



**Sykehus-
produktivitet etter
statlig overtakelse:**

**En nordisk
komparativ analyse**

Sverre A.C. Kittelsen
Frischsenteret & HERO

Jon Magnussen
*NTNU - Norges teknisk-
naturvitenskapelige universitet*

Kjartan S. Anthun
SINTEF Helse

**UNIVERSITETET
I OSLO**
HELSEØKONOMISK
FORSKNINGSPROGRAM
Skriftserie 2007: 1

HERO

Sykehusproduktivitet etter statlig overtakelse:

En nordisk komparativ analyse

Sverre A. C. Kittelsen^a, Jon Magnussen^b og Kjartan Sarheim Anthun^c

Februar

**Helseøkonomisk forskningsprogram ved Universitetet i Oslo
HERO 2007**

^a Frischsenteret - Stiftelsen Frischsenteret for samfunnsøkonomisk forskning & Helseøkonomisk forskningsprogram ved Universitetet i Oslo (HERO)

^b NTNU - Norges teknisk-naturvitenskapelige universitet

^c SINTEF Helse

Kontaktperson: Sverre A.C. Kittelsen, Frischsenteret, Gaustadalléen 21, 0349 Oslo
E-post: sverre.kittelsen@frisch.uio.no

Forord

Norges Forskningsråds evaluering av sykehusreformen av 2002 er dokumentert i rapporten "Resultatevaluering av sykehusreformen". Dette arbeidsnotatet dokumenterer nærmere analysen som er utført om produktivitetsutviklingen i norske somatiske sykehus etter sykehusreformen.

I arbeidet med denne studien har forfatterne deltatt i og trukket på *Nordic Hospital Efficiency Study Group*. Deltakere i gruppen for øvrig har vært Clas Rehnberg, Emma Medin, Unto Häkkinen, Miika Linna, Timo Seppälä og Anette Sjøberg Rød. Deltakerne har bidratt med data og diskusjoner rundt analyser og design. Kjersti Hernæs har vært forskningsassistent ved Frischsenteret og har bidratt med tilrettelegging av data.

Oslo/Trondheim, 15. februar 2007

Sverre A. C. Kittelsen

Jon Magnussen

Kjartan S Anthun

Sammendrag

Ved hjelp av dataomhyllingsanalyse (DEA) beregnes utviklingen i produktivitet i de fire nordiske landene for perioden 1999 til 2004. Dernest analyseres effekten av sykehusreformen på produktiviteten i Norge ved hjelp av en regresjonsanalyse. Datamaterialet består av til sammen 729 sykehusobservasjoner. For sykehusene i Norge og Finland benyttes data for hele perioden 1999-2004, for sykehusene i Sverige for perioden 2001-2004, mens data for Danmark kun har vært tilgjengelig for 2002.

Sykehusene aktivitet måles som tjenesteproduksjon i til sammen seks grupper pasienter, hvor inndelingen baseres på DRG systemet. Ressursbruken måles gjennom bruk av driftskostnader, men korrigeres for lønns- og prisforskjeller mellom de fire landene.

I Norge analyseres årlig utvikling i produktivitet gjennom Samdata-rapportene. Analysene i dette prosjektet viser i grove trekk samme utvikling som resultatene i Samdata, til tross for at det er forskjeller både i operasjonalisering av aktivitet, ressursbruk og i korrigeringer for pris- og lønnsvekst.

Resultatene viser at Norske sykehus, i gjennomsnitt, har et produktivetsnivå i denne perioden som ligger lavere enn tilsvarende i Finland. Forskjellene blir imidlertid mindre mot slutten av perioden. Norske sykehus har i hele perioden et produktivetsnivå som ligger over sykehusene i Sverige.

Effekten av sykehusreformen analyseres gjennom å teste hvorvidt det i perioden etter reformen skjer et positivt eller negativt skift i produktivetsnivået i Norge sammenholdt med endringene i de Sverige og Finland. Ulike modellspesifikasjoner benyttes. Resultatene er robuste ovenfor modellspesifikasjon, og tyder på at sykehusreformen har hatt en statistisk signifikant positiv effekt på produktiviteten i størrelsesorden 3 til 4 prosentpoeng.

1. Bakgrunn

Staten tok 1.januar 2002 over eieransvaret for spesialisthelsetjenesten, herunder alle somatiske sykehus. Sykehusreformen² innebar at eierskapet ble flyttet fra 19 fylkeskommuner til fem regionale helseforetak. De regionale helseforetakene ble organisert etter den geografiske inndelingen til de tidligere helseregionene. Sykehusreformen representerte dermed en resentralisering av spesialisthelsetjenesten, men samtidig innebar foretaksmodellen et element av desentralisering. Mens den fylkeskommunale modellen innebar en delegering av myndighet og ansvar til et lavere politisk nivå, representerer foretaksmodellen en ”dekonsentrering” til et lavere og uavhengig administrativt nivå (Magnussen et al, 2007).

Sykehusreformen ble markedsført med tre klare målsettinger (Ot prp 66): For det første skulle den medføre hardere budsjettstrammer og færre tilleggsbevilgninger, for det andre skulle man i sterkere grad sikre regional likhet i tilgang til spesialisthelsetjenester og, for det tredje, skulle man få økt effektivitet. Effektivitetsforbedringer ble først og fremst antatt å komme som et resultat av fokus på ledelse og etter hvert endringer i funksjonsfordeling og struktur. Begge forholdene kan hver for seg og samlet tenkes å påvirke forholdet mellom ressursbruk og aktivitet i sektoren, og dermed sektorens effektivitet. Parallelt med reformen har det imidlertid også skjedd andre endringer i sektoren. Andelen innsattsstyrt finansiering har endret seg, og i perioden variert mellom 40 og 60 prosent, og det skjer en stadig utvikling i behandlingsteknologi som bl.a. gir seg utslag i en sterk vekst i antall dagbehandlinger. En analyse av konsekvensene av sykehusreformen for sektorens effektivitet er kan derfor ikke uten videre baseres på en sammenlikning av situasjonen før og etter.

