Methodological development and evaluation of 30- day mortality as quality indicator for Norwegian hospitals

Rapport fra Kunnskapssenteret Nr 4–2005

Nasjonalt kunnskapssenter for helsetjenesten

Om rapporten: Dødelighet 30 dager etter innleggelse som kvalitetsindikator for norske sykehus – metodeutvikling og evaluering: Studien er gjennomført for å vurdere hvilken verdi kvalitetsindikatoren "dødelighet innen 30 dager etter innleggelse" kan ha ved somatiske sykehus. Alle landets sykehus har avgitt data til studien for perioden 1997-2001. Vi har målt dødeligheten for tre tilstander; hjerteinfarkt, hjerneslag og hoftebrudd. Disse tilstandene er valgt fordi de er hyppige, alvorlige og fordi behandling kan påvirke utfallet. **Målet med studien** har ikke vært å vise frem og vurdere resultater fra enkeltsykehus eller å sammenlikne sykehus med hverandre. Målet har vært å utvikle en modell for beregning av "30 dagers dødelighet" som kvalitetsindikator, samt å påpeke mangler som bør rettes for at modellen skal kunne brukes som kvalitetsindikator i fremtiden. **Bakgrunn:** Forskjeller i målt dødelighet mellom sykehus avhenger av flere faktorer, ikke bare kvaliteten på den medisinske behandlingen. Andre viktige faktorer er organisatoriske forhold, administrative rutiner ved sykehuset, diagnostisering, kodepraksis og datakvalitet, rutiner *(fortsetter på baksiden)*

Nasjonalt kunnskapssenter for helsetjenesten
Postboks 7004, St. Olavs plass
N-0130 Oslo
(+47) 23 25 50 00
www.kunnskapssenteret.no
Rapport: ISBN 82-8121-006-0 ISSN 1503-9544

nr 4–2005

Nasjonalt kunnskapssenter for helsetjenesten

(fortsettelsen fra forsiden) for innhenting av data og prosesser som pasienten inngår i før, under og etter sykehusoppholdet. **Hovedfunn:** Å fremskaffe sammenliknbare kvalitetsindikatorer basert på dødelighet er en forsknings- og utviklingsprosess. Studien viser at på mange områder tilfredsstiller 30 dagers dødelighet kravene til slike indikatorer, samtidig gjenstår det usikkerhetsfaktorer. Usikkerheten gjør at vi ikke med sikkerhet kan påvise om et sykehus virkelig avviker fra gjennomsnittet i perioden. Dødelighetstallene varierer mellom sykehusene, mest for hjerneslag og hoftebrudd. Studien forteller ikke hvorfor det er slik. Kunnskapssenteret anbefaler at man studerer nærmere hvordan ulike faktorer påvirker dødelighetstallene slik at vi med større sikkerhet kan slå fast om resultatene er knyttet til medisinsk behandlingskvalitet eller andre forhold. Spesielt bør det gjennomføres valideringsstudier for å avgjøre hvilken betydning sykehusenes diagnose- og kodepraksis har for forskjeller i anslått dødelighet.

Title	Methodological development and evaluation of 30-day mortality as quality indicator for Norwegian hospitals
Institution	Norwegian Knowledge Centre for the Health Services (Nasjonalt kunnskapssenter for helsetjenesten)*
Head of centre	John-Arne Røttingen, MD, PhD, Director
Head of project	Jocelyne Clench-Aas, PhD, Dr-es-Sciences
Principal investigators	Jocelyne Clench-Aas, PhD, Dr-es-Sciences, Dag Hofoss, Dr Phil, Ole Morten Rønning, MD, PhD
Authors	Jocelyne Clench-Aas, Jon Helgeland, Tomislav Dimoski, Pål Gulbrandsen, Dag Hofoss, Olaf Holmboe, Petter Mowinckel and Ole Morten Rønning
Data acquisition system development	Tomislav Dimoski
Technical assistance	Reidun Skårerhøgda, Nina Viksløkken Ødegård
ISBN	82-8121-006-0
ISSN	1503-9544
Report number	4 – 2005
Project number	8-176
Number of pages	198
Financial sources	Norwegian Directorate for Health and Social Welfare (Sosial- og helsedirektoratet)

The Norwegian Knowledge Centre for the Health Services is a governmental centre, with a mission to support improvement of health services in Norway. The centre's mission is achieved through supporting decisions about health services by providing expert information and advice founded on knowledge-based summaries, research and development and teaching and presentation in the field of health services.

* In the text, references are made to HELTEF, being the institute responsible for the study, until it was included in the Norwegian Knowledge Centre (NOKC) for the Health Services January 1, 2004.

The Norwegian Knowledge Centre for the Health Services

Oslo, September 2005

TABLE OF CONTENTS

Rapporten i papirutgave er trykt i svart/hvitt og enkelte figurer vil derfor være vanskelige å tyde. Spesielt gjelder dette figurene 5-2, 6-1—6-5, 6-16 og 6-30—6-32. Rapporten med figurer i farger kan lastes ned fra www.kunnskapssenteret.no

1. NORSK SAMMENDRAG (NORWEGIAN SUMMARY)

1.1 BAKGRUNN

I forsøk på å beskrive kvaliteten på behandlingen i sykehus er det utviklet mange kvalitetsindikatorer. Ingen slik indikator er alene et uttrykk for behandlingskvaliteten i et sykehus. Tanken er at man ved å vurdere samlet flere indikatorer som beskriver relevante sider ved sykehusets virksomhet, kan få et inntrykk av generell kvalitet ved sykehuset.

Dødelighet ved hyppige, alvorlige tilstander er en slik kvalitetsindikator. Tidligere ble dødelighet under oppholdet i sykehuset mye brukt, men etter hvert er det alminnelig akseptert at dette målet ikke er godt, dels fordi et mindretall dør under oppholdet og dels fordi det ble ansett som mulig for sykehus å manipulere resultater ved å skrive ut eller overflytte pasienter for tidlig. Derfor er dødelighet innen 30 dager etter innleggelse (30D) av mange foreslått som et bedre mål (1), men også dette er omdiskutert. Bl.a. er det en utfordring at overdødeligheten knyttet til ulike sykdommer skjer med ulik fordeling i tid etter at sykdommen oppstod, og at et felles måletidspunkt neppe er den beste måten å indikere behandlingskvalitet på for alle sykdommer.

Når dødelighet innen 30 dager etter innleggelse (30D) har vært brukt som kvalitetsindikator, har den som regel vært basert på data fra sykehusenes pasient-administrative systemer (PAS), evt. supplert med korrigerende data fra andre kilder utenfor sykehuset. Diskusjonen omkring verdien av 30D som kvalitetsindikator har særlig dreid seg om hvorvidt data i PAS er gode nok til slik anvendelse. I en større amerikansk medisinsk vitenskapelig vurdering av kvalitetsindikatorer for sykehus ble dødelighet under opphold i sykehus for akutt hjerteinfarkt, hjerneslag og hoftebrudd funnet anbefalelsesverdig, dog som tre blant 45 indikatorer av mer enn 200 som ble vurdert.

I 1999 utgav Stiftelse for helsetjenesteforskning (HELTEF) en rapport over dødelighet i norske sykehus (2). Rapporten ble kritisert for flere forhold; noen av de viktigste innvendingene var:

- for dårlig datakvalitet
- ingen korreksjon for at sykehus behandler pasienter som ikke er like syke når de blir innlagt
- ingen korreksjon for muligheten for at data om pasienter fra samme sykehus kan være innbyrdes avhengige (manglende flernivåanalyse)
- ingen korreksjon for at mange pasienter blir behandlet ved flere sykehus i løpet av samme sykdomsepisode

Sosial- og helsedepartementet mente at disse innvendingene kunne imøtekommes, og ga HELTEF, som senere ble fusjonert inn i Nasjonalt kunnskapssenter for helsetjenesten, i oppdrag å utrede hvorvidt dødelighet innen 30 dager etter innleggelse kan fungere som kvalitetsindikator i fremtiden for akutt hjerteinfarkt, hjerneslag og hoftebrudd.

Her må det understrekes at rapporten ikke har som formål å vise frem og vurdere resultater fra enkeltsykehus, men derimot å vise frem og vurdere den metodiske utviklingen frem til en foreslått modell for beregning av 30D og påpekning av visse mangler som bør rettes for at modellen skal kunne brukes. Videre handler rapporten kun om dødelighet innen 30 dager. Muligheten for at andre tidspunkter enn død innen 30 dager etter sykehusinnleggelse er mer relevante mål for de tre tilstandene, er ikke utredet. Valget av de tre tilstandene til å utrede spørsmålet beror på hyppighet, alvorlighetsgrad og at de representerer dels medisinsk, dels nevrologisk og dels kirurgisk behandling.

1.2 MATERIALE OG METODE

Det ble etablert tre medisinske ekspertgrupper med i alt 25 deltakere; en gruppe for hver tilstand. Medlemmene hadde medisinsk, epidemiologisk eller statistisk kompetanse. Gruppene deltok løpende i diskusjoner om planlegging og gjennomføring av studien.

FS-systemet, et system for innhenting av data fra sykehusenes pasientadministrative databaser (PAS), ble brukt. Systemet ble videreutviklet for å koble individdata fra Statistisk sentralbyrå og Dødsårsaksregisteret med data fra PAS, slik at det var mulig å korrigere for sosioøkonomiske forhold i analysene.

1.2.1 Materiale

Samtlige norske somatiske sykehus har frivillig avgitt data til studien. Disse utgjorde 66 enheter, slik vi valgte å dele dem inn (det skjer stadig organisatoriske endringer som er en utfordring når man skal definere et sykehus). Årene som er lagt til grunn er perioden 1997-2001. I løpet av denne perioden skjedde en omdefinering av funksjonen til flere sykehus. Den internasjonale diagnoseklassifiseringen ble endret fra ICD-9 til ICD-10, og kriteriene for diagnosen hjerteinfarkt ble endret. I tillegg ble det gjort endringer i anbefalt behandling for akutt hjerteinfarkt.

Noen pasienter blir lagt inn flere ganger for samme tilstand i løpet av et kalenderår. I analysen har vi kun inkludert første gangs innleggelse for hver av de tre tilstandene i løpet av hvert kalenderår. Med denne begrensningen ble det registrert i alt 176 387 innleggelser i FS-systemet. Av ulike grunner, viktigst her var 1 166 personer som vi ikke fant data om hos Statistisk sentralbyrå, ble kun 174 527 av innleggelsene nærmere vurdert.

Etter råd fra ekspertgruppen for akutt hjerteinfarkt, ble studien avgrenset til å gjelde første gangs hjerteinfarkt. Innleggelser for hjerneslag mindre enn 28 dager etter en tidligere innleggelse for hjerneslag ble også ekskludert, selv om de var de første i et kalenderår (aktuelt for dagene 1.-27. januar). I alt ble 17 155 innleggelser ekskludert fra analysene av disse tre grunnene. Etter dette gjensto 54 095 innleggelser for akutt første gangs hjerteinfarkt, 53 072 for akutt hjerneslag og 50 205 for hoftebrudd. I tillegg ble det ved nærmere gransking funnet omkring 8 000 innleggelser som ikke tilfredsstilte inklusjonskriteriene.

Dersom en pasient i løpet av en episode er flyttet fra et sykehus til et eller flere andre, ble pasienten registrert som innlagt i alle sykehus i løpet av episoden, dog slik at summen av innleggelsene teller som én innleggelse i analysen.

1.2.2 Data

For best mulig å korrigere for ulikheter i dødelighet mellom sykehusene knyttet til egenskaper ved pasientene, innhentet vi data om diagnoser ved tidligere opphold, tilleggsdiagnoser ved innleggelsen som ble analysert, sosioøkonomiske data om pasientene og beregnet avstand fra bolig til sykehuset som innleggelsen fant sted ved. Disse dataene ble brukt som grunnlag for å lage vikarierende variabler (proxy variables) for pasientens allmenntilstand (patient frailty), alvorlighetsgrad ved innleggelsen (disease severity), sosioøkonomisk status og tid fra oppstått sykdom til innleggelse i sykehus.

Vi ønsket også å se nærmere på betydningen av ulike innleggelsesrutiner, tilgang på prehospitale tjenester, kodingsrutiner, utskrivningspraksis og obduksjonspraksis. Dette ble ikke mulig å gjennomføre pga. utilstrekkelige opplysninger fra et flertall av sykehusene.

1.2.3 Statistisk analyse

Vi har brukt logistisk regresjon med død innen 30 dager etter innleggelse som avhengig variabel.

Mange norske sykehus er små, noe som gjør at det kan observeres en betydelig årlig variasjon i dødelighet uten at man kan slutte at det har skjedd en endring i behandlings-kvaliteten. Vi har derfor valgt en statistisk modell som sier mest om resultatet det siste året i rekken (2001 for vårt materiale), men som tar med informasjonen vi har for de fire foregående årene. Fordelene med dette er ikke bare at vi får sikrere konklusjoner også for mindre sykehus, men at vi kan påvise trender for hvert sykehus, slik at sykehus som skiller seg fra andre sykehus i gunstig eller ugunstig retning , kan identifiseres. Ulempen ved metoden er at vi reduserer variabiliteten i materialet, slik at man risikerer at store avvik ved enkeltsykehus enkelte år ikke vil komme frem. Disse avvikene vil likevel være identifisert under forarbeidene til modellen.

I modellbyggingen begynte vi med en modell som inkluderte alle forklaringsvariablene, inkludert variablene som ivaretar sammenheng mellom årstall og sykehusobservasjon og andre såkalte interaksjonsvariabler. Vi foretok deretter en trinnvis eksklusjon av ikke signifikante variabler, der interaksjonsvariablene ble testet først. De sosiodemografiske variablene ble testet samlet, som en pakke.

Som grense for å hevde at en ulikhet bør overses (indifference limit), har vi valgt inntil 20% økning eller reduksjon i log-odds (risikoøkningen knyttet til denne verdien er avhengig av total dødelighet. For hoftebrudd, som har lavest dødelighet, tilsvarer log-odds på 20% nesten 20% økning eller reduksjon i risiko for å dø. For akutt hjerteinfarkt og hjerneslag tilsvarer det om lag 15-16% endring). Som reaksjonsgrense (alert limit) har vi valgt 100% økning eller 50% reduksjon i log-odds i forhold til gjennomsnittet.

1.2.4 Beslutningsregler

Når sykehus skal vurderes ved hjelp av en kvalitetsindikator, finnes det minst tre relevante perspektiver:

- Det ene er perspektivet til dem som kun er interessert i ett sykehus (perspektiv A). For dem er sammenlikning mot gjennomsnitt det viktigste, og bruken av et signifikansnivå på 5% fører bare til feilslutning en gang hvert 20. år

- Det neste perspektivet er myndigheters og eieres (perspektiv B). De må fortløpende fatte beslutninger om flere sykehus, og samme signifikansnivå fører regelmessig til at man feilaktig slutter at sykehus skiller seg ut fra gjennomsnittet. På den annen side vil muligheten for å oppdage et dårlig sykehus, særlig dersom det er lite, være liten hvis man er for konservativ, ved f.eks. å velge et for strengt signifikansnivå
- Det tredje perspektivet (perspektiv C) er hensynet som må tas ved offentliggjøring av resultater der det må vises hvilke sykehus som skiller seg klart fra gjennomsnittet. Her må den totale risikoen for å slutte feil tas i betraktning, noe som må føre til valg av strengere signifikansnivå

Dilemmaet i perspektiv B foreslår vi løst ved å operere med en kort liste over sykehus som trenger å følges opp. Kriteriet for å komme på listen vil være forskjellig avhengig av antall innleggelser per år. På listen kan det befinne seg sykehus som ikke skiller seg signifikant fra gjennomsnittet. En ukritisk leser som ikke tar dette forholdet i betraktning, kan derfor la seg forlede til feilslutninger.

Perspektiv C løses ved bruk av multippel testing.

For å balansere risikoen for å gjøre type I-feil (å slutte at et sykehus er forskjellig fra gjennomsnittet selv om det ikke er det) og type II-feil (å slutte at et sykehus ikke er forskjellig fra gjennomsnittet selv om det er det), beregnet vi styrkefunksjonene for sykehus av ulik størrelse.

1.2.5 Robusthet

Vi har undersøkt hvor robuste modellene er ved å se på hvor stabile konklusjonene er for hvert sykehus ved bruk av forskjellige modeller og ulike datasett. Modellene sammenliknes med grafisk fremstilling og bruk av Pearsons og Spearmans korrelasjonskoeffisienter. En direkte sammenlikning med metodene i den amerikanske vitenskapelige vurderingen (1) er ikke mulig, fordi de brukte lineær regresjon (logistisk regresjon (slik vi gjør) er å foretrekke, men datasettet deres var for stort til logostisk regresjon). Dessuten hadde de 45 kvalitetsindikatorer, mens vi har tre. Med bare tre er det i mindre grad mulig å trekke slutninger om i hvilken grad den enkelte indikator faktisk er et uttrykk for kvalitet.

1.2.6 Kvalitetskontroll av data

Vi gjennomførte kontroll av hyppigheten av feil i de elektronisk innsamlede dataene fra sykehusene. Fra den samlede pasientpopulasjonen i datamaterialet ble det ved hvert sykehus tilfeldig plukket ut 50 pasienter med hver av de tre hoveddiagnosene. Sykehusene ble oppfordret til å la en lege gjennomgå disse 150 journalene for å kontrollere om de opplysningene Kunnskapssenteret hadde hentet ut elektronisk var i samsvar med opplysningene i pasientens journal. Manglende samsvar ble registrert.

1.3 RESULTATER

Dødelighet 30 dager etter innleggelse (30D) for akutt førstegangs hjerteinfarkt var 18,7%, for hjerneslag 17,2% og for hoftebrudd 6,9%. Gjennomsnittsalderen var lavest for hjerteinfarktpasientene, høyest for hoftebruddspasientene. Relativt flere menn var innlagt for hjerteinfarkt, relativt flere kvinner for hoftebrudd, mens kjønnsfordelingen

var jevn for hjerneslag. Andelen innleggelser som førte til flytting mellom sykehus var 9,4 % for hjerteinfarkt, 5,1 % for hjerneslag og 12,4 % for hoftebrudd.

Dødeligheten for hjerteinfarkt og hjerneslag avtok i perioden 1997–2001, mens den var stabil for hoftebrudd.

Kodiagnoser ved den aktuelle innleggelsen, som ikke kunne antas å være komplikasjonsdiagnoser, ble vurdert brukt for å vurdere allmenntilstand og alvorlighetsgrad av sykdom. Mange av disse diagnosene var av begrenset verdi som enkeltopplysninger, fordi for få hadde dem. Vi brukte derfor en enkel tellevariabel, antall slike diagnoser, og antall tidligere innleggelser som vikarvariabler for allmenntilstand. Vi fant klar sammenheng mellom disse to variablene og dødelighet. Sammenhengen var uventet for hjerteinfarkt og hjerneslag, der høyere antall kodiagnoser var forbundet med lavere risiko for død. Den var som forventet for hoftebrudd; høyere antall kodiagnoser var forbundet med høyere risiko for død. En nærmere gransking viste at andelen pasienter med kodiagnoser avhang av lengden på innleggelsen. Det er derfor en tendens til at pasienter med hjerteinfarkt eller hjerneslag som dør tidlig, får færre kodiagnoser enn andre, mens det motsatte er tilfellet ved hoftebrudd, hvor dødeligheten de første dagene er betydelig mindre.

Kodiagnoser fra nåværende opphold egner seg av denne grunn ikke som kovariat med den statistiske metoden vi bruker, der tid fra innleggelse til død ikke blir tatt i betraktning. Analysen benyttet derfor antall kodiagnoser fra tidligere sykehusopphold.

For vurdering av alvorlighetsgraden av tilstanden ved innleggelse, brukte vi et amerikansk, diagnosebasert klassifiseringssystem for hjerteinfarkt og hoftebrudd. Systemet var lite relevant for hjerneslag ifølge ekspertgruppen. For denne tilstanden brukte vi derfor en enkel alvorlighetsinndeling basert på hjerneinfarkt (mindre alvorlig) eller blødning (mer alvorlig). Vi fant en klar sammenheng mellom alvorlighetsgrad av tilstandene vurdert på denne måten, og dødelighet for hjerneslag.

1.3.1 Akutt førstegangs hjerteinfarkt

Dødeligheten er høy ved 20 års alder, lavest ved ca. 40 års alder og stiger så igjen. Først ved ca. 80 år passerer den igjen dødeligheten ved 20 års alder. Kvinner har lavere dødelighet enn menn. Det er redusert dødelighet knyttet til høyere utdanning, høyere inntekt og formue.

Vikarvariablene for allmenntilstand viste at dødeligheten økte med antall tidligere innleggelser og antall tidligere kodiagnoser. Med økende alvorlighetsgrad fant vi økende dødelighet, dog med unntak for det nest laveste trinnet. Dette er ikke i tråd med forventningene og kan skyldes at korte opphold fører til færre kodiagnoser, eller lite presis diagnosekoding i sykehusene.

Å bli flyttet til et sykehus fra et annet sykehus var forbundet med redusert risiko for død.

Det ble funnet en reduksjon i dødelighet gjennom perioden 1997–2001 ved alle sykehus. I 2001 varierte ujustert observert dødelighet mellom 1,6 % og 26,3 % blant de sykehusene som hadde minst 50 tilfeller, men bare fire sykehus hadde dødelighet høyst 12%. Disse sykehusene har imidlertid spesielle funksjoner.

Selv om mange variabler har signifikant effekt i modellen, er det avgjørende om denne effekten har betydning for resultatene på sykehusnivå. Vi fant at:

- å utelate pasienter som var flyttet mellom sykehus, hadde betydelig effekt på to sykehus, men nesten ikke effekt på resten
- enten man har med antall kodiagnoser fra nåværende opphold eller ikke, får det ikke vesentlig betydning for noen sykehus, selv om økt antall kodiagnoser er forbundet med redusert risiko for død
- å ikke bruke resultatene fra 1999 (som var et spesielt år pga. overgang til kodesystemet ICD-10 og endring i infarktdiagnostikken) førte til en del endringer for noen sykehus, dog ikke dramatiske
- å utelate de sosiodemografiske variablene førte ikke til endringer
- å utelate korreksjon for allmenntilstand og alvorlighetsgrad av sykdom fører til betydelig endring for ett sykehus og en del endringer for flere sykehus
- å utelate korreksjon for alder og kjønn fører til betydelige endringer for noen sykehus og en god del endringer for mange sykehus
- å utelate alle opphold kortere enn to dager førte til moderate endringer for de fleste sykehus og store endringer for noen sykehus. Store endringer inntraff for spesialsykehus eller sykehus som hadde stor andel pasienter med svært kort liggetid (mindre enn ett kvarter)

De to sykehusene som affiseres mest av å bruke forskjellige modeller, er spesielle sykehus med en helt annen pasientpopulasjon enn det store gross av sykehus.

Vi konkluderer at det bør korrigeres for alder, kjønn, allmenntilstand og alvorlighetsgrad av sykdom. Bruk av sosiodemografiske variabler synes ikke å være nødvendig. Modellen ser ellers ut til å være rimelig robust, med unntak for sykehus med helt spesielle funksjoner.

Perspektiv A fører til at 16 sykehus vil bli vurdert som forskjellig fra gjennomsnittet. Perspektiv B fører til at 13 sykehus settes på oppfølgingsliste. Perspektiv C vil eksponere kun ett sykehus som forskjellig fra gjennomsnittet.

Nesten ingen av sykehusene vil, i henhold til modellen, havne utenfor det intervallet vi på forhånd definerte som en grense for hva man kan kunne velge å overse. En fordelingsanalyse kan tyde på at sykehusene deler seg i to: en gruppe med omtrent normal dødelighet og en med noe høyere dødelighet. Begge gruppene har liten spredning.

Vi ønsket å teste om det hadde betydning hvorvidt akutt hjerteinfarkt var hoveddiagnose eller bare en av diagnosene ved innleggelse. En sammenlikning av resultatene viste at fire sykehus ble betydelig affisert. Disse hadde enten svært liten eller svært høy andel av hjerteinfarkt som hoveddiagnose.

1.3.2 Hjerneslag

Dødeligheten stiger med alderen, dog noe forskjellig for hjerneinfarkt og hjerneblødning. Dødeligheten ved hjerneblødning er høyere til over 80 års alder, men stiger langsommere med alderen enn dødeligheten ved hjerneinfarkt. Vi observerte ingen kjønnsforskjell i dødelighet. Det er redusert dødelighet knyttet til høyere utdanning, høyere inntekt og formue.

Vikarvariablene for allmenntilstand viste økt dødelighet med antall tidligere innleggelser og antall tidligere kodiagnoser. Pasienter med hjerneblødning hadde som forventet langt høyere dødelighet enn pasienter med hjerneinfarkt. Modelltilpasningen

var ikke så god for tilfeller med høy forventet risiko for død. Dette førte til en grundigere separat analyse av effekten av type hjerneslag (vikarvariabelen for alvorlighetsgrad), og muligheten for interaksjon med andre variabler, spesielt alder.

Å bli flyttet til et sykehus fra et annet sykehus var forbundet med redusert risiko for død.

Den gjennomsnittlige dødeligheten i perioden 1997–2001 var svakt fallende. Utviklingen over tid var imidlertid forskjellig mellom sykehusene, i det mange var stabile og noen faktisk hadde økende risiko for død i perioden. I 2001 varierte ujustert observert dødelighet mellom 8,8% og 28,1% blant de sykehusene som hadde minst 50 tilfeller.

Selv om mange variabler har signifikant effekt i modellen, er det avgjørende om denne effekten har betydning for resultatene på sykehusnivå. Vi fant at:

- å utelate pasienter som var flyttet mellom sykehus, hadde betydelig effekt på ett sykehus, men nesten ikke effekt på resten
- enten man har med antall nåværende kodiagnoser eller ikke, får det ikke vesentlig betydning for noen sykehus, selv om økt antall kodiagnoser er forbundet med redusert risiko for død
- å utelate de sosiodemografiske variablene førte ikke til endringer
- å utelate korreksjon for allmenntilstand og alvorlighetsgrad av sykdom fører til betydelig endring for flere sykehus og en del endringer for de fleste andre sykehus
- å utelate korreksjon for alder og kjønn fører til betydelige endringer for mange sykehus.
- å utelate alle opphold kortere enn to dager, førte til moderate endringer for de fleste sykehus og store endringer for spesialsykehus

Sykehuset som var mest affisert av å utelate flyttede pasienter er spesielt fordi det hadde en helt annen pasientpopulasjon enn de øvrige.

Vi konkluderer at det bør korrigeres for alder, kjønn, allmenntilstand og alvorlighetsgrad av sykdom. Bruk av sosiodemografiske variabler synes ikke å være nødvendig. Modellen ser ellers ut til å være rimelig robust, med unntak for ett sykehus.

Perspektiv A fører til at 20 sykehus vil bli vurdert som forskjellig fra gjennomsnittet. Perspektiv B fører til at 20 sykehus settes på oppfølgingsliste. Perspektiv C vil eksponere sju sykehus som forskjellig fra gjennomsnittet.

En valideringsstudie basert på kliniske data ble gjennomført på et utvalg av 15 sykehus. Canadian Stroke Scale (CSS) ble beregnet på grunnlag av journalinformasjon. Det viste seg at CSS hadde signifikant betydning for dødelighet, som forventet. Ved å sammenligne sykehusene med og uten CSS, ble resultatene signifikant forskjellige for 2 sykehus.

En betydelig andel av sykehusene vil i henhold til modellen havne utenfor det intervallet vi på forhånd definerte som en grense for hva man kan kunne velge å overse. En fordelingsanalyse kan tyde på at sykehusene deler seg i to: en gruppe med omtrent normal dødelighet og en mindre gruppe med høyere dødelighet. Forskjellen mellom beste og dårligste sykehus synes å være betydelig.

1.3.3 Hoftebrudd

Dødeligheten stiger jevnt med alderen i området over 65 år. Vi observerte betydelig lavere dødelighet for kvinner. Det er noe redusert dødelighet knyttet til høyere utdanning, høyere inntekt og formue.

Vikarvariablene for allmenntilstand viste økt dødelighet med antall tidligere innleggelser og antall tidligere kodiagnoser. Med økende alvorlighetsgrad av tilstanden fant vi økende dødelighet.

