnsstatistikken. I noen tilfeller henter vi yrke fra AA-registeret ved mangelfull innrapportering til lønnsstatistikken. Det viser seg å være relativt store avvik mellom hva som ligger i AA-registeret og hva som rapporteres til lønnsstatistikken (se tabell 5). Noe av forskjellen kan forklares i ulike tidspunkt for rapporteringen. Samtidig har lønnsstatistikken tilgang til viktig informasjon om lønn som hjelp i yrkeskodingen. Og foreløpig ser det ut til at det er en relativt større andel av uoppgitte yrker i AA-registeret enn i lønnsstatistikken. Dette gjør samlet sett at vi i utgangspunktet stoler mest på yrkeskodene som innrapporteres til lønnsstatistikken. I og med at yrkeskoden bare spiller noen rolle for tredje kvartal ser vi bare på dette her.

Tabell 5. Fordeling av antall med 1-siffer STYRK i lønnsstatistikken sammenlignet med AA-registeret

Lønnsstatistikken	AA- registeret										
	I alt	1	2	3	4	5	6	7	8	9	X
I alt	126 887	4 403	35 536	38 992	6 507	22 990	87	911	306	16 555	600
1	4 007	3 615	187	72	12	19	.	.	.	94	6
2	38 751	524	33 835	576	41	112	5	5	7	3480	166
3	42 733	153	936	37 157	73	604	6	19	6	3 655	124
4	6 717	13	56	67	6 111	157	.	.	.	302	10
5	26 698	55	181	857	186	21 633	9	52	54	3528	143
6	73	63	.	.	7	.
7	895	.	.	9	5	25	.	796	.	53	3
8	264	7	.	.	230	23	.
9	4 775	6	8	20	21	321	3	11	5	4 367	13
X	1 974	34	331	232	57	110	.	26	.	1046	135

Overensstemmelsen mellom AA-registeret og innrapporterte yrkeskoder i lønnsstatistikken er bra. De eneste to avvikene som er verdt å kommentere er at AA-registeret har om lag 800 flere personer kodet som ledere, og videre at yrkeskode 9 ikke er sammenlignbar i det hele tatt mellom disse to kildene.

4.1.5. 1 Program 5 godkjenning av observasjoner

I program 5 lages en oversikt over godkjente observasjoner og observasjoner som forkastes i statistikken. Hovedtankegangen bak dette kontrollopplegget er som følger:

- Det skal settes romslige grenser for å sikre at observasjoner med antatt korrekte verdier beholdes.
- Grensene skal likevel være satt slik at observasjoner med åpenbart ekstreme verdier, som med stor grad av sikkerhet er feilaktig, holdes utenfor datagrunnlaget.
- Grenseverdiene settes på grunnlag av en rekke forhold, som blant annet erfaring fra tidligere år, kommunikasjon med forhandlingsverden (for eksempel minstelønn innen de ulike tariffområdene) og oppgavegivere, samt statistiske mål på spredning. Grenseverdiene gjennomgås og justeres år for år.
- Underveis i kvalitetssikringsarbeidet sjekkes observasjoner som faller ut med feilkoder, i så stor grad som mulig.

Tabell 6. Antall observasjoner fordelt etter koder for grenseverdier (ok-koder) gjennom 2010

