



UiT

NORGES
ARKTISKE
UNIVERSITET

Handelshøgskolen

Hope for the best, prepare for the worst

DEA-analyser med beste- og verste praksis til kartlegging av effektivitetsnivåer i ortopediske operasjoner ved UNN

—

Maya Åsli & Karoline Davida Kaspara Robertsen

Masteroppgave i økonomi og administrasjon - juni 2017



Innholdsfortegnelse

1	INNLEDNING	1
1.1	BAKGRUNN.....	1
1.2	PROBLEMSTILLING.....	2
1.3	FORMÅL OG AVGRENSNING	3
1.4	OPPGAVENS STRUKTUR.....	4
2	DET NORSKE HELSEVESENET	5
2.1	OPPBYGGINGEN AV NORSK HELSEVESEN.....	5
2.1.1	<i>Universitetssykehuset i Tromsø</i>	6
2.2	INNTEKTER OG FINANSIERING	8
2.2.1	<i>Innsatsstyrt finansiering (ISF)</i>	8
2.2.2	<i>Diagnoserelatert gruppering (DRG)</i>	8
2.2.3	<i>Kostnad per pasient (KPP)</i>	8
3	LITTERATURGJENNOMGANG	9
3.1	TECHNICAL EFFICIENCY IN THE CLINICAL MANAGEMENT OF CRITICALLY ILL PATIENTS	10
3.2	SYKEHUSPRODUKTIVITET ETTER STATLIG OVERTAKELSE: EN NORDISK KOMPARATIV ANALYSE.....	11
3.3	SYKEHUSENES EFFEKTIVITETSUTVIKLING 1992-1999: HVILKE EFFEKTER GA INNSATSSTYRT FINANSIERING?	11
3.4	TIDLIGERE EFFEKTIVITETSANALYSER VED UNN.....	12
4	KAPITTEL 4 – TEORI OG METODE	14
4.1	PRODUKTIVITET OG EFFEKTIVITET	14
4.2	TEKNISK EFFEKTIVITET.....	15
4.3	SKALAEGENSKAPER	15
4.4	DEA – BESTE PRAKSIS.....	17
4.4.1	<i>CCR-modellen</i>	18
4.4.2	<i>BCC-modellen</i>	19
4.5	DEA - VERSTE PRAKSIS.....	21
4.6	LAYERING.....	22
4.7	SENSITIVITETSANALYSE	24
4.7.1	<i>Supereffektivitet</i>	25
4.7.2	<i>AP-outlier metoden</i>	26
4.8	STATISTISK ANALYSE	27
4.8.1	<i>Banker-test på skalanivå</i>	28
4.8.2	<i>Banker-tester på grupper</i>	29
5	DATA OG OPERASJONALISERING AV ANALYSE	30
5.1	DATAINNSAMLING.....	31
5.2	DEFINERING AV UTVALG	31
5.3	OPERASJONALISERING	33
5.3.1	<i>Forskningsmodell</i>	34
5.4	VALG AV VARIABLER	35
5.4.1	<i>Knivtid</i>	37

5.4.2	Forberedelsestid.....	38
5.4.3	Forsinkelse i stuetid.....	38
5.4.4	Forsinkelse i operasjonsstart.....	39
5.4.5	DRG-vekt.....	40
5.4.6	KPP - kostnad per pasient.....	41
5.5	OUTLIERS.....	41
5.5.1	Fjerning av outliere - beste praksis.....	42
5.5.2	Fjerning av outliere - verste praksis.....	44
5.6	DESKRIPTIV STATISTIKK.....	46
6	RESULTATER.....	48
6.1	RESULTATER FRA DATA ENVELOPMENT ANALYSIS - BESTE PRAKSIS.....	49
6.1.1	Total effektivitet (TECRS).....	50
6.1.2	Ren teknisk effektivitet (TEVRS).....	51
6.1.3	Skalaeffektivitet.....	52
6.1.4	Banker-test for skala.....	53
6.2	RESULTATER FRA DATA ENVELOPMENT ANALYSIS - VERSTE PRAKSIS.....	54
6.2.1	Ren teknisk ineffektivitet.....	54
6.3	LAYERING.....	55
6.3.1	Beste praksis.....	55
6.3.2	Verste praksis.....	56
6.4	BANKER-TEST PÅ GRUPPER.....	57
6.4.1	Banker-test grupper beste praksis.....	58
6.4.2	Banker-test grupper verste praksis.....	59
7	DISKUSJON.....	61
7.1	VARIABLENES ANVENDELIGHET.....	61
7.1.1	Svakheter i variablene.....	64
7.2	ANVENDELIGHETEN TIL MODELLEN FOR BESTE PRAKSIS.....	66
7.3	ANVENDELIGHETEN TIL MODELLEN FOR VERSTE PRAKSIS.....	67
7.4	RESULTATENES BIDRAG TIL EFFEKTIVISERINGSARBEIDET.....	68
7.5	BANKER-TESTER PÅ GRUPPER.....	69
7.6	LAYERING-TEKNIKKEN.....	71
8	KONKLUSJON.....	73
8.1	OPPSUMMERING OG KONKLUSJON.....	74
8.2	SVAKHETER VED FORSKNINGEN.....	78
8.3	VIDERE FORSKNING.....	78
	REFERANSELISTE.....	80
	VEDLEGG.....	84
	R-SCRIPT.....	84

Tabelliste

Tabell 1 Utvalg	32
Tabell 2 Input- og outputvariabler	37
Tabell 3 Eksempel fra datasettet til kneproteser	39
Tabell 4 Eksempel på konsekvensen av outlierfjering	43
Tabell 5 Resultat fra outlieranalysene for beste praksis	44
Tabell 6 Resultat fra outlieranalyse for verste praksis	45
Tabell 7 Deskriptiv statistikk variabler i beste praksis	46
Tabell 8 Deskriptiv statistikk variabler verste praksis	47
Tabell 9 Total effektivitet	50
Tabell 10 Resultat total effektivitet	51
Tabell 11 Ren teknisk effektivitet	51
Tabell 12 Resultat ren teknisk effektivitet	52
Tabell 13 Resultat skalaeffektivitet	52
Tabell 14 Banker-test skala	54
Tabell 15 Ren teknisk ineffektivitet	54
Tabell 16 Resultat ren teknisk ineffektivitet	55
Tabell 17 Gjennomsnittlig effektivitetsscore banker-test beste praksis	58
Tabell 18 Banker-test sesongvariasjon beste praksis	59
Tabell 19 Gjennomsnittlig effektivitetsscore banker-test verste praksis	60
Tabell 20 Banker-test sesongvariasjon verste praksis	60

Figurliste

Figur 1 Ortopediske operasjoner ved UNN 2014-2016.....	7
Figur 2 Konstant og variabelt skalautbytte.....	15
Figur 3 Illustrasjon av VRS og CRS.....	21
Figur 4 Beste og verste praksis.....	21
Figur 5 Lagvise effektivitetsfronter i verste praksis.....	23
Figur 6 Lagvise effektivitetsfronter i beste praksis.....	23
Figur 7 Supereffektivitet.....	25
Figur 8 Modell beste praksis.....	34
Figur 9 Modell verste praksis.....	35
Figur 10 Resultat AP-analyse kneprotese.....	45
Figur 11 Resultat DEA beste praksis.....	49
Figur 12 Resultat layering.....	57

Forord

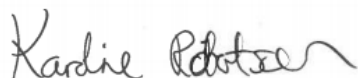
Denne masteroppgaven markerer endepunktet i vår mastergrad i økonomi og administrasjon ved Handelshøgskolen. Produksjonen av oppgaven har vært en lærerik, lang og utfordrende prosess. Oppgaven har gitt oss muligheten til å fordype oss i en problemstilling som vi synes er viktig. Det er spesielt å se hvordan alt arbeidet er materialisert gjennom disse sidene. Det er med stolthet og litt vemod at vi takker for vår tid ved UiT, men først har vi noen andre å takke:

En stor takk rettes først og fremst til vår formidable veileder, universitetslektor Helén Marita Sørensen Holst. Vi setter veldig stor pris på ditt gode humør, din kunnskapsdeling og sakkyndige rådgivning. Takhøyden er stor og dørterskelen lav på kontor 02.308, hvor vi har funnet mye støtte, innsiktsfulle forklaringer og motiverende ord.

Vi vil gjerne også takke Stig Bakken, Morten Espejord og Thomas Krogh ved økonomi- og analysesenteret til UNN, for å ha gitt oss muligheten til å ta fatt på denne spennende oppgaven. Øvrig personale og kompetansepersoner vi har hatt kontakt med på UNN må også nevnes. Vi er takknemlig for at dere har tatt dere tid til oss, og har møtt oss med velvilje og god hjelp som har vært avgjørende for den endelige oppgaven.

Sist men ikke minst ønsker vi å takke Torill i renholdsstaben og Gunn-Heidi i kantina, for at kollokvierommet og hodet alltid har vært klar om morgenen.

Tromsø, juni 2017



Karoline Davida Kaspara Robertsen



Maya Åsli

Sammendrag

Hensikten bak denne oppgaven er å undersøke anvendeligheten til DEA-analyser for kartlegging av effektivitet og ineffektivitet i ortopediske operasjoner ved UNN i 2015. Herunder også undersøke hvordan analysene kan benyttes av sykehuset til planlegging for eventuelle effektiviseringsarbeid.

Gjennom DEA-modeller for beste- og verste praksis estimeres effektivitet og ineffektivitet i fem ulike ortopediske operasjonstyper. Variablene benyttet for ressursforbruk er herunder kostnad per pasient, forberedelsestid, stuetid, forsinkelser i operasjonsstart og forsinkelser i stuetid. Variablen som benyttes for produksjonen er DRG-vekter, hvilket indirekte danner grunnlaget for sykehusets inntektsgenerering.

Resultatene i analysen for beste praksis er at operasjonstypene jevnt over har lave effektivitetsnivåer. Sammenholdt med det faktum at fullbyrdet effektive operasjoner kan gå på bekostning av kvalitet og pasientsikkerhet i operasjonene, anmodes det at sykehuset vil være best tjent på å forbedre effektivitetsnivået blant de mest ineffektive operasjonene. Herunder benyttes modellen for verste praksis til kartlegging av ineffektiviteten i de ortopediske operasjonene. Resultatene fra analysen bekrefter tilstedeværelse av ineffektivitet i operasjonene. Videre benyttes layering-teknikken som kombinerer frontene fra beste- og verste praksis til lagvis inndeling av effektivitetsnivåer. Herunder undersøkes muligheten for å kartlegge trender og tendenser i variablene til de effektive og ineffektive operasjonene. Kostnaden per pasient, forberedelsestid og forsinkelse i oppstartstid anses å være avgjørende for operasjonene som danner effektivitet- og ineffektivitetsfrontene.

Asymptotiske tester, såkalte bankertester for grupper anvendes også for å teste for signifikans i effektivitetsforskjellene mellom sommer- og vintermånedene. Herunder kartlegges et betydelig høyere effektivitetsnivå i sommersesongen for effektivitetsmålene gjort med beste praksis, samt en signifikant ineffektivitet i sommermånedene for effektivitetsmålene gjort ved verste praksis.

Modellene for beste- og verste praksis anses på bakgrunn av denne studien å være anvendelig til effektivitetsanalyser for ortopediske operasjoner ved UNN. Videre tenkes modellen for verste praksis å være mest hensiktsmessig for kartlegging av ineffektivitet og det reelle og forsvarlige forbedringspotensialet i operasjonsaktiviteten. Resultatene fra disse analysene kan benyttes til kartlegging og planlegging av effektiviseringsarbeid.

Programpakken RStudio (Versjon 1.0.136) med tilleggspakken ”Benchmarking”, og Microsoft Office Excel 2011 er benyttet til beregningene i oppgaven.

Nøkkelord: Data Envelopment analysis, beste praksis, verste praksis, effektivitet, banker-test, layering, outliers, operasjoner.

1 Innledning

1.1 Bakgrunn

President for legeforeningen, Marit Hermansen, uttalte seg i februar 2017 på vegne av helsesektoren: *«Vi ønsker å jobbe smartere, og er den én sektor som virkelig har gjort det, er det helsesektoren. Nå har vi kommet til et punkt hvor vi må investere for å klare å effektivisere ytterligere»* (Skihamn, 2017).

Helsebudsjettene blir større for hvert år og har hatt en betydelig økning sett i forhold til den generelle inflasjonen. I 2016 utgjorde helserelevante kostnader, som største post, en andel på om lag 10,5% av brutto nasjonalprodukt her i Norge (Statistisk sentralbyrå, 2017).

Samfunnsmessig står vi også ovenfor realiteten av eldrebølgen, som følger av store fødselskull i etterkrigstiden. Parallelt varsles det om gjennomsnittlig høyere levealder og store tendenser til livsstilssykdommer. Økt etterspørsel og stadig ressurskrevende tjenester medfører at kostnadene trender oppover i den allerede kapitalkrevende sektoren.

I budsjettforliket mellom regjeringspartiene ble sykehusene pålagt et effektiviseringskrav tilsvarende om lag 600 millioner kroner for 2017. Helseminister, Bent Høie, forklarer at de har bevilget mer penger enn befolkningsutviklingen tilsier, samt økt den innsatsstyrte finansieringen. Herunder vil sykehus som ser muligheten til å behandle pasienter på mer effektive måter, nå har muligheten til det (Skihamn, 2016). Sykehusene selv mener effektivitetstaket er nådd og at eneste mulighet for videre effektivisering er gjennom ytterligere investeringer i helse- og omsorg.

Siden effektiviseringsmulighetene avfeies av sektoren, omhandler den politiske diskusjonen i stor grad innsatsfaktorene, gjennom hvor mye midler og ressurser som skal og bør bevilges. Det er likevel slik at sykehusene er blant landets største virksomheter. Helse Nord er eksempelvis landsdelens største arbeidsplass. Sykehusene påvirker hele samfunnet rundt seg, og blir også betydelig påvirket av utviklingen i Norge.

Uavhengig av størrelsen på «pengesekken» har sykehusene et ansvar for å anvende tildelte midler på mest mulig optimalt vis. Bakgrunnen for studien er derfor om denne ensidige debatten, om bevilgninger og antall årsverk, usynliggjør et urealisert effektivitetspotensial

som finnes i sektoren. I studien undersøkes det om en ved hjelp av DEA-analyser kan finne potensiale for økt effektivitet i norske sykehus. Herunder om det er mulig å forbedre utnyttelsen av ressursene for å øke effektivitetsnivåene, slik at en skal kunne møte etterspørselen etter gode og effektive helsetjenester.

Tidligere undersøkelser har hatt utfordringer med å anvende den tilgjengelige dataen til intern benchmarking ved Universitetssykehuset i Nord-Norge. Forutsetningene har helt enkelt ikke vært tilstede for å bruke DEA-analyser til kartlegging av effektivitetsnivåene. For denne oppgaven er intensjonen derfor å teste metodens anvendelighet til effektivitetsanalyser.

1.2 Problemstilling

Formålet med oppgaven er å anvende DEA-analyser til kartlegging av effektivitetsgrad, nærmere bestemt tilstedeværelse av ineffektivitet, på sykehusnivå. For å belyse problemstillingen vil effektivitetsanalysene anvendes på ortopediske operasjoner ved Universitetssykehuset i Nord-Norge. Hensikten er å teste metodens anvendelighet på et ikke-utprøvd datamateriale, herunder finne en velfungerende og reliabel måte for måling av prestasjonsnivå. Intensjonen er å finne ut om metoden er egnet og nyttig som verktøy til denne kartleggingen, slik at resultatene kan legges til grunn ved planlegging-, styrings- og forbedringsarbeid. Herunder er følgende problemstilling formulert:

Er DEA-analyser velfungerende til å kartlegge effektivitetsnivåer ved ortopediske operasjoner ved UNN, og på hvilken måte kan analysene bidra til planlegging av effektiviseringsarbeid?

For en mer presis og ryddig besvarelse av problemstillingen deles denne inn i tre forskningsspørsmål:

1. Er DEA-analyser av typen beste praksis velfungerende til kartlegging av effektivitetsnivået i ortopediske operasjoner ved UNN?

2. Er DEA-analyser av typen verste praksis anvendelig til testing av ineffektivitet blant ortopediske operasjoner ved UNN, og i hvilken grad er denne mer hensiktsmessig enn modellen for beste praksis?

3. På hvilken måte kan UNN anvende resultatene fra analysene gjort med beste- og verste-praksis til kartlegging og planlegging av et eventuelt forbedringsarbeid?

1.3 Formål og avgrensning

Formålet med studien har vært å undersøke om DEA-metoden kan være en hensiktsmessig metode for kartlegging av effektivitetsnivåer, som videre kan bidra i arbeidet med å redusere ineffektivitet. Metoden utprøves med datamateriale fra UNN, og det forsøkes å avdekke effektivitetsnivåene ved utvalgte ortopediske operasjonstyper. Ved å identifisere hvilke operasjoner og operasjonstyper som er mest effektive og ineffektive vil en kunne konstatere og kartlegge områder med potensielt forbedringspotensial. Det tas utgangspunkt et utvalg ortopediske operasjonstyper til analysen. Dette er operasjoner med forholdsvis lav risiko, samt relativt standardiserte prosedyrer.

Det er viktig å presisere at oppgaven undersøker anvendeligheten til metoden, og søker finne hvor eventuelle tiltak kan iverksettes. Det vil ikke være ønskelig eller hensiktsmessig å tilstrebe perfekte effektivitetsnivåer, siden mennesker og helsetjenester er sentrale i produksjonen. I denne analysen produseres kompliserte prosedyrer og behandlinger for pasienter med komplekse sykdomsbilder. I lys av pasientsikkerhet og kvaliteten på behandlingene, vil studien heller utarbeide analyser for det tekniske effektivitetspotensialet. Hvordan og i hvilken grad utbedringer kan gjennomføres overlates til de sakkyndige.

På bakgrunn av tidligere forskning mistenkes tilstedeværelse av ineffektivitet. Sammenholdt med det faktum at et fullkommet effektivitetsnivå i denne typen helsetjenester ikke er hensiktsmessig, vil det være interessant å rette analysen mot ineffektivitet. Det legges vekt på å kartlegge forbedringspotensialet blant de mest ineffektive operasjonene, da en gjennom å forbedre nivåene blant disse muligens vil kunne heve det gjennomsnittlige effektivitetsnivået. Det benyttes derfor både beste-praksis- og verste-praksismodell for å belyse effektivitetsnivåene. Metodene kombineres for å forsøke å avdekke naturlige grupperinger blant operasjonene, slik at en lettere skal kunne se trender blant de ineffektive operasjonene. Dette vil kunne danne grunnlaget for en mer hensiktsmessig effektivisering.

1.4 Oppgavens struktur

Oppgaven innledes med en kort presentasjon av situasjonen i spesialisthelsetjenesten i Norge, og at en av utfordringene innenfor helsesektoren er å få ressursene til å strekke til i en stadig mer krevende sektor. Her fremkommer også motivasjonen bak, samt aktualiteten i temaet for oppgaven og problemstillingen. For å ha tilstrekkelig kunnskap om bransjen og casebedriften presenteres norsk helsevesen i kapittel 2, herunder legges det særlig fokus på Universitetssykehuset i Tromsø. Hensikten er å presentere helsesektorens strukturer og oppbygging, slik at en har god forståelse for de bakenforliggende faktorene før det tas fatt på utprøving av analysene på datamaterialet fra UNN. Kapittel 3 tar for seg tidligere studier og forskning gjort innenfor helsesektoren, som på ulike måter belyser anvendelsen av DEA-analyser for helse- og sykehusenheter. Hensikten med kapittelet er særlig kartlegging av annen forsknings erfaringer fra DEA, gjennom deres valg og argumentasjoner for variabler og modeller.

I kapittel 4 presenteres anvendt teori og metode som danner grunnlaget for å besvare den aktuelle problemstillingen. Dette omfatter i hovedsak teori om måling av effektivitet og DEA-analyser, samt sensitivitets- og statistiske analyser for å avgjøre presisjon i datasett og funn. Videre i kapittel 5 presenteres datagrunnlaget, valget av variabler og modellene som benyttes i analysene. Hensikten er å forklare variablene som benyttes i de ulike modellene, og presentere det endelige datasettet etter at sensitivitetsanalyser er gjennomført.

I kapittel 6 presenteres resultatene fra DEA-analysene med beste- og verste praksis, samt hvordan disse kan benyttes sammen i den såkalte layering-teknikken. I tillegg presenteres resultatene og anvendeligheten til statistiske tester for kartlegging av effektivitetsforskjeller, herunder bankertester. I kapittel 7 diskuteres funnene i analysen og deres anvendelighet for UNN i et eventuelt effektiviseringsarbeid, og konklusjonen følger i kapittel 8, som svar på problemstillingen. Til slutt gjøres det opp noen tanker om svakhetene i analysen og videre hva som kunne løftet studien eller tatt den noen steg videre.

2 Det norske helsevesenet

Staten skal gjennom helse- og omsorgsdepartementet sørge for trygge og gode tjenester til alle i velferdssamfunnet Norge. Det er avgjørende at utgiftsveksten dempes, men på samme tid omhandler dette fundamentale helsetjenester hvor kutt og nedskjæringer verken er ønskelig eller hensiktsmessig. Hva gjør en da, når den tiltakende etterspørselen ikke kan imøtekommes med tilsvarende økning i midler og sysselsetting?

Et viktig steg på veien mot bedre helsetjenester, uten ytterligere kostnadsøkninger, kan være å optimalisere utnyttelsen av tilstedeværende midler og ressurser. Det er nok også derfor fokuset på styrket produktivitet og effektivitet aldri har vært større innen helse- og omsorg. På samme tid er det snakk om store, byråkratiske og kompliserte institusjoner. Spørsmålet blir da hvor en finner potensialet for besparelser, effektivisering og prestasjonsforbedringer. I dette kapitlet gjør vi rede for caset som anvendes i denne studien, gjennom helse- og omsorgssektoren og Universitetssykehuset i Nord-Norge.

2.1 Oppbyggingen av norsk helsevesen

Helse- og omsorgsdepartementet er det øverste helseorganet i Norge, og har det overordnede ansvaret for å gi den norske befolkningen likeverdige helse- og omsorgstjenester (Regjeringen, 2013). Departementet har åtte avdelinger, hvor spesialisthelsetjenesten er en av disse. Begrepet spesialisthelsetjeneste omfatter sykehus, poliklinikker, legespesialister og ambulansetjenester, og disse tjenestene driftes av de ulike regionale helseforetakene i Norge.

Helsetjeneste er et fellesbegrep for alle offentlige og private institusjoner som har som formål å forebygge, diagnostisere og behandle somatisk og psykisk sykdom, tilbud av pleie og omsorg, samt rehabilitering. Helsetjeneste og helsevesen er begrep som ofte brukes om hverandre og som har samme betydning. Det skilles mellom somatisk og psykisk helsetjeneste, hvor somatisk helsetjeneste er behandling av fysiske plager og sykdommer, og psykisk helsetjeneste omfatter psykologiske behandlingsmetoder og rus. Polikliniske tjenester er spesialiserte undersøkelser eller behandling som kan gjennomføres uten at pasienten må legges inn på sengepost (Store norske leksikon, 2009). En sengepost er en avdeling hvor pasienter legges inn for døgnopphold. I 2015 hadde UNN 350 000 polikliniske konsultasjoner som også inkluderer dagkirurgi (UNN, 2016).

En kategori innenfor somatisk helsetjeneste er ortopedi, hvilket er læren om skjelettet og muskulaturen samt skader og sykdommer knyttet til disse. Eksempler på ortopediske sykdommer kan være betennelser, svulster og slitasje, mens skader kan omfatte benbrudd, leddbånd-, muskel- og seneskader (Norsk helseinformatikk, 2016). Behandling kan være kirurgisk, eller ikke-kirurgisk i form av blant annet fysioterapi.

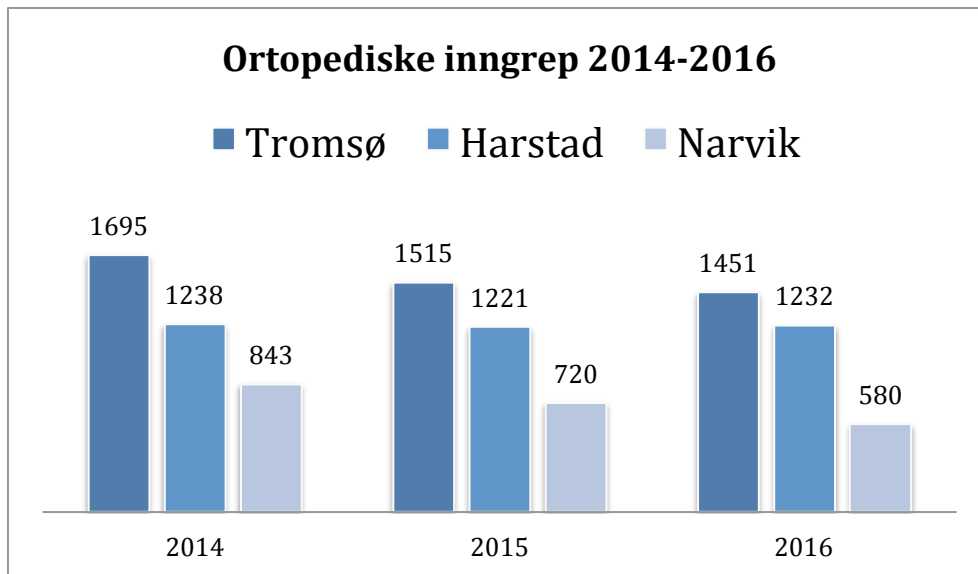
Det er i alt fire regionale helseforetak som har det utøvende ansvaret for behandlingstilbudet innenfor spesialhelsetjenesten, Helse Sør-Øst, Helse Vest, Helse Midt-Norge og Helse Nord. Det utøvende ansvaret henvises til som “sørge-for”-ansvaret, og er fastsatt i spesialisthelsetjenesteloven.

Det geografiske ansvarsområdet til Helse Nord omfatter hele Nord-Norge samt Svalbard. Helse Nord eier fem ulike helseforetak i regionen, og har i tillegg samarbeid med flere private leverandører av helsetjenester, samt privatpraktiserende leger og psykologer. Det overordnede målet til Helse Nord er å tilby de 200 000 innbyggerne i regionen tilstrekkelig og god tilgang til helsetjenester som er trygge og faglig forsvarlig (Helse Nord, 2017).