Å bli flyttet til et sykehus fra et annet sykehus var forbundet med redusert risiko for død.

Dødeligheten i perioden 1997–2001 var stabil, og vi fant ingen vesentlige ulikheter mellom sykehusene når det gjaldt utviklingstrekk. I 2001 varierte ujustert observert dødelighet mellom 3,6 % og 14,3 % blant de sykehusene som hadde minst 50 tilfeller.

Selv om mange variabler har signifikant effekt i modellen, er det viktigste i denne sammenheng om denne effekten har betydning for resultatene på sykehusnivå. Vi fant at:

- å utelate pasienter som var flyttet mellom sykehus hadde betydelig effekt ved to sykehus, men begrenset effekt ved resten
- enten man har med antall kodiagnoser fra nåværende opphold eller ikke, får det ikke vesentlig betydning for noen sykehus, selv om økt antall kodiagnoser er forbundet med redusert risiko for død
- å utelate de sosiodemografiske variablene førte ikke til endringer
- å utelate korreksjon for allmenntilstand og alvorlighetsgrad av sykdom fører til en del endring for de fleste sykehus
- å utelate korreksjon for alder og kjønn fører til moderate endringer for de fleste sykehus

Sykehusene som var mest affisert av å utelate flyttede pasienter er spesielle, med høy andel flyttede pasienter og pasientpopulasjoner ulike de øvrige.

Vi konkluderer at det bør korrigeres for alder, kjønn, allmenntilstand og alvorlighetsgrad av sykdom. Bruk av sosiodemografiske variabler synes ikke å være nødvendig. Modellen ser ellers ut til å være rimelig robust, med unntak for to sykehus med høy andel flyttede pasienter.

Perspektiv A fører til at 19 sykehus vil bli vurdert som forskjellig fra gjennomsnittet. Perspektiv B fører til at 17 sykehus settes på oppfølgingsliste. Perspektiv C vil eksponere fem sykehus som forskjellig fra gjennomsnittet.

En betydelig andel av sykehusene vil, i henhold til modellen, havne utenfor det intervallet vi på forhånd definerte som en grense for hva man kan kunne velge å overse. Forskjellen mellom beste og dårligste sykehus synes å være betydelig.

1.3.4 Andre resultater

Korrelasjon mellom kvalitetsindikatorene

Man kan forestille seg at de enkelte kvalitetsindikatorene er uttrykk for en generell egenskap ved sykehuset, selv om de gjelder tilstander behandlet i forskjellige avdelinger. Hadde vi hatt 45 kvalitetsindikatorer, som i den amerikanske vitenskapelige vurderingen (1), ville det vært relevant å studere dette nærmere. Med tre indikatorer kan

man ikke trekke slutninger i så måte. En positiv og ikke svært forskjellig korrelasjon mellom indikatorene ville imidlertid være et ønsket utgangspunkt. Vi fant nettopp dette, korrelasjonene mellom kvalitetsindikatorene var henholdsvis 0,25 mellom hjerteinfarkt og hjerneslag, 0,18 mellom hjerteinfarkt og hoftebrudd og 0,32 mellom hjerneslag og hoftebrudd.

Empiriske styrkefunksjoner

Med kjennskap til den observerte spredningen i vårt materiale, kan man beregne styrkefunksjoner for de tre diagnosene. Slike styrkefunksjoner fremstår som kurver med antall innleggelser på x-aksen og styrke på y-aksen. Det vil være forskjellige kurver for ulike signifikansnivåer avhengig av om man velger å lete etter ulikheter man ikke bør overse (indifference limit) eller ulikheter det må reageres på (reaksjonsgrense, alert limit).

Kvalitetskontroll av data

Vi fant at våre innsamlede data avvek fra sykehusenes journaler i færre enn 1% av tilfellene mht. tidspunkt for innleggelse, hoveddiagnose og indeksdiagnose (diagnosen som førte til inklusjon i materialet).

1.4 DISKUSJON

I vurderingen av hvorvidt dødelighet innen 30 dager etter innleggelse er anvendelig som kvalitetsindikator i Norge, har vi tatt utgangspunkt i de seks klassene av kriterier brukt i den amerikanske vitenskapelige vurderingen (1). Disse klassene er:

- åpenbar validitet (”face validity”), dvs. det som måles er udiskutabelt viktig og kan påvirkes av behandler eller system
- presisjon, dvs. det som måles må variere tilstrekkelig mye mellom måleenhetene og variasjonen må ikke først og fremst skyldes tilfeldighet eller karakteristika ved pasientene (reliabilitet er ellers vanlig brukt som betegnelse)
- skjevhetsutjevning (”minimum bias”), dvs. man må være i stand til å korrigere for skjevheter knyttet til pasientkarakteristika og mangelfull datakvalitet slik at skjevheter i materialet reduseres til et minimum
- konstruksjonsvaliditet (”construct validity”), dvs. det må finnes empirisk støtte for sammenheng mellom indikatoren og kvalitet, og det bør observeres sammenheng med andre indikatorer for tilsvarende type kvalitet
- gir virkelig kvalitetsforbedring (”fosters real quality improvement”), dvs. kan ikke føre til datamanipulasjon eller incentiver til handlingsmønstre som ikke er i tråd med overordnede verdier og prioriteringer
- kan brukes, dvs. det bør være dokumentert eller argumenteres overbevisende for at indikatoren er et godt supplement til andre indikatorer som er i bruk.

Den åpenbare validitet ligger i at død/overlevelse er det viktigste resultatmål overhodet, at de tre tilstandene vi har studert er hyppige og alvorlige og at medisinsk behandling kan påvirke utfallet av tilstandene.

Vi har nærmet oss kravet om presisjon ved å overveie våre empiriske styrkefunksjoner, og balansen mellom risikoen for å gjøre type I- og type II-feil for de tre tilstandene.

For hjerteinfarkt finner vi et standardavvik på 0,065 (0,18 i det amerikanske materialet) (1). I sykehus med mer enn 100 innleggelser per år er sjansen for å påvise et reaksjonsgrenseavvik nesten 100 % med signifikansnivå på 0,15 %. I sykehus med færre enn 100 innleggelser per år er sjansen for å påvise et reaksjonsgrenseavvik bedre enn 85 % med et signifikansnivå på 5 %. Indikatoren for hjerteinfarkt er den mest presise vi har funnet.

For hjerneslag finner vi et standardavvik på 0,22 (0,32 i det amerikanske materialet) (1). I sykehus med mer enn 100 innleggelser per år er sjansen for å påvise et reaksjonsgrenseavvik over 90% med signifikansnivå på 5%. I sykehus med færre enn 100 innleggelser per år varierer sjansen for å påvise et reaksjonsgrenseavvik mellom 40 % og 90% med et signifikansnivå på 5%. Indikatoren er mindre presis enn den for hjerteinfarkt, men akseptabel dersom man bruker de tre beslutningsperspektivene korrekt.

For hoftebrudd finner vi et standardavvik på 0,19 (0,63 i det amerikanske materialet) (1). I sykehus med mer enn 100 innleggelser per år er sjansen for å påvise et reaksjonsgrenseavvik over 95% med signifikansnivå på 1%. I sykehus med færre enn 100 innleggelser per år varierer sjansen for å påvise et reaksjonsgrenseavvik over 50 % med et signifikansnivå på 5%. Indikatoren er litt mer presis enn den for hjerneslag, og akseptabel dersom man bruker de tre beslutningsperspektivene korrekt.

Ut fra disse observasjonene foreslår vi noen beslutningsgrenser for å sette sykehus på kort liste med behov for oppfølging, der signifikansnivået er avhengig av hvor stort sykehuset er. Vi viser også en tabell over risikoen for å gjøre feilslutninger av type I og type II, knyttet til antall innleggelser per år.

Skjevhetsutjevningen er i litteraturen særlig knyttet til dødelighet under oppholdet i sykehus, et problem vi har eliminert gjennom bruk av 30D. Vi har observert til dels stor variasjon mellom regionene i innleggelsesrater, noe som kan tyde på en seleksjons-skjevhet som vi ikke har kunnet korrigere for. Vi har observert indikasjoner på unøyaktig koding i stort omfang, men siden dette ikke ser ut til å påvirke modellene nevneverdig, er det grunn til å tro at disse unøyaktighetene er forholdsvis jevnt fordelt mellom sykehusene. For sykehus med spesielle funksjoner, noe som indirekte observeres gjennom høy andel av flyttede pasienter, er det grunn til å påpeke at ulike modeller gir ulike resultater og at indikatoren er lite robust. Dette gjelder dog et fåtall sykehus. For øvrig har vi observert robuste resultater dersom modellene inkluderer pasientens alder, kjønn, vikarvariabler for allmenntilstand og alvorlighetsgrad av tilstanden.

For størrelsesorden på skjevhet, har vi beregnet en maksimal forventet skjevhet knyttet til teoretiske betraktninger om ulike feilkilder. Gal eller mangelfull koding knyttet til død ved ankomst til sykehuset synes å være den eneste avgjørende feilkilde for alle de tre tilstandene.

For konstruksjonsvaliditet fant vi riktignok korrelasjon mellom de tre 30D-indikatorene, men dette kan ikke tillegges mye vekt. Ekspertgruppenes vurdering av resultatene i forhold til deres kjennskap til behandlingskvaliteten ved norske sykehus er en tilnærming. Selv om hjerteinfarktindikatoren er den mest presise, var denne ekspert-gruppen i tvil om validiteten, siden noen av resultatene for enkeltsykehus var motsatt av

forventet. Gruppen konstaterte likevel at risikojusteringen gav forventede resultater, det samme gjaldt observert dødelighet. Ekspertgruppen for hjerteinfarkt mente resultatene var gode nok til internt bruk, men frarådde publisering. Ekspertgruppen for hjerneslag hadde små forventninger til resultatene, men ble overrasket over å se at de var i samsvar med deres forventninger om hvilke sykehus som var gode og mindre gode. De godtok at spredningen i dødelighet mellom sykehusene var størst for denne tilstanden, og så dette som et uttrykk for at moderne hjerneslagbehandling trolig ikke er iverksatt i hele landet. Ekspertgruppen for hoftebrudd kom ikke til enighet om resultatene var i samsvar med forventning. Deres viktigste konklusjon var at resultatene kunne tolkes slik at indre-medisinsk oppfølging av pasientene synes å være av større betydning for dødeligheten enn kvaliteten på den kirurgiske behandlingen. Det ble samlet inn opplysninger om rutiner, behandling osv fra sykehusene ved en spørreskjemaundersøkelse. Dessverre var responsraten så vidt lav, spesielt blant de sykehusene som ble vurdert som forskjellig fra gjennomsnittet, at vi ikke fant å kunne bruke disse dataene i den primære analysen. Vi fant indikasjoner på at høyt pasientvolum medførte lavere dødelighet for slag og at antall senger i spesialenhet virket gunstig på dødelighet for hoftebrudd. Ellers er det en tendens til at sykehus med høy dødelighet har pasienter med få diagnoser fra tidligere opphold, eller få tidligere sykehusopphold.

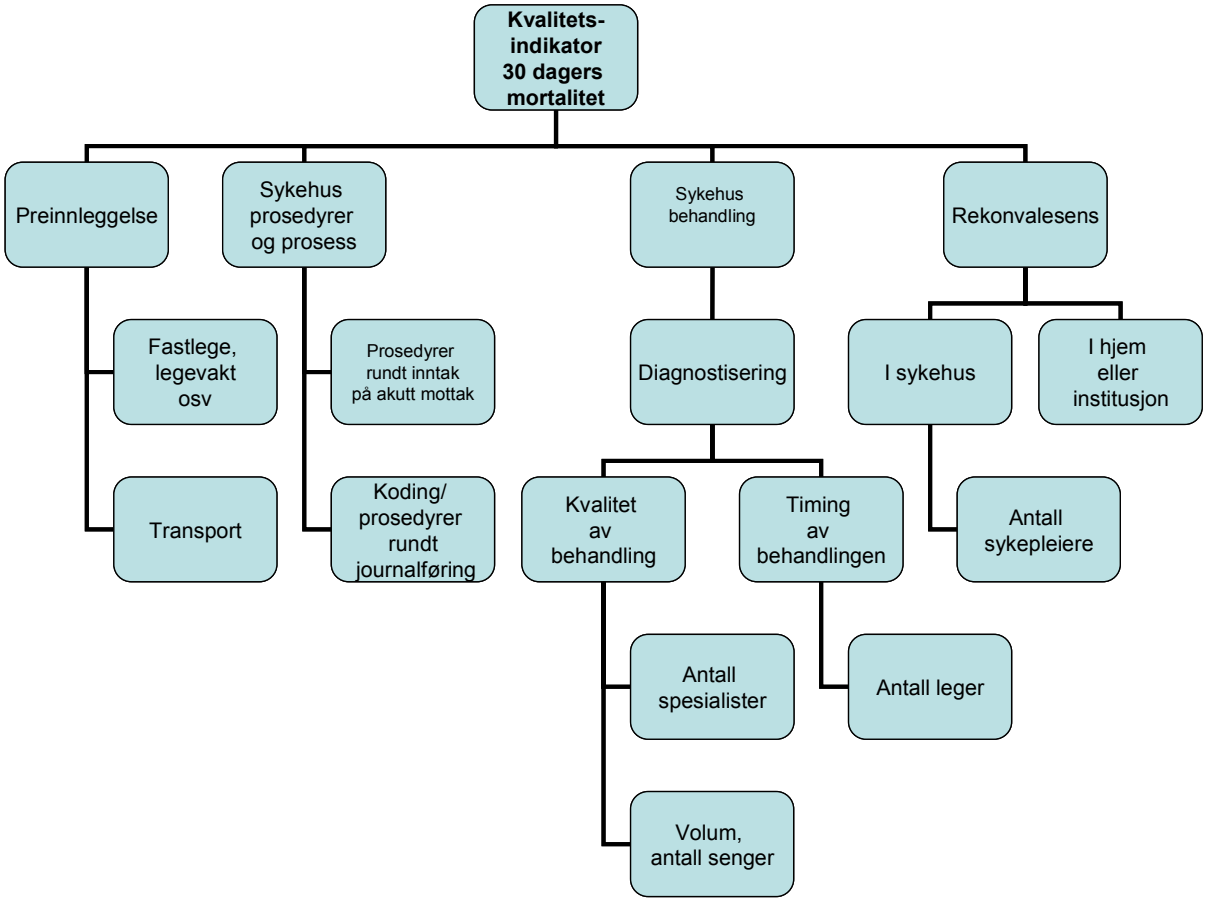
Vi har i den aktuelle undersøkelsen ikke vurdert eller studert om det å bruke indikatoren dødelighet innen 30 dager etter innleggelse er nyttig og faktisk fører til kvalitets-forbedring.

Tilgangen på data er god, men kvaliteten på data er mer tvilsom. Vi har funnet lavere feilrater enn det andre har rapportert fra koderevisjoner ved norske sykehus. Dette skyldes trolig forskjeller i hvilke avvik man har sett etter og hvordan disse er blitt klassifisert. Gjennom våre analyser har vi ikke kunne påvise at den blandede kvaliteten har tungtveiende innflytelse på konklusjonene, men på grunn av begrensningene som ligger i vårt datagrunnlag har vi heller ikke kunnet avkrefte dette. Den mulige feilkilde som har mest alvorlige konsekvenser, er forskjeller i registrering og koding av pasienter som er døde eller dør kort tid etter ankomst til sykehuset.

1.5 KONKLUSJON OG ANBEFALINGER

Nå man evaluerer kvaliteten av sykehusbehandling, er det ikke bare medisinsk behandling som kommer i betrakting, men også administrative rutiner og prosesser, slik som lengden av opphold, tid til operasjon, kapasitet osv., illustrert av figuren nedenfor.

Figur 1-1: Beslutninger og prosesser som påvirker kvalitet og resultater av sykehusbehandling



Forskjeller i 30 dagers dødelighet mellom sykehus kan avhenge av en hvilken som helst beslutning eller prosess i figuren. Det kan også være helt legitime årsaker til at to sykehus har gjort forskjellige valg mht prosedyrer, prosesser or ressursallokering. Faktorer utenfor sykehusets kontroll kan også være av betydning.

Våre hovedkonklusjoner om anvendelse av 30D som kvalitetsindikator er at

1. presisjonen er tilfredsstillende, dvs at vi kan påvise betydelige avvik i kvalitet innen akseptable statistiske feilmarginer
2. det gjenstår usikkerhet når det gjelder skjevhet som følge av ulik diagnose- eller kodepraksis ved sykehusene og manglende kliniske data til risikojustering

3. det bør gjennomføres en valideringsstudie for å kvantifisere betydningen av eventuelle skjevheter

Som kovariater for skjevhetsutjevning anbefales følgende: alder (via splinefunksjoner), kjønn, allmenntilstand (målt ved vikarvariablene antall tidligere innleggelser og antall relevante kodiagnoser fra tidligere innleggelser) samt alvorlighetsgrad (målt ved proxy-variabelen avstand mellom hjem og sykehus, forenklet CCDSS-klassifisering (blødning/ infarkt for hjerneslag) og om pasienten er overført fra et annet sykehus). Sosio-demografiske variable og sivilstatus kan ha betydning i fremtiden, og det er derfor ønskelig å inkludere disse variablene. Fortsatt er det likevel forhold som bør studeres nærmere, og det bør også gjennomføres noen endringer i sykehusene for å bedre data-kvaliteten.

Kunnskapssenteret foreslår:

- Det optimale tidspunktet for registrering av død etter innleggelse utredes for hver av de tre tilstandene
- Det utarbeides en grundigere veiledning i beslutningsanalyse knyttet til indikatoren, basert på statistiske betraktninger om risiko for feil-slutninger
- Data for 2002–2005 innhentes og brukes i en utvidet analyse, der også klinisk informasjon fra journaler inngår, for å se nærmere på robusthet og betydningen av å ha et enhetlig kodeverk (ICD-10) i bunnen, samt å belyse aktuell diagnose- og kodepraksis.
- I samarbeid med sykehus med avvikende resultater (basert på ajourførte datasett) utredes mulige forklaringer på dette, slik at man med større sikkerhet kan slå fast om resultatene er knyttet til medisinsk behandlings-kvalitet, dårlig datakvalitet, organisatoriske eller andre forhold
- Det utredes av om dødelighet ved flere tilstander er egnet som kvalitets-indikatorer
- Det innføres et regelverk som ikke levner tvil om hvordan sykehus skal registrere og kode tilfeller der pasienter dør eller er døde ved ankomst til sykehuset
- Det utredes om fordelene knyttet til en mer presis registrering av allmenntilstand og alvorlighetsgrad for en aktuell tilstand er så store at det oppveier ulempene ved å innføre slike rutiner ved sykehusene

2. ABSTRACT

2.1 INTRODUCTION

Indicators of the quality of health care are often used as a means of evaluation and monitoring trends in health care quality, identifying patients having received varying care and evaluating treatment methods. In this context, a quality indicator is defined as a statistical value, for fixed and current time-periods, indicating how certain processes function or whether specific outcomes have been achieved. One of the suggested outcome quality indicators is probability of death after 30 days (or 30-day mortality (30D), which seems to be the most commonly, used term in the literature, although 30-day case fatality is a more proper term). Evaluating hospital health care quality includes not only evaluating treatment of diseases, but also evaluating administrative routines and processes, such as number of days at hospital, delay to operation, not enough capacity for patients, etc.

A difficulty in using quality indicators is the challenge of comparing hospitals and health care institutions receiving patients with different risk profiles. It is necessary in comparing health care institutions, to account for differences in risk profiles such that hospitals admitting only low risk patients do not compare more favorably than deserved, to hospitals also accepting high-risk patients.

Three disease categories have been selected for evaluation of 30-day mortality as a quality indicator: acute myocardial infarction (AMI), hip fracture and stroke. These three disease categories were chosen as three major causes of death in the Norwegian population.

2.2 METHODS

This study collected data from the Patient Administration System (PAS) for the years 1997-2001, and for the three disease categories acute myocardial infarction (AMI), stroke and hip fracture. The classification of diagnoses changed during the period (1999) from ICD-9 to ICD-10. All hospital admissions for the three categories were collected, for all hospitals in Norway. Only the first admission for the disease in question, per patient, per calendar year, was selected. AMI cases were restricted to the first occurrence (patients with previous diagnosis of 410 at any hospital, since 1994, were removed from the dataset). Admissions for stroke that were less than 28 days from a previous admission were also removed from the dataset, even when being the first in a calendar year.

Information concerning index diagnoses, codiagnoses, procedures and transfers between institutions was collected. The data set was combined with information on socio-demographic status obtained from national statistics from Statistics Norway (SSB). Data that could not be merged with SSB data were removed from the dataset. The sample size prior to disease specific exclusion criteria was 54,095 for AMI, 53,072 for stroke and 50,205 for hip fracture.

The data was analyzed using logistic regression. Analyses were designed to provide an estimate of hospital effects. In evaluating the method, we considered three decision-making perspectives: A) the individual hospital, B) public authorities and policy makers, C) the public. The statistical methods and their associated parameters vary with the three perspectives.

The study design and investigation was strengthened by input from expert groups for each of the disease categories including clinicians, epidemiologists and statisticians. Analyses were designed to indicate if hospitals significantly deviated from the average with respect to 30-day mortality, while accounting for differences in risk profiles.

Hospital effects

Hospital effects are the log-odds of death within 30 days at the various hospitals, compared to the average hospital, controlling for risk adjustment covariates (see chapter 5.6).

An effect of e.g. -0.182 means that this particular hospitals has 10% reduced log-odds for death, while an effect of e.g. 0.41 means that the log-odds of death are 50% greater than in the average hospital, given the levels of the risk adjustment covariates.

Control of data quality was also performed in this study. Fifty patients from each disease category were randomly selected for each hospital. A doctor from the hospital checked by comparing to the journals that the data collected by the FS system was correct. In addition, an independent doctor checked these 50 patients from 15 hospitals for correctness.

2.3 RESULTS

30D for 1st time AMI was 18.7%, for stroke 17.2%, and for hip fracture 6.9%. Average age was less for AMI and highest for hip fracture. Relatively more men were admitted for AMI, and more women for hip fracture. Admissions that resulted in hospital transfers were 9.4% for AMI, 5.1% for stroke, and 12.4% for hip fracture. Mortality decreased over the time period for AMI and stroke, but remained stable for hip fracture.

The general results for the relationships of the risk factors for each disease category are summarized in the following table:

Table 2-1: Summary of relationships of 30-day mortality to risk factors for each disease category.

Covariate (set)	Disease category		
	AMI	**Stroke**	**Hip fracture**
Age	Highest for 20 yrs, lowest for 40 yrs, ↑ from 40 yrs	↑ with age	↑ with age
Sex	Women < men	No sex difference	Women << men
Socio-demographic variables	↑ ↓	↑ ↓	↑ ↓
Hospital transfers	↓ after transfer	↓ after transfer	↓ after transfer
Disease severity	↑ with severity	↑ with severity	↑ with severity
Patient frailty	↑ with frailty	↑ with frailty	↑ with frailty

To estimate the magnitude of possible bias in the results, robustness studies were performed. Hospital effects (on the log-odds scale) were estimated under changed statistical models or with changes in the data set, presumed to exhibit sensitivity for various sources of bias. The hospital effects under changed models or data were compared to the effects from the main analyses. The table below displays two measures of the degree of change. In many of the comparisons, substantial changes occurred mainly for one or two specialized hospitals. Note that a effect magnitude of 0.182 means that the hospital in question has an odds ratio for 30D, with respect to the average across hospitals, of 1.1 (or 0.9).

Table 2-2: Results of testing for the effects of bias for each disease category.

Alternative model/data set	Disease category					
	AMI		Stroke		Hip fracture	
	rank corr.	mean absolute change	rank corr.	mean absolute change	rank corr.	mean absolute change
Transfers removed	0.910	0.031	0.998	0.026	0.850	0.088
Including codiagnoses from present stay	0.989	0.015	0.999	0.070	0.986	0.032
Excluding 1999 data	0.918	0.041	-	-	-	-
Excluding socio-demographics	0.982	0.016	0.994	0.022	0.990	0.021
Excluding severity and frailty	0.936	0.048	0.979	0.068	0.949	0.065
Excluding severity and frailty, age and sex	0.786	0.105	0.941	0.110	0.903	0.099
Excluding stays < 2 days	0.808	0.104	0.926	0.120	-	-

Using the procedures proposed below, hospitals were identified as significantly different from the average, viewed from each of the three decision perspectives. The number of hospitals identified as having performance different from the average, for each decision-making level and disease category is provided in the following table. Deviations in both positive and negative directions are included.

Table 2-3: Number of hospitals with performance differing from average, by decision perspective.

	Decision perspective A (single hospital)	Decision perspective B (public authorities)	Decision perspective C (the general public)
AMI	16	13	1
Stroke	20	20	7
Hip fracture	19	17	5

Risk adjustment was based on data from various administrative databases: distance from home to hospital, socio-demographic data, the number of previous admissions and number of pertinent codiagnoses from previous admissions (for quantifying frailty and comorbidity); disease severity - as measured by the simplified CCDSS (Clinical Criteria Disease Staging System) staging system, and being transferred from another hospital. For stroke, staging was not used. Instead, diagnosis of hemorrhage or infarction was used as severity variable. It would be desirable to improve risk adjustment with clinical variables. Based on our data, we have no conclusive evidence as to the resulting improvement in precision and/ or bias. However, there are indications that the improvement in bias is not likely to be more than moderate.

2.4 CONCLUSION - CAN AND SHOULD 30-DAY MORTALITY BASED ON ADMINISTRATIVE DATA BE USED AS A QUALITY INDICATOR?

The main issue is whether mortality measures based on administrative data are valid indicators of true, hospital-specific mortality, while accounting for presumed bias, resulting from inaccurate coding, diagnostic variability and less than ideal case-mix adjustment. On the one hand, there remains a possibility that the bias of the indicator is large enough to influence the comparison between hospitals in a significant way. On the other hand, the results indicate that there are unacceptably high differences between hospitals. A review of the literature indicates that these differences seem to agree with those reported internationally. It is the role of the public health authorities to weigh the risk of incorrectly exposing hospitals as having poor quality, against the possibility that large apparent discrepancies in mortality reflect a true situation.

The present study is limited by the lack of clinical data and independent validation of diagnoses and codes. Within these limits, we have performed a study of plausible bias magnitudes indicating that unacceptable bias is probably avoided. Still, we feel that the issue is not settled in a satisfactory way. Further study, geared towards resolving the bias question, is recommended.

We have identified some less fundamental issues that need to be addressed: the need for more reliable registration of very early deaths, or choosing a strategy to reduce the sensitivity of 30D mortality to these cases, particularly for acute myocardial infarction, as well as the use of correct decision rules to identify hospitals as performance outliers. It is necessary to finalize the decision rules and their parameters, based on discussion with the various users of the indicators.

Besides bias, the most important criterion is precision. We have shown that the mortality indicator can be used to identify, with good statistical precision, hospitals where the probability of dying is appreciably different from the average.

Further studies should focus on validation of the results using clinical and laboratory data in addition to information from journals and direct communication with hospitals.

The criteria suggested by the HTA (Health Technology Assessment) report (1) were used as a conceptual framework for the evaluation. The results are summarized in Table 2-4 below.

Table 2-4: Evaluation of quality indicators.

Evaluation criterion	Conclusion		
Face validity	The disease categories are major causes of death. It is possible to provide results on a year-by-year basis.		
Precision	We have judged precision (reliability) as the ability to have low type II error probability, while keeping the relevant type I error probability under control. Proper decision rules, based on the user's decision perspective, are to be applied. Error probabilities are low for AMI and hip fracture, and acceptable for stroke. The study group's assessment of precision based on type of quality indicator and disease category[a)]:		
	AMI	Stroke	Hip fracture
	Good	Good	Good
Minimum bias	Without good coverage of clinical data, there will necessarily be some uncertainty whether data quality and risk adjustment is adequate to exclude any case-mix bias in hospital comparisons. However, there were few indications that systematic differences in case-mix did in fact exist between hospitals. Robustness testing resulted in few differences between models. Theoretical sensitivity studies seem to indicate that most kinds of bias are of small to moderate magnitude. It is, however, necessary to investigate further the coding practices for dead on arrival. The study group's assessment of minimum bias based on type of quality indicator and disease category[a)]:		
	AMI	Stroke	Hip fracture
	Acceptable [b)]	Acceptable	Acceptable
Construct validity	There was no clear indication that outlier status for an individual hospital could be explained by hospital characteristics.		
Fosters Real Quality Improvement	The indicator may provide further stimulus to incorrect coding. Otherwise, there are no indications that using this indicator would create incentives that would lead providers to improve performance without improving quality of care.		
Application	The indicator is widely used, and is well documented in the HTA report published by AHRQ. For stroke and hip fracture, there are strong indications that there are substantial differences between hospitals in probability of death after 30 days. A review of the literature resulted in the conclusion that the substantial performance differences found in this study do not run counter to what is known from the literature for AMI or stroke and to a lesser degree hip fracture.		

a) Criteria for evaluation of quality indicators are based on those found in the HTA report published by AHRQ (Agency for Health Care Research and Quality)(1).

b) On the condition that uncertainties concerning coding of dead on arrival is satisfactorily resolved.