Formålet med dette arbeidet er å analysere effekten av sykehusreformen på sektorens produktivitet. Vi har valgt en tilnærming hvor utviklingen i produktivitet i norske sykehus for

² Reformen omfattet mer enn sykehus, men begrepet ”sykehusreformen” har satt seg i dagligtalen, og vil bli benyttet her.

perioden 1999-2004 (tre år før og tre år etter reformen) analyseres parallelt med utviklingen i Finland, Sverige og Danmark. De andre nordiske landene benyttes dermed som "kontrollgrupper". Tilgangen på sammenlignbare data har vært begrenset og i den endelige analysen har vi måttet avgrense oss til å benytte data fra Finland fra 1999-2004, Sverige fra 2001-2004 og Danmark fra 2002, i tillegg til Norge 1999-2004.

Rapporten er disponert som følger. I avsnitt 2 brukes først litt tid på å beskrive forhold rundt design, datainnhenting og datakvalitet. Avsnitt 3 definerer produktivitet og beskriver den metodiske tilnærmingen som er benyttet. I avsnitt 4 beskrives effektivitetsutviklingen i de fire nordiske landene, mens effekten av sykehusreformen diskuteres særskilt i avsnitt 5. Det hele oppsummeres i avsnitt 6. I et vedlegg beskrives nærmere sensitivitetsanalyser på norske data.

2. Design og data

Måling av produktivitet krever informasjon om ressursbruk (innsaksfaktorer/input) og aktivitet (tjenester/output). Flere analyser viser at empiriske produktivetsmål kan være følsomme for operasjonalisering av input og output (se f eks Magnussen, 1996). I den type sammenlikning som foretas her er det utfordringer både mht å benytte enhetlige definisjoner av output og input, men også i forhold til å korrigere for forskjeller i pris/lønnsvekst og nivå. I denne analysen er følgende tilnærming valgt:

Innsatsfaktorbruken er målt som driftskostnader ekskl. kostnader til kapital, forskning og utdanning. Kostnadene er i utgangspunktet målt i løpende priser i hver nasjons valuta. Dette gir utfordringer i forhold til å harmonisere kostnadstallene mellom land, og deflatere dem over tid. Vi har valgt følgende framgangsmåte: Det er konstruert separate prisindekser for tre typer innsatsfaktorer; leger, sykepleiere og andre kostnader. Indeksene for leger og sykepleiere er basert på offisielle lønnsdata for de fire landene, og lønnskostnadene inkluderer alle personellkostnader, det vil si også pensjonskostnader og eventuelle arbeidsgiveravgifter. Indeksen for andre kostnader over tid er basert på den harmoniserte konsumprisindeksen fra

Eurostat. Disse kostnadene er så normalisert til EURO ved hjelp av en kjøpekraftskorrigert prisindeks fra OECD. Vi antar en fordeling mellom de tre typene innsatsfaktorer på henholdsvis 20 % (leger), 50 % (sykepleiere) og 30 % (annet). Dermed konstrueres en indeks med referansepunkt Finland år 2004. Dette tilsvarer en Paasche-indeks med faste kvantumsvekter og med Euro Finland 2004 lik 1. Som andre indekser er Paasche-indeksen kun en tilnærming og gjelder eksakt bare dersom sammensetningen av innsatsfaktorer er konstant. Merk at vi gjennom denne prosedyren regner om en innsatsfaktor som i utgangspunktet er målt i løpende utgifter til en innsatsfaktor som måles i realstørrelser. Tabellene 1-2 viser relativ lønn for leger sykepleiere normalisert mot Finland i 2004. Tabell 3 gjengir den benyttede deflatoren.

Tabell 1: Normalisert lønnsnivå – leger. Inkluderer arbeidsgiveravgift og (stipulerte) pensjonskostnader (Nasjonale kilder; i Norge Statistisk sentralbyrå og Kommunenes sentralforbund)

	1999	2000	2001	2002	2003	2004
Danmark		0.833	0.874	0.913	0.945	0.976
Sverige		0.838	0.882	0.928	0.962	0.978
Finland	0.807	0.782	0.859	0.949	0.943	1.000
Norge	0.836	0.849	0.884	0.920	1.060	1.086

Det er verdt å merke seg at det relative lønnsnivået for leger varierer mellom land og år.

Lønnsnivået i Norge gjør imidlertid ett kraftig hopp fra 2002 til 2003. Dette var et år med særlig høye lønnstillegg for leger i Norge. Tilsvarende gjør lønnsnivået for sykepleiere ett kraftig hopp fra 2001 til 2002.

Tabell 2: Normalisert lønnsnivå – sykepleiere. Inkluderer arbeidsgiveravgift og (stipulerte) pensjonskostnader (Nasjonale kilder; i Norge Statistisk sentralbyrå og Kommunenes sentralforbund)

	1999	2000	2001	2002	2003	2004
Danmark	-	1.169	1.214	1.259	1.300	1.350
Sverige	-	1.120	1.180	1.191	1.232	1.263
Finland	0.811	0.831	0.874	0.905	0.961	1.000
Norge	1.122	1.211	1.195	1.308	1.351	1.400

Tabell 3: Deflator. (Kilde: egne beregninger basert på tabell 1 og 2, samt kjøpekraftsparitetsindekser fra EUROSTAT)

	1999	2000	2001	2002	2003	2004
Danmark		1.09	1.12	1.16	1.19	1.22
Sverige		1.03	1.07	1.09	1.12	1.14
Finland	0.85	0.86	0.90	0.94	0.97	1.00
Norge	1.05	1.09	1.10	1.17	1.23	1.26

Aktiviteten er målt ved bruk av det felles nordiske systemet med diagnoserelaterte grupper, NordDRG. Selv om det er store likheter mellom de nordiske landene i det definisjonsmessige grunnlaget for DRG-systemet³ er det allikevel utfordringer knyttet både til forskjeller i grupperingslogikk og forskjeller i beregnede kostnadsvekter. Vi har, av den grunn, valgt en relativt disaggregert modell med aktivitet definert som DRG-poeng i fem kategorier. Disse er: innlagte pasienter innen medisinske DRG-er, innlagte pasienter innen kirurgiske DRG-er, dagpasienter innen medisinske DRG-er, dagpasienter innen kirurgiske DRG-er, og annen aktivitet. I tillegg er antall polikliniske konsultasjoner med som et eget aktivitetsmål.