	2010			
	1. kv	2. kv	3. kv	4. kv
Godkjente observasjoner	114 026	114 440	116 250	116 992
Lave verdier på lønnsvariablene	2	0	.	.
Avtalt lønn 0 kroner	35	22	.	0
Alder=missing	3	.	0
Heltidsansatte med ialt <11 060 kroner	573	992	280	217
Heltidsansatte, timelønte med fmlonn over 68 445 kroner	4	.	11	6
Heltidsansatte, st1=1 and ialt< 21 060 kroner ...	14	18	16	6
Heltidsansatte, st1 = 2,3 og ialt>115 830 kr	84	86	102	121
Heltidsansatte, st1 = 4,5 og ialt over 78 975 kroner	10	9	10	3
Heltidsansatte, st1 > 5 og ialt over 68 445 kroner	37	27	23	127
Heltidsansatte, Overtid (ot2) over 21 060 kroner	603	103	28	19
Heltidsansatte, tillegg (tl) over 68 445 kroner ...	16	16	4	21
Deltidsansatte med arbtid<=1	139	143	143	153
Deltidsansatte, st1=1 og timefortjeneste>632 ...	9	8	6	9
Deltidsansatte, st1=1 og timefortjeneste<127 ...	3	3	.	3
Deltidsansatte, st1=1 og ialt/(arbtid/37.5) >225 000	3	3	.	4
Deltidsansatte, st1=1 og ialt/(arbtid/37.5) <22 500	38	41	17	22
Deltidsansatte, st1=2,3 og timefortjeneste>527	817	882	954	941
Deltidsansatte, st1=2,3 og timefortjeneste<85 ..	4 075	5 715	2 147	2 336
Deltidsansatte, st1=4,5 og timefortjeneste>264	561	608	671	503
Deltidsansatte, st1=4,5 og timefortjeneste<64 ..	2 730	4 115	1 129	687
Deltidsansatte, st1=7,8 og timefortjeneste>264	3	.	5
Deltidsansatte, st1=7,8 og timefortjeneste<64 ...	41	47	4	10
Deltidsansatte, timefortjeneste>264	151	146	122	233
Deltidsansatte, timefortjeneste<55	723	1 060	167	185
Deltidsansatte, ot2>21 060	653	186	68	44
Deltidsansatte, tl> 68 445	0	.	0
Heltidsdubletter	953	1 251	1 648	185
Deltidsdubletter	1 025	1 391	1 967	287
Hel- og deltidsdubletter	924	1 384	1 111	561
Leger, overtid>105 300	6	0	0	0

Vi ser av denne tabellen at kvaliteten på innrapporterte data bedrer seg gjennom året, i og med at 89% er godkjent i første kvartal og 94,6% er godkjent i fjerde kvartal. Videre ser vi at den gjennomgående største feilkilden er deltidsansatte med svært lav lønn, som utgjorde 5,3% av datamaterialet i første kvartal, og 2,4% i fjerde kvartal.

4.1.5.2 Program 5b, etablering av heltidsekvivalenter

For å best mulig kunne sammenligne lønnsutvikling over tid er vi avhengige av å kontrollere for arbeidstidsendringer, og dette gjør vi ved å regne om alle deltidsansatte til heltidsekvivalenter. I og med at lønnsstatistikken for helseforetakene er basert på full-telling er det ingen vekting inne i bildet, ut over omregningen til heltidsekvivalenter. Heltidsekvivalenter brukes bare i strukturstatistikken for tredje kvartal, i og med at kvartalsindeksen bare inneholder heltidsansatte.

4.1.6. Frafall

Vi har to hovedtyper frafall, enhetsfracfall og partielt frafall. Enhetsfracfall defineres her som frafall av foretak, mens partielt frafall er frafall av personer.

Enhetsfracfall skyldes at foretak ikke har innrapportert til oss, eller at det er feil i datamaterialet vi har fått innrapportert. Under enhetsfracfall av bedrifter kommer også de bedriftene vi har valgt å holde ute manuelt. For helseforetakene for tidsperioden vi ser på var enhetsfracfallet på 0 prosent.

Partielt frafall kan blant annet skyldes mangelfull innrapportering, feil i data eller at personer ikke er blitt innrapportert. Denne type frafall er omtalt i kapittel 4.1.5.1 hvor det er blitt satt minimums- og maksimumsgrenser for de ulike lønnsvariablene. I helseforetakene er det gjennomgående i alle kvartal at de fleste observasjonene som ikke ble godkjent var deltidsansatte med beregnet gjennomsnittlig timelønn under 85 kroner. Disse er i all hovedsak personer som enten mangler arbeidstid, eller som ikke jobbet på tellingstidspunktet.

5. Effekten av kontroll- og revisjonsrutinene

Måling av effekten av kontrollopplegget gjøres her ved å sammenligne datagrunnlaget før og etter at alle SAS-rutinene er kjørt. Dette kapitlet tar for seg hvilken virkning rutinene har på de mest sentrale variablene i lønnsstatistikken.

5.1. Arbeidstid

Arbeidstiden er grunnleggende bestemmende for to forhold: a) om en person er hel- eller deltidsansatt, og b) arbeidstiden benyttes for å beregne månedslønn for en representativ måned. Det legges derfor relativt mye ressurser inn i arbeidet med denne variabelen, og det er et absolutt krav i innrapporteringen at feltet må være utfylt. Mye av kontrollene går av den grunn ut på sjekke sammenheng mellom arbeidstid og lønn.