Limited to PAS data and national statistics, this study recommends the following list of risk adjustment variables:

- age (via spline functions),
- sex,
- patient frailty as measured using the proxies number of previous admissions and number of pertinent codiagnoses from previous admissions,

- disease severity, using the proxies distance from home, simplified CCDSS (Clinical Criteria Disease Staging System) staging and being transferred from another hospital,
- distance from home and socio-demographic data. The predictive value must be weighed against the fact that these data are currently not available in the same time-frame as the PAS data.

3. INTRODUCTION

Indicators of the quality of health care are often used as a means of evaluation and monitoring trends in health care quality, identifying patients having received varying care and evaluating treatment methods.

3.1 WHAT THE INDICATOR MEASURES

A quality indicator is a measurable variable that is used to monitor and assess quality of health care services for fixed and current time-periods. The indicator can assess quality of hospital function both as experienced by patients and using proxies for quality. It should distinguish between structure, process and outcome. It is important to be clear as to whom the indicator is intended for and adapt information for that group.

Mortality indicators, although usually considered as a measure of quality of treatment, also reflect quality in process and structure. There is considerable literature indicating that hospitals vary considerably in essential elements of treatment (3). However, there is also literature to support that probability of death is affected by structure and process through for example for long waiting times for operations, need for moving to other hospitals for better treatment, and shortages in manpower of for example nurses (4-7) or doctors (8).

However, the question is equally important, what can mortality indicators for the chosen disease categories be used for? In addition to be used to measure quality of care for the acute admission, mortality indicators are important research tools to compare, for example, benefits of treatment methods, prevalence of mortality both as a whole and in population subgroups, and the effect of comorbidity and the importance of risk factors.

3.2 FOR WHOM IS THE INDICATOR INTENDED?

The indicator has an important role for the clinical and administrative personnel of hospitals. Mortality indicators are potentially useful and important tools as internal quality indicator.

Quality indicators can also provide information to health care providers and managers, public health policy makers, and health care consumers. The detail and emphasis of the information for each group should be different.

Information intended for the public consumer should be relatively uncomplicated, reflect clearly described elements of quality of care and reflect true options of choice. Mortality from these three diseases almost invariably follows emergency emissions that do not give the patient an element of choice. Should the indicator reflect other elements of quality in the hospital, this is not clear enough to recommend its use as an indicator for the public consumer. However, the general practitioner may benefit from the information provided by mortality indicators on the provider level. For the indicator to

be a tool for the practitioner, the indicator needs to adhere to stringent control procedures to assure that the information is correct on the provider level.

The Ministry of Health and Social Welfare has suggested that quality indicators should be developed for health care structures, processes and outcomes (9). The Ministry of Health and Social Welfare asked HELTEF, one of the predecessors of the Norwegian Knowledge Centre for the Health Services (NOKC), to develop and evaluate the usability and applicability of 30-day mortality. In this context, a quality indicator is defined as a statistical quantity indicating how certain processes function or whether specific outcomes have been achieved.

Different decision-makers may and should sometimes reach different conclusions about the relative performance of hospitals, based on quality indicators. It is in the nature of statistical decision-making, based on data that contain uncertainty, that one's decisions must depend on the scope of one's decision-making. One type of decision-maker is concerned only with a single, specific hospital, and should only guard against possible statistical errors concerning that hospital alone. In practice, this means that significance levels should be set at a suitable, conventional level. Another decision-making role is when one wants to draw conclusions about all the hospitals with negligible statistical uncertainty. It will then be necessary to use multiple testing methods, safeguarding against the possibility of any erroneous conclusion whatsoever. In practice, this is a very strict requirement that is in conflict with the strong desirability of detecting serious deviations in performance. The third decision-making role is a compromise: to identify hospitals with performance that can reasonably be questioned but not necessarily positively and reliably be identified as poor performers.

3.3 CHALLENGES TO BE CONSIDERED

A difficulty in using quality indicators is the challenge of comparing hospitals and health care institutions that receive patients with different risk profiles. It is necessary in comparing health care institutions, to account for differences in risk profiles such that hospitals admitting only low risk patients do not compare more favorably than deserved, to hospitals also accepting high-risk patients. Other considerations are that as progress of research into the cause, pathology and treatment of diseases occur, the definition of the disease can change. This has led to possibly misleading comparisons as to the incidence, prevalence and prognosis of diseases.

In an earlier research project (2), the use of in-hospital mortality as a quality indicator was evaluated. The previous study was the subject of an active discussion. Themes for discussion were: validity of the data set; statistical analysis methods; validity of the outcome measure, in-hospital mortality; and general use of the method. The results of the previous project included the recommendation that using standardized mortality indicators with a longer time-span, such as 30-day would be more useful. It has even been suggested that 30-day is too short (10).

3.4 THIS STUDY

The Directorate for Health and Social Affairs selected three disease categories for the evaluation of the use of mortality as a quality indicator: acute myocardial infarction,

stroke, and hip fracture. These three disease categories were chosen as three major causes of death in the Norwegian population.

In this study, we investigate 30-day mortality rates for acute myocardial infarction, stroke, and hip fracture. In the current work, we have accounted for the comments and suggestions received after an earlier study (2) of the inter-hospital comparison of in-hospital mortality by having:

1. Changed the mortality indicator from in-hospital mortality to 30-day mortality rates,
2. Incorporated expert advisory groups consisting of representatives of different stakeholder groups,
3. Used an extended set of explanatory variables that cover various risk adjustment features that may distinguish the patients at different hospitals,
4. Expanded and changed statistical analysis to include an evaluation of multilevel analysis.

In this discussion we will use the evaluation framework suggested by the Agency for Healthcare Research and Quality, US Dept. of Health and Human Services (1). As quoted in their report the framework includes as follows:

Based on the interviews and a review of the relevant literature, the project team developed an evaluation framework of ideal standards by which to judge quality indicator performance:

- ***Face validity:*** *An adequate quality indicator must have sound clinical and or empirical rationale for its use. It should measure an important aspect of quality that is subject to provider or health care system control.*
- ***Precision:*** *An adequate quality indicator should have relatively large variation among providers that is not due to random variation or patient characteristics.*
- ***Minimum bias:*** *The indicator should not be affected by systematic differences in patient case-mix, including disease severity and comorbidity. In cases where such systematic differences exist, an adequate risk adjustment system should be available based on discharge data.*
- ***Construct validity:*** *The indicator should be supported by evidence of a relationship to quality, and should be related to other indicators intended to measure the same or related aspects of quality.*
- ***Fosters Real Quality Improvement:*** *The indicator should not create incentives or rewards for providers to improve measured performance without truly improving quality of care.*
- ***Application:*** *The indicator should have been used effectively in the past, and/or have high potential for working well with other indicators currently in use.*

Note that precision corresponds closely to the concept of reliability, which is widely used in social sciences and related disciplines.

3.5 STAKEHOLDERS

Quality indicators for health care, as indicated in chapter 4.1, are tools for health practicioners and administration. As such, there is a need, when developing these tools to include stakeholder participation as a vital part of the development process.

The needs of the stakeholders were met in this study through several channels. To assure that some of the different stakeholders could discuss the methodologies being suggested in a collective forum, an expert advisory working group has been established, including:

- medical and scientific experts
- clinicians, doctors, nurses and other health professionals
- health administration officials from the government

Hospital administration is not directly represented; however, the medical experts and clinicians are high enough in the system to be able to assess the needs of the hospital as a whole.

The expert advisory groups were involved in the design of the study and the completion of the protocol. Although NOKC was responsible for the practical implementation of the study, the expert advisory groups were also actively engaged in the interpretation phase. The expert advisory groups consisted of the following individuals:

Table 3-1: Members of the expert advisory groups.

Name	Institution
Acute myocardial infarction	
Gunnar Eriksen	Akershus University Hospital, Lørenskog
Maja Lisa Løchen	University Hospital of Northern Norway, Tromsø
Harald Vik-Mo	St. Olav Hospital, Trondheim
Åsmund Reikvam	Institute of Pharmacotherapeutics, Univ. of Oslo, Oslo
Rune Wiseth	St. Olav Hospital, Trondheim
Stig A. Slørdahl	St. Olav Hospital, Trondheim
Arild Mangskaug	Ullevål University Hospital, Oslo
Eivind Myhre	Sørlandet Hospital, Kristiansand
Bjørn Haug	Helgeland Hospital
Frederic Kontny	Ullevål University Hospital, Oslo
Ottar Nygård	Haukeland University Hospital, Bergen
Stroke	
Halvor Næss	Haukeland University Hospital, Bergen
Eystein Brandt	Innlandet Hospital, Lillehammer
Odd R. Skogen	Ålesund Hospital, Ålesund
Hanne Ellekjær	Levanger Hospital, Levanger
Arve Dahl	Rikshospitalet University Hospital, Oslo
David Russel	Rikshospitalet University Hospital, Oslo
Hip fracture	
Kristian Bjørgul	Østfold Hospital, Fredrikstad

Name	Institution
Norvald Langeland	Buskerud Hospital, Drammen
Olav Røise	Ullevål University Hospital, Oslo
Håkon E. Meyer	Norwegian Institute of Public Health, Oslo
Anders Mølster	Haukeland University Hospital, Bergen
Odd Granlund	Akershus University Hospital, Lørenskog
Emil Mohr	Haugesund Hospital, Haugesund
Cathrine M. Lofthus	Aker Universitetssykehus, Oslo

In addition, professor Nils Lid Hjort at the Institute of Mathematics at the University of Oslo has been a statistical consultant in the project.

This report has been reviewed through an independent peer review process. The reviewers were:

Table 3-2: Report reviewers.

Name	Institution
Kari Nyland	SINTEF
Aage Tverdal	The National Institute of Public Health
Stein Emil Vollset	Section for Epidemiology and Medical Statistics, The University of Bergen
Inger Njølstad	Institute of Community Medicine, University of Tromsø

3.6 Approval by Data Inspectorate and Ethics Committee

The fundamental principle in this investigation is that personal identity is encrypted at two places:

- the hospital
- Statistics Norway (SSB)

Thereby, direct personal identity is removed from the data set, after combining information from the required data registries. The final working data set used in the study has encrypted personal identifying information during the duration of the project (31.06.2012). Thereafter, the encryption keys will be removed, and the data set will be free from direct person identifying information. At that point, the data from SSB will have to be deleted.

The Norwegian Data Inspectorate and the Regional Ethics Committee for Medical Research, Eastern Norway have approved the study protocol.

4. BACKGROUND

4.1 QUALITY INDICATORS

Measuring quality of health care is important for hospitals, physicians, patients, managers and politicians. All need evaluation tools. One such tool is the use of quality indicators (QIs). Tools that represent the quality of services are not easily available. Frequently, proxies for quality evaluate services.

It is important to be precise as to what the quality indicator describes and what it does not. The validity of the underlying data behind the indicator should be tested.

Quality indicators should be:

- Measures that assess a particular health care process or outcome.
- Quantitative measures that can be used to monitor and evaluate the quality of important governance, management, clinical and support functions that affect patient outcomes.
- Measurement tools or flags that are used as guides to monitor, evaluate, and improve the quality of patient care, clinical support services, and organizational functions that affect patient outcomes.

Case fatality, a fixed number of days after admission, is a commonly used indicator. Based on the number of days, after admission, the indicator reflects different parameters. Short-term case fatality (10 days or less) reflects the condition of the patient and medical treatment at the hospital. In-hospital mortality reflects conditions including treatment, structure, and process at the hospital. However, here hospitals with shorter average number of bed-days compare favorably. Mortality a fixed number of days post admission has the advantage that it is the number of deaths within the same number of days that is being compared, however, part of the time may be outside the hospital, and the indicator is as much an indication of quality of after-hospital care as in-hospital care.

Our starting point is that mortality is an important and meaningful dimension when evaluating the results of health care. It can be argued that disease specific mortality cannot reflect the quality of a hospital's treatment of other diseases. Furthermore, if there are major differences in the criteria for admission and diagnosis between hospitals, even disease specific mortality used as an indicator of quality of treatment can be misleading. However, we do not intend to discuss the meaningfulness of selecting mortality as a possible QI further, except for some minor comments in paragraph 4.7.

Mortality after admissions for the three disease categories acute myocardial infarction (AMI), stroke and hip fracture, has been selected for this study by the Directorate for Health and Social Services.

4.2 RISK ADJUSTMENT FOR MEASURING HEALTH CARE OUTCOMES

The concept of "risk" (risk adjustment) is hard to define and includes many factors. Not all can be measured simply and routinely, without extra costs, and their importance may

differ. Different risk adjustment systems operationalize risk differently. Iezzoni (11) lists and discusses two types of factors, variables related to patient physiology and health status – age, sex, clinical instability, main diagnosis, severity-at-admission, and comorbidity – and socio-economic variables – socioeconomic background, ethnicity, functional status, health expectations and life satisfaction.

Risk adjustment is necessary to include in outcomes analysis especially when performing inter-hospital comparisons. The Health Care Financing Administration first published hospital case fatality rates in the USA in 1986. In the institution with the highest case fatality rates, 87.6% died. As that institution turned out to be a hospice, agreement was soon reached that outcomes comparisons should not be made without adjustment for risk adjustment (12). Since then, a large number of risk adjustment systems have been developed.

4.2.1 Variables related to patient physiology and health

4.2.1.1 Risk adjustment for age

Age is an important risk factor for death. In particular, the oldest patients have lower physiological reserves, which mean that they recover much more slowly and/or suffer more treatment complications. Some research argues that age may not make a very big difference. It may not explain more than 1% of the variation in mortality for patients 65 years and older with congestive heart failure and hip fracture, 2% for acute myocardial infarction and 3% for pneumonia (13), Yet, a much used and validated system like APACHE III assigns risk points to age categories (14).

The additional risk implied by age may reflect not only the age-related reduction in physiological capacity, but also differences in treatment. Drugs known to be beneficial for acute myocardial infarction patients (thrombolysis, beta blockers, acetylsalicylic acid and nitrate) have been shown to be administered less frequently to patients over 69 (15). To the degree that age is related to the standard of treatment, and not just to treatment effects, standardizing or controlling for age will make inter-hospital comparisons fairer.

4.2.1.2 Risk adjustment for sex

The fact that women live longer than men, testifies to the potential importance of sex for mortality. Sex-specific physiological factors may be of importance: women metabolize beta-blockers more slowly, and their smaller vessels may be the reason their in hospital case fatality rates for coronary artery bypass graft is higher, even after control for age, left ventricular ejection fraction, comorbidity, and the number of vessels involved (16).

Like age, sex is often not an important risk factor. APACHE III does not include sex as a predictor of death. MMPS (Medicare Mortality Prediction System) does for stroke and pneumonia patients, but not in cases of acute myocardial infarction or congestive heart failure (17). In the RAND study (13) sex predicted 30-day mortality for hip fracture (males were more likely to die), but not for congestive heart failure, acute myocardial infarction, pneumonia or cerebrovascular incidents. Other Norwegian studies also did not find effects of sex (18).

Sex may affect not only treatment results but the administration of treatment, for instance, women have fewer angiographies, PTCAs and CABGs (19), although we do not know of any Norwegian studies. As mentioned above, to the degree that sex is

related to the standard of treatment, and not just to treatment effects, standardizing or controlling for sex will make inter-hospital comparisons fairer.

4.2.1.3 Risk adjustment for disease severity

A stroke may be more or less massive. A myocardial infarction may be mild or heavy. A displaced intracapsular hip fracture may heal more poorly than a non-displaced fracture. Therefore, risk assessment must include controlling for disease severity at admission by some measure of complexity or stage of disease,

As indicated earlier, disease severity at admission can be estimated using a measure of complexity or staging of disease. One staging method, the Clinical Criteria Disease Staging (CCDS) system, was proposed by Gozum et al. (20) and used in this investigation. The system is described below in Appendix 2 (9.2).

4.2.1.4 Risk adjustment for patient frailty (comorbidity)

Many patients, particularly older patients, have multiple health problems. Comorbidity is typically a chronic disease like diabetes, COPD and chronical ischemiae, but can also be an acute condition, as when a myocardial infarction worsens the prognosis for a patient admitted for prostate cancer. Morbidity may heavily affect the probability of a good outcome, and should therefore be adjusted for in comparisons of treatment results.

The codiagnosis that is most closely related to increased mortality from stroke is hypertension (21;22). Increased age increases atrial fibrillation which is a significant factor in increasing the 30-day case fatality rates in older patients (23). Another important codiagnosis in explaining increased mortality of stroke is diabetes (24). Heart disease is another important codiagnosis.

A number of instruments for adjusting risk by comorbidity exists (11). ASA-scores – the Physical Status Classification of the American Society of Anesthesiologists – has been used for decades for preoperative assessment of surgical patients, and Greenfield et al. (25) has developed the Index of coexistent disease (ICED). The Charlson Comorbidity Index has been shown to be a powerful outcomes predictor (26), and even to add to the accuracy of APACHE III-based predictions of in-hospital mortality (27). RAND's comorbidity registrations (13), however, only significantly predicted 30-day mortality for two of the five studied conditions when added to a model containing acute physiologic variables.

4.2.2 Variables describing the patient's socio-demographic condition

For predictions of outcome, it is sometimes important to know not only what kind of disease a patient has, but what kind of patient has the disease. Therefore, risk adjustment often includes adjusting for patient background.

A large body of research has shown that health varies by socio-economic status, both in the Nordic welfare states (28-31) as well as in countries with larger social inequalities (32-37).

Treatment response also varies by socio-economic status. Case fatality rates among cancer patients are 10-15% higher in low-income patients (38). Socio-economic differences in treatment response may reflect not only patients' ability to benefit, but also differences in the treatment they were offered. However, no matter which mechanism is involved, differences in treatment or differential response to the same

treatment, risk adjustment should include controlling for the patient's socio-economic conditions.

Socio-demographic status is not only a question of wealth. Non-material resources must also be taken into account. Much research emphasize the prognostic importance of social support, measured e.g. as having someone to show affection, confide in, hug, understand one's problems, have a good time with, prepare meals for or turn to for suggestions (39). Low scorers on social support also scored lower on physical functioning (40). Likewise, among 1234 acute myocardial infarction patients (risk adjusted), those living alone had a relative risk compared to those not living alone of 1.54 for a new infarction and 1.58 for cardiac death (41). Also, in a study of five year cardiac death, Williams et al. (42) found that controlling for "all known medical prognostic factors", social factors were the most important predictors of death, particularly having a spouse or a confidant: unmarried persons without confidants had a relative risk of 3.34 for cardiac death.

An important factor that may partially explain measured socio-demographic differences is smoking habits. In 28-day mortality studies of myocardial infarction, smoking is a determinant to mortality (18).

4.2.3 Factors associated with hospitalization and diagnosis procedure

4.2.3.1 Risk factors associated with type of hospital and time of admission

Various factors have been associated with premature mortality in studies involving selected diagnoses. Type of hospital has been frequently and significantly associated with quality of care (43;44). Teaching hospitals generally result in lower mortality and better quality of care for elderly patients suffering from acute myocardial infarction (3). Hospital capacity, where this is often geographically associated with less populated areas, led to greater hospital admissions, but not lower death (45). However, in a study of urban and rural hospitals, case fatality rates were lower in the latter, a study where risk adjustment differences were not considered high but could not be ruled out (46) Admission to high volume hospitals led to lower case fatality rates (47;48).

Season of admission was found to be significantly related to mortality from AMI, and significantly affected hospital admissions (49).

4.2.3.1.1 Distance between home and hospital

Distance between home and hospital has been suggested as a proxy for the time from the start of symptoms to arrival in hospital, as this time lag is not readily available in the electronic medical records. The assumption is that most people spend the majority of their time at home; in particular, one expects this assumption to be more relevant the older the patient is. As a proxy, distance between home and hospital has been given as an important reason for differences in risk adjustment between hospitals. How this variable affects mortality is not easy to predict. Given long transport, more of the seriously ill patients would be expected to die before arrival in hospital, thus contributing to a lower case fatality because they are not formally admitted. In addition, very seriously ill and old patients – living in nursing homes – will probably not be sent to hospital at all if the distance is long. On the other hand, patients that survive until they are admitted to the hospitals, could, due to the delay have a poorer prognosis than

surviving patients with shorter traveling distance to hospital. That would increase the probability of dying.

4.3 HOSPITAL MORTALITY, THE METHOD

Mortality is a widely used quality indicator for health care. Using mortality to assess and compare quality in hospitals is however, highly controversial. The quality of the data may not be good enough (50;51), and when comparing hospital quality as evaluated using other indicators, mortality did not always predict hospitals of poor quality (52;53). As indicated previously, biases may be induced in comparing hospitals due to differences in risk adjustment (54-56).

It is important that case fatality rates be risk adjusted for disease severity, comorbidities, and socio-demographic variables (57;58). However, in one study involving AMI, starting from a list of 73 candidate predictor variables, 7 were considered sufficient to be included in the final model (59), whereas in another study 8 were sufficient (60).

4.4 WHICH MORTALITY MEASURE IS VALID?

The literature highlights several methodological problems. In-hospital mortality data is relatively easy to acquire but has its limitations. In-hospital mortality reflects the length of the admission, with hospitals with longer periods of hospitalization having higher mortality (1).

In a comparison between 30-day standardized mortality and in-hospital mortality, differences were not large, although classification of hospitals as statistical outliers differed. There seemed to be no evidence of systematic bias through discharge procedures however (61). One study compared 30-day to 180-days mortality and found the results to be generally the same. The short-term risk should greater reflect health care quality, whereas the 180-days should greater reflect patient characteristics. There was no indication that hospitals with low 30-day mortality were postponing rather than preventing mortality (10;62).

4.5 A SHORT DESCRIPTION OF NORWEGIAN HOSPITALS

Norway has 66 hospitals that treat patients with myocardial infarction, stroke, or hip fracture.

Since 2001, the national government, who took over ownership from the counties, owns almost all Norwegian hospitals. Between 1997 and 2001 the county owned them. Until July 1, 1997, the hospitals were financed through global budgets, from then onwards a mixed model has been applied, varying (40% – 60%) DRG based funding (prospective payment system). Norway has a total population of about 4.5 million. The populations in the counties vary from about 75,000 to about 500,000.

The 19 counties are aggregated into five health regions, see Figure 5-1. The regional health authorities are responsible for planning of specialist health care within the regions, to ensure collaboration between and within counties.

Up until the end of 1997, each hospital has served a specific geographic region, and specifically for acute disorders, there has been very little overlap between hospitals. A few of the 66 hospitals do not treat patients in the acute phase, and some of the hospitals have had their functions redefined during the period of this study.

In Appendix 9.1, we list the participating hospitals and the corresponding aliases used in tables and figures.

5. METHODS

5.1 DATA COLLECTION METHOD

The study uses the FS-system (see 9.3) to make a functioning database using information from the patient administrative systems (PAS) of all hospitals, Statistics Norway (SSB), the Norwegian Causes of Death Register and other registers, including data from laboratory analysis. The data available from the PAS includes such information as age, sex, diagnoses and procedure codes, length of stay, departments, municipality of residence, type of admission (acute or elective) and in-hospital mortality. From Statistics Norway, socio-demographic information, distance between home and hospital of admission, and the exact date and cause of death from the death register was obtained.

The FS-system has to date been used in several multimember studies. This data collection system can semi-automatically collect standardized data about patient stay from any PAS at Norwegian hospitals in anonymous or encrypted format. Patients are uniquely identified at a national level even if they were transferred between or among hospitals for the same disease or for one of the other two diseases that are subject of this study. See Appendix 3 (9.3).

5.2 CASE SELECTION

5.2.1 Defining the cases

All 66 hospitals eligible for the study participated on a voluntary basis.

Establishing 30-day mortality as a year-by-year indicator required distinguishing between patients and admissions or episodes. In the following, when using the term "case", we refer to index admissions, defined as follows:

An index admission (different for acute myocardial infarction, see below) was defined as the *first* occurrence, during one calendar year (in the period 1997-2001), of one of the selected diagnoses (see 5.3.1) for a given patient at a given hospital. To be considered as an index case the hospital stay also had to include treatment (not admissions purely for rehabilitation) and had to be strictly longer than 1 day. An admission is defined as a one-day admission, and thus not qualifying as an index admission when 1) the patient neither is admitted earlier than 0700 nor discharged before 1700, and 2) the admission is not an emergency admission, and 3) the patient has not been transferred.

Patients transferred between hospitals without delay (discharged from the first and admitted to the next hospital within 24 hours), were recorded with index admissions in both hospitals (or even three or four if further transferred to other hospitals).

The list below gives a more detailed specification of the criteria:

- Only one case per patient, per calendar year, per disease category and per hospital.

- For AMI and hip fracture, cases were selected where index diagnosis was recorded either as main diagnosis or as a codiagnosis. For stroke, only admissions where stroke was main diagnosis were selected.
- When multiple admissions occurred for a patient, the first within the calendar year with the diagnosis in question was used; admissions within the same calendar year were considered readmissions.
- We wanted to calculate 30-day mortality for first time acute myocardial infarction admissions only. Only the first occurrence during the period 1997-2001 was defined as an index admission. In addition, we searched all Norwegian hospitals for previous occurrences of the diagnosis of myocardial infarction (410) since 1994. In the case of such occurrences, the admission was not defined as an index admission. This procedure was repeated for all cases. After the implementation of ICD-10, this procedure was changed, as first time myocardial infarction has been given a specific code. Prior occurrences of ICD-10 codes for first time myocardial infarction were thus not searched for (but prior occurrences of the ICD-9 code were always searched for). For stroke patients, an admission within 28 days was considered a prolongation of the first admission. Therefore, for example if a patient admitted in January had been admitted less than 28 days before, in December, the January admission was not considered an index admission.
- Admissions for resurgery related to hip fracture were removed from index admissions.
- Readmissions were all admissions within 12 months of the index admission at the same hospital as the index admission. Readmissions were noted both for diseases related to the index admission diagnosis and other pertinent diagnoses, and included information on main diagnoses, secondary diagnoses and pertinent procedures.
- Data for all patients were included in the initial data set, including those surviving a few hours from time of admission, or arriving dead. However, those coded as dead on arrival were removed prior to data analysis.
- Elective admissions were not considered index admissions.
- For hip fracture admissions, patients younger than 65 years were not included, for AMI and stroke admissions, patients younger than 18 were not included. These were removed prior to analysis.
- Eventually, hospitals with a total of fewer than 100 index admissions for a specific disease category, over the 5-year period, were removed prior to data analysis.
- A random set of 50 patients admitted for stroke, at 15 randomly selected hospitals were selected and retrospectively assigned a severity scale using the Canadian Stroke Scale.

5.2.2 Cases omitted during creation of working file

Cases that contained missing data crucial to analysis, or that presented inconsistent data, have been deleted during the creation of the working files. The data from the hospitals include the hour and minute of admission. However, time of death is registered with date only, hours regarded 00.00. This means that persons who die the same day as

admitted, will have a negative survival (up to 24 hours), unless admitted at midnight sharp. We have added 1 day to the survival time for these cases. Cases showing more than 24 hours negative survival time are regarded inconsistent and have been deleted.

The table below shows the number of cases delivered from the FS system and the number of, and reason for deletion.

Table 5-1: Cases delivered by the FS system.

Number of cases delivered from FS system	**176387**
Disagreement in gender between data from hospital and Statistics Norway	129
Admissions for resurgery related to hip fracture	480
Negative survival time (dead more than 24 hours before admission)	54
Missing person identifying tag	31
Missing data from Statistics Norway	1166
Sum remaining cases	174527

Refinement of the case list according to the criteria listed earlier resulted in the following adjustments:

Table 5-2: Cases remaining in data set.