Grupperingen i DRG baseres på nasjonale versjoner av DRG systemet. Det vil i noen tilfeller være forskjeller mellom landene som gir seg utslag i at man i noen land har valgt å splitte opp noen DRG-er i to eller flere undergrupper⁴. I vår analyse er dette håndtert ved en (vektet) sammenslåing av slike undergrupper slik at vi har fått et felles sett grupper for alle de nordiske landene. I tillegg vil det være ulike tradisjoner for bruk av dagbehandling vs. poliklinikk i de fire landene. I de tilfellene hvor disse lar seg avdekke fra data (eksempelvis er i hovedsak dialyse og rehabilitering registrert som dagbehandling i Norge, men poliklinikk i de andre landene) er data harmonisert.

Til hver DRG er det tilordnet en kostnadsvekt som skal reflektere det relative ressursbehovet til gjennomsnittspasienten i gruppen. Disse vektene kan dermed benyttes i en aggregering av de rundt 500 DRG-ene til et empirisk mer håndterbart antall. Det er imidlertid også betydelige

³ En detaljert beskrivelse av DRG-systemer finnes på <http://www.drginfo.no>

forskjeller mellom landene i de beregnede kostnadsvektene. For å håndtere dette er det i denne analysen benyttet ett felles vektsett, basert på de enkelte landenes nasjonale DRG-vekter for året 2002. For hver DRG er det beregnet en kostnadsvekt som et veid gjennomsnitt av vektene i de fire landene som inngår i analysen.

Ved å bruke et felles vektsett for hele tidsperioden reduserer vi problemer knyttet til at endringer i grupperingslogikk og tekniske vektendringer blandes sammen med endringer i aktivitet. Dette betyr også at endring i kodingsatferd som eventuelt er motivert ut fra endringer etter 2002 ikke vil slå fullt ut. Endring i kodingsatferd som endrer forholdet mellom andel pasienter som havner i komplisert vs. ukomplisert DRG vil allikevel kunne påvirke aktivitetsmålet vårt, men siden vektsettet her er felles for de fire landene vil dette ventelig slå mindre ut enn på et "rent" norsk materiale. Se for øvrig diskusjonen i avsnitt 4, under.

Tabell 4: Antall enheter (sykehus/helseforetak) fordelt på land og år.

	1999	2000	2001	2002	2003	2004	Totalt
Danmark				54			54
Sverige			39	49	49	52	189
Finland	37	39	39	38	39	38	230
Norge	42	42	43	43	43	43	256
Totalt	79	81	121	184	131	133	729

I analysen inngår akuttsykehus med minst to spesialiteter. Psykiatrisk virksomhet er ikke med. Tabell 4 viser antall inkluderte enheter fordelt på land og år.

⁴ I Norge er eksempelvis gruppe 93 – lungebetennelse delt i to undergrupper A og B, med og uten komplikasjoner. I Finland vil alle pasienter tilhøre samme gruppe.

Tabell 5: Gjennomsnittsverdier for realkostnader (input) og tjenesteproduksjonen (outputs) pr sykehus fordelt på land. Forholdstallet mellom tjenesteproduksjon og realkostnader.

	Danmark	Sverige	Finland	Norge	Alle
Innsatsfaktor					
Realkostnader i millioner 2004 EUR	73.89	145.00	69.54	75.71	91.59
Tjenester					
Kirurgiske innlagte DRG	6795	10619	8204	6926	8277
Medisinske innlagte DRG	9074	13411	9113	8825	10123
Kirurgiske dagpasienter DRG	0	877	1298	1474	1154
Medisinske dagpasienter DRG	0	371	104	203	200
Andre DRG	1370	989	646	816	848
Polikliniske konsultasjoner	123611	157305	122864	72014	113992
Tjenester pr millioner 2004 EUR					
Kirurgiske innlagte DRG	92.0	73.2	118.0	91.5	90.4
Medisinske innlagte DRG	122.8	92.5	131.1	116.6	110.5
Kirurgiske dagpasienter DRG	0.0	6.0	18.7	19.5	12.6
Medisinske dagpasienter DRG	0.0	2.6	1.5	2.7	2.2
Andre DRG	18.5	6.8	9.3	10.8	9.3
Polikliniske konsultasjoner	1672.9	1084.8	1766.8	951.2	1244.5

I tabell 5 viser vi deskriptiv statistikk for de innsatsfaktorer og tjenester som inngår i analysen. I tillegg er gjengitt forholdet mellom tjenesteproduksjonen og innsatsfaktorbruk målt som realkostnader.

De svenske observasjonsenheterne er gjennomgående større enn i de andre nordiske landene. Dette skyldes delvis at det ikke alltid har vært mulig å skaffe data til for det enkelte sykehusene i Sverige, slik at en i stedet har vært tvunget til å benytte landsting som observasjonsenhet i enkelte tilfeller. På samme måte er det også noen norske helseforetak der det ikke har vært mulig å fordele dataene på de enkelte fysiske sykehusene. I stedet for de absolutte tallene er derfor antall tjenester pr kostnadsenhet i nedre del av tabell 5 av vel så stor interesse. De svenske sykehusene har et lavere omfang av dagpasienter enn tilfellet er for Norge og Finland. I det danske materialet har det ikke vært mulig å skille ut dagpasienter som egen kategori. Norge og Sverige har et noe lavere omfang av polikliniske konsultasjoner enn de andre landene.

3. Produktivitet – definisjon og metode

Begrepene effektivitet og produktivitet brukes ofte om hverandre, men i samfunnsøkonomisk teori er *produktivitet* forholdet mellom produksjon og ressursbruk. Dette kan måles i absolutt forstand, i vår modell som forholdet mellom tjenesteproduksjonen og realkostnadene, eller i relativ forstand ved sammenligning mellom ulike enheter. Begrepet *effektivitet* er derimot alltid relativt fordi det skal fange opp produktiviteten i forhold til det best mulige for det enkelte sykehuset. Gitt ulike rammebetingelser er det ikke sikkert det er mulig å oppnå samme produktivitet i for eksempel Finland og Norge, selv om sykehus begge steder kan være effektive i forhold til sine muligheter. Vi skal likevel ikke skille skarpt mellom disse begrepene i dette notatet.