2.1.1 Universitetssykehuset i Tromsø

Universitetssykehuset i Nord-Norge er et av helseforetakene som eies av Helse Nord. Sykehuset har som hovedmål å behandle pasienter, utdanne helsepersonell og drive med forskning. Hovedsetet ligger i Breivika i Tromsø, men UNN har også behandlingssteder i Harstad, Narvik, Longyearbyen, Finnsnes og på Storslett. I 2015 hadde UNN 882 årsverk for leger og 1 868 for sykepleiere.

UNN utførte i 2015 til sammen 17 200 kirurgiske operasjoner ved operasjonsstuene i Tromsø, Harstad og Narvik (UNN, 2017), hvorav 3 456 av disse var ortopediske inngrep. Figur 1 viser fordelingen av ortopediske operasjoner i Tromsø, Harstad og Narvik fra 2014 til 2016.



Figur 1 Ortopediske operasjoner ved UNN 2014-2016

UNN er tilknyttet det helsevitenskapelige fakultetet ved UiT - Norges arktiske universitet, og hadde rundt 2 200 praksisopphold for studenter ved sykehusets ulike avdelinger i 2015. Spennet går fra bachelor- til doktogradstudenter innenfor medisin- og helsefag, samt veiledning av leger i spesialisering, heretter referert til som LIS-leger. En LIS-lege er lege som har fullført medisinstudiet og 1,5 år med turnustjeneste, og som har søkt på en spesifikk stilling for spesialisering innenfor et fagfelt. I hovedsak deltar det alltid en LIS-lege på ortopediske inngrep ved UNN i Tromsø, og det hører sjeldenhetene til at én ikke er tilstede. I denne oppgaven tas det derfor utgangspunkt i at det alltid er en LIS-lege tilstede på operasjonene. Planlagte operasjoner refereres til som elektiv operasjon, og disse planlegges 4-6 måneder i forveien. Øyeblikkelig hjelp oppstår ved akutt skadeomfang som må behandles innen 72 timer for å avverge alvorlige helseskader. Disse type operasjonene omtales som ø-hjelp.

Operasjon defineres som en prosedyre som til vanlig utføres i en operasjonsstue, og kan deles inn i to typer, kirurgi og dagkirurgi. Kirurgi brukes når en pasient har vært på sengepost før og/eller etter inngrepet. Dagkirurgi er når pasienten kommer inn på sykehuset og drar hjem igjen samme dag som operasjonen gjennomføres. Dette kan være mindre operasjoner som for eksempel hallux valgus, hvilket er en relativt enkel og standardisert fotoperasjon.

2.2 Inntekter og finansiering

2.2.1 Innsatsstyrt finansiering (ISF)

De regionale helseforetakenes finansiering kan deles i basis- og aktivitetsbasert. Førstnevnte er behovsstyrte rammetilskudd som er hovedkilden til helseforetakenes finansiering.

Aktivitetsbasert er på den andre siden finansiert via innsatsstyring. Somatiske helsetjenester hadde i 2015 en 50/50 fordeling mellom basisbevilgning og aktivitetsbasert finansiering.

Hovedformålet med innsatsstyrt finansiering er å understøtte “sørge-for” ansvaret til de regionale helseforetakene (Helsedirektoratet, 2017b).

2.2.2 Diagnoserelatert gruppering (DRG)

DRG er et pasientklassifiseringssystem som benyttes som grunnlag for beregning av innsatsstyrt finansiering. Ved å klassifisere behandlingen i grupper som er medisinsk meningsfulle og ressursmessig tilnærmet homogene, kan man lage fornuftige klassifiseringsgrupper som beskriver pasientsammensetningen og følgende behandlingsløp. Grupperingen gjøres på bakgrunn av data som registreres i sykehusets pasientadministrative datasystem. Valget av diagnosegruppe baseres på hoveddiagnosen som settes og eventuelle bidiagnoser og prosedyrer eller operasjoner som utføres. I tillegg spiller kjønn og alder inn som sentrale variabler.

Et av hovedprinsippene ved DRG er at kostnadene som oppstår i pasientforløpet skal henføres og beregnes ut fra hvor de har oppstått. Summen av kostnader består altså av den medisinske servicen som er utført, men også administrasjonskostnader er inkludert i tallet i henhold til selvkostprinsippet.

2.2.3 Kostnad per pasient (KPP)

Helse- og omsorgsdepartementet ytret tidlig på 2000-tallet et ønske om å utvikle og etablere en nasjonal KPP-database. Erfaringer fra andre nordiske land viste at bruken av KPP økte kvaliteten på både aktivitets- og kostnadsdata. I 2012 presenterte derfor departementet en nasjonal spesifisering for hvordan de ulike helseforetakene skulle bygge opp kostnad per pasient. Hensikten med en nasjonal standard er å sikre lik praksis på tvers av foretakene, og samtidig kunne sammenligne på tvers av helseforetakene. Årsaken til at Helse- og omsorgsdepartementet ikke har utviklet en KPP-modell på nasjonalt nivå er at helseforetakene er ulike i størrelse og organisasjon, og det tilrettelegges for at foretakene kan

tilpasse sin modell etter behov og struktur. På denne måten vil KPP-modellene være grunnleggende likt på et nasjonalt nivå og kan derfor sammenlignes, samtidig som de ulike foretakene har en modell som passer deres respektive struktur.

3 Litteraturgjennomgang

Som illustrert av Liu, Lu, Lu & Lin (2013) har DEA-metoden på relativt kort tid bemerket seg som et utmerket verktøy til prestasjons- og effektivitetsvurderinger, og blitt en populær metode for effektivitetsanalyser. De påpeker også at DEA har vært mye brukt innenfor offentlig sektor og effektivitetsmåling for helseforetak. Tidligere studier har anvendt metoden på utallige områder, enheter og nivåer, med utgangspunkt i svært forskjellige beslutningsenheter. Til tross for et stort antall studier med DEA-analyse innen helse- og omsorgssektoren, er det relativt få som har undersøkt effektivitet på operasjonsnivå. Majoriteten av studiene undersøker effektivitet på mer overordnede nivåer, blant annet til sammenligning av sykehusstrukturer for å undersøke effekten av reformer og innføringer, eller for prestasjonsmåling av sykehus på tvers av landegrenser.

Hensikten med litteraturgjennomgangen har hovedsakelig vært å få innsikt i metodeanvendelsen. I tillegg har økt kunnskap om den komplekse helse- og omsorgssektoren, samt operasjonsaktiviteter vært essensielt. Særlig vekt har blitt lagt på kartlegging av forhold som bør tas hensyn til i studiens operasjonalisering og i forberedelsene til analysene.

I en artikkel vedrørende fallgruver i DEA-analyser hevder Dyson, Allen, Camanho, Podinovski, Sarrico og Shale (2001, s.245) følgende:

The application of DEA presents a range of issues relating to the homogeneity of the units under assessment, the input/output set used, the measurement of those variables and the weights attributed to them in the analysis. Each of these issues can present practical difficulties in applying DEA [...]

Det blir derfor betydningsfullt å dra lærdom fra andre studiers modellvalg, samt undersøke forarbeidet utført og variablene disse forskningsprosjektene har benyttet til DEA-analyser for effektivitetsvurderinger i sykehus. I dette kapittelet blir det følgelig gjort rede for tidligere forskning med effektivitetsanalyser. Det trekkes frem studier med prestasjonsmålinger gjennomført på sykehus- og operasjonsnivå, samt forskning gjort for UNN spesifikt. En rekke

effektivitetsstudier utført på andre fagområder er gjennomgått, blant annet fiskeri, oppdrett og skipsfart. Disse gjengis ikke i litteraturgjennomgangen, men har blitt benyttet til å tilegne kunnskap om metoden.

3.1 Technical Efficiency in the Clinical Management of Critically Ill Patients

Puig-Junoy (1998) fremmet DEA som verktøy for å kartlegge grad av teknisk effektivitet på pasientnivåer. Studien illustrerer herunder DEAs anvendelighet på mikro- og mesonivå, ved blant annet å benytte pasienter som DMUer. Det legges også stor vekt på en nøkkelbegrensning vedrørende effektivitetsanalyser på helse- og omsorgssektoren. Det diskuteres at en vil ha vanskeligheter med å danne gode mål for output, siden produksjonen innen helsetjenester nærmest utelukkende består av tjenester. Nærmere bestemt ulike behandlinger for å skape bedre helse. Kvalitetsaspektet ved behandlingene er utfordrende å hensynte og det argumenteres for at kvaliteten ofte glemmes når en benytter pasientbehandlinger som mål på aktivitet. Denne begrensningen i effektivitetsanalyser i helsetjenester bør en være ydmyk for, da kvalitetsdimensjonen utvilsomt er viktig i denne bransjen. En vil i følge forskerne ha utfordringer med å finne anvendelige variabler til analysen som gir uttrykk for sykdommers alvorlighetsgrad, vellykkethet i behandlingene, samt påvirkning av livskvalitet.

I Puig-Junoy's studie hevdes det likevel at DEA er den beste rangeringsmetoden for måling av effektivitetsnivået innen sektoren. Dette på bakgrunn av DEAs tilpasningsevner rundt operasjoners kompliserte variabler. Gjennom analysen bevises metodens anvendelighet, både for operasjonsstuers høye antall variabler, men også de komplekse egenskapene til produksjonsvariabler og innsatsfaktorer. I tillegg viser studien at outputvariablene ikke nødvendigvis behøver å være profittorienterte. I studien benyttes data innsamlet for et annet prosjekt, hvilket forsøkte utbedre systemene for kartlegging av pasienters sykdomsbilde. Det var derfor tilrettelagt for en effektivitetsanalyse som kunne hensynte kvalitetsdimensjonen, gjennom et stort utvalg kvalitetsvariabler. Eksempelvis ble det benyttet variabler for sykdommens alvorlighetsgrad ved innleggelse og utskrivelse for de ulike pasientene.

3.2 Sykehusproduktiviteten etter statlig overtakelse: En nordisk komparativ analyse

Kittelsen, Magnussen og Anthun (2007) presenterer en produktivetsstudie av norske sykehus gjennomført med DEA i forbindelse med en større forskning på skandinaviske sykehus. Ressursforbruket måles her på basis av driftskostnader. Tjenesteproduksjonen for seks ulike diagnosegrupper benyttes som mål på aktiviteten, hvor DRG-systemet ble avgjørende for inndelingen. Innsatsfaktorene bygges opp av tre elementer, herunder lønn til leger, lønn til operasjonssykepleiere, samt en post for diverse andre kostnader. Det benyttes seks ulike variabler på outputsiden. De fleste av disse dekomponeres fra produksjonen av henholdsvis DRG-poeng, og deles herunder inn i fem ulike variabler. Blant annet dagpasienter, kirurgiske pasienter og polikliniske konsultasjoner. Det fremheves at DRG-systemet er relativt likt på tvers av de nordiske landene. Konstruksjonen av variablene kan derfor gjøres på samme måte i det større forskningsprosjektet, hvilket fundamentere sammenlignbare resultater.

3.3 Sykehusenes effektivitetsutvikling 1992-1999: Hvilke effekter ga innsatsstyrt finansiering?

Hagen et al. (2001) presenterer i en rapport hvordan effektivitetsnivået var ved norske sykehus gjennom 90-tallet. Dette ble gjort i forbindelse med innføring av innsatsstyrt finansiering på slutten av tiåret, hvilket går ut på at fylkeskommunenes utgifter til behandling av pasienter refunderes av staten. Innføringen var ment å motivere til økt antall behandlinger, uten at dette skulle gå på bekostning av kvalitet eller effektivitet. Datamaterialet til analysene ble i hovedsak innhentet fra SAMDATA, et årlig rapporteringssystem om utvikling og utnyttelse av ressurser i spesialisthelsetjenesten. Modellen for å teste teknisk effektivitet har inputvariabler bestående av årsverk, fordelt på leger og annen arbeidskraft, samt medisinske forbruksvarer som medisiner og lignende. Det argumenteres her for at årsverkene til leger bør være en egen variabel, da de har en sentral rolle i hovedvekten av aktivitetene ved sykehusene. Det benyttes bare én outputvariabel, hvilket er en sammensatt variabel for innleggelser og polikliniske refusjoner. Rapporten konkluderer med signifikant økning i gjennomsnittlig effektivitet for sykehusene, da særlig for region- og sentralsykehusene.

3.4 Tidligere effektivitetsanalyser ved UNN

Universitetssykehuset i Nord-Norge har også vært gjenstand for prestasjons- og effektivitetsanalyser. Gjennom ulike prosjektarbeider har sykehuset blant annet forsøkt å kartlegge og videre også heve prestasjons- og effektivitetsnivået. Dons og Skorstad (2012) gjennomførte en effektivitetsanalyse vedrørende kapasitetsutnyttelsen av operasjonsstuen i en masteroppgave ved Handelshøgskolen i Tromsø. Intensjonen var å benytte DEA som metode for effektivitetsanalysen, men datamaterialet de hadde til rådighet var uten kostnadsvariabler hvilket er utfordrende for en slik analyse. Metoden ble således mer beskrivende, og uten DEA-analyse for effektivitetsmålingen. De benyttet i hovedsak tidsberegninger til kartlegging av effektivitetsnivået i de ulike operasjonsstuen. Studien fant stor grad av ineffektivitet og videre et stort forbedringspotensial. Det påpekes likevel at en DEA-analyse som kunne benyttet DRG-poeng, samt kostnader på operasjonsnivå kunne være interessant.

I en internrapport ved navn «Optimal ressursutnyttelse av operasjonskapasiteten i UNN» (2008), presenterer en styringsgruppe hovedutfordringene og nødvendige tiltak for å bedre den suboptimalt utnyttede operasjonskapasiteten ved UNN. Bakgrunnen for prosjektet var blant annet at UNN vedtok innsparingskrav på nærmere 200 millioner, hvor det var enighet om at stillingskutt ikke kunne være løsningen i kjernevirksomheten til sykehuset. Det ble derfor prekært å optimalisere utnyttelsen for å redusere ressursforbruket i operasjonsaktiviteten. Utredning og statusbeskrivelse baserte seg på måleparametere som ventelister, kartlegging av aktivitet på operasjonsstuen, samt antall gjennomførte og utsatte operasjoner. Det fremkommer ikke hvilke analyseverktøy som eventuelt er benyttet, men mye ser ut til å baseres på kunnskap og erfaringer hos styringsgruppens medlemmer. Det ble også intervjuet eksperter innen felt som operasjonsteam, personlige ventelister for kirurger og omorganisering- og strukturering av øyeblikkelig hjelp kirurgi.

Konklusjonen i internrapporten bygget også på sammenligninger med liknende sykehus som kunne dokumentere betydelig bedre utnyttelse av operasjonskapasiteten, samt tidligere utredninger ved sykehuset som viste til sløsing av tid og ressurser. Styringsgruppen erkjente følgelig at bakgrunnen for den dårlige ressursutnyttelsen besto av flere faktorer, og tiltak ville måtte iverksettes gjennom hele pasientflyten. Operasjoner kan ikke effektiviseres i et isolert prosjekt, herunder tok de for seg flaskehalsen i hele prosessen, fra pasientene henvises til kirurgi til pasientene er utskrevet fra sykehuset. Tiltakene som ble lagt frem var blant annet å

skille ut egne stuer og team til øyeblikkelig hjelp operasjoner, innføre vaksjefer på de ulike stueene for å organisere aktiviteten, samt å sette tydeligere tidsfrister til både pasienter og de ansatte.

I en masteroppgave av Martinussen og Reinholtsen (2013) var UNN en del av utvalget hvor hensikten med oppgaven var å undersøke forholdet mellom prosesskvalitet og effektivitet i somatisk spesialisthelsetjeneste. Det ble beregnet prosesskvalitet, kostnadseffektivitet og teknisk effektivitet med DEA-analyser. Teknisk effektivitet ble analysert med variabler som DRG-produksjon, årsverk til personell og døgnplasser. Prosesskvaliteten ble målt med variabler for henholdsvis epikriser sendt og operasjoner utført innen gitte tidsrammer, samt strykninger og antall korridorpasienter. De kunne ikke påvise korrelasjon mellom kvalitet og effektivitet, men avviser heller ikke at det eksisterer en direkte relasjon da resultatene kan være påvirket av svakheter i metoden eller kompleksiteten i kvalitet- og effektivitetsmåling.

De omtalte studiene viser til DEA som et velfungerende verktøy til analysing av effektivitet innen sykehussektoren, både på overordnet- og mikronivå. Et annet poeng som kommer frem er DEAs fleksibilitet og allsidighet, ved at metoden kan anvendes på svært mange og ulike områder. Studienes valg og erfaringer rundt operasjonalisering av data, samt modeller og benyttede variabler er nyttig i valgene som gjøres for denne studien. Kvalitetsdimensjonen er utvilsomt utfordrende å inkludere i analysen, da tidligere forskning trekker dette frem som en nøkkelutfordring. DRG-poengene anses som velfungerende da disse gir uttrykk for sykdomsbildet, samt prosedyrer og komplikasjoner i behandlingsforløpet. Det legges også merke til at årsverk og annet medisinsk materiell går igjen som kostnadsvariabler blant innsatsfaktorene.

4 Kapittel 4 – Teori og metode

4.1 Produktivitet og effektivitet

Kittelsen & Førsum (2001, s. 23) definerer produktivitet som «forholdet mellom produksjon og ressursbruk» og effektivitet som «forholdet mellom faktisk produktivitet og en norm for best mulig produktivitet gitt de begrensningene en virksomhet driver innenfor».

Begrepet effektivitet kan deles i ytre og indre effektivitet. Ytre effektivitet går ut på å gjøre de rette tingene, mens indre effektivitet handler om å gjøre ting korrekt, som er et spørsmål om hvorvidt produksjonen foregår med bruk av minst mulig ressurser (Kittelsen og Førsum, 2001). Beregningen av produktivitet med én input og én output gjøres med følgende formel:

$$(1) \quad \text{Produktivitet} = \frac{y}{x} = \frac{\text{Output}}{\text{Input}}$$

Produktivitetstallet vil ikke alene si noe om ressursutnyttelsen i selskapet, men kan benyttes til å sette relative mål. Dette er en enkel manøver når bedriften har et én til én forhold mellom input og output, men i virkeligheten benyttes det ofte flere innsatsfaktorer for å produsere et eller flere typer produkter eller tjenester. Totalfaktorproduktivitet (TFP) er en metode for å måle produktivitet som inkluderer alle faktorer i produksjonen (Coelli, Rao, O'Donnell og Battese, 2005). Her vektet variablenes relative betydning i produksjonsprosessen, og produktivitetmålene måles gjennom pris- eller verdivektorer, herunder u_r og v_i , for de ulike variablene. Matematisk uttrykkes det slik:

$$TFP = \frac{\sum_{r=1}^s u_r y_r}{\sum_{i=1}^m v_i x_i} = \frac{\text{vektet sum output}}{\text{vektet sum input}}$$

$$r = (1 \dots, s)$$

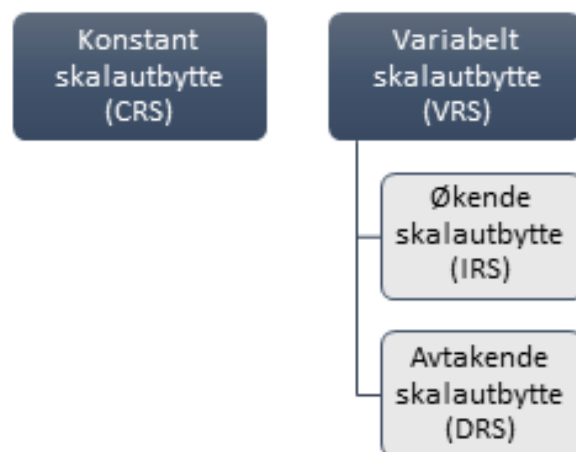
$$(2) \quad i = (1 \dots, m)$$

4.2 Teknisk effektivitet

I følge Farrell (1957), som er den første til å omtale “teknisk effektivitet”, vil forbedringer i effektivitet være en proporsjonal reduksjon av innsatsfaktorene. Han skiller også mellom målinger som er faktorbesparende og produksjonsøkende, herunder inputminimerende og outputmaksimerende problemer (Kittelsen og Førstund, 2001). Videre kan en DMU regnes å være teknisk effektiv når den ikke kan redusere én enhet input, uten å simultant måtte øke minst én av de andre inputene eller eventuelt redusere outputproduksjonen med minimum én enhet. For en outputorientert produksjon vil teknisk effektivitet være når en ikke kan øke produksjonen med én output uten og samtidig redusere produksjonen av minst én annen output, eller eventuelt øke bruken av minst én input. En noe forenklet forklaring er at teknisk effektive enheter ikke sløser med ressursene sine. Teknisk effektivitet forklares også ytterligere i forbindelse med DEA-analyser i kapittel 4.4.

4.3 Skalaegenskaper

Skalaegenskapene til beslutningsenhetene forklarer i hvilken grad det er proporsjonal økning i input når en øker output. Som vist i figur 2 skiller det mellom ulike former for skalautbytte, hvor en hovedsakelig deler mellom konstant skalautbytte (CRS) eller variabelt skalautbytte (VRS).



Figur 2 Konstant og variabelt skalautbytte

Ved den teknologiske forutsetningen i konstant skalautbytte vil det være proporsjonal økning i input og output. Variabelt skalautbytte har ikke dette radiale forholdet, og deles videre inn i økende skalautbytte og avtakende skalautbytte, hvor i henhold til Banker, Charnes og Cooper (1984) et økende skalautbytte vil ha større økning i output enn i input og det motsatte vil være tilfelle ved DRS.

Banker et al. (1984) presenterer følgende matematiske uttrykk:

- (3) Konstant skalautbytte: $f(a * x_1, a * x_2) = a * f(x_1, x_2)$
- (4) Økende skalautbytte: $f(a * x_1, a * x_2) < a * f(x_1, x_2)$
- (5) Avtakende skalautbytte: $f(a * x_1, a * x_2) > a * f(x_1, x_2)$

Skalaeffektiviteten er et mål på hvorvidt selskapene opererer under riktig skala. I tilfellene hvor effektivitetsmålet til analysen er forskjellig når det forutsettes henholdsvis konstant- eller variabelt skalautbytte, vil dette derfor kunne forklares med skala-ineffektivitet.

Effektivitetsmålet i analyser med antakelse om variabelt skalautbytte uttrykker den rene tekniske effektiviteten, mens både teknisk- og skalaeffektivitet inngår i effektivitetsmålet fra analyser med antakelse om konstant skalautbytte. Der en i analyser hvor det forutsettes konstant skalautbytte anmoder at skala ikke påvirker mulig produktivitet, vil en ved å forutsette variabelt skalautbytte hensynta skala. Det matematiske uttrykket dekomponerer den tekniske effektiviteten gjennom å dele gjennomsnittet av effektivitetsmålene på hverandre. Denne beregnes slik:

$$(6) \quad \text{Skalaeffektivitet (SE)} = \frac{TE_{CRS}}{TE_{VRS}}$$

Beslutningsenheten er skalaeffektiv når dette målet er lik 1. Dersom skalaeffektiviteten beregnes å være signifikant forskjellig fra én vil dette indikere at restriksjonen i VRS er avgjørende. Videre identifiseres tilstedeværelse av skalaineffektivitet. Skalaineffektivitet skyldes at observasjonen ikke har optimal størrelsesorden. Skalaeffektivitet kan derfor forklares som et mål for hva enheten har på å tjene på å gjøre justering av skala (Banker et al., 1984)

4.4 DEA – beste praksis

Majoriteten av DEA-studiene legger til grunn effektivitetsberegningen til Farrell (1957). Effektivitetsmålet er da basert på proporsjonal (radial) skalering av ineffektive observasjoner, mot en effektiv front. Farrell diskuterer videre to muligheter for definering av denne produksjonsfronten. Herunder én teoretisk og én empirisk tilnærming. Førstnevnte innebærer at fronten representerer den beste praksis som kan anses med et teoretisk fundament. Den empiriske tilnærmingen tar utgangspunkt i at fronten representerer den beste observerbare praksisen i utvalget. På bakgrunn av kompliserte og komplekse produksjonsprosesser og det faktum at realistiske mål er mer anvendelige, argumenterer Farrell (1957) for den empiriske tilnærmingen. Dette legges også til grunn for DEA-analyser. For DMUene undersøkes det da hvor stor proporsjonal ressursreduksjon som er mulig, uten at reduksjonen går på bekostning av den gitte mengden output. Eventuelt kartlegges størst mulig proporsjonal økning av produksjonen, med de gitte inputnivåene.

Data Envelopment Analysis (DEA) er en deterministisk, ikke-parametrisk metode for vurdering av relativ effektivitet. Gjennom lineær programmering evalueres et sett beslutningsenheter, med tilhørende observasjoner, mot hverandre i form av benchmarking. Beslutningsenhetene vil heretter også refereres til som DMU (Decision Making Unit). Metoden baseres blant annet på forutsetninger om at beslutningsenhetene har lik tilgang på ressurser og at de opererer i forholdsvis like miljøer. De må med andre ord være relativt homogene i produksjonsprosessen.

DMUene rangeres basert på deres relative effektivitetsgrad. De mest effektive enhetene i datasettet regnes som 100% effektive for det respektive utvalget, og konstruerer effektivitetsfronten. Videre vil denne fronten innhulle produksjonsmulighetsområdet hvor de øvrige observasjonene befinner seg. Effektivitetstallet beskriver hvordan den enkelte DMUs effektivitetsnivå er i forhold til det øvrige referansesettet. I likhet med tradisjonell effektivitetsberegning, vil graden av effektivitet besluttes på bakgrunn av avstanden mellom dens observerte nivåer for input og output, samt distansen til effektivitetsfronten. Dette utmåles på en skala fra 0 til 1, hvor de effektive enhetene langs fronten har effektivitetstall 1. De effektive beslutningsenhetene kan sies å ha den mest optimale sammensetningen av innsatsfaktorer og produkter. Implisitt er disse derfor læremestere og gode referanseenheter for de ineffektive DMUene.