AMI		54095
Stroke		53072
Hip fracture		50205
Not index admissions		
Coded as secondary heart attack (ICD-10)	7740	
Previously admitted with infarction and coded (ICD-9) diagnosis 410	6787	
Readmitted for stroke within 28 days	2628	17155
Total		**174527**

5.2.3 Comments on data properties

Duplicates. We have checked the file for duplicate cases defined as identical data on all of the following data:

- Person identifying tag
- Hospital
- Date of admission
- Which of the three diagnosis group the admission is related to (If admitted for two or three diagnosis groups, each group is an individual case and is not regarded as duplicate).

In some cases, it seems that the patient is registered more than once for the same admission. In other cases, it seems that each ward set its own date of discharge if the patient is moved elsewhere in the same hospital, but the new ward keeps the "old" date

of admission. This problem is solved by keeping the case with the longest stay and deleting the rest.

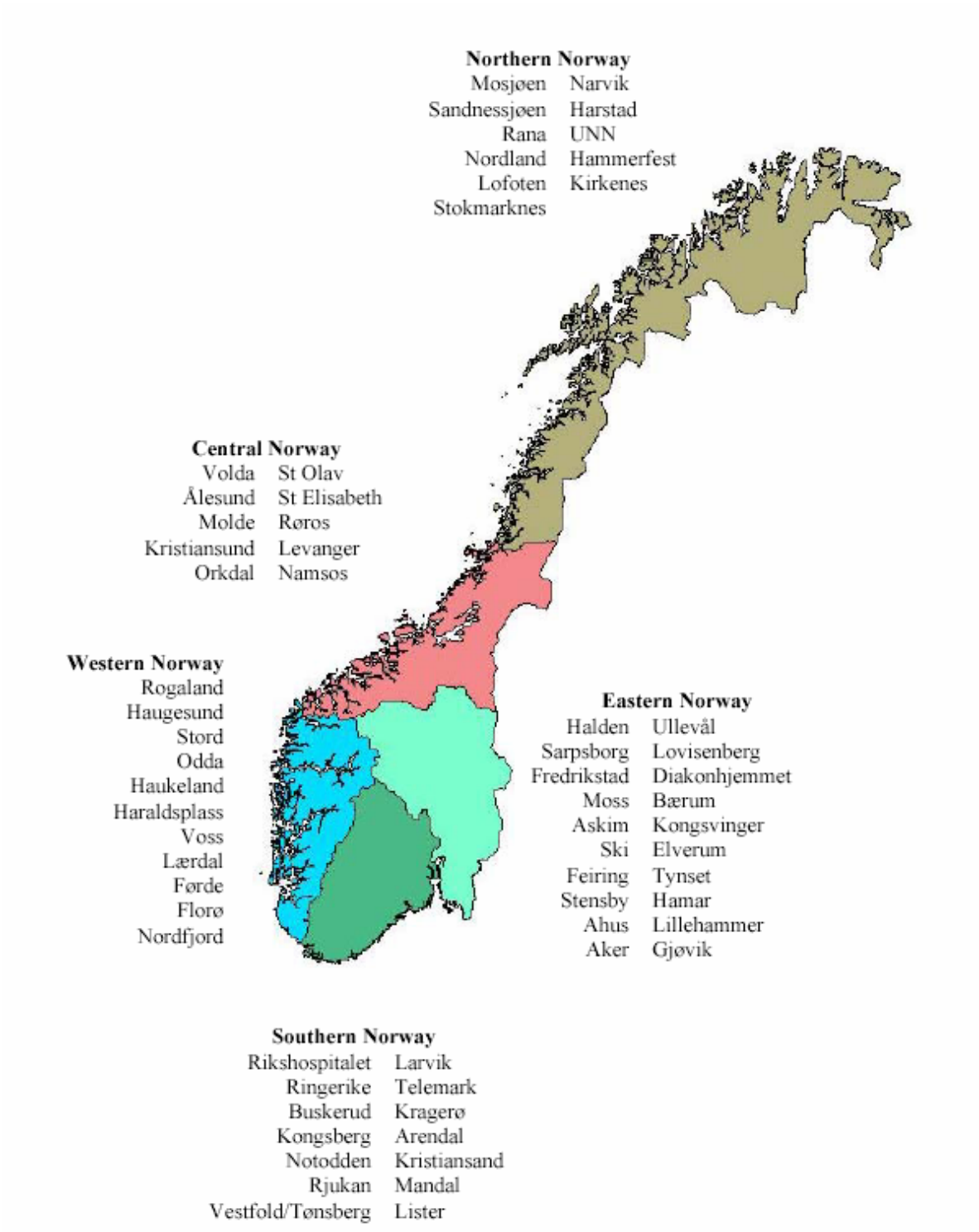
Transfers between hospitals. One disease episode may give rise to (different) index admissions at several hospitals if the patient is transferred. It may, in some cases, be difficult to be certain whether admission to another hospital is a new episode of the same disease. To keep track of chained admissions related to the same episode, we made a cut off point at admission to the second hospital within 24 hours from discharge from the first. This time period is regarded as transportation time. This choice, along with the definition of index admissions, have some consequences for the number of index admissions and the number of movements between hospitals, and called for a redefinition of the gap of time between the two hospitals.

- In some cases the patient is discharged from hospital A on Friday and admitted to hospital B on Monday. Such cases are registered as two index admissions, not as transferal between hospitals.
- Similarly, patients who have stayed at a hospital in December, stayed at home during Christmas, and were readmitted to the same hospital in January, will be registered with two index admissions.
- It seems that time of discharge is not always registered the moment the patient leaves the hospital, but may be registered several hours or even days later. Thus, apparently, the patient may be admitted to the second hospital before leaving the first. In this case, the times are redefined for consistency.

Hospital structure and interdepartmental transfers. This study intended to treat each hospital as a geographically located site, and not as a part of a bigger administrative unit. Some hospitals have geographically dispersed units within the same administrative unit. These units are considered wards of the mother hospital. Further complications arose:

- In some cases, data regarding transfers between hospitals that belong to the same administrative unit, but have different geographical location were inconsistent. We were not able to separate patients transferred from St. Olav to Røros or St. Elisabeth. For the two hospitals in question, these transferred cases were treated as belonging to St Olav.
- For the county of Vestfold, it has not been possible to distinguish data from hospitals in Horten and Sandefjord, from the Hospital of Vestfold. These hospitals were treated as one hospital.

Figure 5-1: Map of health regions and their corresponding hospitals included in this study.



5.3 DEFINITION OF VARIABLES

5.3.1 Diagnoses

The expert groups defined disease categories (defining index admissions) and codiagnosis disease categories, separately for each disease category. This resulted in some definitions, such as diabetes, differing between disease categories. For each index admission, codiagnoses were registered separately for the index admission and for all preadmissions since 1994.

The database was founded on the diagnostic codes used at discharge within the PAS. It would have been advantageous to know both the diagnosis at admission and the diagnosis at discharge, both for main diagnosis and codiagnoses. This would allow identification of any conditions that developed during the stay and that might reflect quality of care. For example, when pulmonary embolism occurs during the hospital stay, it may reflect an effect of the specific admission, and possibly quality of care.

5.3.1.1 Acute Myocardial Infarction

The definition of acute myocardial infarction has changed in later years (63;64;64). A proposed standardization plan has been presented for Europe (65). Small changes in definition can have large effects on the number of patients with the diagnosis of AMI. Blood levels of Troponine T or I (TnT, TnI) as a diagnostic factor can have different thresholds for diagnosis. Decreasing the TnT-threshold to 0.2, 0.1 or 0.03 µg/l has increased the number of diagnoses of AMI in Norway by 17%, 33% and 67% (66;67).

Diagnoses were selected using ICD-10-categorization. The disease classification system changed during the proposed measurement period from ICD-9 to ICD-10. The two systems are not directly compatible. The first year with ICD-10 in effect was 1999.

5.3.1.1.1 Main diagnosis

Table 5-3: ICD codes used in the definition of AMI index cases.

	Disease group	ICD-9	ICD-10
Index diagnosis	Acute heart attack	410	
	Transmural heart attack, 1. time		I21.0, I21.1, I21.2 and I21.3:
	Non-Q attack, 1 time		I21.4
	Unspecified heart attack, 1. time		I21.9
Not index diagnosis	Secondary acute heart attack, transmural or non-Q attack	410	I22.0, I22.1, I22.8 and I22.9:

Transmural 1. time attacks (ICD-10 codes I21.0-I21.3) are assumed to assure the greatest comparability over time and the best comparability between studies, because only these attacks are definitely or almost definitely connected to the ECG criteria for attack. ECG criteria are the only criteria that are unchanging over a longer time. Other

criteria, which are based on symptoms and different attack indicators, will vary considerably from one time point to another. The risk that the database will include first time attack for the same person several times is rather small. For these reasons, we included I21.0-I21.3 in an original data analysis. However, due to insufficient sample size, as these specifications were not possible under ICD-9, and because I21.4 and I21.9 comprised a large part of the total number of first time AMIs, we had to perform the final analyses with I21.0-I21.3 and I21.4 and I21.9.

5.3.1.1.2 Codiagnoses and previous admissions

Information about codiagnoses was collected together with the main diagnosis and also checked for in previous admissions.

The following codes for complications with heart attack were included as codiagnoses:

I23.0, I23.1, I23.2, I23.3, I23.4, I23.5, I23.6 and I23.8

In addition, we collected information on whether or not the following codiagnoses were present:

Table 5-4: The ICD-9 and ICD-10 codes used in defining pertinent codiagnoses. All listed disease groups were used as pertinent diagnoses in previous admissions. Those disease groups in italics indicate disease groups that may be considered as complications when diagnosed in the current admission, and thus omitted.

Pertinent codiagnoses, acute myocardial infarction		
	ICD-10	ICD-9
Diabetes mellitus	E 10 – E 14	250
Hypertension	I 10 – I 15	401 – 405
Angina pectoris and unspecified acute or sub acute ischaemic heart disease	I20, I 24.9	411, 413
Myocardial insufficiency	*I 50*	*428*
Stroke	I 61, I 63, I64	431 – 436
COPD	J 40 – J 44	490 – 496
Malignant tumors	C 00 – C 97	140 – 208
Peripheral arteriosclerosis	I 73	443

5.3.1.2 Stroke

5.3.1.2.1 Main diagnosis

Table 5-5: ICD codes used in the definition of stroke index cases.

	ICD-9	ICD-10
Index diagnosis	ICD-9 categories 431, 434, and 436 were included.	I61 Intracranial hemorrhage I63 Brain infarction I64 Stroke not specified as hemorrhage or infarction

Patients with transient ischemic attacks (TIA) were excluded from this study because practice differs for the hospitalization of these patients.

The ICD categorization changed from ICD-9 to ICD-10 in 1999.

Admissions due to rehabilitation only were not considered as index cases. However, not all hospitals coded rehabilitation consistently.

5.3.1.2.2 Codiagnoses

We collected information on whether or not the following codiagnoses were present either together with the main diagnosis or in previous admissions:

Table 5-6: The ICD-9 and ICD-10 codes used in defining pertinent codiagnoses. All listed disease groups were used as pertinent diagnoses in previous admissions. Those disease groups in italics indicate disease groups that may be considered as complications when diagnosed in the current admission, and thus omitted.

Pertinent codiagnoses, stroke		
	ICD-10	ICD-9
Stroke related case/sequelaes	*I69.1, I69.3 and I69.4*	*438*
Diabetes mellitus	E 10 – E 14	250
Hypertension	I 10 – I 15	401 – 405
Angina pectoris and Unspecified acute or sub acute ischaemic heart disease	I 20, I 24.9	411. 413
Atrial fibrillation	I 48	427.3
Myocardial insufficiency	I 50	428
COPD	J 40 – J 44	490 – 496
Malignant tumors	C 00 – C 97	140 – 208
Peripheral arthrosclerosis	*I 73*	*443*
Deep venous thrombosis	*I80*	*451.1, 451.2*
Urinary infection	*N30 N10-N12, N20.9*	*S950 and S959*
Pneumonia	*J12-J18, J22, J69.0*	*480.9, 481-486, 487*
Pulmonary embolism	*I26*	*415*
Dementia	F00, F01, F02, F03, G30, G310, G312, G319, G328, G910, G912-G919, G937	290, 294.1, 331, 3489, 3498
Fracture	S00-S99	800-904, 910-928, 950-957, 959

5.3.1.3 Hip Fracture

5.3.1.3.1 Main diagnoses

Table 5-7: ICD codes used in the definition of hip fracture index cases.

	ICD-9	ICD-10
Index diagnoses	820 with all subgroups	S72.0, S72.1 and S72.2

All patients with medial fractures (femural neck fractures) were included, together with pertrochanteric as well as subtrochanteric fractures. We did not exclude patients with main or secondary diagnosis of cancer.

We recorded all procedural codes and diagnostic codes for the included patients.

The shift from ICD-9 to ICD-10 during the period under study did not lead to categorization problems for hip fracture. Patients diagnosed with the above-mentioned main categories were selected, both when the diagnoses were assigned as main diagnoses and when they were assigned as codiagnoses. This was especially important for hip fracture, since often the patients develop complications that become the actual cause of death (for example, pneumonia), and possibly, in some cases, receive the complication as main diagnosis code.

5.3.1.3.2 Codiagnoses

We collected information on whether or not the following codiagnoses were present either together with the main diagnosis or in previous admissions:

Table 5-8: The ICD-9 and ICD-10 codes used in defining pertinent codiagnoses. All listed disease groups were used as pertinent diagnoses in previous admissions. Those disease groups in italics indicate disease groups that may be considered as complications when diagnosed in the current admission, and thus omitted.

Pertinent codiagnoses, hip fracture		
	ICD-10	ICD-9
Diabetes mellitus	E 10 – E 14	250
Hypertension	I 10 – I 15	401 – 405
Myocardial infarction, angina pectoris	I 20 – I 25	410 – 414
Pulmonary embolism	I 26	415
Atrial fibrillation	I 48	427.3
Myocardial insufficiency	I 50	428
Stroke	I 61, I 63, I64	431 – 436
Deep venous thrombosis	*I 80*	*451.1, 451.2*
Pneumonia	*J12-J18, J22, j69.0*	*480.9, 481 – 486, 487.0*
COPD	J 40 – J 44	490 – 496
Malignant tumors	C 00 – C 97	140 – 208
Urinary tract infection	T85.7, O08.8 N39.0	599.0

Pertinent codiagnoses, hip fracture		
	ICD-10	ICD-9
Cystitis	N30	595
Pyelonephritis	N10-N12, N20.9	590
Urinary retention	R33	788.2
Complications from orthopedic implant	T84	996.4

5.3.1.4 Socio-demographic factors

Socio-demographic data was collected from basic registers available in Norway concerning the individuals in this study. This information was collected by Statistics Norway. The variables collected were sex, age, number of years of education, income and property/capital for the patients themselves, and years of education, income and property/capital for the spouse.

In this study, for patients living as a couple where both members are still alive:

- joint income was estimated using (sum of incomes)/1.7, in accordance with the definition used by Statistics Norway to correct for that when two people live together they have fewer expenses than living separately;
- joint property/capital was estimated as the sum of the individual property/capitals, and
- joint education was the number of years of education of the member of the couple that was highest.

The Statistics Norway calculated road distances between home and hospital for each of the admissions. Possible extra delays caused by ferries were not considered. In addition, no correction could be made for possible transport by helicopter or plane. This variable was used as a proxy for time from symptom start. Country of birth was used in the absence of data about race or ethnicity.

Some of the variables from Statistics Norway had rather extreme outliers: ”distance in km to hospital”, ”total income for the household”, and ”total property/capital for the household”. To reduce the influence of the outlying data points on the model, these variables were transformed by the logarithmic transformation formula $\ln(x+1)$.

5.3.1.5 Disease severity

Disease severity was estimated using the proxy staging which is described in Appendix 2 (9.2). This study aimed at testing if staging is a successful method to control for disease severity. However, the method applied has only been described and developed using ICD-9 classification. In order to apply this to ICD-10 we needed to translate the coding. This resulted in definitions proposed by the expert groups for stroke, acute myocardial infarction and hip fracture. For stroke, the expert group did not find the staging system satisfactory and chose instead a binary variable classifying stroke case as being a hemorrhage or an infarction.

Patient administrative systems do not have information about severity. Hence, medical records at 15 hospitals were examined to collect data about stroke case-mix. Random selections of patients with index diagnosis stroke were selected from five small county hospitals, five large county hospitals and five large hospitals. A neurological score was

used as a tool to calculate the severity of the stroke for each patient. We used the Canadian Stroke Scale that previously has been validated for retrospective assessment of stroke severity (68;69). Range of the score is 1.5-11. Lower score means more severe strokes.

5.3.1.6 Patient Frailty

Codiagnoses were used as proxies for patient frailty. The total number of codiagnoses listed in the tables of codiagnoses for each disease category was summed up for all previous admissions and the current admission. The sum included only one occurrence of a disease group (e.g. diabetes). If present both previously and currently it would only be counted once. In addition, the disease groups that are italicized in Table 5-4, Table 5-6 and Table 5-8 can be considered as possible results of treatment and are accordingly removed from the current admission. As a complement to using a large number of single covariates, we constructed the following variables, which were eventually the ones used in the final statistical model:

- total number of codiagnoses since 1994;
- total number of pertinent codiagnoses since 1994, but before the index admission (specific for AMI (page 40), stroke (page 41), and hip fracture (page 42); the number was truncated to a maximum of six and weighted inversely by the number of years of registration (i.e. since 1994); Pertinent diagnoses are those codiagnoses identified by the expert groups and listed in tables for each disease category;
- total number of pertinent codiagnoses since 1994 up to and including the index admission, excluding complications during the index admission (This variable was later excluded from the analysis);
- total number of previous admissions since 1994. The number was truncated to a maximum of 25 and weighted inversely by the number of years of registration (i.e. since 1994).

The weighting was introduced because the late admissions would have a much longer period to accumulate previous admissions and diagnoses than admissions from the start of the observation period. Truncation was introduced to limit the influence of the (very few) cases with very high values.

All codiagnoses and earlier admissions since 1994 pertained only to the same hospital as the index admission.

5.3.2 Additional information about hospitals

We collected additional information about hospital routines and characteristics with a questionnaire sent to the hospital management, which was aimed at mapping structural and procedural matters. Questions included admission routines, available expertise, treatment procedures, and more. Mainly because of poor response rate, these data were not used in the primary analysis.

5.3.2.1 Mortality

Date and cause of death were collected from the Norwegian Causes of Death Register. We calculated the number of days from the date of index admission to date of death. If a

patient was transferred between hospitals, the date of admission to the first hospital was used for calculating the number of days until death occurred.

5.4 METHODS OF RISK ADJUSTMENT

The principle of risk adjustment in this report is to build a statistical regression model describing how mortality is influenced by both hospital and the risk adjustment variables. In such a model, the partial effect of the hospital variable describes just the effect of hospital when all other variables are held constant. It follows that only variables that are determined beforehand can be used for risk adjustment.

Thus, risk adjustment is a corollary to the statistical model building. It is immaterial, in principle at least, whether the hospitals are well matched with respect to the risk variables or not.

5.4.1 Logistic regression

The available statistical modeling approaches differ in the choice of mortality measure. Using survival analysis techniques, e.g. Cox proportional hazard analysis, it would be possible to use survival time as the outcome variable instead of just the binary variable survival/death within the specified 30 days period.

It was felt that the possible gain in statistical precision from using survival time did not warrant the loss in model robustness resulting from the more fragile assumptions underlying the Cox proportional hazards model. A logistic regression model was therefore chosen. This conclusion is tied to the relatively short observation period. With a longer period, survival analysis should be considered. Another advantage would be the possibility to treat in a more sophisticated way variables that change during the episode of hospitalization, such as transferring of patients from one hospital to another.

The model contained a moderate number of parameters describing the relative performance of hospitals. Conceivably, these parameters could be viewed as resulting from a random process and modeled as such. In multilevel analysis, this is a common approach.

The choice between models with random or deterministic effects is well known in statistical practice and literature. The decisive factor is the nature of the decision problem facing the statistician. As explained below, the main problem is hypothesis testing about the actual, realized effect parameters. The hypothetical process determining the relative performances of Norwegian hospitals is not relevant as such. Accordingly, the fixed parameter approach is appropriate and is used here. Another reason for fixed parameters is that available software is more practically useful with the large datasets involved.

However, we also have an implicit secondary aim, namely that of giving a correct, overall description of the set of hospitals. Here, the random parameter approach will lead to effect estimates that are biased (towards the overall mean) but with lower overall expected mean squared error. Under this assumption, the estimates can be improved, on the average, by what is known as an empirical Bayes or shrinkage technique. Individual estimates will be pulled in towards the general mean, more so when their sampling variances are large. However, a large sampling error is indistinguishable from a true,

underlying extreme value. This explains why the random parameter model is usually viewed as inappropriate for making inferences about single parameters.

To address properly all the tasks, we chose a two-stage approach: Firstly, we use logistic regression for model building and inference about individual hospital parameters. Secondly, the raw estimates are modeled as random samples from a hypothetical population. The statistical assumptions made in the second stage are not very critical because of the descriptive objective. In particular, the use of a probability distribution for the true effects can be viewed merely as a pragmatic averaging device.

The R statistical system (70) was used for the final statistical modeling. Informal comparisons were made against the SAS system, with parameter estimates from the two systems agreeing to within 0.001 of another.

5.4.2 Hospital effect parameters

Our discussion will be centered on the hospital effect parameters or betas. Formally, they are defined as follows:

The model is

$$\log\left(p_{ij}/\left(1-p_{ij}\right)\right)=\alpha+\beta_i+\gamma' x_{ij}\ ,$$

where p_{ij} is the probability of death within 30 days for the j-th case from hospital i, having covariate vector x_{ij}. The *hospital effects* are the β_i-parameters. They are standardized by the condition

$$\sum_i \beta_i = 0\ ,$$

meaning that β_i measures the effect of hospital relative to the average over all the hospitals. When we speak of e.g. "average performance", we thus mean the average with each hospital counted as one. The relative sizes of hospitals do not enter into this relation. This was felt more appropriate for discussing the population of hospitals than the alternative of patient-weighted averages and parameter standardization. The risk variables we want to adjust for are used as covariates.

5.4.3 Modeling yearly variation

In the statistical analyses, we use data for all available years. The objective, however, is to assess the hospitals' relative performance in the last year of the evaluation period, in this case 2001. This corresponds to the intended use of the quality indicators: based on data up to and including the current year, evaluate the current performance. We expect yearly variation in mortality, so that calendar time must be included in some form as a covariate in the model. In a straightforward logistic regression formulation, we would end up estimating the time average of the hospital effects. If a hospital's performance has a time evolution different from the rest, which a priori is likely to occur, the average effect will obviously be misleading as a measure of current performance.

Accordingly, we formulate the model such that β_i becomes the effect of hospital i in the last year of the period (2001). If "Yeardiff" denotes time in years relative to 2001, we include as covariates Yeardiff, Yeardiff squared and the statistical interaction term between hospital and Yeardiff.

We thus assume that mortalities for all hospitals, on a logistic scale, follow a second order polynomial, with possibly different slopes, and a possibly different intercept in 2001.

Adding a large number of interaction parameters may be regarded as undesirable from a variability point of view. On the balance, it was felt that avoiding bias in the joint effect of time and hospital was the more important concern.

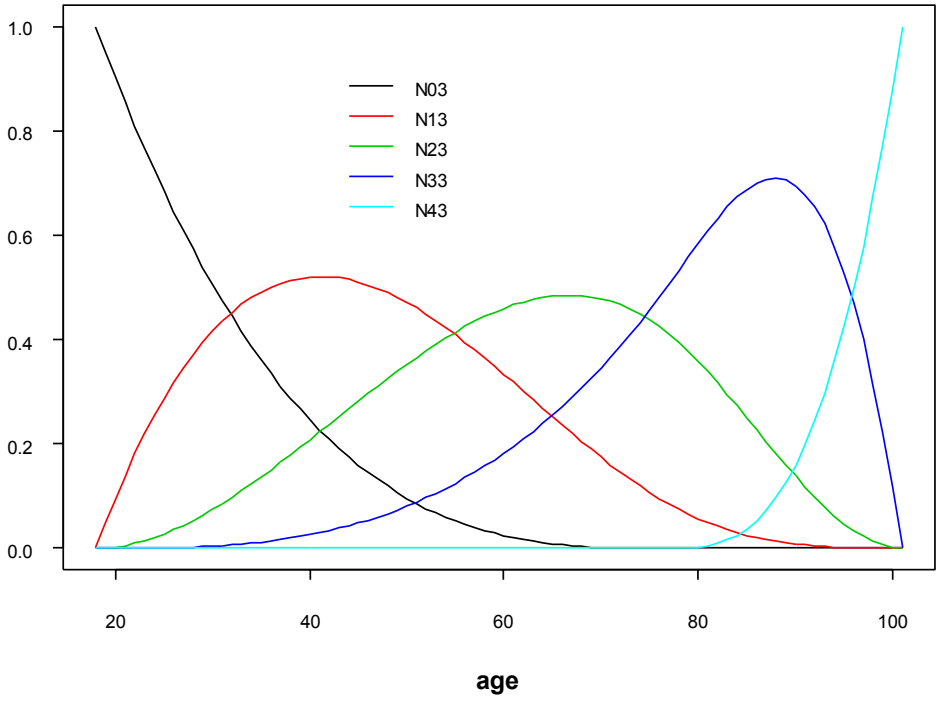
5.4.4 Case Weighting

The moving of patients between hospitals is a problem for the statistical analysis. As explained earlier, each stay is counted as a separate admission, as long as the patient can be assumed to move directly from one hospital to the next. A patient that has been moved will then count for two non-moved in the analysis. To avoid this, we have used a weighted analysis with the relative proportion of the stay at the hospital in question as the case weight.

5.4.5 Splines for flexible fitting of smooth functions

The models require a specification of the functional dependence of log-odds for mortality as a function of age. Since age has a strong effect on mortality, it is important that this functional form is sufficiently flexible so as not to introduce any bias in the model, yet depend on few parameters that have to be estimated. In line with common practice (see e.g. (71) or (72)), we model age dependency by splines. These are piecewise cubic polynomials, with interval boundaries at pre-specified points, the so-called knots. Given enough knots, an arbitrary function can be approximated as closely as desired. Statistical techniques exist for choosing the "best" number of knots. This was not deemed necessary in the present situation. Instead, the knots were fixed a priori. Tests were performed, however, to see if further refinements were necessary. For each disease, the knots were put at the total age intervals and at the median age. For this knot sequence, every spline function can be written as a linear combination of fixed, so-called B-spline basis functions, denoted here by *N03, N13, ... N43*. As these total to one, only the first four are included as covariates in the model. The plot below shows the basis functions for the stroke model. The figure gives an impression of the degree of smoothness and flexibility that can be obtained by combining the basis functions with different weights.

Figure 5-2 : B-spline basis functions, stroke model.



5.5 MODEL BUILDING AND SELECTION STRATEGY

The model building was initiated with a prior model including the explanatory variables listed above, as well as a full set of hospital/year interactions, with year as a categorical variable, and second order terms in frailty/ severity variables.

Starting with the interactions, a stepwise variable exclusion procedure was followed, with likelihood ratio (deviance) tests at a nominal 5% level, until no simpler model could be found (For the sake of uniformity between the three disease models, the whole set of socio-demographic variables was considered as a block with respect to inclusion/ exclusion). This phase was performed first on a reduced data set consisting only of the ten largest hospitals, representing about half of the total number of index admissions. The reason for this was mainly that the sparse data patterns of the smaller hospitals made testing for interactions between hospital and calendar year difficult or impossible, because of model colinearity or by the high variability resulting from the large number of parameters relative to sample size.

At this stage, goodness-of-fit checks on the model were performed, and they were modified in a more informal way, by adding or removing variables that were judged interesting. The resulting models were then applied to the full data sets and checked in the same way. To test the adequacy of the spline functions, they were compared with 3- and 4-interval splines using the AIC criterion.

The final models were checked for outliers and highly influential observations as explained in e.g. (73). We used the following two regression diagnostic tools: Cook's distance and leverage or hat statistic.

The Cook's distance diagnostic for an observation is the standardized difference in the vector of parameter estimates due to deleting the observation, and it can be used to assess the effect of an individual observation on the fitted model.