Tabell 7. Effekt av kontroll og revisjon av avtalt arbeidstid per uke

Første kvartal 2010				
	Antall observasjoner		Gjennomsnittlig avtalt arbeidstid	
	Før revisjon	Etter revisjon	Før revisjon	Etter revisjon
I alt	128 259	114 026	26,62	28,45
Heltidsansatte	63 274	60 938	36,73	36,74
Deltidsansatte	64 985	53 088	16,77	18,94

Andre kvartal 2010				
	Antall observasjoner		Gjennomsnittlig avtalt arbeidstid	
	Før revisjon	Etter revisjon	Før revisjon	Etter revisjon
I alt	132 704	114 440	26,05	28,52
Heltidsansatte	64 437	61 908	36,74	36,74
Deltidsansatte	68 267	52 532	15,97	18,83

Tredje kvartal 2010				
	Antall observasjoner		Gjennomsnittlig avtalt arbeidstid	
	Før revisjon	Etter revisjon	Før revisjon	Etter revisjon
I alt	126 887	116 250	27,64	28,63
Heltidsansatte	65 026	62 903	36,76	36,77
Deltidsansatte	61 861	53 347	18,04	19,04

Fjerde kvartal 2010				
	Antall observasjoner		Gjennomsnittlig avtalt arbeidstid	
	Før revisjon	Etter revisjon	Før revisjon	Etter revisjon
I alt	123 681	116 992	27,34	28,26
Heltidsansatte	62 480	61 775	36,77	36,76
Deltidsansatte	61 201	55 217	17,71	18,75

Som det går fram av tabellen er det flere observasjoner på den ureviderte fila (rådata) enn på den reviderte fila (publiseringsfila) grunnet at alle ikke-godkjente observasjoner her er holdt utenfor etter kriteriene beskrevet i kapittel 4.1.5.1. I

tillegg til dette er også observasjoner helt uten innrapportert arbeidstid holdt utenfor den ferdig reviderte fila. Disse observasjonene registreres som deltidsansatte inntil de sjaltes ut, derav nedgangen i antall deltidsansatte fra urevidert til revidert fil.

Lønnsvariablene

Lønnsvariablene som frigis i den offisielle statistikken er de vi ønsker å måle revisjonseffekten av. Dette gjelder først og fremst månedslønn, som består av avtalt lønn, uregelmessige tillegg og bonus. I tillegg kommer tall for overtidsgodtgjøring, men dette inngår ikke i månedslønna.

Revisjonseffekten på disse variablene defineres her som forskjellen mellom revidert og urevidert verdi, dvs. forskjellen på verdi på endelig publiseringsfil og verdi på rådatafil. Problemet i dette tilfellet er at publiseringsvariablene ikke finnes direkte på rådatafila fordi de ikke innrapporteres direkte fra oppgavegiverne. De er avledede variable som beregnes i løpet av kontrollrutinene, på grunnlag av innrapporterte variable.

For å forenkle vurderingen av revisjonsmålingen tar vi bare for oss ansatte i tredje kvartal i det følgende, med alle lønnsvariabler, med verdier for gjennomsnitt, median, nedre og øvre kvartil, samt minimum og maksimum.

Tabell 8. Effekt av kontroll og revisjon av lønnsvariable. Heltidsekvivalenter. Tredje kvartal. Kroner og antall personer.