Et produktivitetsforholdstall kan være for eksempel polikliniske konsultasjoner pr millioner 2004 EUR, slik den siste linjen i tabell 2 viser, men det tar ikke hensyn til at et sykehus kan ha et høyt forholdstall for enkelte tjenester men lavt for andre. Er det flere tjenester må tjenestene veies sammen. I teorien bør en bruke markedspriser hvis slike finnes, men problemet er at sykehus, som det meste av offentlig sektor ellers, leverer tjenester som i liten grad omsettes i konkurransemarkeder som gir prisinformasjon.

Farrell (1957) foreslo at en i stedet for priser kunne bruke egenskaper ved teknologien. Dersom en kjenner grensen for hva som er mulig å produsere av tjenester ved en gitt ressursbruk vil avveiningen mellom ressursbruken i å for eksempel behandle inneliggende pasienter eller polikliniske konsultasjoner representere et prisforhold som kan veie sammen de ulike tjenestene. Farrell's tekniske effektivitetsmål kan skrives som

$$TE_i = \text{Min} \{ \theta \mid (\theta \mathbf{x}_i, \mathbf{y}_i) \in T \} \quad (1)$$

der $(\mathbf{x}_i, \mathbf{y}_i)$ er input/output-vektoren for en observasjon i , og der T er teknologien eller produksjonsmulighetsområdet. At en input/output-vektor (\mathbf{x}, \mathbf{y}) tilhører

produksjonsmulighetsområdet betyr det at det er mulig å produsere tjenstemengdene i vektoren \mathbf{y} ved å bruke ressursene i vektoren \mathbf{x} . Effektivitetsmålet i (1) er derfor det minste forholdstallet en kan gange innsatsfaktorbruken med og likevel forbli i mulighetsområdet. Effektivitetsmålet vil være et tall mellom 0 og 1 for virkelige observasjoner, og vil representere forholdet mellom nødvendig innsatsfaktorbruk og faktisk innsatsfaktorbruk for et gitt nivå på tjenesteproduksjonen.

Farrells mål for teknisk effektivitet ovenfor er avhengig bl.a. av størrelsen på observasjonen, slik at en kan ta hensynt til stigende eller fallende skalautbytte. Hvis det for eksempel er skalafordeler i sykehusproduksjon vil et lite sykehus ikke være i stand til å produsere like mange tjenester pr ressursenheter som et stort, fordi de teknologiske mulighetene ikke er til stede. Det vil ikke kunne være like produktivt som et stort sykehus, selv om det er fullt ut effektivt. Et mål på teknisk produktivitet vil i stedet basere seg på en sammenligning med maksimalt oppnåelig produktivitet, uansett skala, ved å reskalere både inputs og outputs:

$$TP_i = \text{Min}_{\theta, \lambda} \left\{ \frac{\theta}{\lambda} \mid (\theta \mathbf{x}_i, \lambda \mathbf{y}_i) \in T \right\} \quad (2)$$

DEA-metoden er videreutviklet av bl.a. Charnes et al. (1978) under betegnelsen Data Envelopment Analysis (DEA), og er mye brukt i offentlig sektor. Denne metoden ligger også bak de løpende analyse av sykehusenes produktivitetsutvikling som presenteres i Samdata ⁵. DEA-metoden baserer seg på tre grunnleggende forutsetninger.

For det første antas det a) at de observerte tilpasningene faktisk er mulige. En tilstrekkelig betingelse for dette er at det ikke er målefeil i data. Dette er selvsagt en viktig forutsetning som i mange tilfeller ikke er realistisk, men som spiller mindre rolle for utviklingen i produktivitet over tid enn for nivået på produktivitet. Dernest er metoden basert på b) fri avhenging, dvs at en alltid kan tenkes å bruke mer av en innsatsfaktor uten å redusere

produksjonen, eller produsere mindre av en tjeneste uten å endre ressursbruken. Endelig forutsettes c) konveksitet, som er det samme som å si at dersom en har to mulige tilpasninger, vil også en tilpasning i mellom være mulig. Kombineres disse tre forutsetningene får en et estimat på produksjonsmulighetsområdet:

$$\hat{T}^t = \left\{ (\mathbf{x}, \mathbf{y}) \mid \mathbf{y} \leq \sum_{j \in N^t} \lambda_j \mathbf{y}_j, \mathbf{x} \geq \sum_{j \in N^t} \lambda_j \mathbf{x}_j, \sum_{j \in N^t} \lambda_j = 1, \lambda_j > 0 \right\} \quad (3)$$

Hvor summene løper over alle observasjoner N^t i hvert år t og hvert land l . Input/output-vektorene og teknologiene er indeksert over land og perioder fordi vi ikke ønsker eller trenger å anta at mulighetsområdene er de samme i alle år eller i alle land. I stedet for å basere oss på en enkelt teknologi som referanse kan vi bruke en omhylling av teknologiene for samtlige år og land:

$$\hat{T} = \bigcup_t \bigcup_l \hat{T}^t \quad (4)$$

Vårt empiriske estimat på teknisk produktivitet er da gitt ved innsetting av (4) i (2):

$$\widehat{TP}_i = \text{Min}_{\theta, \lambda} \left\{ \frac{\theta}{\lambda} \mid (\theta \mathbf{x}_i, \lambda \mathbf{y}_i) \in \hat{T} \right\} \quad (5)$$

For å beregne konfidensintervaller for de ulike målene er det benyttet en bootstrappingsteknikk utviklet av Simar og Wilson (1998). Disse metodene gir også en korreksjon for forventningsskjevhet i anslagene på produktivitet. Metodene er basert på en antakelse om at vi kjenner datagenereringsprosessen (DGP) for hvordan virkelige mekanismer har gitt opphav til våre observasjoner. Ved å etterlikne denne DGP i en pseudoverden, der våre opprinnelige observasjoner spiller rollen som virkelighet, får vi generert et sett med pseudoobservasjoner. Denne etterlikningen blir gjentatt i et stort antall iterasjoner (2000 i denne analysen), slik at en får generert en hel fordeling av utfall i simuleringene. Egenskapene til denne fordelingen av observasjoner og tilhørende

⁵ Se Kittelsen & Førsum (2001) for en oversikt over metoden og norske studier.

produktivitetsmål kan så brukes som anslag på de tilsvarende egenskapene ved det opprinnelige utvalget av observasjoner, herunder både forventningsskjevhet og konfidensintervall.