The leverage or hat statistic for an observation is the weight of that observation in the predicted value (on the logit scale) for that observation. It indicates the extremeness of an observation in the space of all the predictor variables. The hat values sum to the number of unknown parameters in the model. Ideally, they should all be close to their mean value. It is common practice to standardize by dividing by the number of observations. In our case, when each hospital can be regarded as having its own intercept, we reason by analogy and standardize by dividing by the number of cases for the hospital in question.

5.6 ASSESSMENT OF INDIVIDUAL HOSPITALS

5.6.1 The three classes of decision-makers

Using quality indicators can be thought of, in a general sense, as making decisions based on inference about relative hospital performance. In statistical inference, we must balance the different error probabilities according to the particular decision task involved. There is the type I error of wrongly stating that a hospital is better or worse than average, and the type II error of not detecting that a hospital is different from the average. With a large number of hospitals, we also have a multiple testing situation where the overall error probability is much greater than the error probabilities for testing individual hypotheses. In our case, we can envision three different types of decision-makers, each with their own objectives and consequences from inferential errors:

A. Those primarily interested in one specific hospital. They include the management and staff of this hospital, local doctors and members of the public in the hospitals natural (geographic) intake area. Their decisions may be e.g. to choose a hospital, to increase staffing in critical areas or to make organizational changes or review treatment routines. This group is testing the hypothesis that a single beta is zero, and must choose significance levels on that basis with a view to the sample size involved.
B. Authorities and regional hospital owners. This group must regularly make simultaneous decisions (e.g. increasing funding or making management changes for poorly performing hospitals, or making hospitals within a region more specialized) about all hospitals or a large group of hospitals. This leads to a possibly large overall type I error probability. On the other hand, being too conservative means that poor performance will go undetected. Thus, there is a delicate balance between the interest of hospital management and staff on one hand, and the public on the other hand. This dilemma can only be resolved by reformulating the decision problem along the lines described below.
C. Those responsible for publishing the hypothetical annual performance review, and must state clearly which hospitals are worse (or better) than average. They are concerned with the overall error probability. This is a standard multiple testing problem.

It will be realized that there is a natural ordering of the groups when it comes to conservativeness in inference (i.e. protecting against type I errors). Testing your local

hospital with 5% level means on the average one error in twenty years, unless it is a bad performer, whereas the same rule for all Norwegian hospitals will lead to perhaps three or four errors each year.

5.6.2 Decision rules

The statistical methods for the three decision problems outlined above are by and large given from the structure of these problems. It is, however, outside the scope of the present investigation to establish and fix the parameters of all decision rules. This would entail careful consideration of the consequences of possible actions and how these consequences should be compared (In decision-theoretic terms, this is called establishment of a loss function). These considerations must ultimately be the responsibility of the actual policy makers, and may depart from the judgments that we have used as a basis for our proposals and recommendations. We do not expect this to result in any major changes.

For decision perspective (A) above, we propose that ordinary significance tests be used, supplemented by confidence intervals if necessary.

For multiple testing (perspective (C) above) we will use studentized multiple contrast tests as described in (74). A hospital effect is declared greater than zero, with multiple P-value p, if its standardized estimate (z value) is greater than the upper p-fractile in a certain null distribution, and analogously for "smaller than zero". This null distribution is the distribution of the maximum of the z-values when all hospital effects are zero.

To resolve the dilemma of decision-makers from perspective (B) described earlier, we will propose the following, pragmatic decision rule:

The objective is to make a shortlist, or follow-up list, of hospitals. A hospital is entered on the list if its beta is significantly different from zero,

Being on the shortlist can obviously not be taken to mean that a hospital is performing badly (or particularly well). Indeed, we know that the list will include some smaller hospitals due to random estimation variability. This is, however, inevitable if we want a reasonable chance of detecting poor performance among the smaller hospitals, as will be demonstrated below. One can think of follow-up as meaning the collection of extra data, reviewing the practices or perhaps pooling data for several years or even several different diseases.

In hypothesis testing, we regard one type of error as being much less desirable than the other type. As described below, we will construct our decision rule on this basis. In practice, this means choosing significance level chosen according to the number of admissions at that hospital. An alternative, that we have not considered here but would fit easily within our framework, would be to use longer time-periods for the smallest hospitals. This would entail a different trade-off between precision and bias.

Though it may not be uncontroversial, we will regard all individual tests as three-decision problems (in the sense of Tukey (75)). In practice, this means that we are only concerned with one-sided error probabilities and view two-sided testing problems as two separate one-sided tests. In the case where the null hypothesis is exactly fulfilled, the stated significance level will be wrong, but it is not easy to see how this would come to be in the first place.

Many statisticians will argue that this is the correct testing procedure in general, but in our case, we feel that the use of a slightly less conservative method is well justified in view of the consequences of not being able to draw conclusions about hospitals with poor performance. It should be noted that as one of the results of this study, we propose significance levels that are lower than the conventional 5% in most cases, meaning that the issue will be largely inconsequential (see 7.2.2.4).

This description of the decision problem is somewhat simplified, and we expect further refinements before the procedure is made operational. One obvious lack in sophistication is that we have not considered the size of hospital effects. Having determined that a hospital is significantly different from the rest is usually taken as the first step in the analysis, the next step being to determine the size of the difference. Confidence intervals can be used for this purpose. However, a situation may arise where we have sufficient information for hypothesis testing, but not enough for producing meaningfully short confidence intervals. We will return to this question below.

5.6.2.1 Indifference and alert limits

For the discussion of precision of statistical procedures, we need to decide on the importance we attach to different parameter sizes. In this, there is an inescapable element of judgment. Anyone who wishes to depart significantly from our choices may have to review our findings in this light.

A hospital effect β_i of $\log(1.2) = 0.182$ means that a patient admitted at hospital has a 20% increased log-odds for dying within 30 days, relative to the average of all hospitals. Arguably, an increased risk of this magnitude is not important in practice. In this report, we will regard this value as the largest hospital effect that can reasonably be ignored from a decision-making perspective, and refer to this value by the term *indifference limit*.

A β_i of $\log(2) = 0.693$ means a 100% increase in log-odds of dying for a patient admitted at hospital i, relative to the average. In our view, this would be a large increase in risk, with great consequences for the patient in question. We will refer to this value as the *alert limit*: a large hospital effect that our evaluation procedure should be able to detect with high probability.

For simplicity, we will also use the same terms for the negative values, i.e. representing better performance than average, when there is no danger of confusion. By *indifference interval,* we will mean the interval (-0.182, 0.182).

In other studies (76), a log-odds value of 1.3 has been used as a threshold value for "higher than acceptable mortality". This corresponds to a β_i of 0.262.

5.6.3 Empirical power functions

In order to balance type I and type II error probabilities, we need to know the power function of the tests for individual hospital effects. This power function will depend on all the actual covariate values in the data set, not only those associated with the admissions at that particular hospital, and are obviously a very complicated, unknown function of all the model parameters. It is in the nature of things that these covariates cannot be known beforehand, and some sort of probability averaging process must be involved to make a suitable power function. This is eased by the fact that the probability

distribution of the admission variables can be assumed to be a rather stable entity, so that historical data can be used.

To design the pragmatic decision rule, we will use the benefits of hindsight to construct an empirical power function. We assume that the probability of stating that a particular hospital beta is greater (smaller) than zero is a function only of the estimation variance for that beta and the true beta value. The estimation variance is assumed to depend only upon the number of admissions for that hospital. We also assume that the estimation variance is inversely proportional to the number of cases. In essence, we construct an average of all power functions for each number of cases. Note that the number of cases is taken to mean the total number of admissions over the period for the particular hospital.

It may be argued that this is an overly simplified way of determining a power function. In our view, the most delicate assumption is that the year-to-year distribution of cases remains sufficiently stable. However, it is not very critical.

5.6.4 Descriptive assessment of hospitals

As explained above, the second stage of the analysis is based on the estimated hospital effects and their estimated covariance matrix. It was assumed that the underlying, unobservable distribution of true betas could be represented as a zero-mean, finite mixture of normal distributions. Disregarding the covariances between betas from different hospitals, assuming the estimation variances as known, maximum likelihood estimates were calculated.

It must be noted here that for all three diseases, the resulting distribution only had one component. Thus, we would have obtained the same results by making the more standard assumption that the true betas were normally distributed with zero mean.

Based on the estimated, underlying distribution and the assumed known estimation variances, second-stage shrinkage estimates of the betas were computed as the conditional expectation of the true betas given the estimates. As an estimate of the true variability of the actually realized hospital betas, the variance of the shrinkage beta estimates can be used. Such procedures are described in e.g. (77).

In general, we may expect shrinkage estimates to be biased in the direction of the overall mean. They will therefore tend to have more concentrated distributions than the true effects. For making formal inference about e.g. the maximum of the hospital effects, other methods (multiple comparison tests) would be more appropriate. However, a small simulation experiment was carried out to give an indication of the amount of bias in the present case. Assuming that the actually observed shrinkage estimates were the true hospital effects, 10000 samples (for each disease category) of raw estimates were drawn from (normal) distributions with the estimated standard deviations. For each sample of raw estimates, a new sample of shrinkage estimates was computed, and its maximum and range recorded. This set of hypothetical samples is assumed to represent the most important part of the true variability in the estimation process leading up to our shrinkage estimates. In the simulated set, the percentages of sample maxima and sample ranges exceeding the observed values were found.

5.6.5 Bias and model robustness

To study the robustness of our models and the stability of the results, we will need to compare results obtained under different models and possibly different data sets. By sensitivity, we will mean the sensitivity of the estimated hospital effects. The change from one model to another will be displayed graphically as a scatter plot of the two sets of beta estimates. In addition, we will compute Pearson and Spearman correlation coefficients as well as the average absolute difference

$$AAD = ave\left|\beta_i^{(1)} - \beta_i^{(2)}\right|$$

and the relative average absolute difference

$$RAAD = \frac{ave\left|\beta_i^{(1)} - \beta_i^{(2)}\right|}{\frac{1}{2}\left(ave\left|\beta_i^{(1)}\right| + ave\left|\beta_i^{(2)}\right|\right)}.$$

In their comprehensive evaluation of quality indicators Davies et al. (1) use a statistical approach which is somewhat different from ours. Firstly, they use linear regression instead of logistic regression. Linear regression is a less satisfactory approach, but was necessitated by the very large data sets involved.

A large number of indicators are studied in Davies et al. (1), making it meaningful to use multivariate techniques to exploit any correlation structure as may exist between the indicators. They are able to reduce estimation noise in individual effect estimates by, in a certain sense, pooling information from correlated indicators. In our case, with only 3 indicators, this is not feasible.

Also, they use a different model for smoothing estimates over time. Their model for time variation assumes a random, but highly correlated, time trajectory for each hospital effect. Though this is formally very different from our model, the two approaches can be regarded as different ways of reaching the same end, namely to remove the random component of the yearly variation.

Because of the differences in data and in models, our results are not directly comparable to those of Davies et al. (1). One would expect, however, that the numbers are of roughly the same magnitude after adjustments as described below. For making informal comparisons, we will refer to their reported values of the following measures:

- Signal Standard Deviation (denoted SSD below). This is a measure of precision that corresponds to the population standard deviation of the shrinkage beta estimates.
- Average Absolute Value of Change Relative to Mean (denoted AACRM below). This is a measure of bias and corresponds to our AAD.

For conversion, we will use the following approximate formulas:

$$\sigma_{S,P} = SSD / p_0(1 - p_0)$$

$$AAD = AACRM / (1 - p_0)$$

Here $\sigma_{S,P}$ denotes the population standard deviation of the shrinkage estimates. We have made use of the assumption that for moderate hospital effects, the effects in the two models are related by the approximate linear relation

$$\theta_i \approx p_0(1-p_0)\beta_i = a\beta_i ,$$

where θ_i denotes effect in the linear model, β_i in the logistic model and p_0 is the mean probability of death.

5.7 QUALITY CONTROL OF DATA

In order to verify the quality of the collected data, a manual data collection study was designed and conducted.

Within each hospital (45 hospitals participated), 50 patients from each disease group were randomly selected from the hospital's patient population for the period 1997-2001. A single A4 printout of a selection of the data gathered, was given to the hospital for comparison with the paper record, regarding: Date and time of admission, date and time of discharge, date of re-admission, index diagnosis, main diagnosis, codiagnoses, procedures connected to the main and index diagnoses. The control is only a verification that the electronic journals and paper records agree on codes used. There were no attempts to check that the codes were clinically correct.

Age and sex were used as control variables to confirm the correctness of the identity of the patient on the form generated from the research data set. However, the identity of the patient was removed to satisfy Norwegian regulations.

Some hospitals do not use paper records anymore. In these hospitals, the A4 printout was compared with the hospital's electronic medical record. A qualified doctor working in the hospital performed the comparison. The doctors were instructed to record the "correct" information on the A4 printout when they found discrepancies.

In addition, at 15 randomly chosen hospitals (stratified according to type of hospital) one medical doctor independently went through all 150 records in the same manner as the hospital had done. The hospitals were not told that they were selected for double checking, until right before the doctor arrived.

6. RESULTS

6.1 DISTRIBUTIONS OF ADMISSIONS AND DEATHS

All figures in this chapter relate to admissions, not to patients. We will sometimes refer to admissions as cases. Because some patients have been admitted to more than one hospital or to the same hospitals in different calendar years, case mortality is lower than patient mortality.

Before the application of exclusion criteria (see 6.5), the number of admissions for *acute myocardial infarction (AMI)* at each hospital, over the 5 year period, ranged from 7 to 2834 with a median of 558 admissions in the 66 hospitals. The numbers of deaths ranged from 2 to 541 with a median of 106.5. After exclusion criteria there were 59 hospitals included in the analyses. In these hospitals, the number of admissions ranged from 151 to 2714 with a median of 607, while the number of deaths ranged from 4 to 522 with a median of 123.

Before the application of exclusion criteria, the number of admissions for *stroke* at each hospital, over the 5 year period, ranged from 9 to 2939 with a median of 618 admissions in 64 hospitals (Two hospitals did not have admissions for stroke). The numbers of deaths ranged from 4 to 441 with a median of 105. In the 59 hospitals included in the analyses, the number of admissions ranged from 107 to 2746 with a median of 681, while the number of deaths ranged from 17 to 435 with a median of 121.

Before the application of selection criteria, the number of admissions for *hip fracture* at each hospital, over the 5 year period, ranged from 1 to 2908 with a median of 484 admissions in 65 hospitals (One hospital did not have admissions for hip fracture). The numbers of deaths ranged from 1 to 211 with a median of 36. In the 57 hospitals included in the analyses, the number of admissions ranged from 111 to 2520 with a median of 472, while the number of deaths ranged from 2 to 193 with a median of 39.

6.2 TOTAL 30-DAY CASE MORTALITY

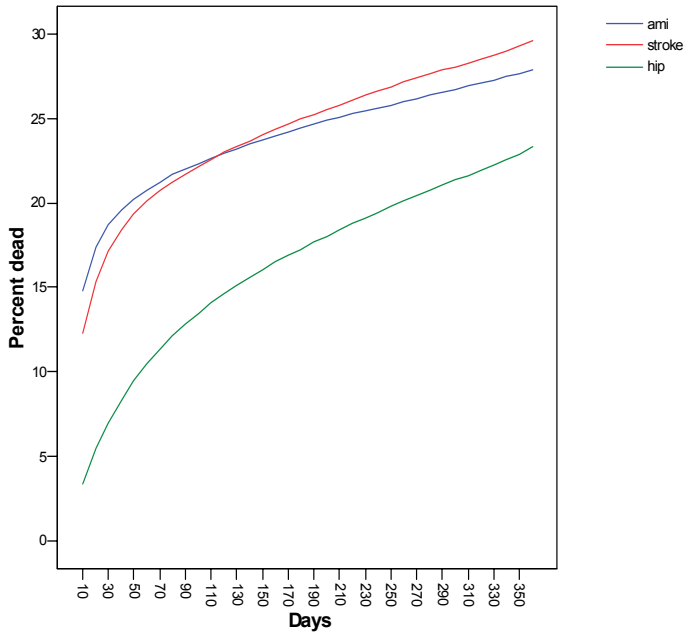
The overall probability of death before 30 days, before exclusion criteria were applied, was 18.7% for acute myocardial infarction, 17.2% for stroke and 6.9% for hip fracture.

Table 6-1 summarizes the number of deaths and case mortality within the different time frames considered in the study.

Table 6-1: Number and percent mortality within the various time frames for each of the three major disease categories. Counts are index admissions. Number of admissions: AMI 54095, Stroke 53072, Hip fracture 50205.

Days after admission	Number of AMI cases dead	Percent of AMI cases dead	SD	Number of stroke cases dead	Percent of stroke cases dead	SD	Number of hip fracture cases dead	Percent of hip fracture cases dead	SD
10	8015	14.8	.36	6518	12.3	.33	1686	3.4	.18
20	9413	17.4	.38	8161	15.4	.36	2742	5.5	.23
30	10132	18.7	.39	9122	17.2	.38	3489	6.9	.25
60	11224	20.7	.41	10676	20.1	.40	5261	10.5	.31
90	11899	22.0	.41	11525	21.7	.41	6465	12.9	.33
120	12407	22.9	.42	12207	23.0	.42	7339	14.6	.35
365	15102	27.9	.45	15730	29.6	.46	11732	23.4	.42

Figure 6-1: Case mortality in 10 day periods up to one year for each of the three disease categories.



6.3 DESCRIPTION OF EXPLANATORY VARIABLES

6.3.1 Socio-demographic variables

6.3.1.1 Age, sex and marital status

The total population is fairly evenly distributed between the sexes, but within each diagnosis there are large differences. Males predominate in admissions for AMI whereas females predominate in admissions for hip fracture. Sex is fairly evenly distributed in stroke patients (see Table 6-2). Similarly, the age distribution between the

disease categories differs substantially. Patients with hip fracture are older, with 93.4% over 60 years whereas only 77.0% of AMI patients are over 60 (Figure 6-2).

Figure 6-2: Age distribution according to disease category.

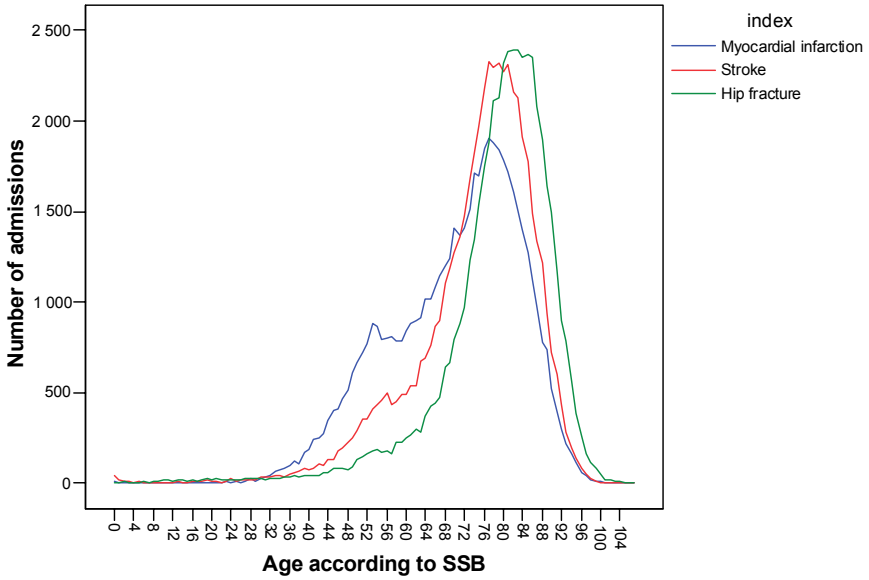


Table 6-2: Distribution of index admissions by disease according to sex and age groups before application of exclusion criteria. n=number of cases.

Sex	Acute myocardial infarction		Stroke		Hip fracture	
	n	Percent	n	Percent	n	Percent
Male	33730	62.4	26674	50.3	14084	28.1
Female	20365	37.6	26398	49.7	36121	71.9
Table Total	54095	100.0	53072	100.0	50205	100.0

Age	Acute myocardial infarction		Stroke		Hip fracture	
	n	Percent	n	Percent	n	Percent
0-9	14	.0	113	.2	82	.2
10-19	9	.0	80	.2	181	.4
20-29	91	.2	165	.3	215	.4
30-39	822	1.5	503	.9	316	.6
40-49	3690	6.8	1475	2.8	671	1.3
50-59	7876	14.6	4172	7.9	1775	3.5
60-69	10239	18.9	7757	14.6	4126	8.2
70-79	16587	30.7	18709	35.3	14632	29.1
80-89	12906	23.9	17532	33.0	22170	44.2
90-99	1842	3.4	2558	4.8	5923	11.8
100 or more	19	.0	8	.0	114	.2
Table Total	54095	100.0	53072	100.0	50205	100.0

As can be seen in Table 6-3, the distribution of both the patients admitted and those that have died of one of the three main disease categories according to marital status varies between the disease categories. These differences are

probably associated with the differences in the population with respect to sex and age.

Table 6-3: Distribution of index admissions by disease according to marital status, before the application of exclusion criteria.

Marital status	Acute myocardial infarction		Stroke		Hip fracture	
	n	Percent	n	Percent	n	Percent
Unknown	79	.1	92	.2	45	.1
Married/ cohabitant	30094	55.6	25184	47.5	15659	31.2
Not married	4637	8.6	5258	9.9	6044	12.0
Divorced or separated	5068	9.4	4351	8.2	3352	6.7
Widowed	14217	26.3	18187	34.3	25105	50.0
Table Total	54095	100.0	53072	100.0	50205	100.0

6.3.1.2 Geographic region of birth and habitation

The distributions of the population according to country of birth (used as a proxy for ethnicity) are shown in Table 6-4.

Table 6-4: Distribution of index admissions by disease according to region of birth.

	Myocardial infarction		Stroke		Hip fracture	
	Count	Percent	Count	Percent	Count	Percent
Norway	52228	96.7	51437	97.1	49157	98.0
Nordic countries	555	1.0	547	1.0	390	.8
Rest of Western Europe, North America and Oceania	504	.9	450	.8	410	.8
Eastern Europe	235	.4	234	.4	96	.2
Latin America	41	.1	29	.1	14	.0
Middle East and North Africa	102	.2	50	.1	24	.0
Rest of Africa	34	.1	39	.1	18	.0
Asia	317	.6	194	.4	51	.1
Table Total	54095	100.0	53072	100.0	50205	100.0

6.3.1.3 Education and income

6.3.1.3.1 Education

The level of education of the patients and/or of the family as a whole is highest for the patients that have been admitted for myocardial infarction, thereafter stroke and lowest for the patients being admitted for hip fracture (see Table 6-5). When the family is considered as a unit, the average levels of education, in number of years, of the spouse having the highest education level are 10.97 (AMI), 10.82 (stroke) and 10.61 (hip fracture).

Table 6-5: Distribution of number of years of education finished by patient or spouse. Numbers are index admissions.

Years of education	Acute myocardial infarction		Stroke		Hip fracture	
	n	Percent	n	Percent	n	Percent
Unknown	707	1.3	808	1.5	680	1.4
.00	99	.2	122	.2	171	.3
6.00	14	.0	6	.0	1	.0
7.00	5	.0	7	.0	3	.0
8.00	15515	28.7	17311	32.6	18685	37.2
9.00	3789	7.0	3200	6.0	2885	5.7
10.00	653	1.2	352	.7	210	.4
11.00	14494	26.8	13895	26.2	12661	25.2
12.00	4370	8.1	4083	7.7	3452	6.9
13.00	6911	12.8	6005	11.3	4996	10.0
14.00	2959	5.5	2961	5.6	2720	5.4
15.00	1121	2.1	1108	2.1	1000	2.0
16.00	529	1.0	399	.8	375	.7
17.00	1216	2.2	915	1.7	562	1.1
18.00	871	1.6	1115	2.1	1026	2.0
19.00	736	1.4	699	1.3	695	1.4
20.00	29	.1	25	.0	24	.0
21.00	75	.1	61	.1	59	.1
22.00	2	.0				
Table Total	54095	100.0	53072	100.0	50205	100.0

6.3.1.3.2 Income

In a similar fashion to the pattern observed for education, income of both the patient and the family as a whole is highest for myocardial infarction patients and lowest for hip fracture patients (see Table 6-6). The average family income, over the period, is: 184,013 NOK (AMI), 167,970 NOK (stroke) and 150,574 NOK (hip fracture).

Table 6-6: Distribution of joint income of the patients and their spouse in increments of 1000 NOK by disease. Numbers are index admissions.

	Myocardial infarction		Stroke		Hip fracture	
	Count	Percent	Count	Percent	Count	Percent
<100	11632	21.5	14243	26.9	17532	35.0
100-200	25089	46.5	25786	48.7	23626	47.1
200-300	11037	20.4	8609	16.3	6156	12.3
300-400	3982	7.4	2645	5.0	1735	3.5
400-500	1212	2.2	810	1.5	528	1.1
500-600	495	.9	374	.7	255	.5
600-700	217	.4	173	.3	110	.2
700-800	94	.2	68	.1	63	.1
800-900	62	.1	57	.1	42	.1
900-1000	37	.1	32	.1	27	.1
>1000	146	.3	149	.3	85	.2
Table Total	54095	100.0	53072	100.0	50205	100.0

6.3.1.3.3 Property/Capital

The trends observed above for education and income are somewhat different here. Personal property/capital, averaged over the period, is highest for stroke patients (375,685 NOK), with AMI patients at the same level (373,945 NOK), and lowest for hip fracture patients (339,468). When looking at family property/capital however, the situation is turned around. AMI patients have the lowest household average property/capital (368,342 NOK), while hip fracture patients have the highest (404,390 NOK). Stroke patients are placed in the middle with an average household property/capital of 392,066 NOK.

Table 6-7: Distribution of the joint property/capital of the patients and their spouse in increments of 1000 NOK by disease. Numbers are index admissions.

	Myocardial infarction		Stroke		Hip fracture	
	Count	Percent	Count	Percent	Count	Percent
<100	10753	19.9	11673	22.0	13051	26.0
100-200	12053	22.3	11290	21.3	10521	21.0
200-300	10234	18.9	9193	17.4	7893	15.7
300-400	6541	12.1	5951	11.2	5242	10.5
400-500	4014	7.4	3989	7.5	3360	6.7
500-600	2696	5.0	2600	4.9	2339	4.7
600-700	1878	3.5	1786	3.4	1676	3.3
700-800	1282	2.4	1362	2.6	1262	2.5
800-900	936	1.7	1031	1.9	861	1.7
900-1000	691	1.3	795	1.5	678	1.4
1000-2000	2256	4.2	2462	4.6	2525	5.0
2000 or more	672	1.2	818	1.5	754	1.5
Table Total	54095	100.0	53072	100.0	50205	100.0

6.3.2 Hospitals, admissions, and deaths

There is a substantial variation in the number of admissions to each of the hospitals. This obviously depends on the size of hospitals. Larger hospitals and those in the eastern part of the country have the greatest proportion of the patients (see Table 6-8 to Table 6-9). Not all patients that have been admitted to a hospital remain at that hospital. For various reasons, patients are transferred to other hospitals. As can be seen in Table 6-10, between 3.7 and 9.8% of patients are transferred, most frequently those admitted for hip fracture.

Table 6-8: Distribution of index admissions according to health region, by disease category.

	Acute myocardial infarction		Stroke		Hip fracture	
	n	Percent	n	Percent	n	Percent
Eastern Norway	18744	34.7	19050	35.9	19855	39.5
Southern Norway	11341	21.0	10512	19.8	10382	20.7
Western Norway	10032	18.5	9208	17.4	8782	17.5
Central Norway	7787	14.4	8562	16.1	6765	13.5
Northern Norway	6191	11.4	5740	10.8	4421	8.8
Table Total	54095	100.0	53072	100.0	50205	100.0

Table 6-9: Distribution of index admissions according to type of hospital, by disease category (Using the research group's own classification).

	Acute myocardial infarction		Stroke		Hip fracture	
	n	Percent	n	Percent	n	Percent
Hospitals with university functions	10654	19.7	9976	18.8	8654	17.2
Larger hospitals without university functions	22926	42.4	22372	42.2	21810	43.4
Minor hospitals	20515	37.9	20724	39.0	19741	39.3
Table Total	54095	100.0	53072	100.0	50205	100.0

Table 6-10: Number of index admissions involving transfers between hospitals, by disease category.