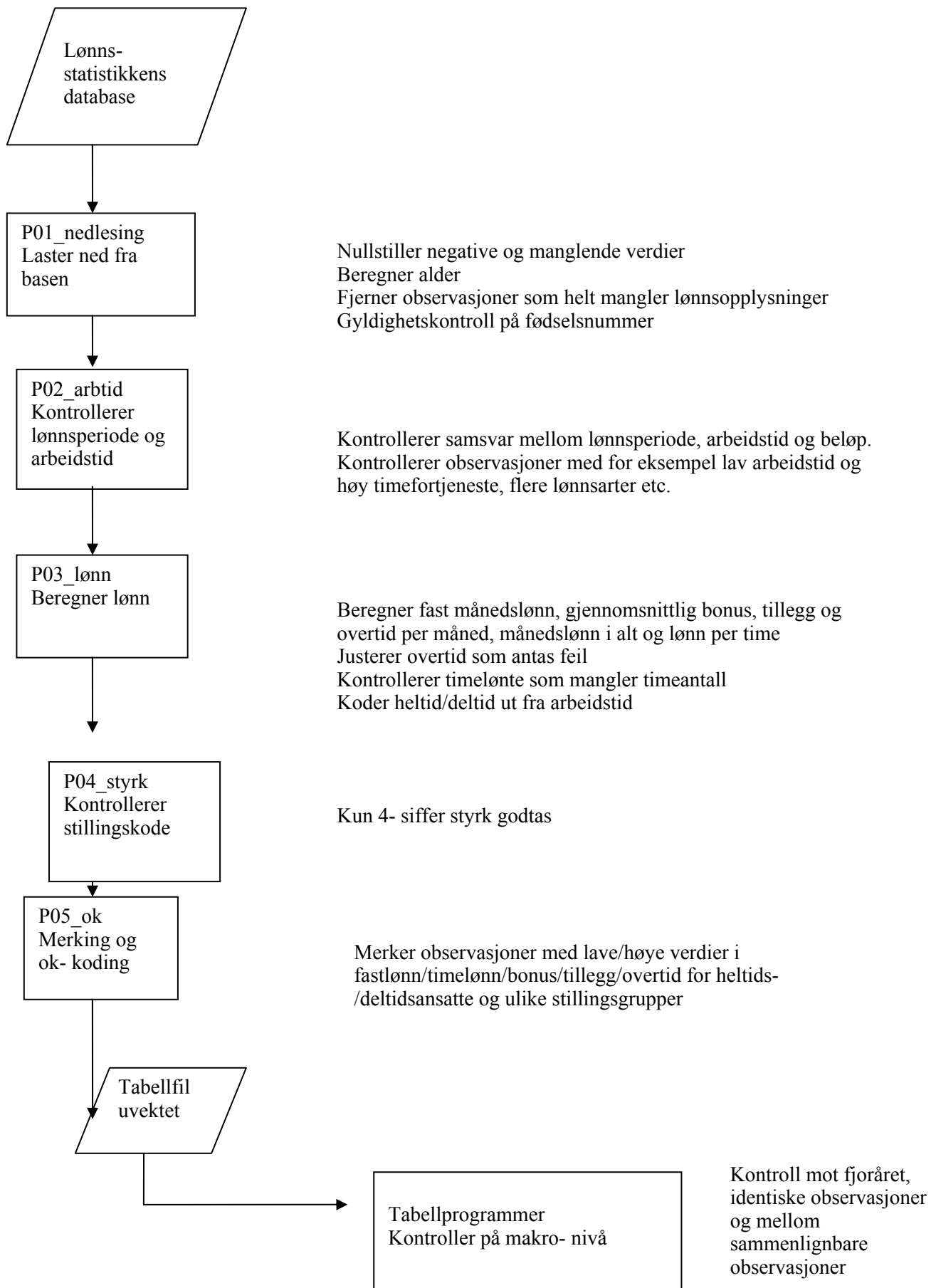
	Revidert			Urevidert			Avvik		
	Revidert	Urevidert	Avvik	Revidert	Urevidert	Avvik	Revidert	Urevidert	Avvik
Antall	116 250	126 887	-10 637						
	Gjennomsnitt			Median					
	Revidert	Urevidert	Avvik	Revidert	Urevidert	Avvik			
Månedslønn	37 789	37 627	162	34 720	34 503	217			
Avtalt lønn	33 409	33 234	175	31 463	31 111	352			
Uregelmessige tillegg	4 379	4 393	-14	2070	1962	108			
Bonus	0	0	0	0	0	0			
Overtid	1 088	1 109	-21	78	76	2			
	1. kvartil			3. kvartil					
	Revidert	Urevidert	Avvik	Revidert	Urevidert	Avvik			
Månedslønn	29 503	28 969	534	40 837	40 767	70			
Avtalt lønn	26 750	26 417	333	35 916	35 875	41			
Uregelmessige tillegg	4	4	0	6 048	5 987	61			
Bonus	0	0	0	0	0	0			
Overtid	0	0	0	1002	1 007	-5			
	Min			Max					
	Revidert	Urevidert	Avvik	Revidert	Urevidert	Avvik			
Månedslønn	8 835	3	8 832	1 023 587	5 340 129	-4 316 542			
Avtalt lønn	1244	0	1 244	162 500	1 256 247	-1 093 747			
Uregelmessige tillegg	3	3	0	990 694	5 308 446	-4 317 752			
Bonus	0	0	0	0	0	0			
Overtid	0	0	0	76 599	95 990	-19 391			

Som det går fram av tabellen gjelder det for alle variablene at de største *kronemessige* endringene skjer i den øvre enden av skalaen, men i og med at revidert månedslønn er høyere enn urevidert månedslønn betyr det at det er et større *antall* observasjoner som er endret i nedre enn øvre del av skalaen.