Effektivitetsmålene i DEA-analyser tar utgangspunkt i de observerbare nivåene innenfor det gjeldende referansesettet. Ressursene er derfor ikke nødvendigvis maksimalt utnyttet av DMUene som danner fronten og anses effektive. Det reelle mulighetsområdet kan være større, slik det nevnes innledningsvis. I noen tilfeller vil det være et teoretisk rom for å øke effektivitetsgraden ytterligere også for de effektive observasjonene. Referansesettet begrenser på denne måten seg selv og utnytter derfor ikke nødvendigvis sitt fulle potensiale, ved at de ineffektive sikter etter nivået til de beste i klassen, herunder er navnet på metoden også beste praksis.

DEA regnes å være spesielt anvendelig når en opererer med flere output og input på samme tid, men også særlig for analyser av organisasjoner hvor en ikke opererer med et profittmotiv. Charnes, Cooper og Rhodes (1978) var de første til å introduserte DEA som metode for måling av effektivitet. I studien knyttet de inn teknikker fra lineær programmering for å kartlegge relativt effektivitetsnivå hos et sett observasjoner, med antakelse om konstant skalautbytte. Metoden kjennes som CCR metoden etter forfatterne, eller CRS-metoden fra engelsk: "Constant Returns to Scale". Senere har Banker, Charnes og Cooper (1984) videreutviklet modellen ved å tilegne variabelt skalautbytte. Etter forfatterene har denne modellen fått tilnavnet BCC-modellen, men omtales også for VRS-modellen på bakgrunn av det engelske navnet "Variable Returns to Scale".

En vil få både en primal- og en dualformulering ved DEA-analyser, siden LP-problemer oppstår i par. Den primale formen presenteres som multiplikatormodellen og dualformen som omhyllingsmodellen. Hvilken modell som velges vil ikke være avgjørende for resultatene, men vil måtte hensyntas ved tolkninger av analysens resultater.

4.4.1 CCR-modellen

Det er mulig å anvende både input- og outputorientering i CCR-modeller. I første omgang presenteres den inputminimerende CCR-modellen, hvor en sikter etter å redusere ressursbruk uten at det går på bekostning av produksjonen. Coelli et al. (2005, s 163) presenterer denne DEA-modellen matematisk på følgende måte:

$$\min \theta_0$$

Når

$$-y_{r0} + \sum_{j=1}^n \lambda_j y_{rj} \geq 0$$

$$\theta_0 x_{i0} - \sum_{j=1}^n \lambda_j x_{ij} \geq 0$$

$$\lambda_j \geq 0$$

$$(r = 1 \dots, s)$$

$$(i = 1 \dots, m)$$

$$(7) \quad (j = 1 \dots, n)$$

θ_0 er en skalar og λ er 1x1 vektor av konstanter. x_i er innsatsfaktoren og y_r er observert output. Beslutningsenhetenes effektivitetsmål beregnes ved å kalkulere verdien på θ_0 , og verdien vil være innenfor en på skala fra 0 til 1, hvor tallet 1 indikerer effektiv frontplassering. Den første restriksjonen indikerer videre et minimalt forbruk av input x_i . De ineffektive vil få en annen beslutningsenhets λ til referanse, mens de effektive vil være deres egen referanseenhets. Størrelsen på λ vil avhenge av hvilket datasett som benyttes. Referanseenhets sier noe om potensialet for inputminimering for de ulike beslutningsenhetene i datasettet.

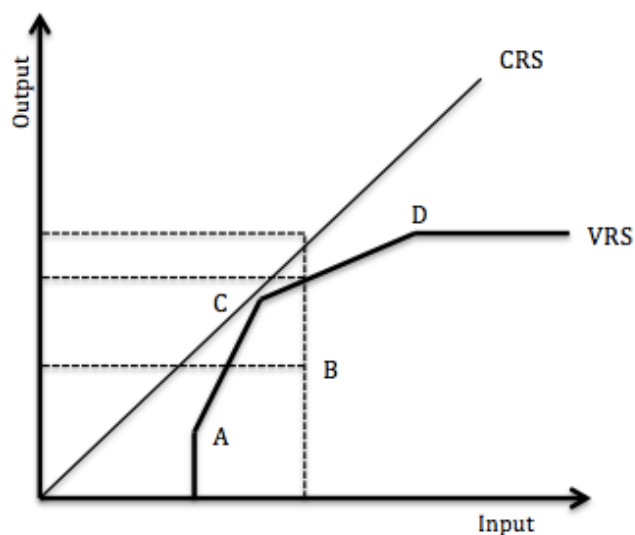
4.4.2 BCC-modellen

CCR-modellens antakelse om konstant skalautbytte er velfungerende når beslutningsenhetene opererer på en optimal skala, men kan også begrense modellens omfang. Når optimal skala ikke lengre er tilfellet vil modellens mål for teknisk effektivitet bli forstyrret av skalaeffektiviteten. Dette vil eksempelvis være tilfellet ved imperfekt konkurranse eller dersom beslutningsenhetene i datasettet er underlagt offentlige- eller finansielle reguleringer. Banker et al. (1984) viser at CCR-modellen relativt enkelt kan tilpasses analyser hvor variabelt skalautbytte kreves, ved at en legger til konveksitetsrestriksjonen. Da vil ikke

fronten lengre være lineær, men omhulle observasjonene tettere og danne et mindre produksjonsmulighetsområde. Dette medfører også at effektivitetsmålene gjerne er større for observasjonene i denne modellen. En inputminimerende BCC-modell presenteres på samme måte som CCR-modellen, men med en restriksjon for summen av λ slik:

$$\begin{aligned}
 & \min \theta_0 \\
 & \text{Når} \\
 & -y_{r0} + \sum_{j=1}^n \lambda_j y_{rj} \geq 0 \\
 & \theta_0 x_{i0} - \sum_{j=1}^n \lambda_j x_{ij} \geq 0 \\
 & \sum_{j=1}^n \lambda_j = 1 \\
 & \lambda_j \geq 0 \\
 & (j = 1 \dots, n) \\
 & (r = 1 \dots, s)
 \end{aligned}
 \tag{8}$$

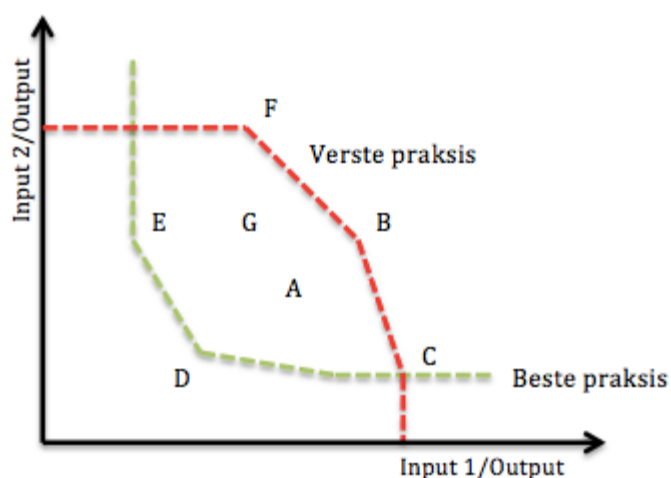
Begrensningen $\sum_{j=1}^n \lambda_j = 1$ uttrykker at summen av alle λ_j skal bli lik 1. Herunder åpner en for variabelt skalautbytte og videre at beslutningsenhetene kan sammenlignes med andre beslutningsenheter av samme størrelse, siden fronten blir liggende tettere på observasjonene. Da en her ikke inkluderer skalaeffektivitet, slik det gjøres med CCR og forutsetning om konstant skalautbytte, vil effektivitetsmålene være noe større enn ved konstant skalautbytte. I figur 3 illustreres effektivitetsfrontene ved henholdsvis CRS og VRS.



Figur 3 Illustrasjon av VRS og CRS

4.5 DEA - verste praksis

“Worst practice DEA”, heretter omtalt som verste praksis ble først presentert av Paradi, Asmild og Simak (2004). Der tradisjonell DEA med beste praksis har som mål å identifisere de effektive beslutningsenhetene i et datasett, er målet med verste praksis å finne de som presterer dårligst. Paradi et. al (2004) bruker en outputmaksimerende modell med variabelt skalautbytte for å illustrere at verste praksis DEA legger til grunn samme matematiske uttrykk som tradisjonell DEA. Det som skiller modellene er at en bytter plass på input- og outputvariablene, for å kunne identifisere de som presterer dårligst. Ved å bytte variablene, og omformulere fra minimering- til maksimeringsproblem, kan man danne en front for verste praksis, hvor en måler de som presterer svakest, se figur 4.



Figur 4 Beste og verste praksis

Liu & Chen (2005) presenterte en matematisk formel for verste praksis med variabelt skalaubytte:

$$\max \theta$$

$$\text{Når}$$

$$\sum_{j=1}^n \lambda_j x_{io} \leq x_{io}$$

$$\sum_{j=1}^n \lambda_j y_{rj} \geq \theta y_{ro}$$

$$\lambda_i \geq 0$$

$$(r = 1, \dots, s)$$

$$(j = 1, \dots, n)$$

$$(9) \quad (i = 1, \dots, m)$$

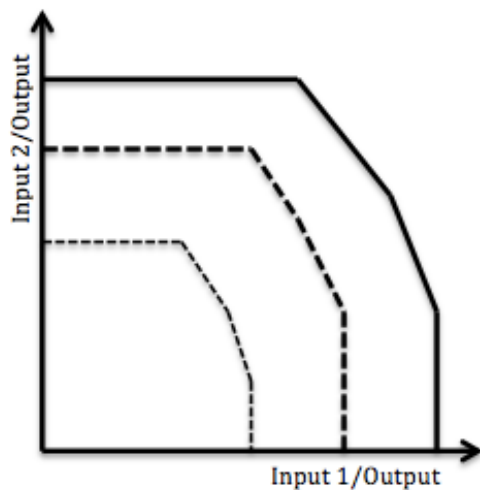
Hvor n = antall DMU, x_{io} = input og y_{ro} = outputs. θ er effektivitetsscoren til den evaluerte DMU. y_{rj} og λ_j er variable vekter som er knyttet til output og input.

Beslutningsenehetene som er mest ineffektive vil få en score lik 1, hvilket i modellen for verste praksis som er outputorientert, er den laveste mulige effektivitetsscoren. Disse DMUene er med på å danne den ineffektive fronten som illustrerer verste praksis, som er illustrert i figur 4. Jo høyere score en DMU får, jo mindre ineffektiv er den i denne sammenhengen.

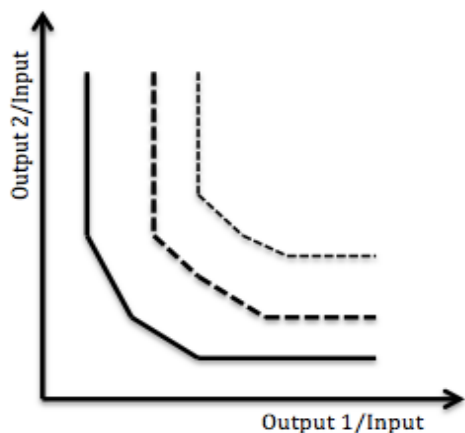
4.6 Layering

Layeringteknikken i DEA-analyser har blant annet vært foreslått av Thanassoulis (1999). Metoden går ut på å stegvis inndele de effektive observasjonene i ulike effektivitetsnivåer. En gjør en sekvensiell ekskludering av frontobservasjonene fra effektivitetsanalysen, for så å gjøre nye analyser uten disse referanseenehetene. For hver nye analyse dannes et nytt

effektivitetsnivå bestående av de effektive observasjonene som danner fronten. På bakgrunn av den lagvise inndelingen i effektivitetsnivåer kalles metoden også for «skrelling av effektivitetsløken», siden en skreller seg innover lag for lag (Paradi et. al, 2004). Paradi foreslår også å benytte teknikken i verste-praksis-analyser. Det tas da utgangspunkt i de mest ineffektive observasjonene, og en gjør stegvis fjerning av de ineffektive frontobservasjonene for å inndelegge de ineffektive observasjonene i grupper basert på deres respektive ineffektivitetsnivå. En vil da få en lagvis fremstilling av ulike effektivitetsnivåene i datasettet, som herunder tilrettelegger for tolkning av bakenforliggende årsaker. Dette kan blant annet gjøres ved å undersøke likhetstrekk mellom observasjonene som ligger innenfor samme effektivitets- eller ineffektivitetsnivå. Da inndelingen skjer på observasjonenes premisser, styrkes muligheten for å avdekke sammenhenger og korrelasjoner mellom observasjoner innen samme effektivitetsnivå. Den lagvise inndelingen av effektivitetsfronter for verste- og beste praksis illustreres i figur 5 og 6.



Figur 5 Lagvise effektivitetsfronter i verste praksis



Figur 6 Lagvise effektivitetsfronter i beste praksis

Paradi et al. (2004, s. 157) benytter også en kombinasjon av best- og worst practice-modellene, og viser at dette åpner for ytterligere effektivitetstolkninger. Metoden gjør det blant annet mulig å avdekke såkalte selv-identifiserende observasjoner. Dette er observasjoner som ligger på fronten i både beste- og verste-praksis analysen, når en undersøker samme datasettet. En vil finne disse observasjonene i et av krysningspunktene mellom frontene, se DMU C i figur 4. Dette kommer av at observasjonen har unike egenskaper som fører til at de kun sammenligner seg med seg selv. Selv-identifiserende observasjoner vil på denne måten utmerke seg i både effektivitets- og ineffektivitetsretningene. Studien poengterer videre at en ikke bør trekke slutninger på bakgrunn av disse observasjonene, da det foreligger stor usikkerhet rundt påliteligheten deres. En bør derfor hensynta disse når en skal vurdere resultatene.

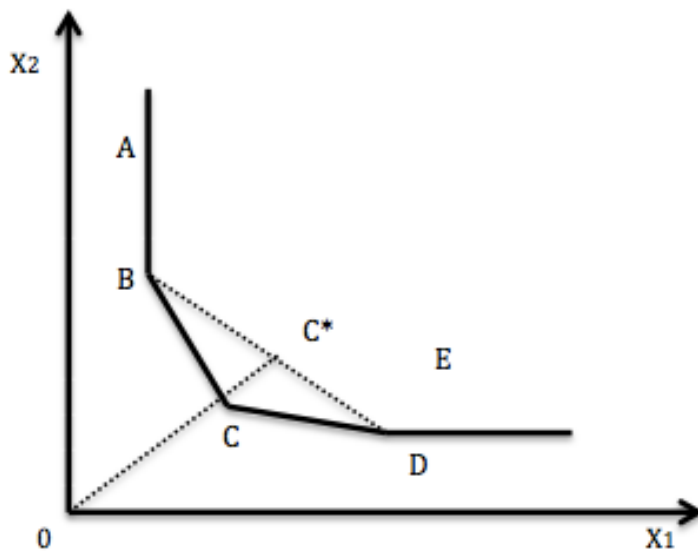
Layering-metoden kan også benyttes til å styrke troverdigheten til effektivitetsmålene. Dersom en observasjon ligger på den første effektivitetsfronten, og på en av de siste ineffektivitetsfrontene, vil dette styrke tolkningen av observasjonen som solid effektiv. I motsatt tilfelle, der observasjonene ligger på den første ineffektivitetsfronten samt en av de lavere effektivitetsfrontene taler dette for en konsekvent ineffektiv observasjon.

4.7 Sensitivitetsanalyse

En outlier er en observasjon som får en usannsynlig høy eller lav effektivitetsscore, hvilket gjør at den ikke kan eller bør sidestilles med resten av observasjonene. Disse kan oppstå på bakgrunn av feil eller mangler i data, atypiske- eller ugyldige observasjoner. Outliers vil påvirke dannelsen av den effektive fronten og dermed også påvirke hvordan beslutningsenheter måles mot hverandre (Bogetoft & Otto, 2011). Videre i dette delkapittelet vil det presenteres to metoder for identifisering av mulige outliere, supereffektivitet og AP-outlier analyse. Metodene formoder ingen fasit for hvilke observasjoner som skal anses som outliere, da dette må vurderes individuelt og med hensyn til analysens formål. Hensikten er å anvende metodene for kartlegging av ekstreme observasjoner som burde betraktes nærmere.

4.7.1 Supereffektivitet

Tanken bak supereffektivitet er at enhetene som defineres som effektiv i beste praksis, likevel har et forbedringspotensiale. Metoden gjør det mulig for DMUene å få en score høyere enn 1 for de som defineres som effektive. De effektive DMUene ligger som nevnt langs fronten og har den høyeste mulige effektivitetsverdien, lik 1. En har i CRS- og VRS-modellen i utgangspunktet ikke tilstrekkelig grunnlag til å forklare hvordan de effektive DMUene presterer i forhold til hverandre. I figur 7 ser en at B, C og D danner den effektive fronten. Punkt C kan bevege seg ut fra fronten, med utgangspunkt i B og D. Ved å dividere verdien på C^* med verdien på C vil en finne ut hvor mye enheten kan øke input med og samtidig ligge på den effektive fronten.



Figur 7 Supereffektivitet

Gitt en forutsetning om konstant skalautbytte og inputorientering har man følgende matematiske uttrykk for supereffektivitet:

$$\begin{aligned}
 & \text{Min } \theta_0 \\
 & \text{når} \\
 & \theta_0 \times x_{i0} \geq \sum_{j=1, j \neq 0}^n \lambda_j x_{ij} \\
 & \sum_{j=1, j \neq 0}^n \lambda_j y_{ij} \geq y_{r0} \\
 & \lambda_j \geq 0, \\
 & j \neq 0 \\
 & (j = 1 \dots, n) \\
 & (i = 1 \dots, m) \\
 & (r = 1 \dots, s)
 \end{aligned}
 \tag{10}$$

Restriksjonen $j \neq 0$ bestemmer at DMU_0 ikke kan benytte seg selv som referanseenheter, som åpner for at en mindre effektiv observasjon kan benyttes som referansepunkt.

Banker og Chang (2006) anbefaler å anvende supereffektivitetsmetoden til identifisering av outliers, fremfor rangering av effektive DMUer. Enheter som har høy supereffektivitet vil kunne være kunstig effektiv, og bør vurderes fjernet fra datasettet på grunn av den signifikante påvirkningen disse har på fronten.

4.7.2 AP-outlier metoden

Minste kvadraters metode er en metode som ofte benyttes for å finne outliers i lineære regresjonsmodeller. Det er utfordrende å få denne metoden til å passe til DEA-analyser vedrørende effektivitet, da fronten i DEA er ikke-parametrisk og minste kvadraters metode

baserer seg på parametriske beregninger. Andrews og Pregibon (1978) presenterte en metode hvor residualene ikke baseres på minste kvadraters metode, og derfor er mer anvendelig for DEA-analyser. Metoden de legger frem er derimot basert på bare én output, hvilket er uheldig med tanke på DEA-metodens store fordel med å hensynta et større antall variabler.

Wilson (1993) videreutviklet AP-metoden og viser hvordan den også kan benyttes i tilfeller med flere variabler. Metoden benyttes for å teste observasjonene gruppevis, og undersøker om noen observasjoner skiller seg fra «klynger» av observasjoner. I motsetning til supereffektivitetsanalysen som konsentrerer seg om effektive observasjoner, tar en også her høyde for de som ikke er effektive. Metoden er matrisebasert og relativt omfattende, da den undersøker alle mulige kombinasjoner av input og output for hver enkelt observasjon. En fjerner stegvis en-og-en observasjon, og videre også grupper med observasjoner for å se den relative endringen i klyngens volum. For de observasjonene som ligger sentralt i klyngen vil volumet sjeldent påvirkes i særlig grad ved ekskludering. I tilfeller der observasjonen som undersøkes ligger lengre borte fra klyngemassen, vil effekten kunne være betydelig større. Målet med AP-outlier analysen er altså å kartlegge hvilke observasjoner som skaper størst endringer ved fjerning, da disse kan skape støy i analysen, og videre må vurderes fjernet fra datasettet. Matematisk kan metoden uttrykkes slik:

$$(11) \quad y = \beta * X + \varepsilon$$

β og ε er variabler for henholdsvis koeffisient- og støy. Den avhengige variabelen uttrykkes gjennom y og X er uavhengige variabler uttrykt i matriseform.

4.8 Statistisk analyse

Som nevnt regnes DEA som en deterministisk metode, hvilket i utgangspunktet er lite anvendelig til testing av hypoteser. Siden konkrete observasjoner er grunnlaget for analysen, vil resultatene ikke kunne påstås allment gyldig, men kun regnes gjeldende for det respektive utvalget. Det er utviklet metoder som tilegner DEA statistiske egenskaper, slik at en likevel kan gjøre tolkninger og bemerke indikasjoner eller antydninger i resultatene. For å oppnå en

basis for å kritisk skille tilfeldige resultater fra reelle stokastiske forhold i DEA-analysen anvendes gjerne asymptotiske tester eller såkalte banker-tester.

4.8.1 Banker-test på skalanivå

Banker og Natarajan (2011) viser hvordan en kan gjøre tester for hvilken skalaforutsetning som bør legges til grunn for datasettet. Testene kan benyttes for å avgjøre hvilken skalaforutsetning som vil være hensiktsmessig for analysen på bakgrunn av skalaineffektiviteten blant observasjonene.

Som vist i formel 6 kan skalaineffektivitet beregnes ved at effektivitetsmålet fra analysen med forutsetning om konstant skalautbytte (θ_{CRS}) divideres på effektivitetsmålet fra analysen med forutsetning om variabelt skalautbytte (θ_{VRS}).

Skalaineffektivitetsverdier lavere enn én indikerer tilstedeværelse av skalaineffektivitet.

Under nullhypotesen om ingen skalaeffektivitet, samt konstant skalautbytte, er effektivitetsmålene lik ved forutsetning om konstant- og variabelt skalautbytte. Nullhypotesen om ingen forskjell i effektivitetsgjennomsnittene kan derfor vurderes ved følgende hypotesetester:

$$(12) \quad \begin{aligned} H_0: \theta_{CRS} &= \theta_{VRS} \\ H_1: \theta_{CRS} &\neq \theta_{VRS} \end{aligned}$$

Dersom en antar eksponentialfordeling med rekkevidde $(1, \infty)$ vil en teste for skala slik:

$$(13) \quad \frac{\sum \theta_{CRS} - 1}{\sum \theta_{VRS} - 1}$$

Verdien evalueres relativt til den halve F-distribusjonen, $F_{2N, 2N}$, med $[2N, 2N]$ frihetsgrader.

Dersom en antar halv-normal fordeling med rekkevidde $(0, \infty)$ vil en benytte følgende uttrykk til å teste for skala:

$$(14) \quad \frac{\sum(\theta_{CRS}-1)^2}{\sum(\theta_{VRS}-1)^2}$$

Verdien evalueres relativt til den halve F-distribusjonen, $F_{N,N}$, med $[N, N]$ frihetsgrader.

Disse tester for anvendelsen av konstant skalautbytte kontra variabelt skalautbytte. Dersom verdiene påviser signifikante forskjeller kan en forkaste nullhypotesen, siden bankertesten påviser ineffektivitet og forskjell mellom forutsetningene for konstant- og variabelt skalautbytte.

4.8.2 Banker-tester på grupper

De asymptotiske egenskapene til ineffektivitetsestimaterne i DEA-analyse kan også anvendes i statistiske banker-tester for å sammenligne to gruppers ulikhet i effektivitet, slik at en kan vurdere om den ene gruppen er signifikant mer effektiv enn den andre.

Banker (1993) foreslår F-tester til testing for signifikante effektivitetsforskjeller mellom to grupper. Videre illustreres det hvordan en kan vurdere om den ene gruppen er signifikant mer ineffektiv enn den andre, så fremst en har et tilfredsstillende stort antall observasjoner i gruppene. Banker-testene forutsetter henholdsvis halv-normal fordeling (THN) eller eksponentiell fordeling (TEX) av ineffektivitetsscorene. Tettheten i fordelingen av effektivitet i gruppen g_1 sammenlignes da med fordelingen til gruppen g_2 , hvor begge gruppene tilhører det samme utvalget N . Ved henholdsvis N_1 og N_2 observasjoner i gruppene g_1 og g_2 , kan det benyttes hypotesetester for å bestemme om det er forskjell i effektiviteten til gruppene:

$$(15) \quad \begin{aligned} H_0: g_1 &= g_2 \\ H_1: g_1 &\neq g_2 \end{aligned}$$

Dersom ineffektiviteten θ , antas å ha eksponentialfordeling (TEX) med rekkevidde $(0, \infty)$ vil en under nullhypotesen om ingen effektivitetsforskjell få:

$$(16) \quad \frac{\sum_{j \in g_1} \binom{(\theta-1)}{N_1}}{\sum_{j \in g_2} \binom{(\theta-1)}{N_2}}$$

Dette evalueres i forhold til kritisk verdi av F med $(2N_1, 2N_2)$ frihetsgrader.

Dersom ineffektiviteten til θ antas å ha halv-normal fordeling (THN), vil en under nullhypotesen om at det ikke er forskjell mellom gruppene, beregne teststatistikken følgende:

$$(17) \quad \frac{\sum_{j \in g_1} \binom{((\theta-1)^2)}{N_1}}{\sum_{j \in g_2} \binom{((\theta-1)^2)}{N_2}}$$

Verdien tolkes i forhold til den kritiske verdien av F-distribusjonen med (N_1, N_2) frihetsgrader.

5 Data og operasjonalisering av analyse

I dette kapittelet presenteres forskningsmodeller, datamaterialet, samt operasjonaliseringen av de utvalgte variablene. Datamaterialet er innhentet på bakgrunn av tematikken i studien og operasjonalisert for å oppnå en mest mulig presis besvarelse på problemstillingen. Det er derfor lagt stor vekt på et mest mulig relevant og valid datamateriale. I første omgang redegjøres det for datainnsamlingen og utvalget med hensyn på analysen som gjennomføres. I neste omgang presenteres variablene og operasjonaliseringen av disse. Til slutt gjennomgås også tester og verktøy benyttet for å sikre at det endelige datautvalget er representativt, samt forskningsmodellene utarbeidet for effektivitetsanalyser av både beste- og verste praksis.