	Acute myocardial infarction		Stroke		Hip fracture	
	n	Percent	n	Percent	n	Percent
None	48983	90.5	50367	94.9	43969	87.6
Once	4871	9.0	2543	4.8	6093	12.1
Twice or more	241	.4	162	.3	143	.3
Table Total	54095	100.0	53072	100.0	50205	100.0

6.3.3 Distance between home and hospital

This variable is a proxy for transportation time to hospital, and is calculated as the shortest distance by road (ferries not included) between the patient's registered residence and the hospital of admission. Some distances are very large, probably resulting from patients being sent to large hospitals in the capital region, or the patients being away from home when getting ill (see Table 6-11).

Table 6-11: Distance between home and hospital in km for hospital admissions, by disease group.

Km	Myocardial infarction		Stroke		Hip fracture	
	Count	Percent	Count	Percent	Count	Percent
<20	31038	57.4	31913	60.1	30677	61.1
20-50	10977	20.3	10784	20.3	10437	20.8
50-100	6437	11.9	6286	11.8	5716	11.4
100-150	2496	4.6	2062	3.9	1848	3.7
150-200	1066	2.0	835	1.6	692	1.4
200-500	1580	2.9	907	1.7	643	1.3
500-1000	345	.6	191	.4	116	.2
> 1000	156	.3	94	.2	76	.2
Table Total	54095	100.0	53072	100.0	50205	100.0

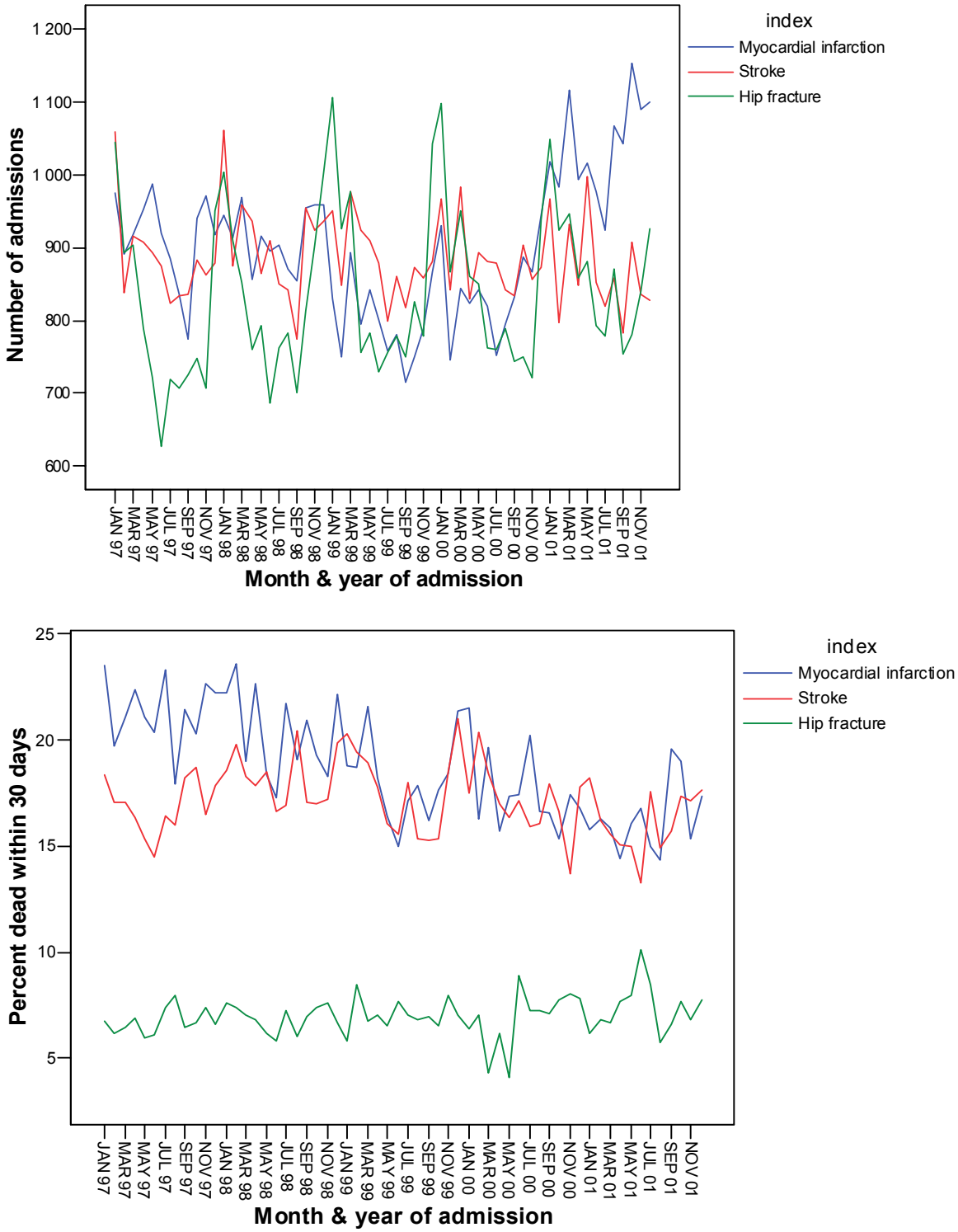
6.3.4 Calendar year

Hospital admission rates are rather even for stroke and hip fracture over the years but increasing for AMI (Figure 6-3). The increase is probably due to changes in definition of AMI using troponine values (see (64;67)). There are seasonal trends, especially for hip fracture, that are seen in Figure 6-4, with lowest admission rates in the late summer/early fall. There exist weekday trends with, not surprisingly, fewest admissions

or discharges on the weekends, the most admissions on Mondays and the greatest number of discharges on Fridays.

30-Day mortality, on the other hand, is decreasing over the five year period for AMI and to some extent for stroke, whereas again it is rather even for hip fracture.

Figure 6-3: Monthly admissions and 30-day mortality.



For stroke, there is a conspicuous pattern of mortality variation depending on the day of week for admission (Figure 6-5). Being admitted during the weekend seems to result in higher mortality.

Figure 6-4: Total admissions by month, for all three disease categories.

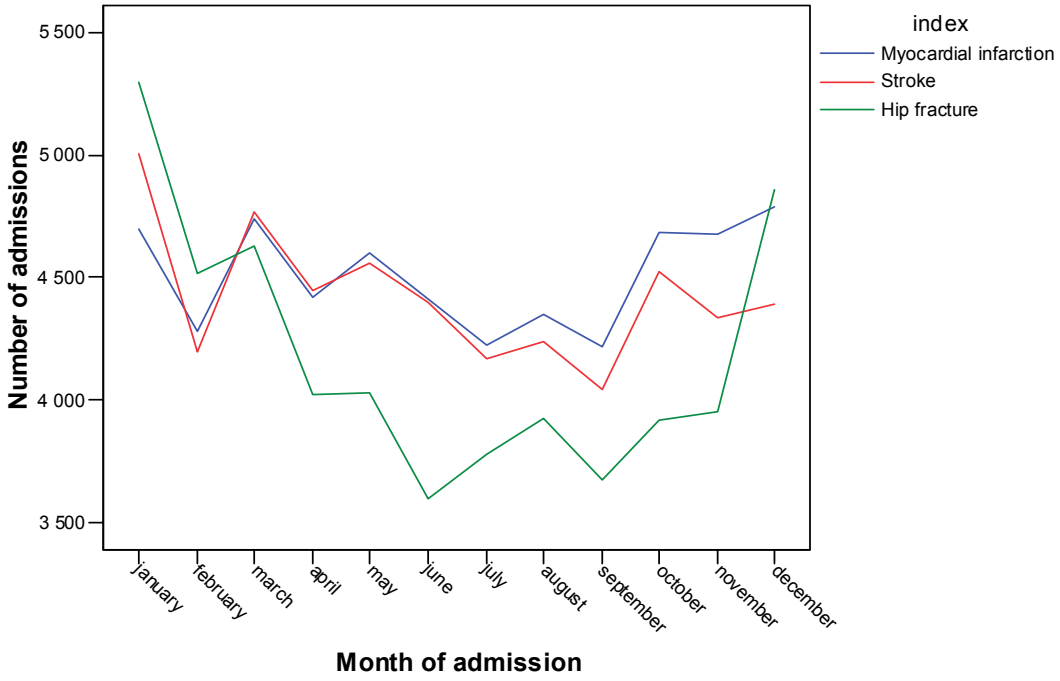
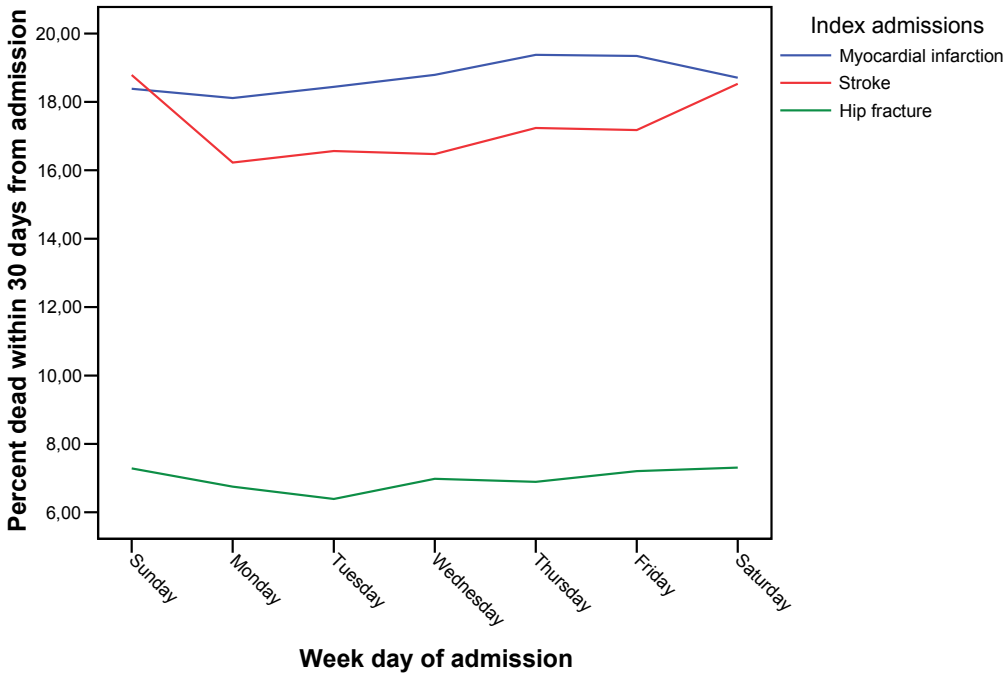


Figure 6-5: Average 30-day mortality by weekday of admission.



6.3.5 Patient frailty (comorbidity)

Two important factors in explaining mortality differences between hospitals are how sick or weak the patient is (patient frailty) and how severe the disease is (disease severity).

Here, number of diagnoses or pertinent codiagnoses was used as a proxy for patient frailty. Both number of codiagnoses at time of admission and previous admissions since 1994 were accounted for. Pertinent codiagnoses are those diagnoses that were selected a priori, by the respective expert groups, as being important to mortality from the disease. Diagnoses that may be a result or complication of treatment are not included for the current admission. This measure has been found to be effective as a proxy (78;79).

As can be seen in Table 6-12, around 70-80% of the patients are admitted with none or one codiagnosis, whereas around 5% of AMI patients and around 10% of the others have three codiagnoses or more. The average number of pertinent codiagnoses was 0.88 for AMI, 1.01 for stroke, and 0.67 for hip fracture. The overall range for all three disease categories at each hospital was 0.5 to 1.2 (excluding main diagnosis).

Table 6-12: Number of pertinent codiagnoses (complications not included) prior to or at admission for each of the three index admission disease categories.

Number of codiagnoses	Acute myocardial infarction		Stroke		Hip fracture	
	Count	Percent	Count	Percent	Count	Percent
0	28888	53.4	20960	39.5	25780	51.3
1	16882	31.2	17763	33.5	12547	25.0
2	6190	11.4	9264	17.5	6762	13.5
3	1749	3.2	3569	6.7	3147	6.3
4	347	.6	1128	2.1	1287	2.6
5	36	.1	323	.6	482	1.0
6	3	.0	55	.1	149	.3
7			10	.0	43	.1
8					6	.0
9					2	.0
Total	54095	100.0	53072	100.0	50205	100.0

6.3.6 Disease severity

Staging for acute myocardial infarction seemed a reasonable proxy for disease severity. The distribution of the patients in the different levels (as defined in 9.2) is presented in Table 6-13 below.

For stroke, the results of an initial analysis including staging were questioned from a clinical point of view. The expert group suggested an alternative classification that was used in the analysis and is presented below.

For hip fracture, higher CCDS stages were associated with increased risk of death, but there were comparatively few cases with these stages.

Table 6-13: Frequency distribution of the various categories of staging.

Acute myocardial infarction	CCDS stage 3.1	38481
	CCDS stage 3.2	5814
	CCDS stage 3.3	6406
	CCDS stages 3.4 - 3.6	1301
	CCDS stages 3.7 – 3.9	2093
Stroke	Infarction	46229
	Hemorrhage	6843
Hip fracture	CCDS stages 1.1 to 2.2	49893
	CCDS stages 2.3 to 3.3	312

6.4 PRELIMINARY DATA ANALYSIS

All covariates were studied in detail. When we included the different codiagnoses in the model, we encountered colinearity problems that in turn led to inflated estimates of the standard errors. When we inspected the codiagnoses more closely, we discovered that they often were present in a few cases only. This in turn led to numerical problems. Furthermore, many of the codiagnoses were highly correlated, which also will lead to colinearity problems.

We also observed that some codiagnoses, expected to be associated with increased risk of death, had the opposite effect in the model. Consequently, we removed all codiagnoses (except dementia) as single variables from the data analyses. As proxy variables, we included total number of pertinent codiagnoses, total number of previous diagnoses and total number of previous admissions since 1994. See 6.8.1 for further discussion. For the various staging variables, some CCDS stages had very few cases, and were subsequently combined into fewer categories.

The correlation structure was examined by the use of eigenvalues and the pairwise correlation matrix. With the reformulated model, there were no colinearities or numerical problems.

6.5 30-DAY MORTALITY AFTER ACUTE MYOCARDIAL INFARCTION

6.5.1 Inclusion/exclusion criteria

”Index admission” refers to the first admission for acute myocardial infarction (ICD-9 codes 410 with all subgroups, ICD-10 codes I21.0, I21.1, I21.2, I21.3, I21.4 and I21.9) in the calendar year. Table 9-5 shows the distribution of the various codes among the index cases. The only reason that a patient may have more than one index admission is when they are transferred, since each hospital will then be counted as a separate index admission. 54095 cases were classified as index admissions from these criteria. Of these, 18.7% were followed by death within 30 days.

Criteria for exclusion were: 1) under 18 years of age; 2) patients dead by accident (external cause of death, violence etc.); 3) either dead on arrival or not emergency; 4) admissions to hospitals with very few cases in the period 1997-2001; 5) admissions to

hospitals with fewer than 100 admissions (total over the period) after using the first 3 exclusion criteria. The final sample size after exclusion was 51432.

In the final data set, a small number of cases had missing values for one or more variables. These cases were included or excluded from analyses according to whether the particular analysis depended on the variables in question. Missing values were almost without exception associated with socio-demographic variables, mainly level of education.

In Appendix 9.1, we list the participating hospitals and the corresponding aliases used in tables and figures.

Table 6-14: Numbers of cases removed for AMI based on the exclusion criteria. Some exclusions relate to more than one criterion.

Exclusion criterion	Excluded hospitals	Excluded cases
Under 18 years of age	-	18
Admissions followed by accidental death	-	20
Neither dead on arrival nor emergency	-	2411
Hospitals with less than 100 admissions	4	130
Hospitals falling under 100 admissions after criteria 1-3 are applied	3	191[a]

a) Remaining number of cases at these hospitals after applying criteria 1-3

Table 6-15: Number of cases in final data set and with complete socio-demographic variables for AMI.

Not Selected	Selected	Selected and with complete socio-demographic variables
2663	51432	50764

6.5.2 Observed mortality

The proportion of patients dead at each hospital varies between years and between hospitals (Table 6-16 and Table 6-17). As can be seen in Figure 6-6, relative 30-day mortality has decreased from 1998 to 2001.

Table 6-16: Acute myocardial infarction. 30-day case fatality for each hospital per year of admission. Percentages not shown if number of admissions is below 50 per year (two hospitals with less than 50 patients/year, all of the five years, are omitted).

Hospital	Year of admission				
	1997	1998	1999	2000	2001
Hsp01	15.4	15.8	25.3	17.9	16.7
Hsp03	15.6	17.0	13.4	16.1	19.8
Hsp04	19.7	18.2		22.0	8.3
Hsp06	20.2	16.0	16.1	12.8	12.1
Hsp07	23.9	23.5	18.3	20.3	22.0
Hsp08	18.2	18.2	17.1	18.3	22.1
Hsp09	26.0	25.5	23.1	16.5	21.3
Hsp10	16.8	17.8	15.8	17.2	16.4
Hsp11	23.6	16.8	20.0	18.7	14.6

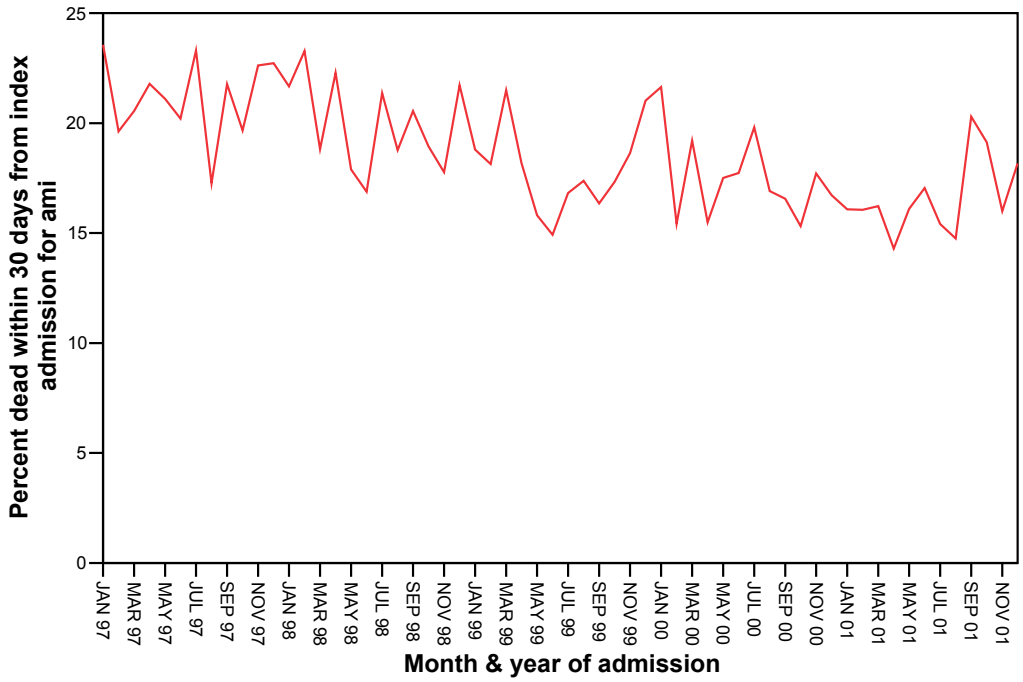
Hospital	Year of admission				
	1997	1998	1999	2000	2001
Hsp12	17.9	22.9	21.7	15.9	20.8
Hsp14	22.1	8.9	19.0	8.1	22.2
Hsp15			17.7	19.0	15.0
Hsp16	18.8	19.9	16.2	13.9	16.2
Hsp17	22.4	20.2	16.1	12.8	14.0
Hsp18	25.0	22.3	26.2	21.5	17.6
Hsp19	22.1	26.0	11.8	23.5	17.3
Hsp20	18.2	18.1	24.1	18.8	13.3
Hsp21	19.6	22.5	19.0	16.5	21.0
Hsp23		13.2			
Hsp24	20.5	20.5	20.3	21.0	19.6
Hsp26	18.7	18.9	21.4	18.6	12.6
Hsp28	19.4	18.2	18.1	21.8	13.8
Hsp29	26.7	26.4	21.9	21.7	26.3
Hsp30	17.5	19.7	11.3	8.3	20.8
Hsp31	24.7	24.2	17.8	15.5	15.7
Hsp32	31.2	28.0	15.6	22.1	16.0
Hsp33					1.6
Hsp34	24.0	20.1	20.9	17.3	18.5
Hsp35	26.4	17.8	20.4	17.7	17.4
Hsp36	17.6	27.8	8.8	17.1	16.7
Hsp37	19.4	20.6	18.0	19.1	17.8
Hsp38	25.4	23.6		18.9	16.9
Hsp39	22.1	21.3	21.7	18.2	17.7
Hsp40	21.4	13.3	17.9	13.0	12.5
Hsp41	20.7	11.3	8.4	6.8	9.0
Hsp42	21.8	28.1			
Hsp43	21.4	20.0	14.7	16.5	13.6
Hsp44	17.2	18.1	10.8	11.8	17.2
Hsp46	17.1	19.2	26.7	16.7	12.7
Hsp47	22.6	27.7	18.4	22.4	14.4
Hsp48	20.3	33.3	18.6	18.3	21.8
Hsp49	18.5	19.5	13.3	15.8	20.0
Hsp50	17.0	18.4	20.9	18.0	14.6
Hsp51	17.9	20.6	23.5	18.4	12.0
Hsp52	13.4	23.5	12.3	18.5	18.8
Hsp53	24.3	26.0	17.6	14.8	19.7
Hsp55	27.2	20.0	17.1	20.8	18.3
Hsp56	23.4	16.3	16.1	17.6	18.3
Hsp57	19.0	18.7	17.8	15.3	18.1
Hsp58	18.0	18.3			
Hsp60	18.9	21.9	16.6	17.4	15.8
Hsp61	23.8	21.5	27.8	23.3	23.0
Hsp62	23.4	18.4	14.3	20.1	21.0
Hsp63	22.2	19.4	11.7	19.5	15.2
Hsp64	20.9	24.4	25.2	28.8	24.6
Hsp65	21.6	21.3	20.0	19.3	23.5
Hsp66	23.5	17.8		18.5	21.0

Table 6-17: Acute myocardial infarction: Number of cases dead within 30 days for each hospital, together with the total number of index cases per year, after applying exclusion criteria.

Hospital	1997		1998		1999		2000		2001	
	Number of AMI cases	Dead within 30 days	Number of AMI cases	Dead within 30 days	Number of AMI cases	Dead within 30 days	Number of AMI cases	Dead within 30 days	Number of AMI cases	Dead within 30 days
Hsp1	228	35	241	38	194	49	179	32	245	41
Hsp3	250	39	235	40	187	25	174	28	197	39
Hsp4	71	14	66	12	42	9	50	11	60	5
Hsp6	262	53	243	39	236	38	242	31	339	41
Hsp7	92	22	81	19	60	11	64	13	59	13
Hsp8	110	20	170	31	117	20	120	22	154	34
Hsp9	100	26	94	24	78	18	79	13	89	19
Hsp10	358	60	314	56	304	48	279	48	373	61
Hsp11	521	123	546	92	540	108	493	92	574	84
Hsp12	84	15	83	19	92	20	82	13	77	16
Hsp14	68	15	90	8	100	19	86	7	81	18
Hsp15	.	.	23	2	164	29	168	32	226	34
Hsp16	218	41	186	37	185	30	202	28	235	38
Hsp17	437	98	406	82	397	64	375	48	485	68
Hsp18	168	42	148	33	107	28	149	32	125	22
Hsp19	77	17	73	19	68	8	68	16	75	13
Hsp20	110	20	105	19	87	21	112	21	120	16
Hsp21	97	19	142	32	79	15	79	13	105	22
Hsp23	43	10	53	7	34	4	48	4	39	7
Hsp24	278	57	273	56	217	44	214	45	271	53
Hsp25	49	8	38	6	29	5	47	8	48	8
Hsp26	198	37	201	38	159	34	156	29	198	25
Hsp28	170	33	154	28	116	21	147	32	159	22
Hsp29	101	27	110	29	64	14	92	20	99	26
Hsp30	57	10	71	14	53	6	60	5	72	15
Hsp31	154	38	120	29	90	16	97	15	159	25
Hsp32	109	34	100	28	77	12	86	19	194	31
Hsp33	.	.	3	0	16	0	23	2	128	2
Hsp34	325	78	329	66	253	53	277	48	297	55
Hsp35	296	78	303	54	269	55	254	45	316	55
Hsp36	74	13	54	15	80	7	76	13	54	9
Hsp37	186	36	180	37	211	38	241	46	314	56
Hsp38	59	15	55	13	40	7	53	10	65	11
Hsp39	480	106	508	108	466	101	559	102	599	106
Hsp40	103	22	90	12	67	12	69	9	96	12
Hsp41	58	12	62	7	131	11	190	13	234	21
Hsp42	124	27	121	34	43	6	.	.	.	.
Hsp43	490	105	496	99	477	70	539	89	712	97
Hsp44	157	27	127	23	93	10	110	13	145	25
Hsp46	82	14	73	14	60	16	66	11	71	9
Hsp47	226	51	390	108	288	53	353	79	354	51
Hsp48	79	16	72	24	59	11	60	11	87	19

Hospital	1997		1998		1999		2000		2001	
	Number of AMI cases	Dead within 30 days	Number of AMI cases	Dead within 30 days	Number of AMI cases	Dead within 30 days	Number of AMI cases	Dead within 30 days	Number of AMI cases	Dead within 30 days
Hsp49	297	55	293	57	181	24	202	32	245	49
Hsp50	253	43	250	46	234	49	222	40	295	43
Hsp51	84	15	102	21	98	23	103	19	75	9
Hsp52	67	9	81	19	73	9	54	10	69	13
Hsp53	202	49	235	61	193	34	189	28	183	36
Hsp55	316	86	499	100	380	65	375	78	465	85
Hsp56	214	50	172	28	205	33	222	39	257	47
Hsp57	575	109	615	115	428	76	463	71	624	113
Hsp58	50	9	60	11	30	5	40	5	43	9
Hsp59	33	9	36	3	31	6	19	3	32	4
Hsp60	212	40	210	46	169	28	178	31	222	35
Hsp61	105	25	93	20	90	25	103	24	126	29
Hsp62	342	80	305	56	265	38	259	52	291	61
Hsp63	126	28	129	25	103	12	113	22	125	19
Hsp64	110	23	127	31	107	27	125	36	138	34
Hsp65	97	21	80	17	90	18	83	16	119	28
Hsp66	102	24	90	16	44	7	54	10	81	17

Figure 6-6: 30-day case fatality of AMI as a function of month and year of admission (percent).



6.5.3 Distribution of covariates

Summary measures of the distribution of risk adjustment covariates are shown in Table 6-18 below. The table shows minimum, maximum and standard deviation between hospitals (i.e. the various statistics for the hospital means), as well as the overall mean and standard deviation, for each covariate. For categorical covariates, the indicator variables for each category (except the first (reference) category) are used. Apparently, most of the variability is on the individual rather than on the hospital level, (a notable exception is distance between home and hospital).

Table 6-18: Distribution summaries for AMI covariates. Categorical variables in percentage units.

Variable	Between hospital min	Between hospital max	Between hospital std. dev	Overall std.dev	Overall mean
Age	58.94	73.78	2.51	13.39	70.01
Sex: female	24.74	48.76	3.90	48.45	37.65
Marital status: not married	5.12	17.05	2.38	27.92	8.52
Marital status: divorced or separated	2.73	18.15	3.10	29.24	9.44
Marital status: widowed	10.00	35.15	4.03	44.06	26.36
Maximum education in household (years)	9.92	13.63	0.64	2.74	10.97
Natural logarithm of: household income + 1	11.65	12.27	0.12	0.89	11.92
Natural logarithm of: household property/capital + 1	10.93	12.79	0.25	2.10	12.30
Natural logarithm of: distance from home to hospital + 1	0.31	2.78	0.49	0.91	1.12
Number of preadmissions regardless of cause, weighted	0.02	0.32	0.06	0.47	0.24
Number of pertinent codiagnoses at previous admission, weighted	0.02	0.11	0.02	0.18	0.08
CCDS stage 3.2	3.09	50.00	6.16	30.84	10.64
CCDS stage 3.3	2.35	22.21	3.97	32.35	11.87
CCDS stages 3.4 to 3.6	0.00	4.11	0.87	15.10	2.33
CCDS stages 3.7 to 3.9	0.00	6.33	1.37	19.41	3.92
Transferred from another hospital	0.88	14.22	2.73	21.61	4.91

6.5.4 Model selection

Starting with a model having interactions between hospital and year (viewed as a categorical variable), we performed a stepwise backward model selection. The largest model (M6 in Table 6-20 below) was only fitted to the subset of data consisting of the ten largest hospitals, as explained in 5.5. With this exception, the final model selection was performed on the complete data set.