Ved å se på forskjellen mellom revidert og urevidert verdi for lønnsvariablene fordelt etter yrke, vil man få et bedre bilde av effekten av kontroll- og revisjonsopplegget. For å forenkle tar vi i tabell 9 bare for oss månedslønn.

Tabell 9. Effekt av kontroll og revisjon av månedslønn etter yrke. Heltidsekvivalenter. Tredje kvartal. Antall personer og kroner.

	Antall uvektede observasjoner			Gjennomsnittlig månedslønn		
	Urevidert	Revidert	Avvik	Urevidert	Revidert	Avvik
Totalt	126 887	116 250	-10 637	37 627	37 789	162
Administrative ledere	4 007	3 861	-146	49 830	50 341	511
Adm. dir.	242	227	-15	84 178	86 070	1 892
Ledere i små foretak	2	2	0	45 941	45 941	0
Akademiske yrker	38 751	36 078	-2 673	46 628	46 819	191
Leger	12 022	11 144	-878	61 188	61 330	142
Psykologer	2 724	2 558	-166	41 338	41 947	609
Spesialsykepleier	19 000	17 627	-1 373	39 311	39 553	242
Høyskoleyrker	42 733	38 617	-4 116	34 223	34 523	300
Sykepleiere	25 254	22 171	-3 083	35 273	35 677	404
Bioingeniører	4 070	3 831	-239	32 686	33 082	396
Radio- og audiografer	2 288	2 163	-125	32 785	33 168	383
Fysio-, ergoterapeuter	2 320	2 165	-155	32 095	32 336	241
Sosionomer, barnevernspedagoger, vernepleiere	4 569	4 205	-364	34 310	34 378	68
Kontoryrker	6 717	6 273	-444	27 613	27 555	-58
Sekretærer	5 804	5 392	-412	26 930	26 879	-51
Salg-/service-/omsorgsyrker	26 757	24 047	-2 710	30 203	30 032	-171
Legesekretærer	2 406	2 199	-207	26 979	26 744	-235
Hjelpepleiere	16 008	14 284	-1 724	31 039	30 791	-248
Håndverkere	890	870	-20	28 232	28 154	-78
Operatører/sjåfører	259	258	-1	25 949	25 943	-6
Yrker uten krav til utdanning	4 730	4 380	-350	26 136	26 072	-64
Renholdsarbeidere	3 768	3 500	-268	26 078	26 014	-64
Ikke klassifisert	2 043	1 866	-177	34 115	33 128	-987



Figurregister

Antall innkvitterte observasjoner og antall observasjoner på Oracle-tabellen pr dato, første kvartal 2010	7
Antall innkvitterte observasjoner og antall observasjoner på Oracle-tabellen pr dato, andre kvartal 2010.....	7
Antall innkvitterte observasjoner og antall observasjoner på Oracle-tabellen pr dato, tredje kvartal 2010.....	8
Antall innkvitterte observasjoner og antall observasjoner på Oracle-tabellen pr dato, fjerde kvartal 2010.....	8

Tabellregister

Tabell 1.	Kombinasjoner av fødselsnummer og d-nummer i helseforetakene 2010	10
Tabell 2.	Status etter kontroller og endringer.....	11
Tabell 3.	Oversikt over overtidstimer og overtidsbetaling	13
Tabell 4.	Fordeling av observasjoner på 1-siffer yrkeskode gjennom 2010	13
Tabell 5.	Fordeling av antall med 1-siffer STYRK i lønnsstatistikken sammenlignet med AA- registeret	14
Tabell 6.	Antall observasjoner fordelt etter koder for grenseverdier (ok-koder) gjennom 2010.....	15
Tabell 7.	Effekt av kontroll og revisjon av avtalt arbeidstid per uke	16
Tabell 8.	Effekt av kontroll og revisjon av lønnsvariable. Heltidsekvivalenter. Tredje kvartal. Kroner og antall personer.....	17
Tabell 9.	Effekt av kontroll og revisjon av månedslønn etter yrke. Heltidsekvivalenter. Tredje kvartal. Antall personer og kroner.....	18