5.1 Datainnsamling

Datasettet består av informasjon fra samtlige ortopediske operasjoner gjennomført ved UNN i 2015 og anses for øvrig å konkret representere virkeligheten. Videre er datamaterialet av kvantitativ art, samt sekundært ved at det hentes ut fra UNNs interne informasjons- og pasientdatasystemer. Det er også lukket, i den forstand at informasjonen er innsamlet uavhengig og i forkant av studien. Sammenholdt med at analysemetoden er teknisk og regelbundet, taler dette for liten grad av datapåvirkning. Data vedrørende operasjoner er gjerne av sensitiv art, blant annet på grunn av strengt pasient- og personvern. Pasientenes identitet og annen sensitiv informasjon er sikret ved at UNN selv har anonymisert datasettet til analysen.

Datagrunnlaget er av betraktelig stor størrelse, som sammen med analysens ekstensive karakter, taler for høy generaliseringsgrad. Resultatene fra analysene kan likevel ikke regnes som generaliserbare, da DEA-analyser måler relativ effektivitet. Resultatene vil derfor være subjektivt gjeldende for det spesifikke datasettet. Dette er også på bakgrunn av at ulike typer operasjoner har svært ulike karakteristikk, noe som også vil være gjeldende for ulike avdelinger, sykehus, samt demografiske og økonomiske forhold. Metoden og fremgangsmåten vil derimot kunne regnes som overførbare, siden en anser analysene som velfungerende for datamaterialet i studien.

Et stort observasjonsgrunnlag, med mange DMUer er også fordelaktig for å fange opp trender og tendenser i datasettet. Følgelig sikres også representativiteten i analysegrunnlaget. Spesielt høyt- eller lavt ytende enheter vil med stor sannsynlighet plukkes opp, hvilket er avgjørende for dannelsen av frontlinjen, diskriminering mellom DMUer og analysen. Derfor er identifisering av feilkilder og ikke-representerbare observasjoner vektlagt særlig. Dette utdypes nærmere i kapittel 5.5 vedrørende outliers. Det ses videre tendenser til at tidligere forskning baserer seg på tilnærmet like datasett og variabler med sammenfallende elementer, hvilket taler for tilfredsstillende internreliabilitet.

5.2 Definerings av utvalg

Oppgaven tar utgangspunkt i ortopediske operasjoner gjennomført ved UNNs avdelinger i Tromsø, Harstad og Narvik i 2015. Det ble gjennomført 3456 ortopediske inngrep, hvorav 2381 av disse var elektive operasjoner, altså omtrent 67 prosent.

Hver pasient blir tildelt et unikt identifikasjonsnummer. I noen tilfeller vil én pasient gjennomgå flere ortopediske inngrep. Disse er det utfordrende å skille mellom i analysen, da begge er knyttet til samme identifikasjonsnummer. Dette utgjorde totalt 263 operasjoner, og det ble vurdert nødvendig å fjerne samtlige for å unngå ufordelaktige påvirkninger på resultatene i analysen. Utelukkelsen anses som uproblematisk da disse operasjonene var jevnt fordelt mellom de ulike operasjonstypene og det heller ikke kunne sees noen avgjørende endringer i utvalget. Datasettet er likevel av relativt stor størrelse og etter denne utplukkingen gjenstår 2118 elektive operasjoner.

Det skilles mellom 46 ulike operasjonstyper innenfor ortopedisk kirurgi. Antallet gjennomførte operasjoner varierer fra 1 til 713. På bakgrunn av problemstillingen vil det hverken være nødvendig eller hensiktsmessig å inkludere alle operasjonene i datasettet. Det er derfor plukket ut fem ulike operasjonstyper som benyttes i analysene, disse presenteres i tabell 1 under.

Tabell 1 Utvalg

	Elektiv	Ø-hjelp	Totalt
<i>Artroskopi kne</i>	176	8	184
<i>Artroskopi skulder</i>	175	2	177
<i>Hofteprotese primær usementert</i>	94	3	97
<i>Kneprotese primær</i>	159	0	159
<i>Hallux valgus</i>	199	4	203
<i>Sum</i>	<i>803</i>	<i>17</i>	<i>820</i>

Utvalget består følgelig av fem ulike operasjonstyper, med et gjennomført antall mellom 94 og 199 operasjoner i 2015. Av disse er det to proteseoperasjoner, to artroskopiske inngrep, samt en forholdsvis enkel fotoperasjon kalt hallux valgus. Hofteprotesene har utelukkende blitt utført i Tromsø, mens kneproteser har blitt utført i Harstad og Tromsø. De resterende gjennomføres på alle tre lokasjonene.

Som det fremkommer av tabell 1 ovenfor er antallet ø-hjelpsoperasjoner relativt lavt for samtlige av disse operasjonstypene. Disse blir ekskludert fra utvalget, da det anses å være mer

hensiktsmessig å fokusere på forbedringspotensialet hos operasjonene som planlegges i forkant.

Artroskopiske inngrep går ut på at kirurgen går inn i et ledd, som regel skulder, kne eller albue, med kamera for å undersøke skadeomfanget, men også gi behandling. Videre i oppgaven vil artroskopisk kne med minisk henvises til som a-skopi kne, og artroskopi skulder omtales som a-skopi skulder. En protese er en kunstig kroppsdel som erstatter en skadet eller manglende kroppsdel, hvorav de vanligste er hofte- eller kneprotese. En hofteprotese kan settes inn, og festes enten med eller uten bruk av spesialsement. I oppgaven er det tatt utgangspunkt i de usementerte hofteproteseoperasjonene, som heretter omtales som hofteprotese. Kneproteseoperasjoner av typen primær omtales videre i oppgaven som kneprotese. Hallux valgus er en feilstilling av stortåen som gjør at den vris innover i retning lilletåen, hvilket fører til at leddet hovner opp på innsiden av foten (norsk helseinformatikk, 2013). Hallux valgus er en vanlig diagnose hos både kvinner og menn, og behandlingen er relativt standardisert og lett å gjennomføre.

5.3 Operasjonalisering

I det bedriftsøkonomiske perspektivet er lønnsomhet et overordnet mål på sikt.

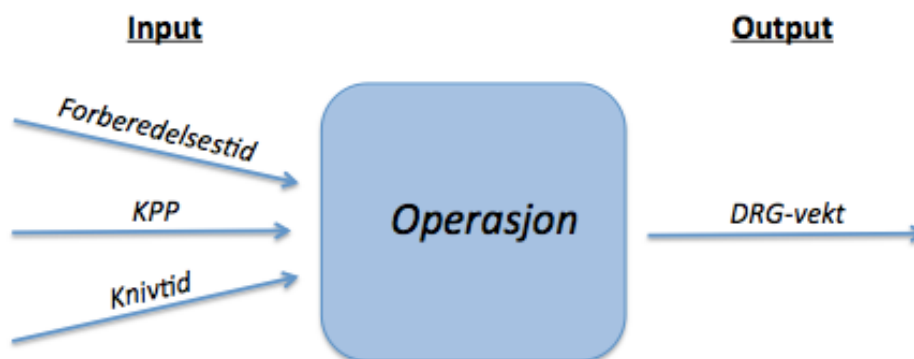
Profittmaksimering skapes gjennom økt produksjon og inntekter, eller redusert ressursbruk for å dempe kostnadssiden. Overføres dette til effektivitetsanalyser og beslutningsenheter i DEA vil en kunne si at målet er enten minimering av input eller maksimering av output. Da tid og ressurser er kostbare vil effektivisering av produksjonsprosessene ha stor innvirkning på lønnsomheten. Inputorientering benyttes når en søker å redusere inputelementene proporsjonalt, uten å endre produksjonsmengden av output. Ved en outputorientering ønskes proporsjonal økning i output, uten at dette medfører endringer i inputmengden.

Universitetssykehuset i Nord-Norge er lagt til grunn i denne studien, og som offentlig institusjon pålegges de spesielle oppgaver innenfor deres ansvarsområder. Profitt og lønnsomhet er ikke nødvendigvis målsetningen, da en ved offentlige sykehus ikke genererer inntekt gjennom salg. Inntektene innløper som statlig støtte og bevilgninger gjennom tilskudd eller refusjoner for den respektive produksjonen. Output trenger derfor ikke være varer eller tjenester som en har til hensikt å tjene penger på, men er i dette tilfellet en produsert tjeneste som gir grunnlag for inntekten gjennom statlige refunderinger. Utgangspunktet for analysen

er derfor å begrense ressursforbruket, inputsiden, så mye som mulig samtidig som outputsiden holdes konstant. På denne måten vil en kunne undersøke om det er muligheter for å møte myndighetenes effektiviseringskrav.

5.3.1 Forskningsmodell

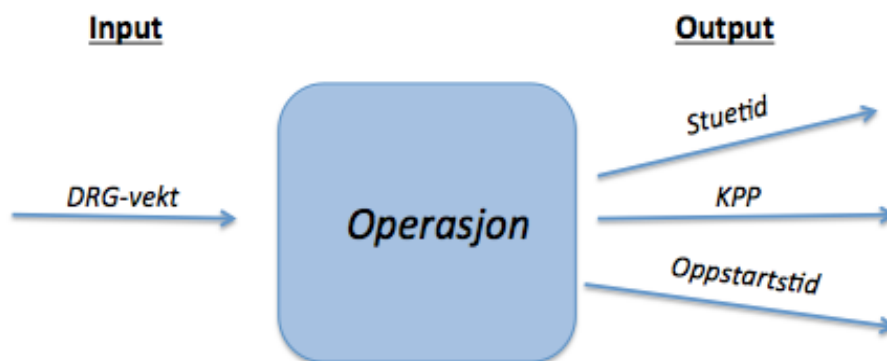
Forskningsmodellen for beste-praksis er inputorientert og søker å finne DMUene som presterer best. Disse anses som en transformator, som benytter ulike innsatsfaktorer for å produsere produkter eller tjenester. I dette tilfellet representerer innsatsfaktorene ressursforbruket til operasjonene, mens produksjonen er DRG-poeng som indirekte genererer inntekter til sykehuset. En har som hensikt å minimere kostnadene og tidsforbruket uten at DRG-vekten påvirkes. Sistnevnte inkluderer faktorer for blant annet sykdomsbildet til pasienten, prosedyrer og den medisinske behandlingen. Intensjonen er ikke å motivere til overbehandling for å øke DRG-vektene, følgelig er outputsiden i modellen ikke ønskelig å påvirke i analysen. Dette er heller en variabel som tar høyde for disse faktorene, men som bør få være upåvirket av effektivisering. Forskningsmodellen for beste praksis er derfor bygget opp slik:



Figur 8 Modell beste praksis

Modellen for verste praksis er en speiling av beste-praksismodellen, en bytter side på variablene og får derfor en outputorientert modell. Videre kartlegges også de ”verste” observasjonene med høyeste grad av ineffektivitet. DRG-vekten skal fortsatt holdes urørt, mens en har outputvariabler som er mer hensiktsmessig til måling av verste-praksis. Kostnad per pasient benyttes også i denne modellen, men nye variabler benyttes for tidsaspektet. Her trekkes det inn mål på forsinkelser, både før og under operasjonene. Dette utdypes i neste

delkapittel. Disse reflekterer dårlig ressursutnyttelse og bidrar til å fremheve de verste prestasjonene blant operasjonene. Følgelig ser modellen for verste praksis slik ut:



Figur 9 Modell verste praksis

I analysen søkes det å finne områder med forbedringspotensial på bakgrunn av høye nivåer for ineffektivitet. Siden analysen ikke hensyntar kvalitet og omtanke for menneskene involvert, er det ikke dermed sagt at resultatene viser reelt forbedringspotensial. Variablene som benyttes i modellen er hensiktsmessig og anvendelig til å måle en ren og teknisk effektivitet. Det anerkjennes at modellen ikke inkluderer dimensjoner for kvalitet, helsegevinster eller økt livskvalitet. Det er viktig å påpeke at disse aspektene er vel så viktig som effektivitet, og at en i praksis ikke kan eller bør effektivisere operasjonsaktiviteten på bekostning av disse. Sykehusets mål er sammensatt og i helseproduksjonen vil helsegevinst og sikker omsorg for mennesket komme i første rekke.

5.4 Valg av variabler

DEA-analyser vil ikke kunne diskriminere mellom effektive og ineffektive enheter på en god måte dersom forholdet mellom antall DMU og summen av antall variabler er lavt (Avkiran, 1999). Ekspertene strides om hvor mye større observasjonssettet bør være. Eksempelvis presenterer Golany og Roll (1989) en tommelfingerregel med minimumskrav på dobbelt antall observasjoner i forhold til variabler for input og output. Dette forbigår denne studien med god margin.

Variabler som benyttes i DEA-modeller kan deles i to kategorier: kontrollerbare- og ikke-kontrollerbare. Sistnevnte er faste og/eller eksogene variabler som ikke kan påvirkes av de ulike beslutningsenhetene, og derfor heller ikke i analysen. I studien tas det utgangspunkt i kontrollerbare variabler til analysene. Relevans for det aktuelle temaet er særlig viktig i valget av variabler, siden disse på best mulig måte skal beskrive det oppgaven har som hensikt å undersøke. Siden variablene er kontrollerbare vil de også fundamentere et godt analysegrunnlag for eventuelle forbedringsarbeid.

DRG-vektene og kostnad per pasient er mer sammensatte variabler. Det er mulig å både kontrollere og påvirke disse, og på samme tid vekkes mer sykehuspolitiske spørsmål og diskusjoner rundt blant annet kvaliteten på behandlingene. Elementer innen begge variablene vil være kontrollerbare, andre kan regnes som mer indirekte kontrollerbare. Dette utdypes nærmere i kapittel 5.4.5.

For at variablene skal svare til hensikten med forskningsmodellene, er det benyttet ulike variabler til de ulike modellene. Dette på bakgrunn av at beste praksis modellen har til hensikt å beskrive hvor effektiv de beste operasjonene er, mens en med modellen for verste-praksis ønsker å kartlegge de som presterer dårligst. I modellen for beste praksis er variablene derfor forberedelsestiden og selve knivtiden benyttet i de enkelte operasjonene. Disse vil reflektere den faktiske tidsbruken før og under operasjonene, og særlig interessant er tiden brukt av de mest effektive observasjonene. Til verste-praksismodellen er det derimot valgt variabler som måler forsinkelse i oppstartstiden fra det planlagte tidspunktet, samt mertid brukt på operasjonsstua mot det som var planlagt tidsbruk. Disse variablene vil beskrive forsinkelser i forkant av operasjonene og merforbruk av stuetiden, for lettere å identifisere de operasjonene som har størst forbedringspotensiale.

Som det fremkommer av tabell 2 under, er kostnad per pasient og DRG-vekt variabler som brukes i begge modellene.

	<i>Beste praksis</i>	<i>Verste praksis</i>
DRG-vekt	✓	✓
KPP	✓	✓
Knivtid	✓	
Forberedelsestid	✓	
Forsinkelse i stuetid		✓
Forsinkelse i operasjonsstart		✓

Tabell 2 Input- og outputvariabler

5.4.1 Knivtid

Knivtid regnes som tidsforbruket på selve operasjonsaktiviteten, fra kirurgen gjør det første kuttet til operasjonssåret er sydd igjen. I det innhentede datamaterialet fra UNN er det gjort et skille mellom planlagt og faktisk knivtid, der sistnevnte er tiden som ble benyttet under den gjennomførte operasjonen. Faktisk knivtid anvendes som variabel i analysen og regnes som en innsatsfaktor i modellen for beste praksis.

Det er flere faktorer som kan påvirke knivtiden. UNN er et universitetssykehus, og vil derfor ha leger og studenter under opplæring. Minimum én LIS-lege (lege i spesialisering), skal være tilstede ved alle operasjonene for mest mulig læringsutbytte. Det kan antas at knivtiden derfor vil avhenge av hvor langt i opplæringsforløpet denne kirurgen er kommet. Det vil også kunne dukke opp elementer som ikke har kommet fram under forundersøkelsene, hvilket gjør operasjonen mer omfattende enn først antatt. I noen tilfeller opplever en også det motsatte, nemlig at operasjonen er mindre komplisert enn forventet.

Sykdomsbildet, kompleksiteten og den allmenne tilstanden til pasienten vil også variere, hvilket også er avgjørende for knivtiden. Dette reflekteres ikke gjennom variabelen for

knivtid, men vil være tatt hensyn til i DRG-vekten. Dette gjør det mulig og finne mulige forklaringer på forsinkelser ved nærmere undersøkelser. Dersom en operasjon er mer komplisert enn antatt vil kirurgene kunne bruke ekstra tid, som igjen er en kvalitetsfaktor som ikke foreskrives i denne analysen. Operasjonsarbeidet skal også kvalitetssikres, fordi god kirurgisk behandling er formålet med operasjonen framfor å gjennomføre raskest mulig. Knivtiden vil derfor kunne påvirkes både negativt og positivt på bakgrunn av mange ulike faktorer.

5.4.2 Forberedelsestid

Forberedelsestid er tiden fra pasienten trilles inn til operasjon, til anestesipersonalet begynner med anestesen. Den faktiske forberedelsestiden anvendes som inputvariabel i beste praksis-modellen. Tidsbruken kan her påvirkes av godt forarbeid og planlegging av tiden før operasjonen. Dersom personalet som er ansvarlig for å gjøre pasienten kjent med prosedyrene har tatt seg tid til å forklare og forberede pasienten, vil ofte forberedelsestiden kunne være minimal. Det samme er gjeldende for forundersøkelser. Kontrolleres det at alle nødvendige utredninger og tester er gjort i forkant av operasjonen, vil en ikke behøve å bruke tid på dette ved selve inngrepet og videre oppta stuetid. Tiden vil også kunne påvirkes av pasientens sykdomsbilde, da enkelte diagnoser eller tilstander krever ekstra oppmerksomhet i forkant av operasjonene. Da dette er et rent tidsmål vil ikke forklaringsfaktorer hensyntas i denne variabelen, men dette hensyntas i større grad av DRG-vekten.

5.4.3 Forsinkelse i stuetid

Stuetid er tidsrommet fra en pasient trilles inn til operasjon og frem til etterarbeidet er gjennomført og pasienten kan trilles ut igjen. Innbakt i denne variabelene er forberedelsestid, anestesi, knivtiden og etterarbeidet. Ved å beregne differansen mellom planlagt og gjennomført stuetid avdekkes det om denne spesifikke operasjonen brukte mer eller mindre tid enn planlagt. Variabelen henvises til som forsinkelse i stuetid, og da den beregnes ved å subtrahere planlagt stuetid fra den faktiske tidsbruken vil en også kunne ha “negativ forsinkelse”. Dette er tilfellet når operasjonsaktiviteten har tatt kortere tid å gjennomføre enn det som var planlagt. Det kan være viktig å bemerke seg at en i denne analysen ikke vil kunne si noe om hva tidsavviket skyldes. Tanken er heller å kunne kartlegge om det er den spesifikke operasjonene eller andre faktorer som skaper forsinkelser. Dette knyttes da sammen med neste variabel, differanse forsinkelse i oppstartstid, som vil kartlegge om

heftelsen har skjedd i forkant av den aktuelle operasjonen. Variabelen anvendes som outputvariabel i modellen for verste praksis og forsinkelse i stuetid beregnes på følgende måte:

$$\text{Forsinkelse i stuetid} = \text{Faktisk stuetid} - \text{planlagt stuetid}$$

I tabell 3 under vises et eksempel fra datasettet mottatt fra UNN. Den planlagte og faktiske stuetiden er oppgitt, og brukt til å beregne planlagt stuetid i minutter og faktisk stuetid i minutter.

Tabell 3 Eksempel fra datasettet til kneproteser

Nr.	Kostnad operasjon	Planlagt stuetid	Planlagt stuetid i minutter	Faktisk stue inn	Faktisk stue ut	Faktisk stuetid i minutter	Faktisk stuetid
1	13 481	02:44	164	11:16	14:00	164	02:44
2	26 093	02:44	164	07:45	10:50	185	03:05
3	25 505	02:44	164	12:40	16:11	211	03:31
4	16 250	02:44	164	11:43	14:09	145	02:25
5	17 499	02:44	164	08:05	11:13	188	03:08

5.4.4 Forsinkelse i operasjonsstart

I datamaterialet fra DIPS registreres også tidspunktet kalt stue inn, dette er tidspunktet når pasienten er klargjort av anestesipersonalet og trilles inn til operasjon. Tidspunktet er i analysen omdøpt til tidspunktet for operasjonsstart. Variabelen forsinkelse i operasjonsstart, beregnes ved å finne differansen mellom det planlagte- og det faktiske tidspunktet pasienten ble trillet inn til operasjonen. Hensikten med denne variabelen er å inkludere en indikator på

om operasjonen starter senere enn planlagt tidspunkt. Dette tilrettelegger for å kunne skille mellom denne operasjonenes mer- eller mindreforbruk og forsinkelser som skjer uavhengig av denne spesifikke operasjonen. Variabelen brukes som outputvariabel i modellen for verste praksis, og beregnes på følgende måte:.

$$\text{Forsinkelse i oppstartstid} = \text{Faktisk stue inn} - \text{planlagt stue inn}$$

Her vil en kunne få både positiv og negativ verdi på variabelen, avhengig av om planlagt stue inn er større eller mindre enn faktisk stue inn. Dette gjelder også for differanseberegningen av stuetid. En positiv verdi antyder et merforbruk av tid og en negativ verdi vil vise til at aktiviteten har vært gjennomført raskere enn forventet.

5.4.5 DRG-vekt

Diagnoserelaterte grupper er en del av beregningen som gjøres innenfor innsatsstyrt finansiering. DRG-poengene tilsvarer ressursene som brukes for å behandle en gjennomsnittspasient innenfor de ulike operasjonstypene, og er således et uttrykk for relativ ressursinnsats for en diagnoserelatert gruppe (Helsedirektoratet, 2017a).

Kostnadsvekter benyttes som et grunnlag for beregning av refusjoner i innsatsstyrt finansiering til de regionale helseforetakene. I tillegg er de ment som et verktøy for måling av aktivitet og produktivitet ved foretak som tilbyr somatiske helsetjenester (Helsedirektoratet, 2017c). Kostnadsvektene er en relativ størrelse som tar utgangspunkt i hva et opphold i en DRG-gruppe kostet i gjennomsnitt, sammenlignet med en gjennomsnittlig pasient.

Gjennomsnittsvekten er 1,0 og har en fast stykkpris som justeres årlig av Helsedirektoratet. I 2015 var enhetsprisen pr DRG-poeng 41 462 kr (Helsedirektoratet, 2014). Enhetsprisen multipliseres med 0,5 siden ISF dekker 50% av behandlingstkostnadene. Summen av DRG-vekter regnes som DRG-poeng og danner grunnlaget for refusjonsberegningen. Beregningen kan illustreres gjennom formelen:

$$\text{Refusjon ISF} = \text{DRG-poeng} * 0,5 * \text{Enhetsprisen for DRG}$$

DRG-vekt er beregnet for hver enkelt pasients behandlingsforløp og illustrerer hele behandlingen, ikke bare operasjonen. Vekten er en indikasjon på hvor komplisert behandlingsløpet har vært for hver enkelt pasient.

Hensikten med studien er som nevnt å undersøke potensialet for effektivisering, herunder kostnadsbesparelser gjennom å utnytte tildelte midler på best mulig vis. Sykehuset produserer indirekte DRG-poeng for hver pasient som behandles. Disse beregnes nøye og systematisk da hvert DRG-poeng genererer refusjoner til sykehuset. Utfordrende pasienter, operasjoner og prosedyrer medfører høyere poengsum, som igjen er avgjørende for inntektsgenereringen.

Gjennom å benytte DRG-poengene som variabel i modellen tar en implisitt høyde for hvor komplisert og utfordrende hver operasjon er. Selv om en også har andre elementer enn bare operasjonsaktiviteten innbakt i variabelen, tenkes likevel å gi uttrykk for forholdet mellom de ulike operasjonene.

5.4.6 KPP - kostnad per pasient

Kostnad per pasient knytter sammen hvilken behandling en pasient har mottatt og hva denne spesifikke behandlingen har kostet (Helsedirektoratet, 2017d). Følgelig kan en betrakte behandlingsløpet til en pasient som en prosess, bestående av flere delprosesser med ulike medisinske tjenester og prosedyrer. Det er gjennom KPP mulig å kostnadsberegne disse deltjenestene separat, for så å koble de entydig til pasienten. Kostnadene i KPP består av blant annet tjenestekostnader tilhørende personell i behandlingsprosessen og eventuelle forbruksvarer hvilket kan summeres til en unik kostnad per pasient.

Kostnaden per pasient benyttes som en variabel innsatsfaktorer i analysen, og reflekterer kostnadene som knyttes direkte til den enkelte operasjonen som undersøkes. Denne variabelen består følgelig av blant annet lønn til hovedkirurg, anestesi, og eventuelle forbruksvarer som proteser og liknende.

5.5 Outliers

I oppgaven er det benyttet flere metoder for å identifisere mulige outliers i datasettene. Gjennomsnitt og standardavvik blant annet er beregnet for alle input- og outputvariablene, og ved unormale verdier har utvalget vært undersøkt for feilkilder eller spesifikke DMUer med

verdier som skiller seg særlig fra de andre. Dersom det har dukket opp ekstreme verdier i de statistiske analysene for variablene bør datasettet undersøkes for potensielle outliers slik at disse kan ekskluderes fra analysen. Dette vil sikre et mer representativt utvalg og videre også en pålitelig studie.

Supereffektivitetsmetoden benyttes i både beste- og verste-praksis modellene. Intensjonen med denne analysen er å undersøke for ekstreme verdier. En DMU som har en unaturlig høy supereffektivitetsverdi bør undersøkes som en potensiell outlier. Videre tas det også stilling til om den enkelte observasjonen bør fjernes fra datasettet.

I noen tilfeller er det registrert null minutter i forberedelsestid og knivtid, hvilket anses som lite sannsynlig, og kan være en mulig registreringsfeil. Variabler innenfor beste praksis som har null minutter i forberedelsestid eller knivtid blir derfor plukket ut for å minimere sjansen for feilkilder. En verdi på null i variablene differanse stuetid og stue inn vil forekomme når det ikke er differanse mellom planlagt tidspunkt og reelt tidspunkt. Dette er ikke tilfelle for noen av operasjonene, og det blir derfor ikke plukket ut noen fra verste praksis.