Hovedresultatene for produktivitet rapporteres i neste avsnitt. I denne studien er vi likevel ikke primært interessert i en beskrivelse av de enkelte sykehusenes produktivitet, men heller hvordan sykehusreformen har påvirket denne produktiviteten. I avsnitt 5 vises derfor resultatene av en regresjonsanalyse der de beregnede produktivitetstallene er avhengige variable, og der reformen inngår sammen med andre potensielle forklaringsvariable. Først gis imidlertid en kort beskrivelse av forskjellene mellom de fire landene.

4. Nordiske forskjeller

Tabell 6 gjengir de gjennomsnittlige bootstrapkorrigerte⁶ produktivitetmålene for de fire landene for perioden 1999-2004.

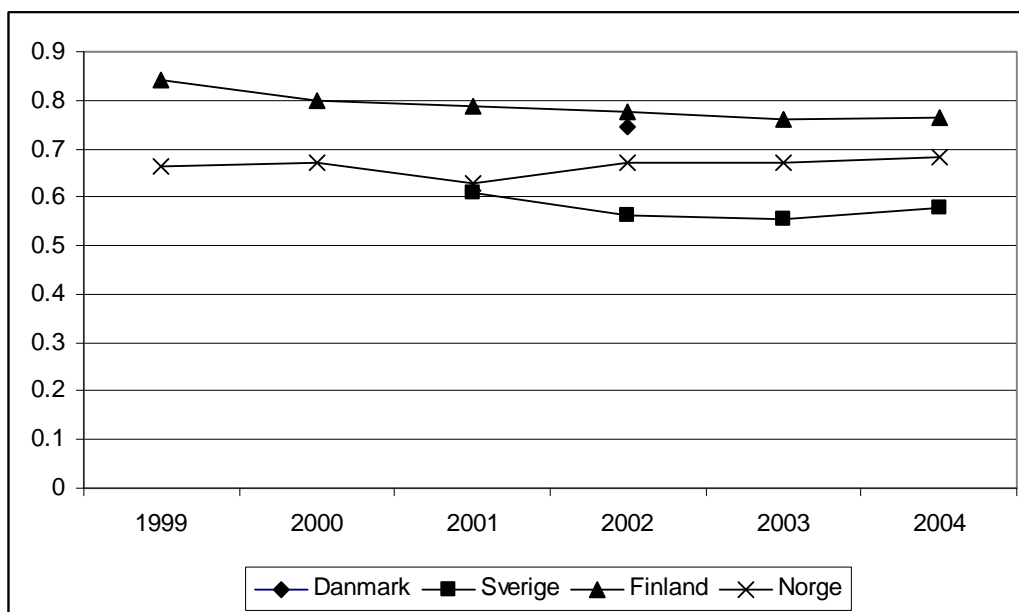
Tabell 6: Bootstrapkorrigerte produktivitetsmål. (95 % konfidensintervall i parentes).

	1999	2000	2001	2002	2003	2004
Danmark				0,747 (0,724-0,766)		
Sverige			0,611 (0,601-0,619)	0,561 (0,552-0,568)	0,553 (0,544-0,561)	0,578 (0,561-0,594)
Finland	0,842 (0,819-0,863)	0,799 (0,777-0,819)	0,789 (0,774-0,801)	0,774 (0,760-0,788)	0,760 (0,742-0,777)	0,763 (0,747-0,777)
Norge	0,663 (0,653-0,671)	0,670 (0,661-0,678)	0,627 (0,619-0,636)	0,670 (0,659-0,680)	0,673 (0,660-0,683)	0,683 (0,667-0,696)

De samme resultatene er også vist i figur 1. Analysene indikerer flere forhold. For det første synes produktivitetssnivået i Finland å ligge jevnt over nivået i Norge gjennom hele denne perioden. I Häkkinen et al (2006) ble produktiviteten i Finland og Norge i 1999 sammenliknet, og resultatene av den analysen er i tråd med de forskjeller som framgår her. Vår analyse indikerer imidlertid at Norge i perioden fra 1999 til 2004 har nærmet seg nivået i

Finland i betydelig grad. Nivået i Norge ligger på sin side noe over nivået i Sverige. Samtidig indikerer figuren at en negativ utvikling tidlig i perioden gradvis snur til en bedring de siste to årene. Forskjeller mellom år er imidlertid bare i mindre grad signifikante.

Figur 1: Gjennomsnittlig produktivetsnivå – Bootstrapkorrigerte tall.



Utviklingen i Norge, slik den presenteres her er i grove trekk lik den utviklingen som presenteres i Samdata. Unntaket er perioden 2000-2001, hvor vår modell gir et kraftigere produktivetsfall enn rapportert i Samdata. For perioden 2001-2004 er imidlertid resultatene sammenliknbare. Det er allikevel viktig å være klar over hvorfor vår modell skiller seg fra Samdata.

For det første deflateres kostnadstallene i vår analyse gjennom en indeks basert på lønnsvekst for yrkesgruppene leger og sykepleiere. Samtidig omregnes driftskostnadene til Euro og korrigeres for prisforskjeller mellom de nordiske landene. Dette betyr at endringen i kostnader

⁶ Bootstrapkorrigerte produktivetsmål er mål korrigert for forventningsskjevhet ved hjelp av

fra ett år til et annet vil avvike fra endringene i Samdata, hvor kostnadene deflateres med prisindeks for kommunalt konsum fram til 2003 og med indeks for offentlige helse og velferdstjenester fra 2004.

For det andre benyttes i vår analyse et noe mer *disaggregert aktivitetsmål* enn i Samdata. Mens Samdata benytter to typer aktivitet; DRG-poeng og polikliniske kostnader, benyttes her fem typer DRG-aktivitet og polikliniske konsultasjoner. Som vist i Pedersen (2006) vil valg av aktivitetsmål kunne påvirke resultatene.

For det tredje er det forskjeller i behandling av dagopphold og poliklinisk aktivitet i denne analysen og i Samdata. I vår analyse er betydelige dagaktiviteter knyttet til dialyse (DRG 317), rehabilitering (DRG 462) og strålebehandling (DRG 410) behandlet som poliklinisk aktivitet. Dette er nødvendig for å gjøre norske data sammenliknbare med data fra de andre nordiske landene. Mens Samdata benytter (estimerte) polikliniske inntekter som et mål på poliklinisk aktivitet, benytter vi i vår analyse antall polikliniske konsultasjoner.