Note that we do not make any a priori assumptions on the yearly variation of performance. Both smooth and sudden changes are possible, and the hospitals may evolve independently over time.

In each step of the selection process, variables or (predefined) sets of variables were removed from the model when the standard log-likelihood (deviance) test, comparing the resulting model to the previous one, was non-significant at the 5% level. When the test was significant, the selection process stopped.

After inspection of the estimated coefficients for all three diseases, ethnicity appeared non-significant. Ethnicity was then taken out of the group of socio-demographic variables and treated separately.

The backwards selection process is summarized in the tables below.

Table 6-19: Variable names and type for myocardial infarction (see 5.4.5).

Variable set	Variable name	Type of variable	Description
	yearfactor	Categorical	The years 1997-2001 as factor
	yeardiff	Continuous	Year – 2001 (a negative number)
	hospital	Categorical	Hospital code 1…66
Age/gender	N03	Continuous	B-spline basis function (of age)
	N13	Continuous	B-spline basis function (of age)
	N23	Continuous	B-spline basis function (of age)
	N33	Continuous	B-spline basis function (of age)
	sex	Categorical	Distribution of the sexes
Socio-demographic	marital	Categorical	Marital status
	maxedu	Continuous	Maximum education in household
	logincome	Continuous	Natural logarithm of: household income + 1
	logworth	Continuous	Natural logarithm of: household property/capital + 1
	ethnicnor	Categorical	Born in Norway (Yes/No)
Frailty	preadmiss	Continuous	Number of preadmissions regardless of cause
	prenumbdiagn	Continuous	Number of pertinent codiagnoses at previous admission
Severity	logdistance	Continuous	Natural logarithm of: distance from home to hospital + 1
	CCDS stage	Categorical	Staging for severity
	moved in	Categorical	Transferred from another hospital

Table 6-20: Description of candidate models used.

Model name	Variables included in model	Interactions
M6	age/gender set, socio-demographic set, frailty set, severity set, second order terms in preadmiss and prenumbdiagn	Yearfactor:hospital
M7	age/gender set, socio-demographic set, frailty set, severity set, yeardiff, second order terms in yeardiff, preadmiss and prenumbdiagn	(Yeardiff and Yeardiff squared):hospital
M8	age/gender set, socio-demographic set, frailty set, severity set, yeardiff, second order terms in yeardiff, preadmiss and prenumbdiagn	Yeardiff:hospital
M9	age/gender set, socio-demographic set, frailty set, severity set, yeardiff, second order terms in yeardiff and preadmiss	Yeardiff:hospital
M10	age/gender set, socio-demographic set, frailty set, severity set, yeardiff, second order terms in yeardiff and preadmiss	
M11	age/gender set, socio-demographic set, frailty set, severity set, yeardiff, second order term preadmiss	
M17	age/gender set, socio-demographic set, excluding ethnicnor, frailty set, severity set, yeardiff, second order term in preadmiss	
M31	age/gender set, frailty set, severity set, yeardiff, preadmiss squared	

Table 6-21: Analysis of deviance for AMI model selection.

Data set	Model name	Resid..Df	Resid. Deviance	Df	Deviance change	P-value	AIC change
subset	M6	20429	16874.47				
subset	M7	20449	16898.46	-20	23.99	0.243	-16.01
subset	M8	20458	16914.44	-9	15.98	0.067	-2.02
subset	M9	20459	16915.19	-1	0.75	0.385	-1.25
subset	M10	20468	16930.71	-9	15.52	0.078	-2.48
subset	M11	20469	16932.16	-1	1.45	0.228	-0.55

subset	M17	20470	16933.05	-1	0.89	0.346	-1.11
subset	M21	20471	16973.15	-1	40.10	<1e-4	38.10
all	M7	50562	41560.86				
all	M10	50679	41690.26	-117	129.40	0.204	-104.60
all	M11	50680	41691.12	-1	0.86	0.354	-1.14
all	M17	50681	41691.41	-1	0.28	0.594	-1.72
all	M31	50687	41824.44	-6	133.03	<1e-4	121.03

The stepwise procedure selected as the final model M17, which does not include interactions between hospital and year. Using the minimum AIC criterion leads to the same conclusion. All 50764 cases with complete (non-missing) data were used in fitting this model. Of these, 2 had weights exactly zero.

The order of the splines used for age effect was tested by comparing the AIC of the final model with the AIC using degree 4 splines (i.e. two internal knots, placed at the 33% and 67% percentiles 65 and 78). With an increase in AIC by 1.46, the degree 4 splines were rejected.

Two measures of the final model's prediction ability were computed: area under the receiver operating characteristic, or C statistic, and the optimal rate of correct classification.

Table 6-22: AMI final model – measures of prediction ability.

C statistic	0.74
Classification rate	0.82

Hosmer-Lemeshow goodness-of-fit statistics were computed for the final model (see Table 6-23). These values are statistically significant. It is widely accepted, however, that goodness-of-fit tests for models with very large data sets must be viewed with caution. The reason is that they become very sensitive to small, irrelevant discrepancies in the model, and may lead to unnecessarily complex models that are hard to interpret. In this case, the observed chi-square statistic would be highly likely if the average absolute discrepancy between modeled and true 30-day mortality probabilities is about 0.04, measured on the logistic (beta) scale. It is hard to see that any resulting errors in hospital could approach even this magnitude. We will accordingly view the model fit as acceptable.

Table 6-23: Hosmer-Lemeshow goodness-of-fit statistics for final AMI model.

Data set	C statistic	df	P-value
Subset	21.01	8	0.0071
All	22.43	8	0.0042

To check for influential observations, various influence measures were computed for the final model. The deviance residuals ranged from -2.10 to 3.20, which is no cause for

concern. The largest leverage hat statistic was 0.085. A set of 31 observations showed high or very high hat values, above 0.02. As noted earlier, it is reasonable to standardize hat values by the number of cases per hospital. Looking at cases with hat value greater than 0.005 and standardized hat value greater than 2, we found another set of 775 cases with moderately high leverage. Both sets belong to hospital Hsp55. There does not seem to be any great difference between these sets and the rest of the data, except for a somewhat lower median age. Hsp55 is among the hospitals with greatest rate of transferred patients, implying that it is reasonable to expect cases from this hospital to have increased influence on the model, via the variable "moved in".

The maximum Cook's distance was 0.0055, with 27 cases above 0.0005. These cases had a very high mortality. Hospitals with a small number of admissions accounted for 21 cases. Among these, 6 came from one of the smallest hospitals, Hsp59, which also turned out to have a relatively high uncertainty in its effects estimate. As noted above, the leverage of cases from small hospitals will tend to be high. The influential cases were judged to be acceptable and therefore retained in the model.

6.5.5 Covariate parameter estimates

The parameter estimates for risk adjustment covariates are displayed in the following table:

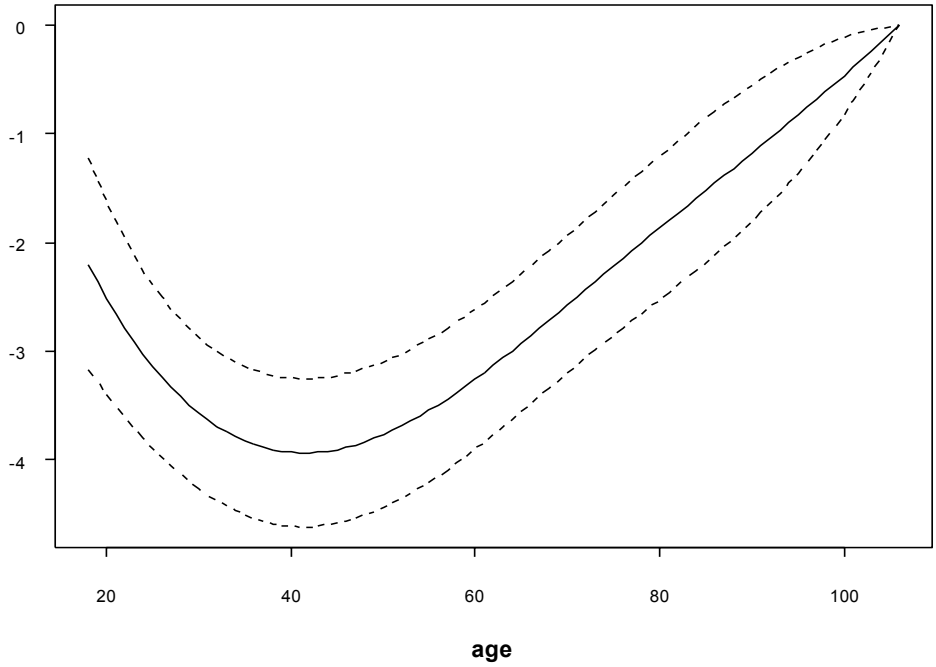
Table 6-24: Final AMI model – parameter estimates.

Model term		Estimate	Std. Error	Z value	Two-sided p-value
(Intercept)		2.502	0.401	6.24	<1e-4
N03	age function	-2.202	0.497	-4.43	<1e-4
N13	age function	-5.211	0.481	-10.84	<1e-4
N23	age function	-2.732	0.255	-10.71	<1e-4
N33	age function	-0.873	0.419	-2.08	0.037
Yeardiff	=year-2001	-0.077	0.0091	-8.45	<1e-4
female sex		-0.077	0.028	-2.76	0.00586
maxedu		-0.012	0.0052	-2.27	0.023
logincome		-0.073	0.021	-3.41	0.00066
logworth		-0.040	0.0070	-5.80	<1e-4
married/cohabitant		-0.040	0.023	-1.71	0.087
not married		0.181	0.034	5.31	<1e-4
divorced or separated		-0.046	0.039	-1.16	0.245
widowed		-0.095	0.026	-3.72	0.00020
logdistance		-0.042	0.016	-2.56	0.011
preadmiss		0.298	0.050	5.92	<1e-4
preadmiss squared		-0.044	0.013	-3.45	0.00056
prenumbdiagn		0.360	0.078	4.63	<1e-4
moved in		-0.504	0.113	-4.44	<1e-4
CCDS stage 3.2		-0.282	0.023	-12.16	<1e-4

Model term		Estimate	Std. Error	Z value	Two-sided p-value
CCDS stage 3.3		0.088	0.013	6.48	<1e-4
CCDS stage 3.4 to 3.6		0.274	0.016	16.68	<1e-4
CCDS stage 3.7 to 3.9		0.326	0.011	29.25	<1e-4

The dependency of mortality on age through the spline functions N03,..., N33 is shown in the figure below.

Figure 6-7: Final AMI model – effect of age.



Mortality first decreases, reaches a minimum around 40 years, and then increases strongly with age. The number of previous diagnoses and the number of previous admissions have a moderately strong effect on mortality. Relatively smaller increases in mortality follow from never having been married or cohabitating.

Being moved into the hospital has a fairly large effect with negative sign, i.e. acting to reduce risk. The likely explanation is that on the balance, transfers tend to come after the initial period of high mortality. It is thus the less severely ill patients that tend to be transferred. The socio-demographic variables have relatively small, but significant, effects. They act, as would be expected, to reduce mortality with increasing education, income or fortune.

The effect of CCDS stage is more difficult to interpret. Note that this variable is entered in the model with Helmert contrasts, meaning that the numbers shown measure the difference between the actual category and the mean of the previous (i.e. lower stages) categories. This parameterization is better suited to examining effects that are believed to be increasing, compared to the usual parameterization where each category is contrasted with a fixed reference category. However, the two parameterizations are

equivalent in the sense that the statistical model remains the same, and one set can be computed from the other. Generally, mortality is steadily increasing when progressing through the stages, as one would expect. However, being in stage 3.2 actually reduces mortality as compared to stage 3.1. This is contrary to the clinical interpretation of the staging system, and is probably a manifestation of the diagnosis timing effects discussed in 6.8.1.

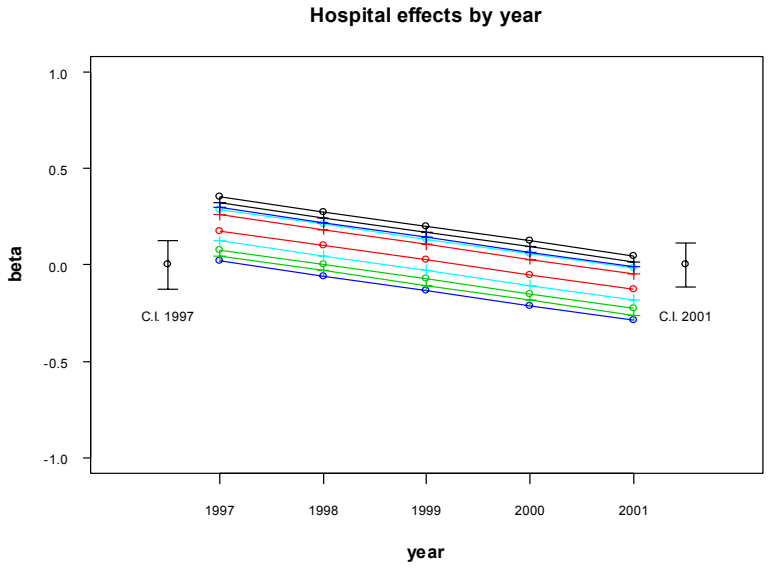
The number of previous admissions and previous diagnoses increase the risk of dying. The reason for the second order term in the number of previous admissions is that this variable does not act linearly on mortality, but has an effect that increases more and more slowly as the variable increases.

6.5.6 Variation of hospital performance over time

An important question in assessing the quality indicators is whether they show stability over time or have large temporal variability with no reasonable explanation. In the model, temporal variability not explained by risk adjustment covariates is either assigned to random variation or incorporated in the year/hospital part of the model.

The figure below illustrates the effects of calendar year and hospital. For clarity, only the ten largest hospitals are shown.

Figure 6-8: Effect of hospital and year for the ten largest hospitals. Confidence intervals are indicated for the start and end years only. These must be superimposed on every curve to display their uncertainty.



Note the relatively large uncertainty, shown only for the start and end of the observation period. As is evident from the statistically significant, negative coefficient of Yeardiff (Table 6-24), there is an overall downward trend in mortality during the period.

From the calculations of residual autocorrelations reported in section 6.10, it follows that there is no indication of any monthly or weekly pattern of variation except random fluctuations.

6.5.7 Model robustness

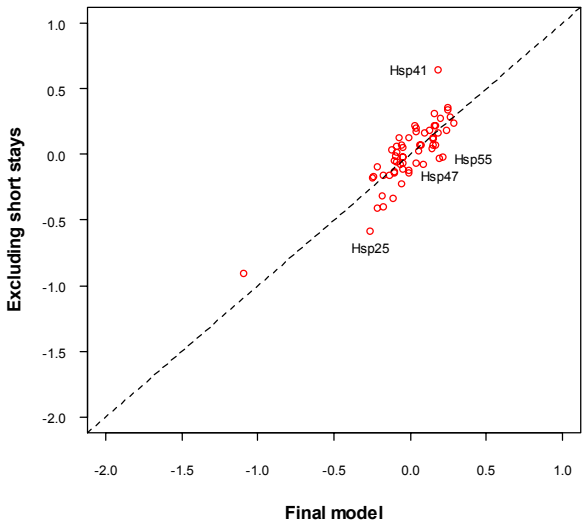
The number of codiagnoses (the sum of pertinent codiagnoses from previous admissions plus those under the current admission, but excluding those that may be considered an effect of treatment) was initially used as a proxy for patient frailty. When included in the model, this variable had a parameter estimate of -0.21 and thus acting to reduce risk, significant at the 0.1% level, which obviously is contrary to expectation. The variable, which is discussed further in section 6.8.1 below, was excluded from the final analysis.

We were concerned about the robustness of our analysis with respect to three factors:

- inclusion/exclusion of transferred patients
- inclusion/exclusion of all 1999 data. This was the first year with ICD-10 coding, which conceivably might have led to inconsistent coding
- inclusion/exclusion of the risk adjustment covariate "number of codiagnoses excluding complications"
- inclusion/exclusion of short admissions (<2 days). Some of our concerns about data quality apply to short stays in particular, see 6.9.2 below.

To examine the robustness of our final model, we compared the hospital effect estimates of the final model with estimates obtained after varying the above factors. The figures below show the results. The hospitals with apparently large changes are labeled.

Figure 6-9. Robustness of AMI model. Hospital effects with stays <2 days excluded against effects from final model

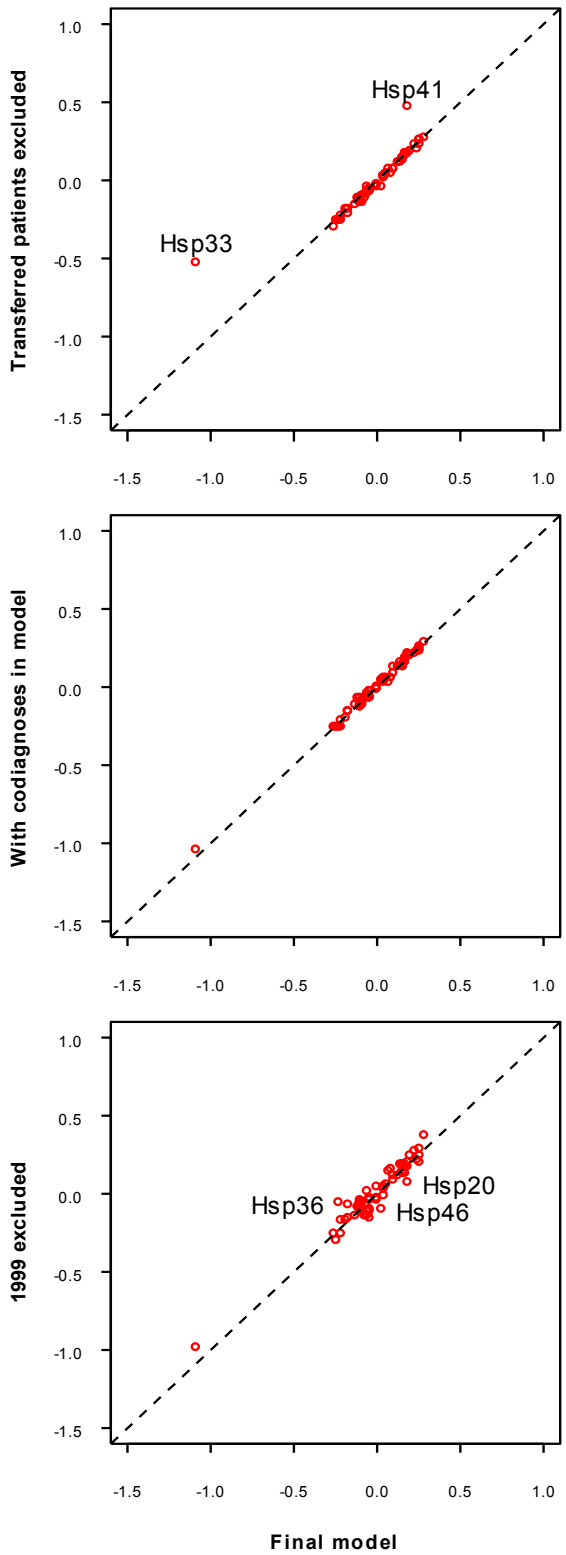


Hospital effects

Hospital effects are the log-odds of death within 30 days at the various hospitals, compared to the average hospital, controlling for risk adjustment covariates (see chapter 5.6).

An effect of e.g. -0.182 means that this particular hospitals has 10% reduced log-odds for death, while an effect of e.g. 0.41 means that the log-odds of death are 50% greater than in the average hospital, given the levels of the risk adjustment covariates.

Figure 6-10: Robustness of final AMI model. Hospital effects from alternative models against effects from final model.



Without transferred patients, the effects for Hsp33 and Hsp41 increased by 0.57 and 0.28 respectively. For Hsp33, this is in fact smaller than the estimated effect's standard deviation under the final model, but for Hsp 41 it amounts to 1.75 standard deviations. The proportions of transferred patients are among the largest in these hospitals, which are highly specialized and have relatively few direct admissions for AMI. For the other hospitals, differences were very small.

Inclusion of codiagnoses in the model leads to very small changes in the effect estimates.

When data from 1999 were excluded, hospital effects showed fairly small changes. The largest changes occurred for Hsp36, with an increase by 0.17, and Hsp46 and Hsp20 with decreases of 0.14 and 0.12 respectively. The change for Hsp36 is slightly above one standard deviation, for the other two it is slightly less.

Excluding stays lasting less than 2 days, we see that Hsp33 and Hsp41 again display sensitivity of effects. Hsp47 and Hsp55 will be discussed further in 6.9.3. Although the effect of Hsp25 exhibits a large change, the change is consistent with its standard deviation.

The table shows the statistical measures of the changes (see 5.6.5).

Table 6-25: Robustness of final AMI model. Statistical measures of changes in hospital effects.

Measure	Transferred patients excluded	Model without codiagnoses	1999 excluded	Short stays excluded
AAD	0.031	0.015	0.041	0.104
RAAD	0.213	0.103	0.283	0.650
correlation	0.910	0.996	0.963	0.830
rank correlation	0.992	0.989	0.918	0.808

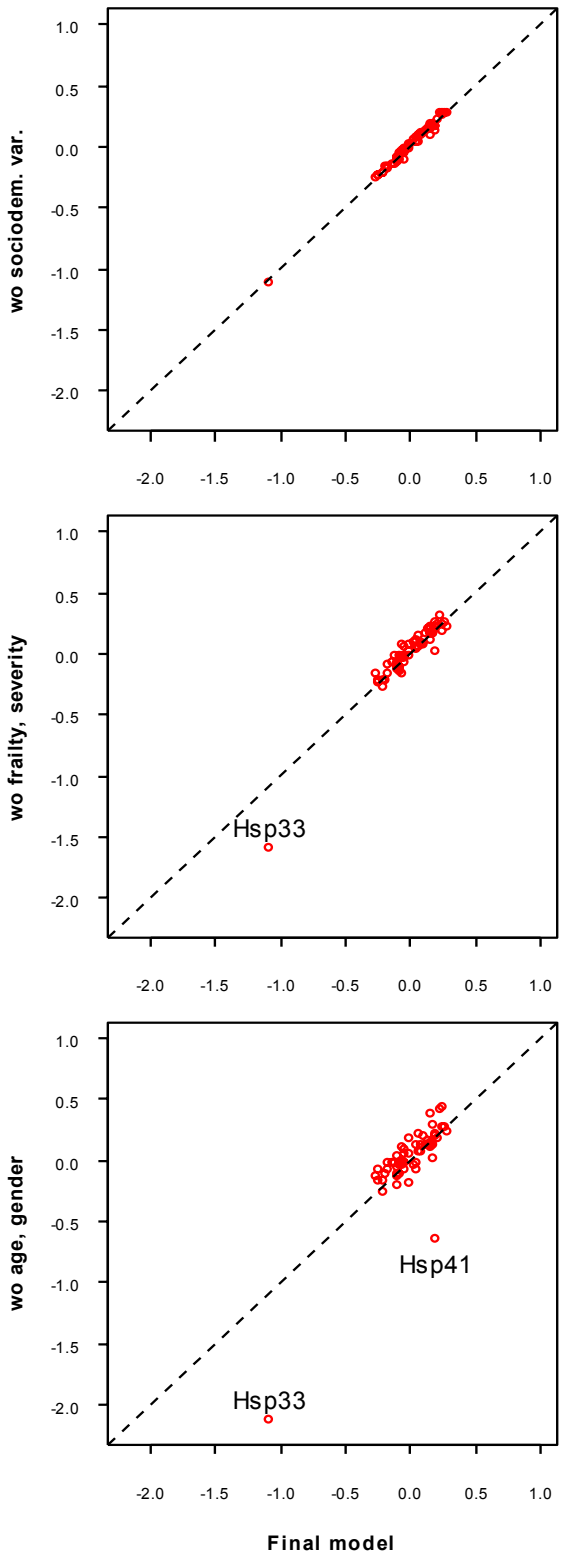
We concluded that the robustness of estimated hospital effects was acceptable, with the provision that for Hsp33 and Hsp41, results were highly dependent on inclusion/exclusion of transferred patients, and must therefore be regarded as unreliable.

6.5.8 Comparison of results with and without risk adjustment

To estimate the magnitude of the bias in the risk adjustment, we compute various bias indicators, following (1). We have investigated the change in hospital effect estimates when successively smaller sets of risk adjustment covariates are used. As explained in 6.9.1, we may regard these changes as plausible upper bounds for the true bias.

The figure below shows estimates from our final model plotted against those of the reduced models. First, we remove socio-demographic variables, then both socio-demographic and frailty/severity variables, and last, also age and gender, leaving no risk adjustment covariates in the model.

Figure 6-11: Comparison of results for AMI with and without risk adjustment. Hospital effects from reduced models against effects from final model.



The overall magnitude of the changes are very small when socio-demographic variables are excluded, fairly small when frailty and severity variables are also excluded, and moderate when age and gender are excluded in addition.

Exceptions are Hsp33 and Hsp41. They both have specialized roles compared to the other hospitals, and it is not unexpected that their effect estimates are particularly sensitive to changes in risk adjustment. These hospitals differ markedly from the others in the distribution of covariates and typically receive patients with lower mortality.

Quantitative measures of the change in effects are shown in the table below (see 5.6.5).

Table 6-26: Indicators of bias in AMI risk adjustment. Statistical measures of changes in hospital effects.

Measure	Socio-demographic set excluded	Frailty/severity set also excluded	Age/gender set also excluded
AAD	0.016	0.048	0.105
RAAD	0.108	0.321	0.659
correlation	0.995	0.958	0.834
rank correlation	0.982	0.936	0.786

The overall conclusion is that assessment of relative hospital performance is not very sensitive to the exclusion of socio-demographic variables, a finding that was also seen elsewhere (80). Excluding frailty and severity indicators results in somewhat larger changes in estimated hospital effects. The additional changes due to exclusion of age and gender are of the same order.

6.5.9 Hospital effects

To judge the magnitude and importance of the estimated hospital effects (defined in 5.4.2), the three decision procedures described earlier were applied:

- the single hospital rule: testing one hospital by the standard test. A one-sided p-value, if less than 10%, is reported
- the follow-up list rule: testing each hospital by the standard test, but with level according to the number of cases. Listed hospitals are reported
- the multiple hypotheses testing rule: all effects are tested simultaneously by a multiple testing procedure. A simultaneous p-value, if less than 10%, is reported.

Hospital effects

Hospital effects are the log-odds of death within 30 days at the various hospitals, compared to the average hospital, controlling for risk adjustment covariates (see chapter 5.6).

An effect of e.g. -0.182 means that this particular hospitals has 10% reduced log-odds for death, while an effect of e.g. 0.41 means that the log-odds of death are 50% greater than in the average hospital, given the levels of the risk adjustment covariates.

The table below shows the shrinkage and raw estimates, the raw estimate standard deviation, the decision procedure results and number of cases per hospital.

Table 6-27: AMI final model – hospital effect estimates (Positive estimate indicates increased mortality).