Observasjonene som er fjernet fra analysen er fortsatt reelle operasjoner. Uavhengig av om de uvanlige verdiene skyldes registreringsfeil eller unormale forhold i forbindelse med operasjonen, vil det selvfølgelig være mulig at også disse har ineffektivitetsnivåer som kunne vært utbedret, herunder forbedringspotensiale. Disse er fjernet fra datasettet på bakgrunn av at de ikke er representative for observasjonsmassen. Det kunne vært gjennomført egne analyser for disse observasjonene, men i dette tilfellet fokuseres det på det som anses som gjennomsnittlige operasjonene hvilket er i flertall.

5.5.1 Fjerning av outliere - beste praksis

Under a-skopi kne ble det identifisert en potensiell outlier med en supereffektivitet på 2,1716. I tillegg består datasettet av én DMU med null minutter i forberedelsestid. Begge disse fjernes etter kriteriene som er satt for definering av outliere. Innenfor operasjonene a-skopi skulder avdekkes det gjennom den deskriptive statistikken på variablene et unormalt høyt standardavvik for forberedelsestiden brukt på a-skopi skulder. Dette skyldtes at en DMU hadde en unormal høy verdi på forberedelsestiden brukt, 1432 minutter. Dette anses som lite reelt og ikke representativt for datasettet for øvrig, og den ekskluderes derfor fra datasettet.

Når statistikken beregnes på nytt blir verdiene betraktelig normalisert, hvilket bekrefter at beslutningsenheten skiller seg fra de andre, se tabell 4.

Tabell 4 Eksempel på konsekvensen av outlierfjering

DMU 92	Før	Etter
<i>Gjennomsnitt</i>	20,88	12,77
<i>Standardavvik</i>	107,79	10,45
<i>Størst</i>	1 432	40
<i>Minst</i>	0	0

I den innledende analysen av hofteproteser identifiseres det ingen DMUer med ekstreme verdier, men én DMU med 0 minutter i knivtid. Denne fjernes fra datasettet.

Blant kneproteser identifiseres det to DMUer med 0 minutter i forberedelsestid, og disse fjernes også. I datasettet tilhørende hallux valgus blir det identifisert én potensiell outlier, med en supereffektivitetsscore på 3,2368. Dette kan antas å være på grunn av en høy DRG-vekt, på henholdsvis 3,514, hvilket er langt over gjennomsnittet på 0,7297. Denne blir derfor også ekskludert fra datasettet.

På bakgrunn av de ulike outlieranalysene plukkes det ut 7 observasjoner, hvilket antas å redusere støy i datasettet. For øvrig er antallet som ekskluderes minimalt i sammenligning med det totale antallet observasjoner. Det anses derfor å ha svært liten påvirkning på representativiteten i analysen. En oversikt over resultatet fra outlieranalysene som er gjort i beste praksis kan sees i tabell 5.

Tabell 5 Resultat fra outlieranalysene for beste praksis

Beste praksis	Statistikk	Supereffektivitet	Ugyldig verdi	Resterende antall
<i>A-skopi kne</i>	0	1	1	174
<i>A-skopi skulder</i>	1	0	0	174
<i>Hofteprotese</i>	0	0	1	93
<i>Kneprotese</i>	0	0	2	157
<i>Hallux valgus</i>	0	1	0	198

5.5.2 Fjerning av outliers - verste praksis

I datasettet for både a-skopi kne og kneproteser identifiseres det ingen potensielle outliers. Innenfor a-skopi skulder identifiseres derimot tre outliers hvor en av disse har en supereffektivitetsscore på hele 6,5370. Dette kan tenkes å skyldes et merforbruk av stuetid og stor forsinkelse i operasjonsstart, men observasjonen har også datasettets høyeste KPP på henholdsvis 84 097 kr. De to andre har også høyt merforbruk i begge tidsvariablene, og en KPP over gjennomsnittet. Alle tre DMUer utelukkes fra datasettet.

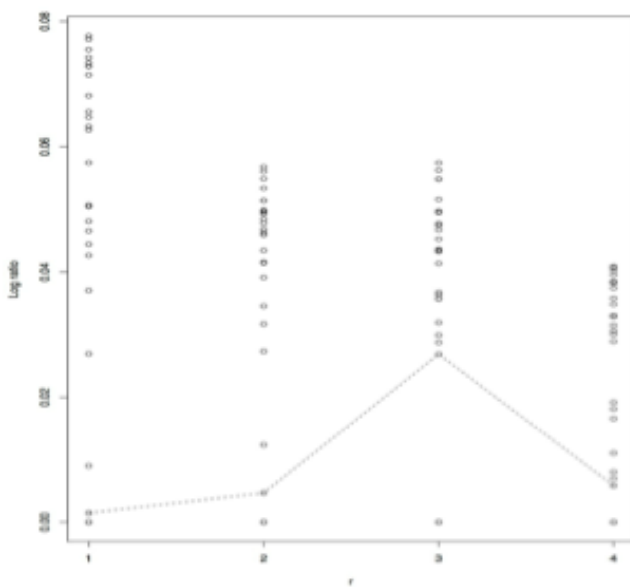
I datasettet for hofteproteser identifiseres syv mulige outliers. Samtlige av disse har en særdeles høy supereffektivitet, fra 42,8459 til 246,5965, og plukkes derfor ut.

Blant hallux valgus-operasjonene ble det høyeste antallet med potensielle outliers identifisert, hele 12 stykker utmerket seg. Supereffektivitetsverdien varierer mellom 4,2119 og 14,0635. Samtlige observasjoner plukkes ut for å forhindre støy i datamaterialet.

Gjennom den deskriptive statistikken ble det avdekket standardavvik i variablene forsinkelse i stuetid og forsinkelse i operasjonsstart, som begge hadde ekstreme verdier. Ved nærmere undersøkelse viste det seg og være en gruppe med DMUer som forårsaket dette og ikke en enkelt DMU. På bakgrunn av dette ble det gjennomført outlieranalyse med AP-metoden for å se nærmere på grupperinger, som ikke nødvendigvis har ekstreme verdier.

Under er et eksempel fra AP-analysen på kneproteser presentert. Den viser 3 observasjoner som bør undersøkes nærmere. I tabellen til høyre kan en se at DMU 20, 59 og 78 bør undersøkes nærmere. Dette resultatet stemmer overens med det som ble oppdaget i den deskriptive statistikken, og samtlige av variablene som ble identifisert med AP-metoden

plukkes ut. I noen tilfeller ble DMUer identifisert som mulig outlier i både supereffektivitet og AP-analysen. Dette er med på å forsterke mistanken om at den aktuelle DMUen kan være en outlier. Resultatene fra outlieranalysene kan sees nærmere i tabell 6 under.



	V1	V2	V3	V4	V5
1	78				0,8901
2	20	59			0,7897
3	20	59	78		0,7028
4	16	20	59	78	0,6406

Figur 10 Resultat AP-analyse kneprotese

Tabell 6 Resultat fra outlieranalyse for verste praksis

Verste praksis	Statistikk	Supereffektivitet	AP	Resterende antall
<i>A-skopi kne</i>	0	0	1	175
<i>A-skopi skulder</i>	0	3	2	170
<i>Hofteprotese</i>	0	7	1	86
<i>Kneprotese</i>	0	0	3	156
<i>Hallux valgus</i>	1	12	1	185

5.6 Deskriptiv statistikk

Tabell 7 og 8 viser gjennomsnittet, standardavviket og største og minste verdi av input- og outputvariablene brukt i beste praksis og verste praksis. KPP og DRG-vekt er som nevnt variabler som benyttes i begge analysene gjennomført i oppgaven, og her vil verdiene være det samme. Som det fremkommet av tabellen er det store variasjoner innenfor de ulike typene operasjoner. Tabellene under viser aktuelle tall for datasettene etter at outliere er fjernet. For å gjøre fremstillingen oversiktlig er variablene for beste praksis og verste praksis presentert i forskjellige tabeller.

Tabell 7 Deskriptiv statistikk variabler i beste praksis

A-skopi kne	<i>Input</i> KPP	<i>Input</i> Forberedelsestid	<i>Input</i> Knivtid	<i>Output</i> DRG-vekt
<i>Gjennomsnitt</i>	21 536	15	21	0,350
<i>Standardavvik</i>	10 132	9	14	0,046
<i>Størst</i>	67 055	70	96	0,445
<i>Minst</i>	4 117	2	4	0,025
A-skopi skulder	<i>Input</i> KPP	<i>Input</i> Forberedelsestid	<i>Input</i> Knivtid	<i>Output</i> DRG-vekt
<i>Gjennomsnitt</i>	22 251	21	43	0,665
<i>Standardavvik</i>	11 096	108	28	0,282
<i>Størst</i>	84 097	1 432	139	1,939
<i>Minst</i>	5 765	1	9	0,032
Hofteprotese	<i>Input</i> KPP	<i>Input</i> Forberedelsestid	<i>Input</i> Knivtid	<i>Output</i> DRG-vekt
<i>Gjennomsnitt</i>	24 792	23	62	3,090
<i>Standardavvik</i>	11 651	13	14	0,580
<i>Størst</i>	80 378	66	137	5,290
<i>Minst</i>	6 840	1	35	0,120

Kneprotese	<i>Input</i> KPP	<i>Input</i> Forberedelsestid	<i>Input</i> Knivtid	<i>Output</i> DRG-vekt
<i>Gjennomsnitt</i>	22 768	30	93	2,831
<i>Standardavvik</i>	10 139	17	19	0,127
<i>Størst</i>	62 590	102	151	4,408
<i>Minst</i>	5 958	1	54	2,821
Hallux valgus	<i>Input</i> KPP	<i>Input</i> Forberedelsestid	<i>Input</i> Knivtid	<i>Output</i> DRG-vekt
<i>Gjennomsnitt</i>	21 864	17	45	0,727
<i>Standardavvik</i>	9 334	18	32	0,457
<i>Størst</i>	52 088	138	202	3,514
<i>Minst</i>	5 648	1	9	0,036

Tabell 8 Deskriptiv statistikk variabler verste praksis

A-skopi kne	<i>Input</i> DRG-vekt	<i>Output</i> KPP	<i>Output</i> Stuetid	<i>Output</i> Operasjonsstart
<i>Gjennomsnitt</i>	0,350	21 536	-19	34
<i>Standardavvik</i>	0,046	10 132	20	130
<i>Størst</i>	0,445	67 055	64	398
<i>Minst</i>	0,025	4 117	-94	-367
A-skopi skulder	<i>Input</i> DRG-vekt	<i>Output</i> KPP	<i>Output</i> Stuetid	<i>Output</i> Operasjonsstart
<i>Gjennomsnitt</i>	0,665	22 251	-17	112
<i>Standardavvik</i>	0,282	11 096	36	112
<i>Størst</i>	1,939	84 097	101	419
<i>Minst</i>	0,032	5 765	-93	-126

Hofteprotese	<i>Input</i> DRG-vekt	<i>Output</i> KPP	<i>Output</i> Stuetid	<i>Output</i> Operasjonsstart
<i>Gjennomsnitt</i>	3,090	24 792	4	7
<i>Standardavvik</i>	0,580	11 651	25	100
<i>Størst</i>	5,290	80 378	84	323
<i>Minst</i>	0,120	6 840	-40	-328
Kneprotese	<i>Input</i> DRG-vekt	<i>Output</i> KPP	<i>Output</i> Stuetid	<i>Output</i> Operasjonsstart
<i>Gjennomsnitt</i>	2,831	22 768	9	72
<i>Standardavvik</i>	0,127	10 139	29	119
<i>Størst</i>	4,408	62 590	89	405
<i>Minst</i>	2,821	5 958	-112	-195
Hallux valgus	<i>Input</i> DRG-vekt	<i>Output</i> KPP	<i>Output</i> Stuetid	<i>Output</i> Operasjonsstart
<i>Gjennomsnitt</i>	0,727	21 864	-8	120
<i>Standardavvik</i>	0,457	9 334	42	117
<i>Størst</i>	3,514	52 088	155	420
<i>Minst</i>	0,036	5 648	-67	-212

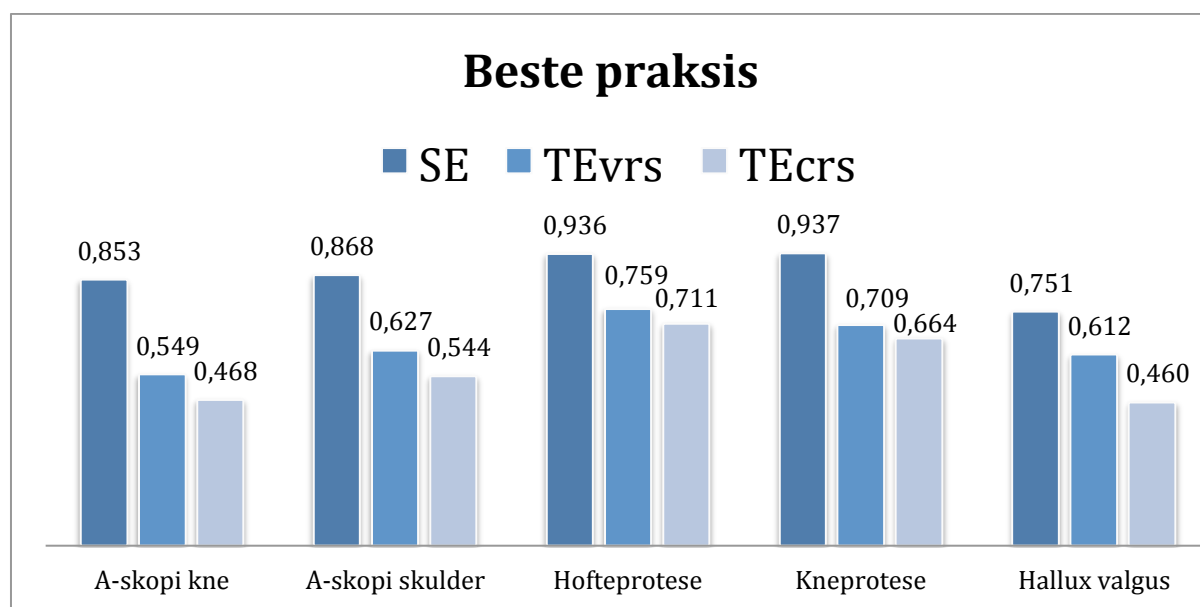
6 Resultater

Dette kapitlet har til hensikt å redegjøre for resultatene for de ulike analysene gjennomført i oppgaven. Kapitlet innledes med en gjennomgang av resultatene fra beste praksis analysen for de ulike operasjonstypene. Det presenteres resultater for både variabelt og konstant skalautbytte og gjøres en kort drøftelse av skalaeffektivitetens betydning for den videre analysen. Videre presenteres resultatene fra bankertesten for skala, hvilket er benyttet for å teste for signifikante forskjeller mellom analysene gjort med forutsetninger for henholdsvis variabelt- og konstant skalautbytte. Herunder redegjøres det også for valget av skalaforutsetning for den videre analysen. Ytterligere identifiseres og undersøkes observasjonene som danner fronten med ineffektive beslutningsenheter, gjennom modellen

for verste praksis. Deretter illustreres også hvordan analysene for beste- og verste praksis kan benyttes sammen, og hvordan en kan teste for reelle forskjeller mellom effektivitetsnivåene. Resultatene fra banker-tester på gruppenivå, hvilket tester signifikansen i forskjellene mellom de ulike gruppene av observasjoner illustreres deretter. En kombinasjonen av frontlinjene til beste- og verste praksis anvendes i layering-teknikken, hvor det er kartlagt ulike effektivitet- og ineffektivitetsnivåer. DMUene i de ulike lagene er undersøkt for forskjeller- og likheter i ressursutnyttelse. Herunder presenteres det hvordan layering-metoden er anvendt til å kartlegge fellestrekk for de respektive effektivitetsinndelingene.

6.1 Resultater fra Data Envelopment Analysis - beste praksis

Figur 11 sammenfatter resultatene fra DEA-analysen. Det gjøres her et skille på resultatene fra analysen med teknologiforutsetning om konstant skalautbytte (TE_{CRS}) og analysen med antakelse om variabelt skalautbytte (TE_{VRS}). Dette for å kunne gjøre en vurdering vedrørende hensyn til skalaeffektivitet i analysen. Videre i delkapittelet vil de ulike resultatene forklares nærmere. Søylen med SE viser den målte skalaeffektiviteten, TE_{VRS} viser ren tekniske effektivitet og total effektivitet illustreres i søylen TE_{CRS} .



Figur 11 Resultat DEA beste praksis

Som det fremkommer av figur 11 har hofteproteseoperasjonene det høyeste målte effektivitetsnivået under både CRS og VRS. Gjennomsnittlig total effektivitet (TE_{CRS}) målt for hofteprotesene er 71,1%. Relativt nært ligger også kneproteser med 66,4% effektivitet. Både a-skopi kne og hallux valgus blir målt til å ha effektivitetsverdier under 50%.

6.1.1 Total effektivitet (TE_{CRS})

Beregning for den totale tekniske effektiviteten ble presentert i kapittel 4.4.1, og beregnes ved antakelse om konstant skalautbytte. Resultatene er presentert i tabell 10, som viser gjennomsnittlig resultat for hver av operasjonstypene.

Tabell 9 Total effektivitet

	Antall DMU	Antall effektive CRS	% antall effektive
<i>A-skopi kne</i>	174	8	4,60%
<i>A-skopi skulder</i>	175	9	5,17%
<i>Hofteprotese</i>	92	8	8,60%
<i>Kneprotese</i>	156	7	4,49%
<i>Hallux valgus</i>	198	5	2,51%

Som en kan se av tabell 9 er antallet effektive operasjoner innenfor hver gruppe ligger mellom 5 og 9 stykker, som prosentvis utgjør er en lav andel effektive observasjoner sett i sammenheng med det totale antallet observasjoner i datasettet. Hofteproteser er operasjonstypen med flest effektive frontobservasjoner i utvalget, med et prosentvis mål på 8,60% av DMUene. Blant operasjonene av typen hallux valgus er kun 2,51% av det totale antallet DMUer målt til å være effektive ved CRS-forutsetning. Dette gjenspeiles også i resultatene fra DEA-analysen for konstant skalautbytte vist i tabell 10, hvor hofteproteser har høyest gjennomsnittlig effektivitetsscore på 0,711. Videre kommer det også frem at hallux valgus har den laveste effektiviteten i denne analysen, 0,460.

Tabell 10 Resultat total effektivitet

	Gjennomsnitt	Standardavvik	Minst
<i>A-skopi kne</i>	0,468	0,199	0,087
<i>A-skopi skulder</i>	0,544	0,205	0,021
<i>Hofteprotese</i>	0,711	0,182	0,027
<i>Kneprotese</i>	0,664	0,145	0,440
<i>Hallux valgus</i>	0,460	0,197	0,045

6.1.2 Ren teknisk effektivitet (TE_{VRS})

Tabell 11 viser antall effektive DMUer med variabelt skalautbytte innenfor hver av de fem operasjonstypene. Da frontlinjen for denne rene teknisk effektiviteten omhyller beslutningsenhetene nærmere enn ved konstant skalautbytte, er det naturlig at effektivitetsverdiene er litt høyere og at flere av DMUene vurderes noe mer effektive i VRS-analysen. Operasjonene av typen hallux valgus har med forutsetning om variabelt skalautbytte 13 DMUer langs effektivitetsfronten. I tillegg har hofteproteser en liten økning i antall effektive DMUer, fra 8 DMUer på CRS-fronten til 13 effektive operasjoner på VRS-fronten. Kneproteser har den laveste prosentandelen med effektive DMUer med henholdsvis 5,77%. Det kan også sees at hofteproteser har den høyeste antallet med effektive DMUer på fronten med forutsetning om variabelt skalautbytte, herunder 13,98% av observasjonene.

Tabell 11 Ren teknisk effektivitet

	Antall DMU	Antall effektive VRS	% antall effektive
<i>A-skopi kne</i>	174	13	7,47%
<i>A-skopi skulder</i>	175	16	9,14%
<i>Hofteprotese</i>	92	13	13,98%
<i>Kneprotese</i>	156	9	5,77%
<i>Hallux valgus</i>	198	13	6,53%

Som det kan sees i tabell 12 er hofteproteser også operasjonstypen som er beregnet å ha høyest gjennomsnitt i effektivitetsverdiene for variabelt skalautbytte, med et estimat på 0,760.

Den minste effektivitetsverdien blant hofteproteseoperasjonene måles til 0,502. Dette er den høyeste minste-verdien. Herunder kan det tolkes at massen av hofteproteseoperasjoner ligger nærmere effektivitetsfronten, enn eksempelvis a-skopi kne som har minste effektivitetsverdi på 0,201 og antakelig en større variasjon. Til tross for at operasjonstypen kneproteser har det laveste målte antall for effektive operasjoner, 5,77%, har operasjonstypen en av de høyeste effektivitetsgjennomsnittene i analysen. Standardavvikene er lavest for hofteproteser og kneproteser, hvor spredningen derfor er litt mindre mellom de ulike effektivitetmålingene, enn for de andre operasjonstypene. Standardavviket for de øvrige operasjoner er relativt jevnt og ligger omkring 0,2.

Tabell 12 Resultat ren teknisk effektivitet

	Gjennomsnitt	Standardavvik	Minst
<i>A-skopi kne</i>	0,549	0,202	0,201
<i>A-skopi skulder</i>	0,627	0,203	0,252
<i>Hofteprotese</i>	0,760	0,143	0,502
<i>Kneprotese</i>	0,709	0,144	0,466
<i>Hallux valgus</i>	0,612	0,191	0,235

6.1.3 Skalaeffektivitet

På bakgrunn av at det kartlegges effektivitetsforskjeller mellom de ulike skalaforutsetningene er det er testet for tilstedeværelse av skalaeffektivitet i observasjonene. Resultatene presenteres i tabell 13.

Tabell 13 Resultat skalaeffektivitet

	Gjennomsnitt	Standardavvik	Minst
<i>A-skopi kne</i>	0,853	0,987	0,43
<i>A-skopi skulder</i>	0,868	1,009	0,08
<i>Hofteprotese</i>	0,937	1,270	0,05
<i>Kneprotese</i>	0,937	1,002	0,95
<i>Hallux valgus</i>	0,751	1,032	0,19

Skalaeffektivitet er som nevnt et mål på effektiviteten som henger sammen med observasjonens størrelse og skalautnyttelse. Skalaeffektiviteten i analysen er lik for hofteproteser og kneproteser, hvor begge har en utmålt verdi på 0,937. Dette kan tolkes som at operasjonene av hofte- og kneprotese i liten grad er signifikant forskjellig fra 1, hvilket videre indikerer relativt stor grad av skalaeffektivitet. Operasjoner av typen hallux valgus derimot, har en gjennomsnittlig skalaeffektivitet på 0,751, hvilket er signifikant forskjellig fra 1 og indikerer større tilstedeværelse av skalaineffektivitet og uheldig størrelsesorden.

Konstant skalautbytte vil kunne benyttes for proteseoperasjonene, men skala kan se ut til å ha betydning for de øvrige effektivitetsmålingene. Det er usikkerhet rundt hva skala representerer i denne sammenhengen, når det ikke er bedrifter som benchmarkes, men operasjoner. Fagområdet er komplisert og det vil kreve mye innsikt og kompetanse for å uttale seg om hvordan operasjoner eventuelt skalerer opp eller ned, samt hva som er en god størrelsesorden. Det gjøres en bankertest for skala, for å undersøke hvilken skalaforutsetning som vil være mest hensiktsmessig i den videre analysen.

6.1.4 Banker-test for skala

Samtlige av de kalkulerte verdiene i testen for skala er større enn den kritiske verdien slik det presenteres i tabell 14, og det påvises følgelig en signifikant forskjell mellom konstant- og variabelt skalautbytte. Nullhypotesen om at gjennomsnittlig effektivitetsscore er lik for de ulike forutsetningene og kan forkastes, da det kartlegges tilstedeværelse av skala ineffektivitet. Dette taler altså for en ekstra restriksjon og forutsetning om variabelt skalautbytte, slik at en hensyntar skalaen og fronten omhyller observasjonene så nært som mulig. I litteraturgjennomgangen er det heller ikke funnet annen forskning som gjør videre drøftelser rundt skala for operasjoner, så noen nærmere utredelse anses ikke hensiktsmessig. Det antas derfor variabelt skalautbytte for de videre analysene.

Tabell 14 Banker-test skala

	TEX			THN		
	Kalkulert verdi	Konf.int 95%	Kritisk verdi	Kalkulert verdi	Konf.int 95%	Kritisk verdi
<i>A-skopi kne</i>	1,1793	19	0,54114	1,3216	161	0,54424
<i>A-skopi skulder</i>	1,2220	19	0,54995	1,3850	161	0,55161
<i>Hofteprotese</i>	1,2014	19	0,54574	1,4871	161	0,56275
<i>Kneprotese</i>	1,1527	19	0,53546	1,2649	161	0,53731
<i>Hallux valgus</i>	1,3936	19	0,58222	1,7719	161	0,58931

6.2 Resultater fra Data Envelopment Analysis - verste praksis

Når det kommer til effektivitetsanalyser med hensyn på verste praksis, er det i realiteten de som er best på å være dårlige en har til hensikt å måle. Altså de mest ineffektive som kalkuleres å ha de ”verste” effektivitetsverdiene. Det er beslutningsenhetenes bruk av dårlige outputs, herunder ressursutnyttelse, som presenteres i dette delkapittelet.

6.2.1 Ren teknisk ineffektivitet

I analysen med modellen for verste-praksis er det hofteprotese som kommer best ut, med kun én ineffektiv operasjon, av totalt 88. A-skopi skulder og kneproteser har de høyeste prosentandelene for antall ineffektive DMUer i datasettet, på henholdsvis 5,81% og 5,13%.

Tabell 15 Ren teknisk ineffektivitet

	Antall DMU	Antall effektive VRS	% antall ineffektive
<i>A-skopi kne</i>	175	6	3,43%
<i>A-skopi skulder</i>	172	10	5,81%
<i>Hofteprotese</i>	88	1	1,14%
<i>Kneprotese</i>	156	8	5,13%
<i>Hallux valgus</i>	186	9	4,84%

I tabell 16 er resultatene for den rene tekniske ineffektiviteten presentert. Hallux valgus er operasjonstypen som har lavest gjennomsnitt på 1,97. Operasjonstypen har også det laveste standardavviket, hvilket i dette tilfellet betyr at gjennomsnittet til hallux valgus ligger tettest opp mot ineffektivitetsfronten. Følgelig er hallux valgus også den operasjonstypen med høyest tilstedeværelse av ineffektivitet. A-skopi kne er operasjonstypen som scorer høyest for teknisk ineffektivitet, med 3,17. Dette tilsier at gjennomsnittet for operasjonene av typen a-skopi kne ligger lengre fra ineffektivitetsfronten, og er best på å unngå dårlige outputs.