Endelig benyttes i denne analysen et annet *vektsett* enn i Samdata. Mens Samdata baseres på norske DRG-vekter har vi i denne analysen benyttet et felles nordisk vektsett basert på et veid gjennomsnitt for de fire landene i 2002.

På denne bakgrunn er det ikke overraskende at modellene gir noe ulike resultater.

Sensitivitetsanalyser viser at det i stor grad er bruk av et fast vektsett basert på 2002 data, kombinert med en felles nordisk kostnadsdeflator som synes å forklare forskjellen mellom vår modell og Samdata for perioden 2000-2001. Samtidig er det viktig å være klar over at 2001 (og til dels 2002) er spesielle år ift regnskapsførte kostnader. Dels skyldes dette at kostnader som vanligvis belastes regnskapet for januar påfølgende år i 2001 ble utgiftsført på kalenderåret, dels at deflatoren for 2001 i Samdata er sterkt påvirket av en sterk lønnsvekst i

bootstrappingmetoden i Simar & Wilson (1998).

offentlige virksomheter utenfor sykehusene. Dette betyr at produktivitetstall for 2001 og 2002, uavhengig av modell, må tolkes med en viss varsomhet. Vi kommer tilbake til dette forholdet i diskusjonen i neste avsnitt.

5. Har sykehusreformen hatt effekt?

Hensikten med å foreta en komparativ studie er å benytte de andre nordiske landene som en kontrollgruppe. I analysen vil vi imidlertid også ha behov for å skille mellom endringer i produktivitet som skyldes reformen og endringer som skyldes andre forhold. Slike forhold vil dels være knyttet til strukturelle endringer som ikke er begrunnet i reformen, dels endringer i behandlingsformer og behandlingsteknologi og dels endringer i incentivstrukturer som ikke er begrunnet i reformen. Vi skal fokusere på:

- endringer i finansieringssystem, operasjonalisert gjennom endring i andel stykkprisfinansiering
- større strukturelle endringer, operasjonalisert gjennom en dummyvariabel for år
- sykehusspesifikk heterogenitet, operasjonalisert gjennom dummyer for sykehus

I tillegg inkluderes variable som dels kan korrigere for målefeil i produktivitetmålene våre, dels kan inkludere elementer av teknologisk endring. Vi gjennomfører dermed en analyse hvor forskjeller i de beregnede målene på produktivitet forklares med utgangspunkt i følgende variable:

- Pasientsammensetning (CMI) inkluderes fordi vi ikke er trygge på om justering for pasientsammensetning gjennom DRG-systemet fullt ut fanger opp all variasjon i materialet vårt. Variabelen er definert som gjennomsnittlig antall DRG-poeng pr pasient, og vil øke med pasienttyngde.
- Avvik fra forventet liggetid (LOS_A) beregnes som forskjellen mellom faktisk liggetid og den liggetiden vi ville ha forventet dersom alle sykehusene hadde samme

liggetid innen den enkelte DRG. Denne variabelen øker med økende liggetid.

Variabelen vil dels fange opp målefeil i output målene, dels teknisk endring.

- Aktivitetsbasert finansiering (ABF) inkluderes for å måle effekten av varierende incentiver i finansieringssystemet. Her er det kun variasjon over tid i Norge og Sverige. Verken Finland eller Danmark har i denne perioden hatt en ABF-ordning.
- Vi inkluderer en tidsvariabel for å fange opp tidsspesifikke utviklinger som ikke fanges opp av de øvrige variablene (eks: teknisk endring). Tidsvariabelen er i analysene gitt tre ulike utforminger: En ren lineær trend, en kvadratisk trend, og en årsspesifikk effekt håndtert gjennom en binær ("dummy") variabel. Her rapporteres kun modellen med en dummy-variabel for år⁷.

Deskriptiv statistikk for variablene er vist i tabell 7.

I vår analyse vil effekten av sykehusreformen framkomme som et skift i produktivitetsvariabelen i forhold til etablert trend. Det er derfor ikke uvesentlig hvordan en operasjonaliserer reformen. Vi har valgt to varianter. I *modell I* fanges reformen opp gjennom en dummyvariabel for de norske sykehusene for årene 2002, 2003 og 2004. Her forventes altså et skift i produktivitet allerede første år. I *modell II*, fanges reformen opp gjennom en dummyvariabel for årene 2003 og 2004. I tillegg inkluderes her dummyvariable for de norske sykehusene for årene 2001 og 2002 separat. Denne modellen har to fordeler i forhold til modell 1. Den åpner for at effekten ikke nødvendigvis inntreer umiddelbart, og den unngår problemene knyttet til fordelingen av kostnader mellom 2001 og 2002.

⁷ En kvadratisk trend synes å være en bedre spesifisering enn en lineær trend, men gir små forskjeller fra bruk av årsummier for modellens øvrige variable.

Tabell 7: Deskriptiv statistikk - forklaringsvariable

	ABF	CMI	LOS-A
1999	0.27 (0.25)	0.79 (0.09)	0.93 (0.11)
2000	0.26 (0.25)	0.79 (0.08)	0.93 (0.11)
2001	0.29 (0.27)	0.83 (0.09)	0.96 (0.12)
2002	0.22 (0.31)	0.85 (0.10)	0.98 (0.12)
2003	0.33 (0.34)	0.85 (0.10)	0.97 (0.12)
2004	0.27 (0.30)	0.82 (0.09)	0.97 (0.13)

Datamaterialet er et ubalansert panel for perioden 1999 til 2004. Analysene som presenteres her er med utgangspunkt i de bootstrapkorrigerte produktivitetmålene. Analysene er kjørt med både fixed-effekt (OLS, modell IA og IB) og random-effect (GLS- modell IIA og IIB). Fixed effect vil tilsvare en modell med faste sykehusspesifikke konstantledd⁸. Modellen utnytter kun variasjon innen hvert sykehus over tid. Alle egenskaper som er konstante over tid for det enkelte sykehus fanges da opp av sykehusdummiene, for eksempel alle nivåforskjeller mellom landene. Random effect modellene utnytter både variasjonen innen og mellom sykehusene, og vil dermed generelt gi mer presise estimat enn en fixed effect modell. Dersom vi har utelatte variable som korrelerer med våre forklaringsvariable vil imidlertid en random effect modell kunne gi forventningskjevne estimat. Resultatene av regresjonsanalysen presenteres i tabell 8⁹.