Tabell 16 Resultat ren teknisk ineffektivitet

	Gjennomsnitt	Standardavvik	Størst
<i>A-skopi kne</i>	3,17	2,05	12,82
<i>A-skopi skulder</i>	2,18	1,16	7,70
<i>Hofteprotese</i>	2,46	1,47	8,44
<i>Kneprotese</i>	2,13	1,23	8,55
<i>Hallux valgus</i>	1,97	0,96	6,54

6.3 Layering

Det ble kjørt fire runder med layering, gjennom en kombinasjon av beste- og verste praksis analysene. For hver runde blir de effektive frontobservasjonene identifisert og plukket ut av datasettet, før neste analyse gjennomføres Slik er det dannet fire grupper for hver av modellene, hvor DMUene som ligger ytterst i begge retningene inndeles i totalt åtte effektivitet- og ineffektivitetsgrupper. På bakgrunn av argumentasjonen for skala i forrige delkapittel, er layering-analysene gjort med forutsetning om variabelt skalautbytte. For å vise metodens anvendelighet på observasjoner av denne typen, samt hvordan den kan bidra til effektiviseringsarbeid, presenteres layeringmetoden på operasjonstypen a-skopi kne.

6.3.1 Beste praksis

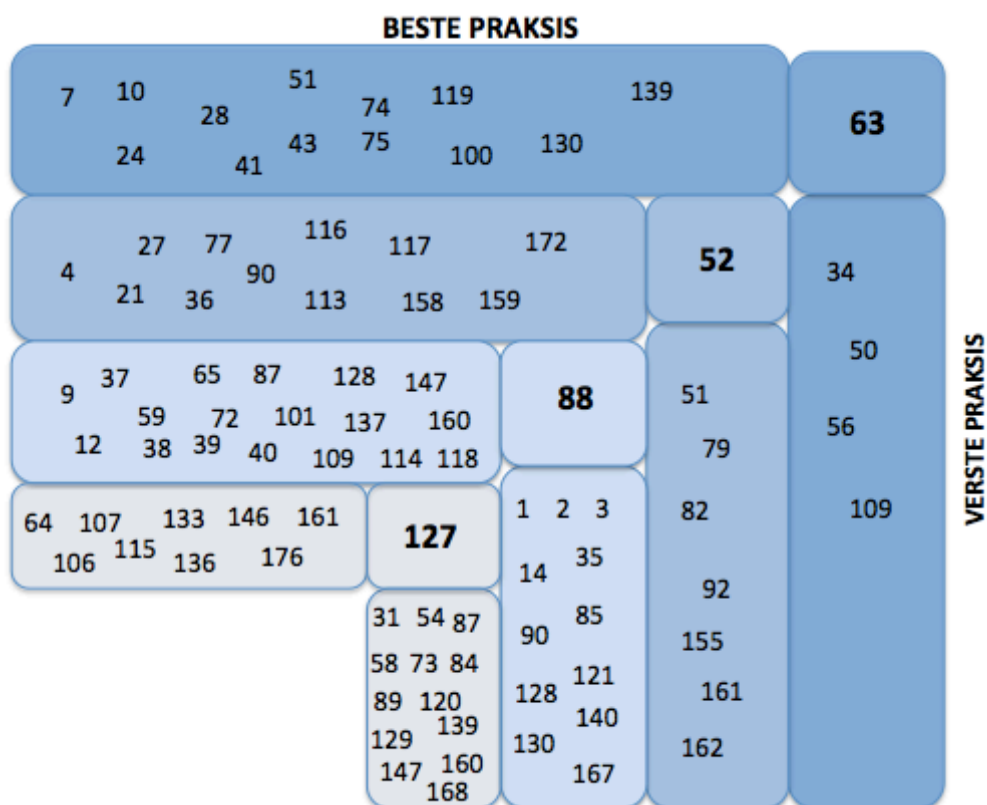
I første runde ble det identifisert 14 effektive DMUer. Hovedtrekkene ved disse var at de hadde en lav verdi på kostnad per pasient, samt lav forberedelsestid. I andre runde, når frontobservasjonene fra første analyse er trukket ut, ble det identifisert 13 effektive

frontobservasjoner, disse hadde de samme hovedtrekkene i ressursanvendelsen som operasjonene i den første runden. I runde nummer tre identifiseres det hele 20 effektive observasjoner langs frontlinjen. Her sees det et en endring i tendensene til variablene hos DMUene, sammenlignet med hvordan variablene var for observasjonene i de to første rundene. På dette nivået har observasjonene hovedsakelig høy kostnad per pasient, samtidig varierer forberedelsestiden mellom kort og lang. I siste runde har de effektive fortsatt en høy kostnad per pasient, men ingen tydelige tendenser i de andre variablene. Samlet sett viser analysen at det kan være en sammenheng mellom størrelsen på variablene kostnad per pasient og forberedelsestid, når en estimerer effektivitetsverdien til de mest effektive observasjonene i beste-praksis. Resultatene presenteres i figur 12.

6.3.2 Verste praksis

Ved layering i modellen for verste praksis ble det observert en gradvis økning i antallet frontobservasjoner for hver analyserunde. Her identifiseres de mest ineffektive DMUene, front for front. Første runden ble det identifisert 6 DMUer som danner fronten av ineffektive enheter. Felles for disse var betydelige forsinkelser i variabelen for sen operasjonsstart. Det sees derimot ingen tydelig sammenheng mellom observasjonenes KPP-verdier. I andre runde og på frontlinje nummer to ble det identifisert 8 ineffektive DMUer. Disse har også et fellestrekk for høy grad av forsinkelser i operasjonsstart, men de har imidlertid også høye kostnader per pasient. I tredje runde av analysen dannet 13 DMUer den ineffektive fronten. Her kan det se ut til at en DMU skiller seg fra de resterende på ineffektivitetsfronten. Denne DMUen har høy verdi på kostnad per pasient, samt noe overforbruk av den planlagte stuetiden. De øvrige observasjonene på den tredje fronten hadde en lav verdi på kostnad per pasient, men store forsinkelser i operasjonsstart. Den fjerde frontlinjen består av hele 14 ineffektive observasjoner, og også her skiller DMUene seg i ressursforbruket. Den ene gruppen i denne fronten har høye KPP-verdier og overforbruk av stuetid. Den andre gruppen i dette ineffektivitetslaget har også høye KPP-verdier, men er derimot mer forsinket i oppstartstiden. Resultatene fra layering i verste praksis kan også sees i figur 12.

For hver runde i layering-analysen ble det oppdaget en DMU som plasserte seg på fronten for både beste og verste praksis. Disse ligger i kryssningspunktet mellom effektivitetsfrontene.



Figur 12 Resultat layering

Hensikten med å identifisere self-identifiers er ikke nødvendigvis at disse er irrelevante eller dårlige observasjoner i analysen. De kan være like reelle, samt ha forbedringspotensiale på lik linje med de andre observasjonene. Det anses ikke som hensiktsmessig å bruke tid på å fjerne disse observasjonene, da disse hovedsakelig kun er forstyrrende for seg selv. Det vil kunne være fordelaktig å være oppmerksomme på disse, siden observasjonene identifiserer seg med seg selv og er deres egne referanseenheter. Konklusjoner og slutninger bør derfor ikke gjøres for disse DMUene. Det kan også være nødvendig med egne analyser for disse for nøyere kartlegging av deres respektive nivåer effektivitet eller ineffektivitet.. Dette vil være spesielt gjeldende i analyser med antakelse om variabelt skalautbytte (Paradi et. al, 2004).

6.4 Banker-test på grupper

Det er gjennomført banker-test på gruppenivå for å kartlegge om det er signifikante forskjeller i effektivitetsnivået innen de ulike operasjonstypene. Det er grupper av operasjoner innenfor ulike operasjonstyper som måles mot hverandre i testen, og operasjonstypene som

anvendes til banker-testene er a-skopi skulder og hofteproteser. Gruppeinndelingen er gjort på bakgrunn av sesong, herunder sommer- og vintermånedene.

6.4.1 Banker-test grupper beste praksis

I første omgang ble det forsøkt å påvise signifikant forskjell i effektiviteten til a-skopi skulder når DMUene var inndelt i grupper for henholdsvis vinterhalvår og sommerhalvår. Det ble observert en relativt stor forskjell i gjennomsnittlig effektivitet for disse gruppene, hvor 8 prosent skilte effektivitetsgjennomsnittet til sommergruppen og vintergruppen. Bankertesten viste derimot ingen indikasjon på at effektivitetsforskjellene var signifikante, og en har derfor ikke grunnlag til å påstå at disse kan forklares med annet enn tilfeldigheter.

I neste omgang ble det derfor gjennomført tester på grupper inndelt for færre måneder, herunder vinter- og sommermånedene. Det er da gjennomført bankertester for både operasjoner av typen a-skopi skulder og for hofteproteseoperasjoner i hver sin omgang. Operasjoner gjennomført i desember, januar og februar danner den éne gruppen. Den andre gruppen består av operasjoner gjennomført i juni, juli og august. Gruppens effektivitetsnivå måles mot hverandre for å undersøke om forskjellene skyldes tilfeldigheter, eller om det kan konstateres signifikante forskjeller. Det skilles også mellom effektiviteten som er kartlagt ved henholdsvis beste praksis og verste praksis. Nullhypotesen som testes for gruppene er at det ikke er forskjell mellom effektivitetsnivået i gruppene, og der den kalkulererte verdien overstiger den kritiske verdien som er utregnet vil en kunne forkaste nullhypotesen.

Innledningsvis kunne det konstateres at effektivitetsmålene fra beste praksis viser en høyere gjennomsnittlig effektivitet for både hofteproteser og a-skopi skulder i sommermånedene, kontra vintermånedene. Dette er illustrert i tabell 17.

Tabell 17 Gjennomsnittlig effektivitetsscore banker-test beste praksis

	Gjennomsnittlig effektivitetsscore totalt	Gjennomsnittlig effektivitetsscore sommer	Gjennomsnittlig effektivitetsscore vinter
<i>Hofteprotese</i>	0,7763	0,8365	0,7510
<i>A-skopi skulder</i>	0,6683	0,7172	0,6402

Tabell 18 Banker-test sesongvariasjon beste praksis

	TEX				THN			
	Kalkulert verdi	Kritisk verdi	0,025	0,975	Kalkulert verdi	Kritisk verdi	0,025	0,975
<i>Hofteprotese</i>	0,9115	0,4769	0,0256	39	0,9148	0,4858	0,0015	647,789
<i>A-skopi skulder</i>	0,4516	0,3111	0,0256	39	0,4285	0,3690	0,0015	647,789

I tabell 18 presenteres resultatene fra bankertestene på operasjoner gjennomført i sesongene vinter og sommer. Den kalkulerede verdien for hofteproteseoperasjonene, ved antakelse om eksponentialfordeling, er 0,9115 og er følgelig høyere enn den kritiske verdien som er 0,4769. Under antakelse om halv-normal fordeling er det tilsvarende forskjell mellom kalkulert og kritisk verdi, herunder 0,9148 mot en kritisk verdi på 0,4858. For hofteprotese påvises det derfor med bankertesten en signifikant forskjell mellom sommer- og vinteroperasjonene. Det kan derfor antas med 95 prosent sannsynlighet at disse to operasjonstypene er bedre på ressursutnyttelse i sommermånedene.

Det måles også større kalkulerede verdier enn kritiske verdier for a-skopi skulder, ved både eksponential- og halvnormal fordeling. Dette viser at en også her kan påvise signifikant forskjell mellom sesongene sommer og vinter, og at operasjonene er betydelig mer effektive i sommermånedene.

6.4.2 Banker-test grupper verste praksis

Det er også undersøkt om en kunne påvise signifikante forskjeller mellom sommer- og vintersesongen på effektivitetsmålinger gjort med verste praksis. Her vil en da måle signifikans for forskjellene i grad av ineffektivitet. Nullhypotesen er følgelig at det ikke er forskjell i ineffektivitetsmålene for disse to sesongene. Gjennomsnittsverdiene for ineffektiviteten målt i verste praksis er illustrert i tabell 19.

Tabell 19 Gjennomsnittlig effektivitetsscore banker-test verste praksis

	Gjennomsnittlig effektivitetsscore totalt	Gjennomsnittlig effektivitetsscore sommer	Gjennomsnittlig effektivitetsscore vinter
<i>Hofteprotese</i>	2,2139	2,1757	2,3704
<i>A-skopi skulder</i>	2,3788	1,1776	2,3989

Når en sammenligner effektivitetsmålene fra analysen gjort med modellen for verste praksis kan det sees at sommermånedene har høyere gjennomsnittlig ineffektivitet, da disse ligger nærmere verdien 1. Dette gjelder for både hofteproteser og a-skopi skulder. Altså er vintermånedenes effektivitetsgjennomsnitt bedre når en benytter effektivitetsmålene fra verste praksis, fordi flertallet av disse operasjonene ligger lengre unna ineffektivitetsfronten. Herunder kan det sies at operasjonene på vinteren er bedre til å unngå dårlig ressursutnyttelse. Det gjennomføres bankertester for å undersøke om en kan påvise at disse forskjellene ikke skyldes tilfeldigheter. Resultatene fra bankertesten på de to sesongene for verste praksis presenteres i tabell 20.

Tabell 20 Banker-test sesongvariasjon verste praksis

	TEX				THN			
	Kalkulert verdi	Kritisk verdi	0,025	0,975	Kalkulert verdi	Kritisk verdi	0,025	0,975
<i>Hofteprotese</i>	0,3598	0,2646	0,0256	39	0,2488	0,2945	0,0015	647,789
<i>A-skopi skulder</i>	0,5517	0,3556	0,0256	39	0,3694	0,3477	0,0015	647,789

For hofteproteseoperasjonene kan det konstateres signifikant forskjell mellom sesongene dersom det antas eksponentialfordeling. Ved halv-normalfordeling vil en derimot ikke kunne påvise signifikante forskjeller, siden den kalkulerede verdien ligger under den kritiske verdien kan ikke nullhypotesen om ingen forskjell mellom sesongene forkastes. For operasjonene av typen a-skopi skulder derimot, påvises signifikant forskjell ved begge fordelingene. Det vil derfor kunne konstateres signifikante forskjeller mellom sommer- og vintersesongen for a-skopi skulder med 95 prosent sannsynlighet.

7 Diskusjon

Hensikten med denne studien er å illustrere DEA-analysers anvendelighet til effektivitetsmålinger på operasjonsnivå ved UNN. Da det er ønskelig å kartlegge tilstedeværelsen av ineffektivitet og på best mulig måte kartlegge områdene med forbedringspotensial. For å kunne svare på anvendeligheten til DEA-analyser av typen beste- og verste praksis vil det innledningsvis drøftes rundt modellenes kvalitet. Variablenes egenskaper er avgjørende for denne diskusjonen og vurderes herunder. Videre gjøres det sammenligninger med valg av variabler i annen forskning, samt deres diskusjoner rundt variablenes anvendelighet. Avslutningsvis gjøres det en drøftelse rundt hvordan analysene, samt resultatene fra disse, vil kunne bistå UNN i deres arbeid med å øke effektivitetsnivået i operasjonsaktivitetene. Dette gjøres gjennom utprøving av layering-teknikken, samt bankertester på grupper.

7.1 Variablenes anvendelighet

Det er benyttet to ulike modeller i analysen. Disse måler effektivitetsnivået til observasjonene i datasettet, med hensikt å finne de mest effektive eller ineffektive observasjonene. I modellen for beste praksis benyttes tre inputvariabler: KPP, forberedelsestid og knivtid. Det benyttes en outputvariabel: DRG-vekt. I modellen for verste praksis er DRG-vekt benyttet som inputvariabel og KPP, differanse stue inn og differanse stuetid er variabler output. Da variablene som anvendes i modellene for beste- og verste praksis overlapper hverandre, vil diskusjonen rundt variablene være samlet. Dernest følger diskusjonen vedrørende de ulike modellens bidrag.

KPP-verdi er som nevnt et mål for kostnad per pasient. Dette var ferdigutviklet for datamaterialet i analysen i januar 2017, og har vært avgjørende for å kunne innlemme kostnader i analysen. Variablen som benyttes bygges opp av den enkelte pasientens operasjonskostnader, hvilket bidrar i modellene som anvendelig kostnadsindikator. Denne verdien er kun gjeldende for operasjonsaktiviteten. Manglende kostnadsvariabler var en uttalt hindring for analysen til Dons og Skorstad (2012), hvor DEA-analyser viste seg å være utfordrende å gjennomføre på grunn av manglende indikatorer for kostnadsbildet. KPP-verdien påvirkes av behandlingene som pasienten mottar og knytter sammen alle direkte- og indirekte kostnader som er tilknyttet den enkelte pasientens operasjon. Dersom pasienten har en infeksjon, behov for ekstra medisiner eller spesialproteser vil dette være uttrykt i KPPen til

gjeldende pasient. Således anses KPP-verdi å være anvendelig, men også en solid variabel for å inkludere kostnadsbildet.

Kostnadsindikatorene som benyttes i annen forskning bygger i stor grad på årsverk til leger og annet personell. I tillegg har blant annet Kittelsen et al. (2007) benyttet øvrige driftskostnader som kostnadsindikator. Dette menes å være både hensiktsmessig og velfungerende i analyser med intensjoner om å måle overordnet effektivitetsnivå, som også var tilsiktet i studien til Kittelsen et al. (2007). Andre studier har benyttet ulike overordnede kostnadsindikatorer, som for medisinsk forbruk for sammenligning mellom ulike sykehus sine prestasjoner. Dette ble også vurdert til denne oppgaven, men ved å benytte KPP, er samtlige av de overnevnte kostnadene innlemmet. Det er også et mer presist mål for kostnadsforbruket på pasientnivå enn en eventuell dekomponering av andre overordnede kostnader ble vurdert til å være. Siden KPP-verdien som anvendes i denne analysen er gjeldende for hver enkelt operasjon, legges det til rette for en mer særskilt effektivitetsanalyse. Det er således også tilrettelagt for DEA-analyser på et lavere enhetsnivåer.

DRG-vekt er i utgangspunktet en verdi som representerer ressursene som anvendes til å behandle en pasient. Denne vekten fremstiller implisitt pasientens behandlingsforløp, og tar høyde for hvor komplisert og utfordrende den enkelte pasientens sykdomsbilde og behandling er. Inkludert i vekten er blant annet faktorer for sykdommens alvorlighetsgrad, prosedyrer gjennomført og hvilket medisinsk utstyr som er benyttet. Disse DRG-poengene legges til grunn når det bestemmes hvor mye av de medgåtte kostnadene sykehuset får refundert. DRG-vektene som produseres av sykehuset gjenspeiler derfor sykehusets forventede inntekt, tilnærmet likt som produksjonen av produkter for salg vil gjøre i en produksjonsbedrift. Herunder kan det argumenteres at DRG-vekten er en velfungerende outputvariabel, til tross for at den også representerer ressursbruk. I Kittelsen et al. (2007) benyttes også DRG-poeng som outputvariabel som mål på produksjonen. I deres studie utforskes effektivitet mer overordnet for sykehus, og de har derfor anvendt flere DRG-vekter som output for å inkludere alle pasientgruppene i sykehusene. Andre outputvariabler er også argumentert å være velfungerende, eksempelvis har Hagen et al. (2001) benyttet en sammensatt variabel som tar høyde for polikliniske refusjoner i forhold til antall innleggelser. DRG-vekten vil være avgjørende for refusjonene til UNN, og da midlene utmåles på bakgrunn av sykehusets produserte DRG-vekter, kan disse variablene anses som to sider av samme sak.

Hver operasjon har ulik kompleksitet, utfordringer og behandlinger som DRG-vekten også gir uttrykk for. Dette vil en derfor også kunne ta høyde for i analysemodellen. Kittelsen et al. (2007) påpeker også at DRG-systemet er relativt likt for de skandinaviske landene, og vekten tas blant annet i bruk i denne studien for å gjøre forskningen sammenlignbar med annen forskning. Det gjør også at analysemodellene i større grad er anvendbar på tvers av landegrenser.

KPP-verdi og DRG-vekten er variabler som benyttes i begge modellene. Det benyttes derimot forskjellige variabler for å innlemme tidsforbruket i de ulike modellene. I modellen for verste praksis anvendes differansen mellom planlagt- og anvendt stuetid, samt heftelse i operasjonsstart. Disse vil gi uttrykk for forsinkelser eller merforbruk av tid i forbindelse med operasjonene. Ved å anvende to ulike variabler som forklarer forsinkelser er det ønskelig å gjøre et skille mellom operasjonene som blir forsinket underveis og operasjoner som starter senere på grunn av allerede oppstått forsinkelse. Oppstartstidvariabelen hensyntar forholdet før operasjonsaktiviteten, herunder om forsinkelsen skyldes forhold i forkant av den gjeldende operasjonen. Variabelen for forsinkelse i stuetiden vil forklare tidsoverskridelser under den undersøkte operasjonen.

Forsinkelser og merforbruk av tid er negative faktorer i effektivitetsbildet. Intensjonen med å benytte disse i modellen for verste praksis er å kunne belyse hvor lite effektiv de ineffektive er i realiteten. Andre variabler kunne vært dannet for å ta hensyn til tidsoverskridelser, eksempelvis knivtid eller tidsbruken på etterarbeidet. Bakgrunnen for valget av oppstartstid for operasjonen og stuetid var at de skulle dekke mest mulig av operasjonsaktiviteten. Den ene skulle kunne gi uttrykk for forsinkelser som er forårsaket allerede før operasjonen. Den andre tidsvariabelen er ment å kunne ta hensyn til selve operasjonsaktivitetens merforbruk fra planlagt tidsbruk. Det er noen få operasjoner som får negative verdier i denne variabelen, og dette skjer på bakgrunn av at operasjonen starter før planlagt tid eller har vært raskere enn forventet. Det anses lite hensiktsmessig å diskutere operasjonene som starter før planlagt tid, siden dette hører til sjeldenheten, samt at variablene har til hensikt å måle ineffektivitet. Det nevnes likevel at disse variablene ikke var problematiske for analysen, og heller ikke ga nevneverdig utslag på effektivitetsmålet til gjeldende operasjoner.

De tidsbaserte variablene som benyttes i modellen for beste praksis er forberedelsestid og knivtid. Begge disse er faktisk brukte tidsmål, i motsetning til de foregående som er

differansemål. Den aktuelle forberedelsestiden og knivtiden er også tenkt å være mer nøytrale i forhold til effektivitetsmåling. Variablene for faktisk tidsforbruk vil være velfungerende når en skal kartlegge det generelle effektivitetsbildet blant observasjonene. Disse hensyntar to viktige momenter i forbindelse med operasjonen, herunder forberedelsene og selve aktiviteten. Hverken Hagen et al. (2001) eller Kittelsen et al. (2007) anvender variabler som omhandler tidsaspekter i sine analysemodeller. Effektivitetsanalysene i disse studiene er derimot gjort med hele sykehus som beslutningsenheter. Variabler for tidsforbruk og nødvendigheten rundt disse er dessverre ikke diskutert i noen av studiene, men det kan tenkes at det ble ansett som lite hensiktsmessig å vurdere hele sykehusets tidsbruk samlet. Det observeres at disse heller anvender variabler som omfatter ulike former for driftskostnader som innsatsfaktorer. I operasjonssammenheng kan tidsaspektet tenkes å ha vesentlig påvirkning på effektiviteten, og i hvert fall være vesentlig når det kommer til ressursforbruket. Forsinkelser skaper også store synergier, hvor konsekvensene kan være betydelige forsinkelser eller strykninger for andre operasjoner. Det er derfor ansett som fordelaktig å inkludere tidsvariabler i modellene for denne oppgaven.

7.1.1 Svakheter i variablene

I henhold til Dyson et al. (2001) burde de inkluderte variablene være av betydning for hele produksjonsprosessen, og i tillegg være felles for alle DMU-ene. Faktorer som det ikke tas høyde for i analysemodellen er blant annet LIS-leger og medisinstudenters deltakelse i operasjonene. LIS-leger skal i hovedsak være delaktig i hver operasjon, mens studenter har mer samlingsbaserte besøk på operasjonsstuene. Da deres deltakelse i operasjonene sannsynligvis vil gi utslag i effektivitetsmålingen ville det vært mer optimalt for analysens pålitelighet å kunne inkludere denne faktoren. Det er foreløpig utfordrende å benytte variabler for å ta hensyn til disse, da det ikke er laget en overordnet oversikt over hvilke praktikanter som deltar i operasjonene og eventuelt hvor erfarene og delaktige disse er. Følgelig har en heller ikke en sikker indikator for hvor mye undervisningsdelen påvirker effektivitetsnivået.

Tidlig i spesialiseringsforløpet vil LIS-legene være observatører, mens hovedkirurgene utfører mesteparten av operasjonen. Erfarne leger i spesialisering vil derimot kunne gjennomføre store deler av prosedyrene. Dette vil naturligvis gi utslag på effektivitet, da undervisningsaktivitet antakelig har stor betydning for tidsbruk og lengde på operasjonene. Kanskje særlig med tanke på opplæring og kvalitetssikringen som må gjøres av

hovedkirurgen i forbindelse med operasjonene. Det vil likevel ikke påvirke effektivitetsmålingene i analysen direkte, da det faktisk brukte tidsaspektet måles. Dette vil heller kunne betraktes som en mulig forklaringsfaktor på de ulike effektivitetsnivåene som beregnes.

Det kunne også vært fordelaktig å inkludere variabler i modellene som tok hensyn til kvaliteten i behandlingene, men som det påpekes av Puig-Junoy (1998) er dette en velkjent utfordring når det kommer til effektivitetsanalyser på sykehusnivå. I denne artikkelen presenterer Puig-Junoy at DEA-analyser på sykehus vil ha vanskeligheter med å ta hensyn til kvalitet gjennom velfungerende variabler, da det sjeldent måles kvantitativt på slike faktorer. Problematikken baseres hovedsakelig på at sykehusets produksjon består i helsetjenester, som er svært komplekse og sammensatte tjenester. Effektivitetsanalysene innen helse- og omsorg tar derfor i sjeldent høyde for kvalitet, forbedret livskvalitet eller sykdomsforbedring. I nevnte studie, forsøker Puig-Junoy å ivareta kvalitetsaspektet ved å benytte variabler for blant annet sykdommens alvorlighetsgrad ved innleggelse, sammenholdt med reduksjonen i alvorlighetsgraden ved utskrivelse. Dette hevdes å være velfungerende for å ivareta kvalitetsdimensjonen. Til gjengjeld ble effektivitetsstudien gjennomført for sykehus som allerede hadde vært anvendt i en annen studie som undersøkte kvalitetsaspekter ved behandlingen. Dataen fra kvalitetsstudien er derfor innhentet spesifikt til måling av kvalitetsfaktorer, herunder blant annet sykdomsbildet og eventuell forbedring ved utskrivelse. Dette datamaterialet fikk Puig-Junoy låne i forbindelse med effektivitetsanalysen. Til denne oppgaven, hvor effektivitetsanalysen gjøres på operasjonsaktiviteten ved UNN, er det ikke funnet slike kvalitetsvariabler.

Som nevnt tidligere kan en svakhet også være at DRG-vektene er at den dannes på grunnlag av hele behandlingsforløpet. Til tross for at operasjonene er en stor del av dette, vil likevel en andel av denne vekten forklares av elementer som ikke er relevant for analysen. Modellen ville kunne vært mer reliabel dersom en hadde hatt mulighet til å identifisere operasjonsdelen av vekten. For videre også å anvende bare operasjonsdelen av DRG-vekten som outputvariabel i analysen. På grunn av måten DRG-vekten bygges opp per i dag er dette dessverre ikke mulig.

DRG-vekten vil likevel være et mål på hvor utfordrende og kompleks behandlingene har vært i forhold til hverandre. Forholdet mellom de ulike operasjonenes DRG-vekter vil

sannsynligvis være relativt likt som forholdet gjennom for hele behandlingsforløpet. I stor grad anses variabelen derfor å være gjeldene for forholdene i de ulike operasjonene. Dersom en pasient har hatt en infeksjon i forbindelse med sykehusinnleggelsen eller krever ekstra oppfølging på grunn av bisykdommer som for eksempel diabetes eller demens vil dette også være gjeldende for operasjonsdelen av behandlingsforløpet. Variabelen anses derfor som rettvise for flertallet av operasjonene, selv om de sannsynligvis her er presentert på en høyere skala. Sammenholdt med det faktum at annen forskning også benytter DRG-vekter eller tilsvarende variabler for output, formodes variabelen og ytterligere også modellen å være troverdig for majoriteten av observasjonene.

Modellenes variabler er valgt på bakgrunn av studiens formål og hensikt. Målet har vært å undersøke graden av effektivitet på operasjonsnivå gjennom DEA-analyser, for å kunne kartlegge et eventuelt forbedringspotensial. Variablene benyttet i modellene anses som de mest fordelaktige på bakgrunn av datamaterialet som er til rådighet. Det kan kanskje tenkes å være mer fordelaktig med DRG-verdier som er gjeldende kun for operasjonsaktiviteten. Da disse i dag representerer hele behandlingsforløpet, vil en ha vanskeligheter med å påstå at resultatene er ubetinget riktige og representative. Det kan likevel argumenteres for at de store linjene i resultatet vil påvirkes i mindre grad av denne svakheten. I tillegg vil det være interessant å kunne kontrollere for undervisningsaktiviteten i operasjonene. Da undervisningen kan ha relativt stor påvirkning på tidsaspektet i operasjonene. LIS-leger skal være tilstede ved alle operasjoner. Dette betyr at selv om en ikke kan forklare i hvilken grad operasjonene er påvirket av undervisning, vil analysens effektivitetsmål være reelle. Ineffektiviteten som måles er derfor tilstede. Svakheten består i at en ikke kan forklare hvilke av de lave effektivitetsverdier som har oppstått på grunn av undervisning. Dette anses derfor ikke som en hindring for effektivitetsanalysene, og resultatene vurderes derfor å være troverdige og reelle mål for effektiviteten i operasjonsaktiviteten.

7.2 Anvendeligheten til modellen for beste praksis

I første omgang ble det gjort en analyse for kartlegging av beste praksis, herunder operasjonene med de beste effektivitetsnivåene. Av analysen kommer det frem at effektivitetsnivåene er generelt lave for en stor andel av de ortopediske operasjonene. Majoriteten av operasjonene har lav effektivitet, og gjennomsnittlig effektivitetsnivå ligger mellom 46 prosent og 71 prosent med forutsetning for konstant skalautbytte. Ved forutsetning

om variabelt skalautbytte utmåles de ulike operasjonene til å ha en effektivitetsverdi med gjennomsnitt mellom 55 prosent og 76 prosent. Medianen til effektivitetsverdiene til askopioperasjonene på kne er blant annet bare 51 prosent og det er kun 36 av observasjonene som har en effektivitetsverdi på over 70 prosent.

Når det kom til analysen av skalaeffektivitet ble det identifisert stor forskjell mellom de ulike operasjonstypene. Det gjøres en kort diskusjon i resultatkapittelet 6.1.4 vedrørende utfordringene rundt skala og hva det vil utgjøre i operasjonssammenheng. I denne analysen, og med operasjoner som observasjonsenheter, er det usikkerhet knyttet til hva skala er og hvordan dette kan anvendes. Siden det er utfordrende å svare på slike spørsmål uten inngående innsikt og kompetanse i helse- og medisinfaget, gjøres det ingen nærmere drøftinger rundt skaleringen til operasjoner. Det forutsettes variabelt skalautbytte videre i analysene og ren teknisk effektivitet blir utforsket og kartlagt.

Effektiviteten som kartlegges med modellen for beste praksis anses å være troverdig, men effektivitetsnivåene er som nevnt ikke utpreget høy for de ortopediske operasjonene. Et viktig moment er også at operasjoner ikke nødvendigvis bør ha til sikte å være fullkommen effektiv, på bakgrunn av blant annet kvalitetsaspektet og pasientsikkerhet. Det vil ikke nødvendigvis være hensiktsmessig å forsøke å tilegne de øvrige operasjonene like egenskaper som de operasjonene som ligger langs den effektive frontlinjen, da dette vil kunne gå på bekostning av kvaliteten i behandlingene. Sammenholdt med lave nivåer for effektivitet blant operasjonene, gjennomføres det videre analyser som i større grad er rettet mot ineffektivitet. Analysen kan således anses å være anvendelig til å måle effektivitet, men kanskje ikke hensiktsmessig når analysen blir stående alene. Argumentasjonen for dette er således at sykehuset kan ha et ønske om og for effektive operasjoner, men at forbedringsarbeidet for å heve det totale effektivitetsnivået i operasjonene er best tjent ved å fokusere på å heve effektivitetsnivået til de «verste» yterne. Modellen for verste praksis er derfor utforsket, for å kunne ha best mulig anlegg til kartlegging og forbedring av de ineffektive observasjonene.

7.3 Anvendeligheten til modellen for verste praksis

Modellen for verste praksis inkluderer tidsvariabler som tar hensyn til forsinkelser både før og under operasjonene. Disse variablene er benyttet for å fremheve negative effektivitetsegenskaper i effektivitetsmålingen. I tillegg benyttes også DRG-vekt og KPP-

verdi på lik linje som i modellen for beste praksis. Effektivitetsmålene for verste praksis viser, ved antakelse om variabelt skalautbytte, gjennomsnittlige ineffektivitetsverdier mellom 1,97 og 3,17. Her kommer det også til uttrykk at mange observasjoner er lite effektiv, gjennom høye ineffektivitetsmål. Det ble identifisert 6 observasjoner langs ineffektivitetsfronten, hvilket er «læremestere» for de andre observasjonene. Det kan ses tendenser til at de som måles til å ha beste effektivitetsverdier i modellen for beste praksis, også er de observasjonene som har laveste ineffektivitetsverdier i modellen for verste praksis. Modellene kan på denne måten si å bekrefte hverandre. Det er derimot en liten forskjell på hvilke observasjoner som regnes å være mest ineffektiv i modellene for beste- og verste praksis. Det kan tyde på at tidsvariablene vedrørende forsinkelser og merforbruk i tid, har gitt litt andre målinger for de ineffektive observasjonene i verste praksis.

Gjennom denne analysemodellen dannes en god oversikt over hvilke operasjoner som er «best» på å være ineffektiv, altså operasjonene med lavest nivåer for effektivitet. At disse ineffektive observasjonene er såkalte læremestere, vil kanskje være en banal påstand for ineffektive operasjoner, sammenlignet med læremestrene fra beste praksis som er de mest effektive operasjonene. Likevel kan de ineffektive tenkes å være rollemodeller for hvordan en operasjon ikke bør være. Tanken er at UNN vil kunne benytte analyser for verste praksis til å kartlegge de mest ineffektive operasjonene, og videre forsøke å heve effektivitetsnivåene til disse. Dersom en trekker paralleller til pareto-prinsippet, vil kanskje 80 prosent av ineffektiviteten skyldes 20 prosent av operasjonene. Tanken er i hvert fall at det er hensiktsmessig å kartlegge de mest vitale operasjonene for forbedringsarbeidet, slik at innsatsen ilegges der det vil være mest formålstjent. Sannsynligvis vil det være mest hensiktsmessig å forbedre de mest ineffektive, for å være sikker på at en ikke påvirker kvaliteten i behandlingene. En kan følgelig dra nyttig lærdom av disse ineffektive læremestrene, og analysen for verste praksis anses som mer hensiktsmessig for å danne grunnlag til forbedringsarbeid.

7.4 Resultatenes bidrag til effektiviseringsarbeidet

Både modellen for beste- og verste praksis anses som tilfredsstillende, med relativt solide og velfungerende variabler som er relevant for analysens motiv. Siden modellene kan vurderes å være reliable, vil også effektivitetsmålene tolkes som pålitelige. Først og fremst kan sykehuset anvende DEA-analysene til å gjøre tolkninger vedrørende tilstedeværelsen av ineffektivitet i

operasjonsaktivitetene. Videre vil en også kunne gjøre seg opp en mening om hvor forbedringspotensialet befinner seg.

Effektivitetsnivåene som kartlegges i denne analysen er generelt lave, og et relativt stort forbedringspotensial påvises for effektivitetsnivået til alle operasjonstypene som undersøkes. I hvilken grad det er ønskelig å forbedre effektivitetsnivået er derimot uvisst. Dette på bakgrunn av at ivaretagelse av pasientsikkerhet, og at en sørger for kvalitet i operasjonsaktivitetene, utvilsomt er det viktigste hensynet for UNN. Det er ikke ønskelig at effektivisering skal gå på bekostning av kvaliteten i behandlingene. Derfor er det vanskelig å ta stilling til hvor stor del av ineffektiviteten som potensielt kan og bør forbedres. Operasjonene med relativt høye effektivitetsnivåer vil likevel kunne sies å ha mindre av dette realistiske og hensiktsmessige forbedringspotensialet enn de operasjonene med høy grad av ineffektivitet. Det argumenteres følgelig for at sykehuset vil være best tjent med å fokusere på forbedringsarbeid blant operasjonene med de laveste effektivitetsnivåene. Det er følgelig håp for gode effektivitetsnivåer, gjennom å forbedre de operasjonene med størst grad av ineffektivitet.

7.5 Banker-tester på grupper

En stor del av å tilrettelegge for forbedringsarbeid vil være å kartlegge hvor dette arbeidet vil være best tjent. For å kartlegge om forskjellene i effektivitet innenfor de ulike operasjonstypene er signifikant eller om de skyldes tilfeldigheter, er det benyttet asymptotiske tester, herunder bankertester. Det ble funnet store forskjeller i effektivitetsnivåene for de ulike operasjonstypene. Herunder er det ønskelig å undersøke om det finnes effektivitetstrender i operasjonene innen de ulike operasjonstypene. Testene ble gjort for to ulike operasjonstyper, henholdsvis a-skopi på skulderleddet og hofteproteser, for å utforske hvordan UNN kan planlegge et eventuelte forbedringsarbeid. Det anses ikke som hensiktsmessig å teste for forskjeller på tvers av de ulike operasjonstypene, da disse vil ha ulike forutsetninger.

I første omgang ble det forsøkt å påvise signifikant forskjell i effektiviteten på vinterhalvåret og sommerhalvåret for operasjonstypen a-skopi skulder. Banker-testen viste ingen indikasjon på dette, og en har derfor ikke grunnlag til å påstå at effektivitetsforskjellene kan forklares med annet enn tilfeldigheter. Operasjonene av typen a-skopi skulder måtte således anses å være like effektive i disse to halvårene. Det ble likevel funnet en relativt stor forskjell i gjennomsnittlig effektivitet for disse gruppene, hvor hele 8 prosent skilte vinter- og

sommerhalvåret. Det er til gjengjeld testet på store grupper i denne bankertesten, og hele datasettet for a-skopi skulder var inkludert i halvårsgruppene. Det undersøktes derfor videre om en kan påvise mer signifikante forskjeller dersom en begrenser gruppene. I neste omgang ble det derfor gjennomført tester på grupper inndelt for færre måneder, herunder vinter- og sommermånedene.

Det ble gjennomført bankertest for både operasjoner av typen a-skopi skulder og for hofteproteseoperasjoner i hver sin omgang. Dette ble først gjort for effektivitetsverdiene fra analysen av beste-praksis. For hofteproteseoperasjonene konstateres det gjennom testen en signifikant forskjell mellom vinter- og sommermånedene. Hofteoperasjonene er herunder signifikant mindre effektive i vintermånedene enn i sommermånedene. Dette tilsier at for denne operasjonstypen er det bedre ressursutnyttelse i sommermånedene enn i vintermånedene. En kan på bakgrunn av bankertesten også konstatere med 95 prosent sannsynlighet at effektivitetsforskjellene er reelle, og følgelig er det lite sannsynlig at disse skyldes tilfeldigheter.

For a-skopi skulder konstateres det også gjennom bankertestene at det er signifikante forskjeller i effektivitet mellom de respektive sesongene. Det måles betydelig høyere effektivitetsnivåer for disse operasjonene på sommeren enn på vinteren. Det er følgelig også sannsynlig med 95 prosent sikkerhet at forskjellene er virkelige og at de ikke styres av tilfeldigheter. Videre vil en også kunne anta at det vil være mest hensiktsmessig for sykehuset å fokusere på å heve effektivitetsnivået til disse operasjonstypene for vintermånedene. Årsakene til effektivitetsforskjellene har en ikke grunnlag for å diskutere på bakgrunn av bankertestene da det vil kreves ytterligere analyser for kartlegging av mer bakenforliggende årsaker. Det kan likevel dannes et bilde av hvilke operasjoner som kan prioriteres til nærmere undersøkelser, slik at en kan finne avgjørende faktorer for effektivitetsnivået til de respektive operasjonene.

For ineffektivitetsgjennomsnittene målt med verste praksis sees det også tendenser til at det er forskjell mellom vinter- og sommersesongen. Gjennomsnittene for vinter- og sommermånedene differerer noe mindre enn de gjorde i målingene med beste praksis. Siden effektiviteten måles gjennom forskjellig modellorientering for beste- og verste praksis er det vanskelig å uttale seg om betydningen av dette. Det testes derfor for signifikans i effektivitetsforskjellene til sesongene fra målene gjort med modellen for verste praksis. For

hofteproteseoperasjonene, med forutsetning om halvnormal fordeling, kan en ikke påvise en signifikant forskjell med bankertesten. Det kan en derimot under eksponentialfordeling. Det påvises også signifikante forskjeller under begge fordelingene for a-skopi skulder. Det bekreftes følgelig at operasjonsaktiviteten er betydelig mer ineffektiv om sommeren, og at operasjonene gjennomført på vinteren er bedre på å unngå dårlige outputs.

Sykehuset vil på bakgrunn av disse funnene kunne fokusere på å heve ineffektivitetsnivået i sommermånedene for å redusere det totale ineffektivitetsnivået. Gjennom å få bekreftet et forbedringspotensial for ineffektivitet i sommermånedene, vil en kunne gjøre sesongfokusering på effektivitetsforbedring. Siden det konstateres signifikante forhold i bankertestene, vil en også ha bedre grunnlag for å avgjøre hvilke operasjoner det er verdt å undersøke nærmere i kartleggingsarbeidet. Følgelig vil UNN kunne ha et sikrere grunnlag for forbedringsarbeidet.

7.6 Layering-teknikken

En annen metode for å danne bedre grunnlag til et eventuelt forbedringsarbeid er gjennom å anvende layering-teknikken. Metoden benyttes for å kartlegge grupper av effektive og ineffektive observasjoner, som igjen kan undersøkes for likhetstrekk i anvendelse av de ulike variablene. I analysen ble det kartlagt fire effektivitetslag for de mest effektive- og de mest ineffektive observasjonene. For hvert effektivitetslag undersøkes observasjonene for likhetstrekk i variablene.

I det første laget for beste praksis identifiseres 14 operasjoner, hvor samtlige har lave KPP-verdier og kort forberedelsestid. I lag nummer to, bestående av 13 operasjoner, kan de samme tendensene sees. Dette kan tyde på at lave kostnader og kort forberedelsestid er avgjørende faktorer for å være i gruppen med de mest effektive operasjonene. De var lite fellestrekk for de øvrige variablene i de to første lagene, og disse kan kanskje derfor tolkes å være mindre forklarende for de høyeste effektivitetsnivåene. Effektivitetsfront nummer tre består av hele 20 observasjoner. Her kan det også sees et skifte i variablene til frontobservasjonene. Der operasjonene i de to ytterste effektivitetslagene hadde lav KPP og lav forberedelsestid, har disse observasjonene hovedsakelig en høy verdi for KPP og forberedelsestider som er både korte og lange. Det er derfor få fellestrekk som kan konstateres mellom dette laget og de to første. Det kan være et mulig funn at KPP og forberedelsestid er mest avgjørende for

effektivitet. Det faktum at observasjonene på den tredje effektivitetsfronten har høy kostnad per pasient og varierende forberedelsestider, kan videre tilsi at det også er mulig for observasjoner å bli forholdsvis effektive til tross for høye KPP-verdier eller lengre forberedelsestid til operasjonene.

Teknikken anvendes også på modellen for verste praksis. Den første ineffektivitetsfronten ble dannet av 6 operasjoner, hvor samtlige hadde betydelige forsinkelser i oppstartstidspunktet for operasjonen. I ineffektivitetsfront nummer to kartlegges det 8 operasjoner, hvilket også har høy grad av forsinkelse i operasjonsstarten. Ytterligere finner en også at samtlige av disse har høye KPP-verdier. Da forsinkelse i oppstartstidspunkt for operasjonene går igjen i alle observasjonene i disse to lagene, kan dette også tenkes å være den mest avgjørende variabelen for ineffektivitet. Høy KPP-verdi kan også være betydningsfullt, selv om dette ikke ble gjeldende for alle operasjonene i den første ineffektivitetsfronten. I den tredje ineffektivitetsfronten preges operasjonene av lave verdier på variabelen for KPP, men store forsinkelser i oppstartstiden. I den fjerde ineffektivitetsfronten finner en også en tydelig trend på høye KPP-verdier, men bare noe forsinkelse i operasjonsstart og stuetid. Følgelig underbygges den tydelige tendensen om at forsinkelse i oppstartstidspunkt er avgjørende for høy ineffektivitetsscore. Det kan også sees tendenser til at høy KPP-verdi medfører høyere målt til ineffektivitet, men det er også DMUer med lavere KPP-verdier som måles å være ineffektive.

Gjennom kartlegging av fellestrekk blant både de effektive og de ineffektive observasjonene, vil en få en stødigere fundament i forbedringsarbeidet. Basert på funnene i denne analysen vil en minimering av kostnader og forsinkelser i operasjonsstart kunne gi større økning av effektivitetsnivået, enn en ville fått ved å fokusere på de andre variablene. Da høye verdier for KPP- og forsinkelse i oppstartstid var en gjennomgående trend hos de ineffektive.

Observasjonene i de ytterste effektivitetslagene for beste praksis hadde tendenser til lav verdi for kostnad per pasient, samt kort forberedelsestid. Dette tyder på at disse variablene er avgjørende for at operasjonene skal være blant de mest effektive.

Som nevnt tidligere er nok ikke optimal effektivitet et mål som verken er realistisk eller hensiktsmessig for alle operasjonene. Denne typen helsetjenester vil kanskje ikke kunne ventes å ha optimale effektivitetsnivåer, da en først og fremst må ivareta pasientene og påse at operasjonene gjennomføres med hensyn til kvalitet og sikkerhet. At KPP-verdien er lav blant

de effektive operasjonene, kan se ut til å bekreftes av den høye KPP-verdien til operasjonene som måles til å være mest ineffektive. Dette kan være en grei forklaringsfaktor å undersøke nærmere.

Om tendensene som identifiseres i denne layering-analysen skyldes tilfeldigheter eller om funnene representerer faktiske forhold, er det ikke grunnlag for å konstatere. Selv om noen av disse trendene kommer klart frem i de ulike lagene, er antallet observasjoner som finnes på fronten forholdsvis lite. Det vil følgelig være vanskelig å uttale seg om graden av gyldighet i disse funnene. Effektivitetsmålene er også relative og gjelder i hovedsak den enkelte operasjonen. Det kan likevel være kunnskapsberikende for UNN å undersøke hvordan forholdene var for operasjonene som kartlegges i de ulike lagene, og utforske i hvilken grad en kan påvirke disse faktorene. Gjennom analysen dannes relativt klare tendenser i variablene til de ulike operasjonene, hvilket kan være verdifullt for planleggingen av et eventuelt forbedringsarbeid. Dersom disse trendene også kan påvises gyldig, vil denne typen analyser kunne anses som et velfungerende verktøy også for kartlegging av bakenforliggende årsaker til de målte effektivitetsnivåene.

Layeringteknikken er videre også bekreftende for funnene som gjøres i modellene for beste og verste praksis. Da observasjonene som befinner seg på frontene i modellen for beste praksis, er blant de som får dårligst verdier for ineffektivitet i modellen for verste praksis. Dette underbygger antakelsen om at modellene er velfungerende og at effektivitetsmålene er relativt sikre.

8 Konklusjon

I det avsluttende kapitlet vil det oppsummeres og trekkes konklusjoner på bakgrunn av funn og diskusjonen i kapittel 7. Analysens formål har vært å vise DEA-analyse som anvendelig verktøy til kartlegging av effektivitetsnivå i ortopediske operasjoner ved UNN, samt å avdekke mulige anvendelser for sykehuset i arbeidet med å heve effektivitetsnivået. Gjennom å benytte én modell for beste praksis og én for verste praksis, kartlegges både effektivitet- og ineffektivitetsnivået i operasjonene. Metodene kombineres også gjennom layering-teknikken, for å forsøke å kartlegge trender i effektivitetsnivåene og danne et grunnlag til forbedringsarbeid. På bakgrunn av at DEA er en ikke-parametrisk metode, anvendes i tillegg bankertester for å teste signifikansen i resultatene gjennom gruppetester.

8.1 Oppsummering og konklusjon

Norske sykehus mener midlene som bevilges til sektoren ikke er tilstrekkelig for å møte det stadig økende behovet for helsetjenester. Staten på sin side mener sykehusene får mer penger enn noen gang, og hevder sykehusene selv må ta ansvar gjennom å bli flinkere til å utnytte ressursene. Helsetjenester av god kvalitet er elementært i velferdssamfunnet, og løsningen kan ikke ansees å være nedbemanning eller store kostnadskutt som går på bekostning av kvaliteten på tjenestene. Siden en betydelig del av sykehusbudsjettene bindes opp i operasjonsaktiviteten, er det ønskelig å kartlegge forbedringspotensialet i operasjonene.

I utprøvingen av DEA-analyser med modellen for beste praksis ble det anvendt to variabler for tidsforbruk og kostnad per pasient som innsatsfaktorer. Tidsvariablene gir gode indikasjoner på varigheten av operasjonene, og variasjonen i tidsbruken til de ulike operasjonene. Kostnad per pasient er en et nyutviklet kostnadsestimat ved UNN, som viser seg å svært velfungerende for analyser på pasientnivå. Der tidligere kostnad- og effektivitetsstudier har benyttet deler av driftskostnadene og lønn til personell som variabel for medgåtte kostnader, har en nå tilgang på et mer presist og nøyaktig mål på kostnadene knyttet til den spesifikke pasienten. Når DRG-vektene, hvilket er benyttet som output, også er skikket for å inkludere kompleksitet og behandlingskomplikasjoner ser modellen ut til å være valid og funksjonell til målingsformålet. Dette til tross for at DRG-vektene gjelder for hele behandlingsforløpet og at faktorer som undervisning og praksisleger ikke kan tas hensyn til. Da DRG-vekter også kan bekreftes som gode mål på produksjonen gjennom annen litteratur, vurderes også effektivitetsmålene å være reliable i denne modellen for beste praksis.

Variablene som benyttes i analysen inkluderer ikke kvalitetslementer som helsegevinst eller pasientsikkerhet. Dette er elementer som analysen derfor ikke har dekning for å svare på. På samme tid er det også tvilsomt at målet for operasjonsaktiviteten er og eller bør være fullkommen effektivitet. På bakgrunn av blant annet kompleksiteten i operasjoner og ivaretagelse av pasientsikkerhet, er det nok verken ønskelig eller hensiktsmessig for sykehus å tilstrebe en fullt effektiv operasjonsaktivitet.

I hvilken grad ineffektiviteten er mulig å forbedre uten at dette går på bekostning av pasientsikkerhet og kvaliteten i tjenestene, vil det kreves mer innsikt og kompetanse i fagområdet for å uttale seg om. Formålet med oppgaven var å undersøke om DEA-analyser er velfungerende til kartlegging av det rene effektivitetsnivået i operasjonene, samt i hvilken

grad det kan bidra til å kartlegge hvor eventuelle effektiviseringstiltak kan iverksettes. Analysen svarer derfor ikke på hvordan og i hvilken grad utbedringer skal gjennomføres. Herunder måler modellen for beste praksis det en har som hensikt å måle i denne oppgaven, effektivitetsnivået til de ulike ortopediske operasjonene.