⁸ Disse rapporteres ikke.

⁹ For å gi CMI og LOS_A en lettere tolkning er disse skalert til 100.

Tabell8: Effekt av sykehusreformen. Fixed effect (modell IA og IIA) og random effect (modell IB og IIB). Koeffisienter og t-verdier.

	Modell IA	Modell IIA	Modell IB	Modell IIB
Reform	0.043**	0.033**	0.040**	0.032**
	5.30	3.18	4.83	3.08
Norge2001		-0,028**		-0,031**
		2.51		2.71
Norge2002		0.022*		0.014
		1.86		1.18
CMI	0.0010*	0.0010*	-0.007	-0.008*
	1.75	1.71	1.57	1.75
LOS-A	-0.0027**	-0.0029**	-0.0032**	-0.0034**
	6.18	6.58	8.22	8.65
ABF	-0.0062	0.0068	-0.079**	-0.064**
	0.21	0.21	3.90	3.11
Constant	0.9031**	0.9185**	1.1238**	1.1427**
	12.36	12.59	18.66	18.97
D2000	-0.023**	-0.020**	-0.018**	-0.018**
	3.01	3.02	2.56	2.56
D2001	-0.047**	-0.034**	-0.044**	-0.030**
	7.12	4.13	6.25	3.40
D2002	-0.065**	-0.056**	-0.051**	-0.040**
	8.71	6.71	6.58	4.67
D2003	-0.068**	-0.062**	-0.057**	-0.052**
	8.83	7.51	7.16	6.08
D2004	-0,058**	-0.051**	-0,056**	-0.050**
	9.02	6.54	7.26	6.05
Rho	0.905	0.905	0.820	0.820

** p< 0.05 * p<0.10

En Hausmann-test gir grunnlag for å forkaste en hypotese om random effect for modell I, men forkaster ikke for modell II. Samtidig ser vi at de fire modellene gir sammenliknbare resultater, med unntak for effekten av ABF, som varierer mellom modellene, både mht styrke og signifikans.

Effekten av sykehusreformen er positiv og i størrelsesorden 3,2 til 4,3 prosentpoeng. Vi har kjørt ytterligere sensitivitetsanalyser basert på alternative spesifikasjoner av output, uten at dette endrer resultatene. En større grad av aggregering av outputkategorier gir imidlertid en sterkere reformeffekt. Analysen er også gjennomført på det norske materialet alene, også da

med sammenlignbart resultat, bortsett fra at reformeffekten er noe svakere og mister sin statistiske signifikans fordi antall observasjoner (og dermed frihetsgrader) reduseres betraktelig. Vi tolker dette som om at utviklingen i produktivitet i Norge, med unntak av reformen, ikke skiller seg vesentlig fra utviklingen i produktivitet i våre naboland. Dette viser også at den nordiske sammenligningen har vært fruktbar i og med at resultatene fremstår som klarere i statistisk forstand enn om analysen hadde vært utført kun på norske data.

Effekten av pasientsammensetning (CMI) er ikke signifikant. Det ser dermed ikke ut til at endringer i pasientsammensetning innen sykehusene i denne perioden har påvirket effektivitetsmålene.

Effekten av avvikende liggetid er negativ og signifikant. En avvikende liggetid på 10 % vil i vår modell lede til en produktivitetsnedgang på i størrelsesorden 2,5 til 3,5 prosentpoeng. Dette er som forventet; sykehus med lang liggetid vil være mindre produktive enn sykehus med kortere liggetid.

Effekten av innsatsstyrt finansiering vil i denne modellen være effekten av en endring i ISF sats. Effektene er små og ikke signifikante i modellen med fixed effect, og tolkningen av dette blir at endring av ISF satsen i denne perioden i liten grad har påvirket sykehusenes produktivitet. Modellen med random effect gir noe overraskende negativ sammenheng mellom graden av stykkprisfinansiering og produktivitet. Siden dette estimatet er basert på en kombinasjon av en tverrsnitts- og en tidsserieeffekt kan fortegnet imidlertid skyldes de finske (effektive) sykehusene uten stykkprisfinansiering .

Årsdummiene er alle signifikante og negative. En alternativ spesifisering med en tidstrend gir tilsvarende resultat.

6. Konklusjon

Vi konkluderer på bakgrunn av disse analysene at sykehusreformen har gitt en positiv effekt på produktiviteten i sektoren. Størrelsen på effekten anslås til å være i området 3 til 4 prosentpoeng. Analysen forteller ikke hvilke aspekter ved sykehusreformen som er utslagsgivende, men ved å kontrollere for enkelte nøkkelvariable viser analysen at denne reformeffekten ikke virker via endret pasienttyngde (CMI), endret liggetid, eller endringer i ISF-satsen. Bruk av de nordiske landene som kontrollgruppe har gitt resultatene den nødvendige statistiske utsagnskraften og dessuten vist at reformeffekten ikke skyldes endringer i teknologi eller andre endringer over tid som er felles for de nordiske land. I så måte er reformeffekten robust.

Samtidig tyder våre analyser på at produktivitetsnivået i sektoren i Norge er noe under nivået i Finland og over nivået i Sverige i den perioden denne analysen omfatter. Her må det imidlertid tas større forbehold knyttet til kvalitet og sammenliknbarhet på data.

Referanser

- Charnes, A.; Cooper, W.W. and Rhodes, E. "Measuring the efficiency of decision making units." *European Journal of Operational Research*, 1978, 2(6), pp. 429-44.
- Farrell, M.J. "The measurement of productive efficiency." *Journal of the Royal Statistical Society*, 1957, 120, pp. 253-81.
- Häkkinen U, Liinna M and Magnussen J: Comparing efficiency in the Nordic countries. The case of Norway and Finland. *Health Policy*, 2006 3 (77) pp 268-278
- Kittelsen, S.A.C. and Førsund, F. "Empiriske forskningsresultater om effektivitet i offentlig tjenesteproduksjon." *Økonomisk forum*, 2001, (6), pp. 22-29.
- Magnussen J: "Efficiency Measurement and the Operationalisation of Hospital Production." *Health Services Research*, 31 1996 pp 21-37

Magnussen J, Kaarbøe O and Hagen T P: "Centralised or decentralised? A case study of Norwegian hospital reform" *Social Science and Medicine – in press*

Ot prp 66 (2001-2002): "Om lov om helseforetak m.m. Statens trykningskontor, Oslo

Pedersen M: "Har effektiviteten i somatiske sykehus økt i perioden 1999-2004? Betydning av ulike modellspesifikasjoner og metodeforutsetninger. Sintef rapport STF A60, 2006

Samdata somatikk –sektorrapport 2005. SINTEF Helse, Trondheim

Simar, L. and Wilson, P.W. "Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models." *Management Science*, 1998, 44, pp. 49-61.