Effektivitetsnivået måles relativt lavt for ortopediske operasjoner. Sammenholdt med det faktum at kvalitetsaspektet er utfordrende å ivareta på en tilfredsstillende måte i DEA-analysene, anmodes modellen å være anvendelig men noe mindre hensiktsmessig for operasjoner. Det forsøkes derfor å utforske ineffektiviteten i operasjonene. På bakgrunn av at disse ”verstingene” kanskje i større grad vil kunne forbedre effektivitetsnivået, uten at effektiviseringen går på bekostning av kvalitet og pasientsikkerhet, forsøkes ineffektiviteten i operasjonene kartlagt.

Videre er derfor modellen for verste praksis benyttet. I denne modellen måles det hvor ineffektiv operasjonene er, gjennom å snu på modellen fra beste praksis og anvende ”negative outputs”. KPP-verdi og DRG-vekt er også benyttet som variabler i denne modellen, da de også her anses å være hensiktsmessig til å måle henholdsvis kostnader og produksjonen for operasjonene. Det benyttes derimot variabler for tidsforbruk som anses å ha negativ påvirkning på effektiviteten. Herunder er forsinkelse i operasjonsstart og forsinkelse i stuetiden anvendt som dårlige outputs i modellen for verste praksis. Det måles da på hvor mye forsinkelser og kostnader de mest ineffektive operasjonene har hatt. Dette for å kartlegge operasjonene som er de verste av de verste, og har høy grad av ineffektivitet. Denne modellen anses videre også som skikket til måling av ren teknisk ineffektivitet, men også her er det et viktig poeng at kvalitet i behandlingen og helsegevinsten som produseres ikke reflekteres direkte i modellen. Det måles således på ren ineffektivitet i produksjonen, og ikke på resultatet eller kvaliteten til operasjonene. Det viser seg gjennom analysene at de ortopediske operasjonene har høye nivåer for ineffektivitet, altså lav effektivitet. Gjennom å kartlegge de mest ineffektive operasjonene, som er frontlinjeeenheter, identifiseres læremestre for hvordan operasjonene kan unngå å være ineffektiv. Videre også det mulige forbedringspotensialet.

Norske sykehus er som nevnt tidligere pålagt effektiviseringskrav av myndighetene, og tanken er at de vil kunne benytte en kombinasjon av disse effektivitetsanalysene til kartlegging av forskjeller og likheter mellom ulike operasjonstyper og -grupper. Ved å benytte analysene til å danne et bilde av hvor forbedringspotensialet befinner seg, vil de

kunne ha et grunnlag for forbedringsarbeidet. Kombinasjonen av modellen for beste- og verste praksis som anvendes i layering-teknikken, vises anvendelig til kartlegging av lagvise inndelinger av de effektive- og ineffektive operasjoner. Når en setter sammen resultatene på denne måten vil en kunne benytte kunnskapen om de beste og de verste yterne blant operasjonene til å utforske hvor forbedringspotensialet befinner seg, samt hva som kan endres på for å heve effektivitetsnivåene. Blant de ortopediske operasjonene tenderer dette å være kostnad per pasient, forberedelsestid og forsinkelse i operasjonsstarten.

Det er også forsøkt å påvise effektivitetsforskjeller innen de ulike operasjonstypene, slik at forbedringsarbeidet vil kunne spisses ytterligere. Det ble kartlagt en klar effektivitetsforskjell mellom på vinter- og sommermånedene, og siden DEA er en ikke-parametrisk metode ble bankertester benyttet for å undersøke signifikansen i disse forskjellene. Det ble testet for forskjeller mellom sesongene for både operasjoner av typen a-skopi skulder og hofteproteseoperasjoner. Gjennom bankertestene på effektivitetsmålene fra beste praksis bekreftes det signifikant forskjell i effektiviteten til begge operasjonstypene. Disse har begge en mer hensiktsmessig ressursutnyttelse om sommeren enn på vinteren, hvilket taler for at sykehuset har størst forbedringspotensial for operasjonene som gjennomføres på vinteren. Her vil det være hensiktsmessig å ta lærdom fra operasjonene som er effektive i sommermånedene, da disse har en mer hensiktsmessig ressursutnyttelse.

Det ble også kartlagt effektivitetsforskjeller mellom vinter- og sommermånedene i effektivitetsmålene fra modellen for verste praksis. Gjennomsnittet tydet på større grad av ineffektivitet i sommermånedene enn for vintermånedene, hvilket også ble bekreftet som signifikante forskjeller gjennom bankertester. En kan følgelig med overveiende sannsynlighet si at operasjonene på vinteren er bedre på å unngå ”dårlige” outputs enn operasjonene i sommermånedene. Herunder er det størst grad av ineffektivitet i sommermånedene. UNN kan benytte denne informasjonen i forbedringsarbeidet gjennom å ta lærdom av disse operasjonene. Det vil kanskje være aktuelt å undersøke hvilke forhold som er avgjørende for å unngå dårlige outputs for operasjonene. På denne måten vil sykehuset også kunne ha bistand i undersøkelse av mulige tiltak for å gjøre operasjonene mindre ineffektive.

For sykehuset vil det følgelig være hensiktsmessig å anvende DEA-analyser av både beste- og verste praksis for kartlegging av effektivitetsnivået i operasjonsaktiviteten. Disse modellene vil være et stødig hjelpemiddel til å kartlegging. Særlig på bakgrunn av de nye

kostnadmålene de har tilgjengelig i KPP-verdiene. Analysene måler på ren teknisk effektivitet, og tar ikke høyde for kvalitetsaspekter eller pasientsikkerhet i operasjonene. I hvilken grad det kartlagte effektivitetspotensialet er realistisk og hensiktsmessig å forbedre kan ikke fastslås på bakgrunn av analysene. Videre er det derfor også lagt vekt på ineffektivitet, da det antas at ineffektive operasjoner har større reelt forbedringspotensial enn de operasjonene som vurderes til å ha bedre effektivitetsnivåer. Hensikten er således at sykehuset kan sikte etter og ta lærdom av de beste og mest effektive operasjonene i forbedringsarbeidet, men legge innsatsen i å forbedre de verste.

Dette kan blant annet gjøres ved å anvende analysene til kategorisering og gruppering av operasjonstypene, slik at en lettere vil kunne avdekke tendenser i ressursanvendelsene til de ulike operasjonene. Da DEA måler relative effektivitetsnivåer, vil disse i første omgang være gjeldende for den undersøkte operasjonen. Dette understreker det faktum at en ikke bør stole blindt på effektivitetsnivåene og trendene som kartlegges, men heller benytter metoden til et hjelpemiddel i kartleggingen.

DEA-analyser gjennom både beste- og verste praksis vurderes derfor å være anvendelige for måling av effektivitet i ortopediske operasjoner ved UNN. Likevel kan verste praksis tenkes å være mer hensiktsmessig på bakgrunn av lave effektivitetsmålinger og for ivaretagelse av kvalitetsaspektet. De kan benytte metoder for kategorisering og gruppering av effektive- og ineffektive operasjoner, for å kartlegge forholdene som er avgjørende for de ulike effektivitetsnivåene. På bakgrunn av at analysene vurderes å være valide og troverdige, kan det også se ut til at det er rom for å forbedre effektiviteten i operasjonsaktiviteten ytterligere. Herunder vil det kunne være mulig å imøtekomme i hvert fall deler av effektiviseringskravet fra regjeringen. Analysene er velfungerende for å kartlegge i hvilken grad det eksisterer ineffektivitet, og hvor dette befinner seg. Til tross for at en ikke får besvart hvilke tiltak som bør gjøres, eller har grunnlag for å gjøre slutninger vedrørende hva som er reelt og hensiktsmessig forbedringspotensial. Det kan likevel formodes at det er mulig å heve effektivitetsnivået til operasjonene med de verste effektivitetsnivåene. Herunder at en sikter etter de beste operasjonene, og forbedrer de verste – hope for the best, prepare for the worst.

8.2 Svakheter ved forskningen

Kvalitetsaspektet ved behandlingene er utfordrende å ta i betraktning og det argumenteres for at kvaliteten ofte glemmes når en benytter pasientbehandlinger som mål på aktivitet. Kvalitet og pasientsikkerhet er viktige aspekter ved operasjoner og analysemodellen bør nok ideelt kunne ta hensyn til dette. En vil delvis ha tatt høyde for sykdomsbildet og kompleksitet i analysen, gjennom å innlemme DRG-vektene, men en vil likevel ikke ha indikasjon på hvor vellykket helsebehandlingene er og videre også helsegevinst eller reduksjon i sykdom. For øvrig anerkjennes det at DRG-vekten er gjeldene for hele behandlingsforløpet til pasientene, og at en dermed hensyntar mer enn bare operasjonsaktiviteten ved å benytte disse som variabler. Likevel anses vektene å være relativt rettvise for forholdet mellom operasjonene, og det er derfor ikke ansett som en avgjørende svakhet for troverdigheten til analysen.

En ideell DEA-analyse ville nok kunne tatt hensyn til disse kvalitetsfaktorene. Likevel er målet med analysen er i dette tilfelle å kunne estimere ren effektivitet, og kvalitetsaspektet har derfor blitt litt tilsidesatt i modellen. Det er likevel viktig å vurdere pasientenes interesser og påse at forbedring av effektivitet ikke går på bekostning av sikkerhet og kvalitet i operasjonene.

Effektiviteten måles også relativt i analysene. Hver DMU sees på som en transformeringsboks, som en beregner det relative effektivitetsnivået til. Informasjonen og målene som finnes gjennom analysene kan derfor ikke brukes direkte i strategi eller forbedringsarbeid. Metodene benyttes heller til å finne ut hvilke observasjoner som bør undersøkes nærmere.

8.3 Videre forskning

Studien viser DEA-som en velfungerende metode til effektivitetsanalyser på operasjoner. En er avhengig av velfungerende variabler som reflekterer ressursforbruk og produksjon på en god måte i analysene. Dette kan synes å være relativt godt sikret i denne oppgaven. Likevel kunne modellen vært utbedret og forskningen styrket ved å finne passende variabler for å hensynta undervisningsaktiviteten i forbindelse med operasjonene. Dette, samt DRG-veker som kun gjelder operasjonene vil kunne danne en sikrere og mer pålitelig analyse. Følgelig foreslås det å utbedre modellene så langt dette lar seg gjøre på et senere tidspunkt.

Noe som kunne løftet studien er en undersøkelse av bakenforliggende årsaker til sesongforskjellene som kartlegges. Det kan for eksempel tenkes at årsaken til mer ineffektivitet i sommermånedene er bemannings- eller kapasitetsforskjeller på grunn av ferieavvikling. Vær og vind vil kanskje ha innvirkning på effektivitetsnivåene. Et såkalt nullføre kan tenkes å medføre flere og mer utfordrende hofteskader i vintermånedene, og det kan kanskje gjøres koblinger til at bestemor klatrer etter kakeboksen når julen nærmer seg. Bakenforliggende årsaker til de ulike effektivitetsnivåene kan eksempelvis undersøkes i en to-stepsanalyse. Herunder vil en kunne kartlegge i hvilken grad leger i spesialisering eller været er forklaringsfaktorer for de ulike effektivitetsmålingene.

Et annet element som ikke er utforsket i denne analysen er om effektivitetsforskjellene kan skyldes behandlingsstedet for operasjonene. Da ortopediske operasjonsstuer er fordelt på tre ulike lokasjoner. Herunder kunne det eksempelvis vært gjort en regresjonsanalyse for å forsøke å påvise effektivitetsforskjeller mellom de ulike operasjonsstedene Tromsø, Harstad og Narvik. Det kan tenkes at stedet for operasjonene er en forklarende faktor som har signifikant innvirkning på effektivitetsscoren, herunder kunne en undersøkt hva en har å lære av de ulike operasjonslokasjonene.

Videre tar denne analysen utgangspunkt i ortopediske operasjoner, og det kunne vært en idé å undersøke effektiviteten i andre operasjonstyper for å teste anvendeligheten til DEA-analysene ytterligere. For øvrig vil forskningen kunne løftes ved at en undersøker effektivitetsnivåenes utvikling over tid. Utarbeidelsen av kostnadsestimatet KPP var ferdigstilt i januar 2017, men da bare for operasjoner gjennomført i 2015. Det er disse dataene denne studien har belaget seg på til kostnadsvariabler, men etter hvert som det utvikles KPP-verdier også for andre år vil studien kunne videreutvikles til å gjelde flere tidsperioder.

Det kunne også vært interessant å undersøke i hvilken grad slakk er tilstede for de ineffektive operasjonene og om operasjonene har mulighet til å utbedre slakk i inputene for paretooptimalitet og herunder også en utbedret effektivitet. Videre kunne også skala vært undersøkt nærmere. Det er funnet lite drøftelser rundt hva skala utgjør i operasjonsammenheng, og en studie av dette vil derfor være et viktig bidrag for effektivitetsanalyser på dette området.

Referanseliste

Andrews, D. F., & Pregibon D. (1978). Finding the Outliers that Matter. *Journal of the Royal Statistical Society*, 40(1). 85-93. Hentet fra: <https://www.jstor.org/>

Avkiran, N. K. (1999). An application reference for data envelopment analysis in branch banking: helping the novice researcher. *International Journal of Bank Marketing*, 17(5), 206-220. Hentet fra: <http://www.emeraldinsight.com/>

Banker, R. D. (1993). Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation. *Management Science*. 39(10). 1265-1273. Hentet fra: <http://www.dynamed.com/>

Banker, R. D., Chang, H. (2006). Super-efficiency procedure for outlier identification, not for ranking efficient units. *European Journal of Operational Reserach*, 175(2), 1311-1320. Hentet fra: <http://www.sciencedirect.com/>

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, 30(9), 1078-1092. Hentet fra: <http://www.jstor.org/>

Banker, R. D., & Natarajan, R. (2011) Statistical tests based on DEA efficiency scores. I W. W. Cooper et. al (Red), *Handbook on data envelopment analysis*. 273-295. New York: Springer US

Bogetoft, P., Otto, L. (2011). *Benchmarking with DEA, SFA and R*. New York: Springer US

Charnes, A., Cooper, W. W., Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429-444. Hentet fra: <http://www.sciencedirect.com/>

Coelli, T. J., Rao, D. S. P., O'Donnell C. J., Battese, G. E. (2005). *An introduction to Efficiency and Productivity Analysis* (2.utg). New York: Springer US

Dons, M. F. & Skorstad, M. (2012). *Utnyttelsen av operasjonskapasiteten ved UNN: en effektivitetsanalyse av bruken av operasjonsstuen på UNN for 2010 og 2011*.

(Masteroppgave) Hentet fra: <http://munin.uit.no>

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of operational research*, 132(2), 245-259. Hentet fra: <http://www.sciencedirect.com/>

Farrell, M. J. (1957). The measurement of productive efficiency. *The journal of the royal statistical society series A-general* 120(3), 253-290. Hentet fra: <https://www.jstor.org/>

Golany, B. & Roll, Y. (1989). An Application Procedure for DEA. *Omega*, 17(3), 237–250. Hentet fra: <http://www.sciencedirect.com/>

Hagen, T. P., Iversen, T., & Magnussen, J. (2001). *Sykehusenes effektivitetsutvikling 1992-1999: Hvilke effekter ga innsatsstyrt finansiering?* (Helseøkonomisk forskningsprogram Rapport 2001:5). Oslo: Universitetet i Oslo

Helsedirektoratet. (2017a). Aktivitet og effektivitet i somatiske sykehus: DRG- og pasientindeks. Hentet 13.mars 2017 fra https://statistikk.helsedirektoratet.no/HtmlViewer.ashx?Dd_ContentId=aa71f53e-8a15-427c-b945-e7697d5285fd&Dd_ContentType=Dashboard

Helsedirektoratet (2014). Innsatsstyrt finansiering 2015. Hentet 28.mars 2017 fra <https://helsedirektoratet.no/Lists/Publikasjoner/Attachments/775/Innsatsstyrt-finansiering-2015-regelverk-IS-2230.pdf>

Helsedirektoratet. (2017b). Innsatsstyrt finansiering (ISF) og DRG-systemet. Hentet 6.mars 2017 fra <https://helsedirektoratet.no/finansieringsordninger/innsatsstyrt-finansiering-isf-og-drg-systemet>

Helsedirektoratet. (2017c). Kostnadsvekter og kostnad per pasient (KPP). Hentet 3.mars 2017 fra <https://helsedirektoratet.no/finansieringsordninger/innsatsstyrt-finansiering-isf-og-drg-systemet/kostnadsvekter-og-kostnad-per-pasient-kpp>

Helsedirektoratet. (2017d). Nasjonal spesifisering for KPP-modellering. Hentet 14.mars 2017 fra <https://helsedirektoratet.no/retningslinjer/nasjonal-spesifisering-for-kpp-modellering-2012-begreper-og-metoder>

Helse Nord. (2017). Om oss. Hentet 2.mars 2017 fra <https://helse-nord.no/om-oss>

Internrapport fra UNN (2008). *Optimal ressursutnyttelse av operasjonskapasiteten i UNN*.

Kittelsen, S.A.C., Førsum, F.R. (2001). Empiriske forskningsresultater om effektivitet i offentlig tjenesteproduksjon. *Økonomisk forum nr 6*, 22-29. Hentet fra: <http://www.frisch.uio.no/>

Kittelsen, S. A., Magnussen, J., & Anthun, K. S. (2007). *Sykehusproduktiviteten etter statlig overtakelse: En nordisk komparativ analyse* (Helseøkonomisk forskningsprogram Rapport 2007:1). Oslo: Universitetet i Oslo

Liu, F. F., Chen, C. (2005). Identifying Bank Failures with Two-stage Data Envelopment Analysis in the Worst-case Scenario: The Case of Taiwan Banks. *WSEAS Transactions on Information Science and Applications*, Vol.9(3), 93-102 Hentet fra: <http://www.wseas.org/>

Liu, F. F., Chen, C. (2008). The worst-practice DEA model with slack-based measurement. *Computers and Industrial Engineering* 57(2) 496-505. Hentet fra: <http://www.sciencedirect.com/>

Liu, J. S., Lu, L. Y. Y., Lu, W-M., Lin, B. J. Y. (2013). A survey of DEA applications. *Omega* 41(5), 893-902. Hentet fra: <http://www.sciencedirect.com/>

Martinussen, J. & Reinholdtsen, I. (2013). *Kvalitet og effektivitet i norsk helsevesen. En ikke-parametrisk og statistisk tottrinnsanalyse av somatisk spesialisthelsetjeneste*. (Masteroppgave) Hentet fra: <http://munin.uit.no>

Norsk helseinformatikk. (2013). Hallux valgus. Hentet 4.mai fra <https://nhi.no/sykdommer/muskelskjelett/legg-ankel-fot/hallux-valgus-operasjon/>

Norsk helseinformatikk. (2016). Ortopedisk utredning. Hentet 22.februar 2017 fra

<https://nhi.no/sykdommer/muskelskjelett/undersokelser/ortopedisk-utredning/>

Paradi, J. C., Asmild, M., Simak, P. C. (2004). Using DEA and Worst Practice DEA in Credit Risk Evaluation. *Journal of Productivity Analysis* 21, 153-165. Hentet fra:

<https://link.springer.com/>

Puig-Junoy, J. (1998). Technical Efficiency in the Clinical Management of Critically Ill Patients. *Health Economics*. 7(3), 263-277. Hentet fra: <http://onlinelibrary.wiley.com/>

Regjeringen. (2013). Ansvarsområder for Helse- og omsorgsdepartementet. Hentet 7.mars 2017 fra <https://www.regjeringen.no/no/dep/hod/dep/ansvarsomraader/id446/>

Skiphamn, S. S. (2016, 5. desember). Legeforeningen om budsjettet: – Et gedigent løftebrudd av Høie. Hentet fra: <http://www.vg.no/>

Skiphamn, S. S. (2017, 7. februar). Norske sykehus klarer ikke holde budsjettene. Hentet fra: <http://www.vg.no/>

Statistisk sentralbyrå. (2017). Helseregnskap, 2016. Hentet 20. mars 2017 fra

<https://www.ssb.no/nasjonalregnskap-og-konjunkturer/statistikker/helsesat>

Store norske leksikon. (2009). Poliklinikk. Hentet 6.mars 2017 fra

<https://sml.snl.no/poliklinikk>

UNN. (2016). Nøkkeltall 2012 - 2015. Hentet 6.mars 2017 fra <https://unn.no/Documents/Om-oss/N%C3%B8kkeltall%202012-2015.pdf>

UNN. (2017). Om oss. Hentet 6.mars 2017 fra <https://unn.no/om-oss>

Wilson P. W. (1993). Detecting Outliers in Deterministic Nonparametric Frontier Models with Multiple Outputs. *Journal of Business & Economic Statistics*, 11(3). 319-323. Hentet fra:

<https://www.jstor.org/>

Vedlegg

r-script

```
#Inputminimerende DEA – beste praksis

#Definere datasett

data<-hofteprotese

#Definere variabler
xvar <- c(2,3,4)
x1 <- data[,xvar]
yvar <-c(5)
y1 <- data[,yvar]
x=as.matrix(x1)
y=as.matrix(y1)

vrs<-dea(x,y,RTS="vrs", ORIENTATION="in")

#Varierende skalautbytte
M1vrs<-dea(x,y,RTS="vrs",ORIENTATION="in")
M1svrs<-sdea(x,y,RTS="vrs",ORIENTATION="in")
Evr<-eff(M1vrs)
Esvrs<-eff(M1svrs)

#Konstant skalautbytte
M1crs<-dea(x,y,RTS="crs",ORIENTATION="in")
M1scrs<-sdea(x,y,RTS="crs",ORIENTATION="in")
Ecrs<-eff(M1crs)
Escrs<-eff(M1scrs)

out<-(cbind(Evr, Ecrs, Escrs))
print(out)
write.csv2(out,file="hofteprotese.csv")
```

```

#Outputmaksimerende modell

#Definere datasett
data<-askopi_skulder

#Definere variabler
xvar <- c(3)
x1 <- data[,xvar]
yvar <-c(5,6,7)
y1 <- data[,yvar]
x=as.matrix(x1)
y=as.matrix(y1)

vrs<-dea(x,y,RTS="vrs", ORIENTATION="out")

#Varierende skalautbytte
M1vrs<-dea(x,y,RTS="vrs",ORIENTATION="out")
M1svrs<-sdea(x,y,RTS="vrs",ORIENTATION="out")
Evrs<-eff(M1vrs)
Esvrs<-eff(M1svrs)

#Konstant skalautbytte
M1crs<-dea(x,y,RTS="crs",ORIENTATION="out")
M1scrs<-sdea(x,y,RTS="crs",ORIENTATION="out")
Ecrs<-eff(M1crs)
Escrs<-eff(M1scrs)

out<-(cbind(Evrs, Ecrs, Escrs))
print(out)
write.csv2(out,file="askopi_skulder.csv")

```

```

#AP og Banker

data<-hallux_valgus

xvar <- c(3)
x1 <- data[,xvar]
yvar <-c(2,6,7)
y1 <- data[,yvar]
x=as.matrix(x1)
y=as.matrix(y1)

#Benchmarking: outlieranalyse AP
tap <- outlier.ap(x,y, NDEL=4)
print(cbind(tap$imat,tap$rmin), na.print="", digit=2)
outlier.ap.plot(tap$ratio)

# for å få skrevet resultatet av outlieranalysen til excelfil:
outl<-cbind(tap$imat,tap$r0)
write.csv2(outl,file="outl_hallux_valgus.csv")

#BANKER test – benchmarking
#test for skala, eller ekstra variabler

xvar <- c(2,3,4)
x1 <- data[,xvar]
yvar <- c(5)
y1 <- data[,yvar]
x1=as.matrix(x1)
y=as.matrix(y1)
E1<-eff(dea(x1,y,RTS="crs",ORIENTATION="in")) #En må være crs den andre vrs
xvar <- c(2,3,4)
x1 <- data[,xvar]
yvar <- c(5)
y1 <- data[,yvar]
x2=as.matrix(x1)
y=as.matrix(y1)
E2<-eff(dea(x2,y,RTS="vrs",ORIENTATION="in"))

```

```
#AP og Banker fortsettelse
```

```
xvar1 <- c(3)  
E1 <- data[,xvar1]  
xvar2 <- c(2)  
E2 <- data[,xvar2]
```

```
#test tex-exponential
```

```
TEX <- sum(E1-1) / sum(E2-1)  
TEX #testverdi  
qf(.95, 2*length(E1), 2*length(E2)) #kritiske verdi  
pf(TEX, 2*length(E1), 2*length(E2))
```

```
#test thn-halvnormal
```

```
THN <- (sum((E1-1)*(E1-1)))/(sum((E2-1)*(E2-1)))  
THN  
qf(.95, length(E1), length(E2))  
pf(THN, length(E1), length(E2))
```

```

#Banker-test for forskjell på grupper

data<-hofteprotese_bp_banker_res

xvar1 <- c(2)
E1 <- data[-c(1:32),xvar1]
E2 <- data[-c(33:44),xvar1]

TEX <- (sum(E1-1)/length(E1)) / (sum(E2-1)/length(E2))
TEX
qf(.025, 2*length(E1), 2*length(E2))
qf(.975, 2*length(E1), 2*length(E2))
pf(TEX, 2*length(E1),2*length(E2))

THN <- sum((E1-1)*(E1-1))/length(E1) / (sum((E2-1)*(E2-1))/length(E2))
THN
qf(.025, length(E1), length(E2))
qf(.975, length(E1), length(E2))
pf(THN, length(E1), length(E2))

#Banker-test for skala

data<-halluxvalgus

xvar1 <- c(3) #crs
E1 <- data[,xvar1]
xvar2 <- c(2) #vrs
E2 <- data[,xvar2]

#test tex-exponential
TEX <- sum(E1-1) / sum(E2-1)
TEX #testverdi
qf(.95, 2*length(E1), 2*length(E2)) #kritiske verdi
pf(TEX, 2*length(E1),2*length(E2))

#test thn-halvnormal
THN <- (sum((E1-1)*(E1-1)))/(sum((E2-1)*(E2-1)))
THN
qf(.95, length(E1), length(E2))
pf(THN, length(E1), length(E2))

```