Vedlegg – sensitivitetsanalyse Norge

Analysene i avsnitt 4 er basert på én av flere mulige modellspesifikasjoner. I dette vedlegget vises mer detaljerte analyser av utviklingen i produktivitet for de norske sykehusene. Følgende modeller er spesifisert:

	Utvalg	Deflatering	Vekter	# Output
A Grunnmodell	Nordisk	Nordisk	Nordisk 2002	MD1,KD1,MD2,KD2,OUT, A
B 3-input	Nordisk	Nordisk	Nordisk 2002	D1,D2,OUT
C 2-input	Nordisk	Nordisk	Nordisk 2002	INP,OUT
D Kun norske observasjoner	Norsk	Nordisk	Nordisk 2002	INP,OUT
E Norske vekter - 2002	Norsk	Nordisk	Norsk 2002	INP,OUT
F Norske vekter – kjedet	Norsk	Nordisk	Norsk - kjedet	INP,OUT
G Norsk deflatering	Norsk	Norsk KK	Norsk - kjedet	INP,OUT
I Samdata	Samdata	Norsk KK	Norsk - kjedet	INP,OUT

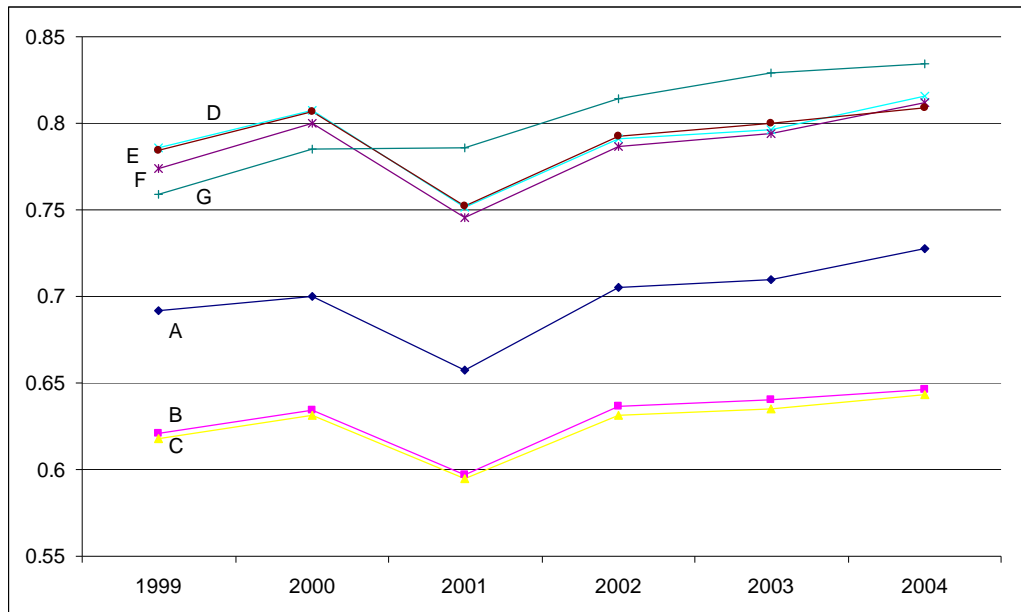
Hvor:

M = medisin K = kirurgi D1 = døgnopphold D2 = dagopphold
INP = innlagte (=D1+D2) OUT = polikliniske konsultasjoner A = annet

Gangen i dette modelloppsettet er som følger. Med utgangspunkt i basismodellen A, som er den samme som rapporteres i avsnitt 4 og 5 ovenfor, spesifiseres modellene B og C med en sterkere aggregering av aktiviteten. I modell B oppheves skillet mellom medisin og kirurgi, og det skilles kun mellom døgnopphold, fagopphold og poliklinikk. I modell C betraktes så alle innlagte (døgn- og dagopphold) samlet. Modell D tilsvare modell C, men analysen baseres kun på data fra norske sykehus. Modell E tilsvare modell D, men vektsettet som benyttes i aggregeringen av opphold til DRG-poeng endres fra å være et felles-nordisk vektsett til å kun baseres på norske kostnadsvekter. Modell F tilsvare modell E, men her benyttes årsspesifikke (kjedede)

kostnadsvekter. Modell G tilsvarer modell F, men deflatering skjer her med norsk prisindeks for kommunalt konsum.

Figur V1 viser utvikling i perioden for disse 6 ulike modellene.



Overgangen fra modell A til modellene B og C viser hvordan en aggregering av tjenestene reduserer produktivetsanslagene med ca 0,07. Modellene D,E og F får derimot et ca 15 prosentpoengs påslag i produktivetsanslagene fordi den norske sykehusene ikke lenger sammenlignes med de mer effektive observasjonene fra de andre nordiske landene, i praksis de beste finske sykehusene.

For øvrig er utviklingen påfallende parallell med unntak av modell G for perioden 2000-2001. Mens de andre modellene gir som resultat et produktivetsfall i denne perioden er nivået det samme i 2001 som i 2000 i modell G. Vi velger å konkludere med at de produktivetsanalysene som her er kjørt er rimelig robuste mht så vel aggregering, vektsett og deflatering.

For å undersøke robustheten av reformeffekten har vi også kjørt regresjoner på produktivitetsestimatene fra modellene B-G, tilsvarende de som er vist i tabell 8 for modell A. Reformeffekten er gjennomgående noe høyere i de aggregerte modellene B og C. Derimot fører utelukkelsen av de andre nordiske landene i modellene D-G til færre observasjoner og dermed et tap av frihetsgrader som gjør at anslagene ikke er statistisk signifikante. Reformeffekten er således fortsatt robust, men det har vært nødvendig med den nordiske kontrollgruppen for å gi resultatet statistisk utsagnskraft.