Hospital	Shrinkage estimate	Raw estimate	Raw estimate Std. Error	P-value (one-sided)	Listed for follow-up	P-value (multiple testing)
Hsp1	-0.026	-0.046	0.086			
Hsp3	-0.055	-0.101	0.090			
Hsp4	-0.014	-0.053	0.163			
Hsp6	-0.004	-0.007	0.085			
Hsp7	0.023	0.067	0.137			
Hsp8	0.044	0.095	0.105			
Hsp9	0.067	0.171	0.122	0.081		
Hsp10	-0.060	-0.094	0.073	0.099		
Hsp11	-0.005	-0.007	0.057			
Hsp12	0.015	0.042	0.132			
Hsp14	-0.080	-0.242	0.139	0.040	yes	
Hsp15	-0.033	-0.081	0.119			
Hsp16	-0.023	-0.042	0.091			
Hsp17	0.032	0.047	0.066			
Hsp18	0.131	0.263	0.099	0.00377	yes	
Hsp19	0.025	0.079	0.143			
Hsp20	0.072	0.184	0.122	0.066		
Hsp21	0.062	0.154	0.119	0.098		
Hsp23	-0.042	-0.214	0.198			
Hsp24	0.109	0.175	0.076	0.011	yes	
Hsp25	-0.051	-0.260	0.198	0.094		
Hsp26	-0.022	-0.042	0.093			
Hsp28	-0.065	-0.135	0.102	0.093		
Hsp29	0.105	0.255	0.117	0.015	yes	
Hsp30	-0.049	-0.183	0.162			
Hsp31	0.070	0.154	0.108	0.076		
Hsp32	0.088	0.203	0.112	0.035	yes	
Hsp33[a)]	-0.024	-1.088	0.654	0.048	yes	
Hsp34	0.084	0.129	0.072	0.037		
Hsp35	0.100	0.158	0.074	0.017	yes	
Hsp36	-0.066	-0.235	0.156	0.066		
Hsp37	0.035	0.060	0.083			
Hsp38	-0.001	-0.004	0.160			
Hsp39	0.125	0.166	0.056	0.00163	yes	
Hsp40	-0.069	-0.216	0.142	0.064		
Hsp41[a)]	0.050	0.187	0.162			
Hsp42	0.087	0.288	0.149	0.026	yes	
Hsp43	-0.082	-0.113	0.060	0.030		

Hospital	Shrinkage estimate	Raw estimate	Raw estimate Std. Error	P-value (one-sided)	Listed for follow-up	P-value (multiple testing)
Hsp44	-0.042	-0.101	0.117			
Hsp46	0.010	0.033	0.151			
Hsp47	0.128	0.191	0.068	0.00258	yes	
Hsp48	0.085	0.250	0.136	0.033	yes	
Hsp49	-0.049	-0.083	0.082			
Hsp50	-0.038	-0.064	0.081			
Hsp51	-0.017	-0.048	0.130			
Hsp52	-0.032	-0.106	0.150			
Hsp53	0.081	0.144	0.086	0.047		
Hsp55	0.158	0.222	0.062	0.00018	yes	0.021
Hsp56	-0.059	-0.104	0.085			
Hsp57	-0.064	-0.085	0.057	0.067		
Hsp58	-0.036	-0.171	0.189			
Hsp59	-0.010	-0.071	0.236			
Hsp60	0.023	0.042	0.089			
Hsp61	0.067	0.161	0.116	0.082		
Hsp62	-0.032	-0.051	0.075			
Hsp63	-0.071	-0.170	0.115	0.070		
Hsp64	0.110	0.240	0.106	0.012	yes	
Hsp65	0.040	0.101	0.120			
Hsp66	-0.015	-0.045	0.140			

(a) Results are sensitive to exclusion of transferred patients

The raw estimates have estimation error standard deviations ranging from 0.057 to 0.65, with a median of 0.11. Comparison of the above results with the raw case fatality rates show that hospitals with effects significantly different from zero, with few exceptions stand out by having a combination of low model standard estimate and high (or low) case fatality rate.

The shrinkage adjusted hospital effects (shrinkage estimates) have a population standard deviation of 0.065, as compared to the raw estimates' population standard deviation of 0.21. The first figure below shows estimated probability densities for both the raw and shrinkage estimates. We have indicated the indifference interval (i.e. the interval around zero where the odds of dying deviate no more than 10% from the average) as well as the alert limits (i.e. the points where the odds of dying deviate from the average by a factor of two), cf. section 5.6.2.1. In the second figure, the same data have been translated to an absolute risk scale. The hospital-specific probability of death refers to a hypothetical average patient, and is not necessarily equal to the average mortality for that hospital.

Figure 6-12: Estimated probability densities for hospital effect estimates.

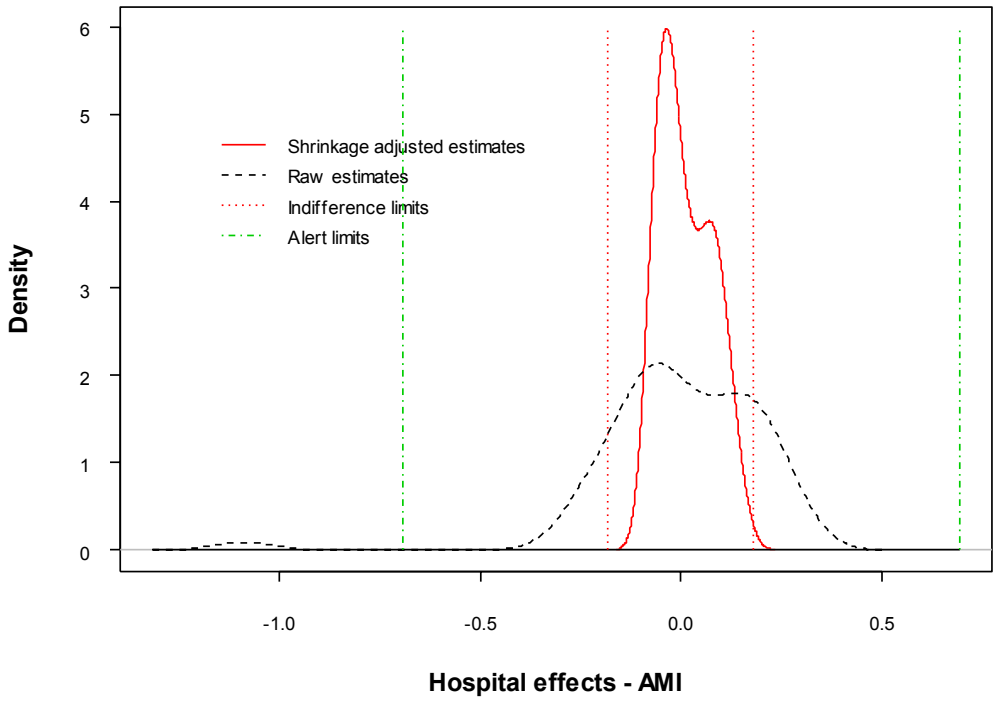
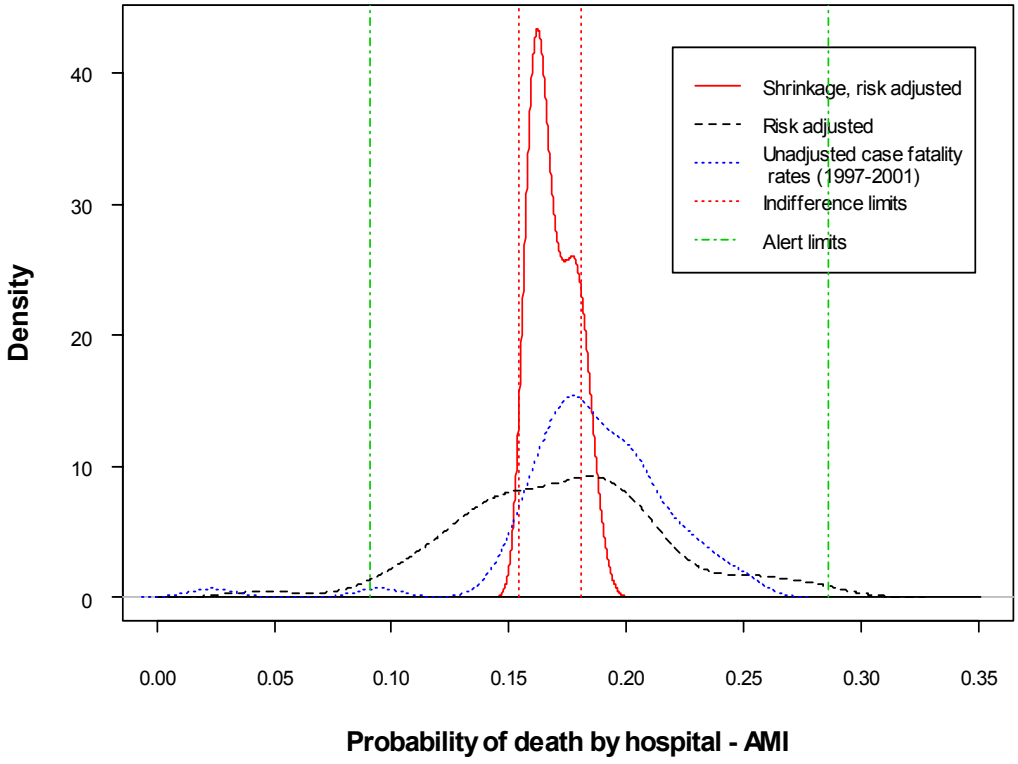


Figure 6-13: Estimated probability densities for hospital-specific probability of death



For reasons outlined earlier, the distribution of the shrinkage estimates must be regarded as the more accurate, overall representation of the set of Norwegian hospitals with respect to AMI 30-day mortality, despite their bias towards zero. Almost all of the

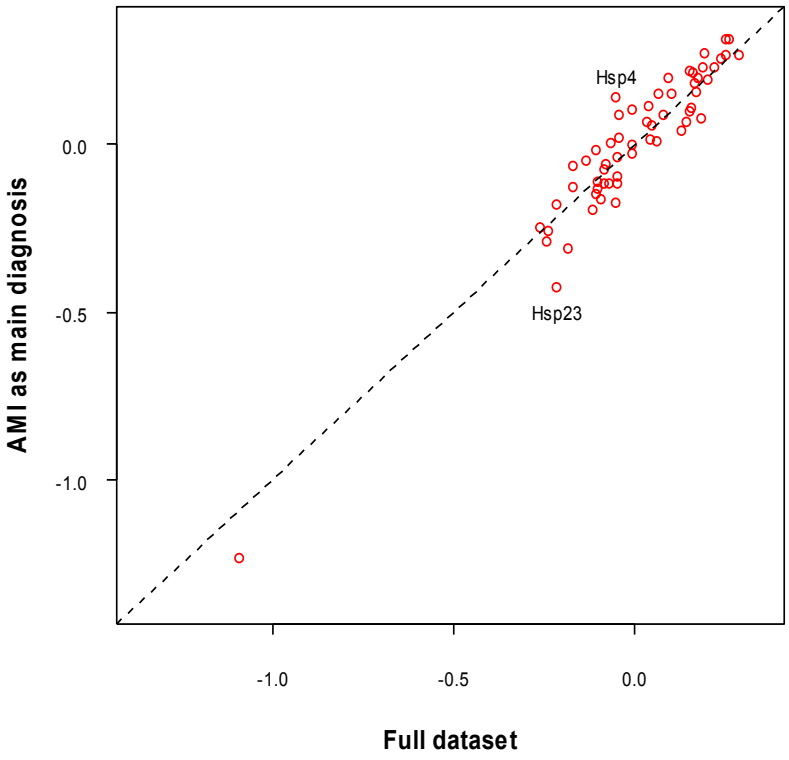
distribution is concentrated inside the indifference interval (-0.182, 0.182) (see 5.6.2.1). There is apparently a minor subset of hospitals with effects clustered somewhat above 0.

The raw estimates show a much larger spread, due to estimation variance. Based on the simulation experiment described in 5.6.4, we found that if the shrinkage estimates reported above were the actual, true effects, the chance of obtaining a sample maximum as small as we have observed here (0.158), is only 5%, and similarly for the sample range. We may thus conclude that the true spread in hospital effects is very likely to be larger than in the diagram above.

6.5.9.1 Restricting analysis to AMI as main diagnosis

The expert group suggested that there may be important clinical differences in acute myocardial infarction when it is recorded as main diagnosis or secondary diagnosis. Therefore the analysis was redone with only those cases that were recorded as main diagnosis, the proportion of which varied between hospitals from 0.47 to 0.94, with a mean of 0.85. There is a financial incentive to code AMI as secondary diagnosis, and this may be the reason for the somewhat surprisingly large spread. The figure shows estimated hospital effects for the new data set plotted against those of the full data set.

Figure 6-14: Change in hospital effects when AMI is main diagnosis.



The largest changes in hospital effects were -0.21 and 0.19, for Hsp23 and Hsp4 respectively. The changes are roughly one standard deviation. The AAD was 0.0576 and the rank correlation coefficient 0.926. The hospital effects have a slightly increased spread when data are restricted.

The conclusions regarding significance testing were strongly affected for four hospitals. For three of these, the proportion of cases with AMI as main diagnosis was either in the

top or bottom decile. A small proportion means that a relatively large part of the cases was excluded, so that changes in results are not surprising. A large proportion may also indicate a hospital with coding practices differing from the rest.

6.6 30-DAY MORTALITY AFTER STROKE

6.6.1 Inclusion/exclusion criteria

”Index admission” refers to the first admission for stroke (ICD-10 codes I61, I63, I64, or ICD-9, 431, 434 or 436 as main diagnosis) in the calendar year. Table 9-5 shows the distribution of the various codes among the index cases. One patient may have more index admissions if she/he has had new incidents of stroke in different years, or if she/he has been transferred, since each hospital will then be counted as a separate index admission. Patients readmitted within 28 days are not counted as admitted for a new stroke but counted as readmission for the same stroke as the index stroke. 53072 cases were classified as index admissions for stroke from these criteria. Of these cases 17.2% were followed by death within 30 days.

Criteria for exclusion were: 1) under 18 years of age, 2) patients dead by accident (external cause of death, violence etc.), 3) either dead on arrival or not an emergency case, and 4) admissions to hospitals with very few cases in the period 1997-2001. Hospitals with less than 100 cases after using the first three exclusion criteria were dropped. The final sample size was 49974 cases.

In the final data set, a small number of cases had missing values for one or more variables. These cases were included or excluded from analyses according to whether the particular analysis depended on the variables in question. Missing values were almost without exception associated with socio-demographic variables, mainly level of education.

In Appendix 9.1, we list the participating hospitals and the corresponding aliases used in tables and figures.

Table 6-28: Numbers of cases for stroke removed based on the exclusion criteria. Some exclusions relate to more than one criterion.

Exclusion criterion	Excluded hospitals	Excluded cases
Under 18 years of age	-	165
Admissions followed by accidental death	-	35
Not emergency or dead on arrival	-	2764
Hospitals with less than 100 admissions	5	70
Hospitals falling under 100 admissions after criteria 1-3 are applied	2	150[a]

a) remaining cases at these hospitals after applying criteria 1-3

Table 6-29: Number of cases in final data set and with complete socio-demographic variables for stroke.

Not Selected	Selected	Selected and with complete socio-demographic variables
3098	49974	49363

6.6.2 Observed mortality

Table 6-30 shows the variation in proportion of dead within 30 days between years at each hospital. Relative 30-day mortality has been rather stable, but there is an indication of a small decline towards the end of the period (Figure 6-15).

Table 6-30: Stroke. Variation 30-day case fatality for each hospital per year of admission. Percentages not shown if number of admissions is below 50 per year (Four hospitals with less than 50/year patients, each of the five years, are omitted).

Hospital	1997	1998	1999	2000	2001
Hsp01	11.1	16.3	28.6	24.3	23.4
Hsp03	20.2	22.7	19.7	18.5	17.8
Hsp04		32.1	27.6	15.4	23.9
Hsp05	16.0				
Hsp06	11.8	16.0	17.0	15.8	13.1
Hsp07	22.3	17.0	19.8	18.4	14.2
Hsp08	18.5	27.3	20.5	20.9	15.1
Hsp09	22.0	26.5	20.3	13.2	12.9
Hsp10	16.3	14.9	13.5	14.9	20.5
Hsp11	14.8	16.0	14.7	12.3	10.8
Hsp12	21.7	37.9	13.4	28.8	16.7
Hsp14	15.5	23.7	27.6	14.1	14.8
Hsp15			16.0	13.8	13.1
Hsp16	16.4	10.7	12.4	20.1	11.0
Hsp17	19.6	20.1	13.3	14.8	14.1
Hsp18	21.6	17.3	19.8	17.6	19.3
Hsp19	13.2	28.1	15.3	18.5	27.0
Hsp20	12.7	14.9	17.9	14.4	15.1
Hsp21	25.3	25.3	25.2	30.6	15.4
Hsp24	24.3	25.8	19.8	19.3	20.4
Hsp25		6.8			22.0
Hsp26	14.3	18.6	15.9	15.8	11.6
Hsp28	12.3	19.4	17.6	12.8	19.6
Hsp29	17.5	17.5	16.7	14.0	10.0
Hsp30	17.7	27.8	21.1	15.1	21.5
Hsp31	23.5	23.8	29.6	20.0	21.8
Hsp32	17.3	22.1	27.1	20.3	28.1
Hsp34	26.3	19.8	22.4	21.8	16.9
Hsp35	21.6	19.4	20.4	15.4	15.8
Hsp36	18.3	14.9	10.9	15.1	13.5
Hsp37	13.0	16.6	18.3	20.1	19.6
Hsp38	8.3			8.2	20.0
Hsp39	15.1	14.6	16.0	16.6	13.2
Hsp40	19.6	17.7	11.5	14.9	14.3
Hsp41			1.5		

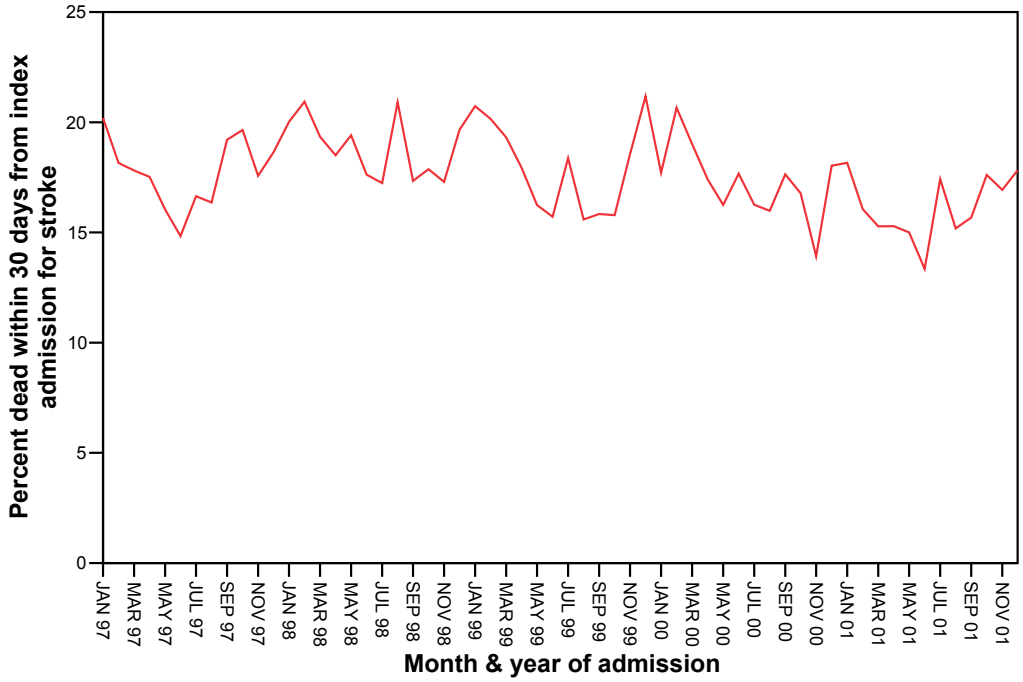
Hospital	1997	1998	1999	2000	2001
Hsp42	21.8	24.2			
Hsp43	13.2	17.4	16.5	15.5	19.0
Hsp44	18.9	24.6	20.4	13.4	8.8
Hsp46	15.8	12.7		14.5	16.4
Hsp47	24.5	20.4	18.7	21.4	14.2
Hsp48	27.1	16.4	14.7	19.0	16.4
Hsp49	18.8	18.0	18.3	19.6	23.4
Hsp50	17.0	17.6	17.9	18.0	15.3
Hsp52	18.5		16.0	20.3	17.2
Hsp53	14.6	13.6	17.6	12.5	9.1
Hsp55	22.2	20.6	21.3	21.8	15.3
Hsp56	16.1	16.7	12.4	18.7	10.6
Hsp57	14.6	14.1	16.9	12.4	12.5
Hsp60	15.8	25.1	19.3	18.1	24.9
Hsp61	16.3	21.4	22.4	23.9	17.3
Hsp62	24.3	29.6	21.3	18.5	15.7
Hsp63	19.2	15.6	12.3	13.8	13.1
Hsp64	19.4	15.6	27.6	26.7	27.3
Hsp65	13.5	11.1	18.0	15.6	22.7
Hsp66	25.0	17.2	14.0	19.2	19.7

Table 6-31: Stroke: Number of cases dead within 30 days for each hospital, together with the total number of index cases per year, after applying exclusion criteria.

Hospital	1997		1998		1999		2000		2001	
	Number of stroke cases	Dead within 30 days	Number of stroke cases	Dead within 30 days	Number of stroke cases	Dead within 30 days	Number of stroke cases	Dead within 30 days	Number of stroke cases	Dead within 30 days
Hsp1	243	27	295	48	168	48	177	43	205	48
Hsp3	203	41	207	47	208	41	195	36	214	38
Hsp4	40	11	56	18	58	16	52	8	67	16
Hsp5	75	12	5	1	15	2	8	4	4	1
Hsp6	203	24	206	33	188	32	209	33	213	28
Hsp7	94	21	112	19	96	19	98	18	113	16
Hsp8	92	17	132	36	151	31	182	38	159	24
Hsp9	109	24	98	26	118	24	114	15	116	15
Hsp10	270	44	249	37	274	37	308	46	283	58
Hsp11	548	81	562	90	539	79	568	70	529	57
Hsp12	69	15	87	33	82	11	66	19	66	11
Hsp14	71	11	76	18	76	21	71	10	88	13
Hsp15	.	.	33	5	181	29	196	27	191	25
Hsp16	268	44	244	26	249	31	224	45	245	27
Hsp17	515	101	512	103	540	72	494	73	504	71
Hsp18	148	32	173	30	162	32	159	28	187	36
Hsp19	53	7	64	18	85	13	54	10	63	17
Hsp20	71	9	87	13	84	15	97	14	86	13
Hsp21	83	21	99	25	107	27	98	30	117	18
Hsp23	43	11	40	13	34	8	42	7	50	15

Hospital	1997		1998		1999		2000		2001	
	Number of stroke cases	Dead within 30 days	Number of stroke cases	Dead within 30 days	Number of stroke cases	Dead within 30 days	Number of stroke cases	Dead within 30 days	Number of stroke cases	Dead within 30 days
Hsp24	202	49	229	59	237	47	264	51	240	49
Hsp25	43	9	59	4	49	5	48	10	50	11
Hsp26	224	32	220	41	207	33	240	38	241	28
Hsp28	162	20	160	31	199	35	219	28	158	31
Hsp29	143	25	103	18	102	17	121	17	120	12
Hsp30	62	11	72	20	57	12	73	11	79	17
Hsp31	170	40	147	35	142	42	130	26	170	37
Hsp32	75	13	86	19	59	16	69	14	96	27
Hsp34	228	60	227	45	250	56	238	52	266	45
Hsp35	301	65	346	67	294	60	325	50	292	46
Hsp36	60	11	67	10	64	7	53	8	52	7
Hsp37	146	19	151	25	175	32	244	49	224	44
Hsp38	72	6	45	6	26	6	61	5	60	12
Hsp39	503	76	514	75	550	88	523	87	507	67
Hsp40	56	11	62	11	78	9	74	11	70	10
Hsp41	47	5	47	7	65	1	46	3	35	1
Hsp42	101	22	95	23	43	14	1	0	.	.
Hsp43	560	74	603	105	594	98	504	78	421	80
Hsp44	164	31	142	35	142	29	134	18	170	15
Hsp46	76	12	55	7	44	5	69	10	55	9
Hsp47	208	51	372	76	364	68	370	79	330	47
Hsp48	59	16	73	12	68	10	63	12	55	9
Hsp49	240	45	250	45	224	41	189	37	171	40
Hsp50	224	38	239	42	268	48	289	52	242	37
Hsp51	26	1	24	2	40	12	33	9	44	13
Hsp52	65	12	44	8	50	8	64	13	64	11
Hsp53	192	28	176	24	165	29	160	20	175	16
Hsp55	221	49	432	89	385	82	385	84	391	60
Hsp56	218	35	216	36	186	23	198	37	180	19
Hsp57	378	55	397	56	396	67	412	51	449	56
Hsp58	34	8	43	10	20	2	36	6	25	3
Hsp59	24	6	26	6	29	8	20	7	25	6
Hsp60	203	32	191	48	187	36	182	33	169	42
Hsp61	166	27	201	43	170	38	159	38	191	33
Hsp62	218	53	199	59	240	51	286	53	331	52
Hsp63	130	25	154	24	130	16	145	20	122	16
Hsp64	165	32	147	23	116	32	131	35	110	30
Hsp65	104	14	108	12	128	23	122	19	110	25
Hsp66	80	20	87	15	86	12	78	15	76	15

Figure 6-15: 30-day case fatality of stroke as a function of month and year of admission (percent).



6.6.3 Distribution of covariates

Summary measures of the distribution of risk adjustment covariates are shown in Table 6-32 below. The table shows minimum, maximum and standard deviation between hospitals (i.e. the various statistics for the hospital means), as well as the overall mean and standard deviation, for each covariate. For categorical covariates, the indicator variables for each category (except the first, reference category) are used. Apparently, most of the variability is on the individual rather than on the hospital level (a notable exception is distance between home and hospital).

Table 6-32: Summaries for stroke covariates. Categorical variables in percentage units.

Variable	Between hospital min	Between hospital max	Between hospital std. dev	Overall std.dev	Overall mean
Age	60.59	79.25	2.55	11.76	74.56
Sex: female	40.83	62.20	3.57	50.00	49.91
Marital status: not married	5.13	19.24	2.89	29.52	9.65
Marital status: divorced or separated	2.70	16.77	2.98	27.43	8.19
Marital status: widowed	14.59	43.11	4.16	47.61	34.72
Maximum education in household (years)	9.58	13.68	0.74	2.78	10.82
Natural logarithm of: household income + 1	11.59	12.23	0.12	0.90	11.83

Natural logarithm of: household property/capital + 1	11.03	13.01	0.27	2.17	12.22
Natural logarithm of: distance from home to hospital + 1	0.31	2.28	0.43	0.85	1.03
Number of preadmissions regardless of cause, weighted	0.12	0.42	0.06	0.48	0.30
Number of pertinent codiagnoses at previous admission, weighted	0.01	0.16	0.03	0.21	0.12
Hemorrhage	8.65	40.00	4.07	33.57	12.94
Transferred from another hospital	0.00	43.75	6.70	20.99	4.62

6.6.4 Model selection

Starting with a model having interactions between hospital and year (i.e. with a separate intercept for each year for each hospital), we performed a stepwise backward model selection. The largest model (M6 in Table 6-34) was only fitted to the reduced data set consisting of the ten largest hospitals. With this exception, the final model selection was performed on the complete data set.

Note that we do not make any a priori assumptions on the yearly variation of performance. Both smooth and sudden changes are possible, and the hospitals may evolve independently over time.

In each step of the selection process, variables or (predefined) sets of variables were removed from the model when the standard log-likelihood (deviance) test, comparing the resulting model to the previous one, was non-significant at the 5% level. When the test was significant, the selection process stopped.

After inspection of the estimated coefficients for all three diseases, ethnicity appeared non-significant. The socio-demographic variables were then regrouped so that ethnicity could be tested separately.

The resulting candidate model was checked with a Hosmer-Lemeshow goodness-of-fit test, showing poor fit, particularly for cases with high, predicted mortality. Since type of stroke has such a large effect, it was suspected that the lack of fit was related to this variable. Splitting the data set by type of stroke, and comparing separate models for each part, led to the consideration of models with interaction terms particularly between type of stroke and age (the other variable with large effect). The selection process was therefore repeated with a new set of models extended with interaction terms.

The backward selection process for the extended a priori model is summarized in the tables below.

Table 6-33: Variable names and types.

Variable set	Variable name	Type of variable	Description
	yearfactor	Categorical	The years 1997-2001 as factor
	yeardiff	Continuous	Year – 2001 (a negative number)
	hospital	Categorical	Hospital code 1…66

Variable set	Variable name	Type of variable	Description
Age/gender	N03	Continuous	B-spline basis function (of age)
	N13	Continuous	B-spline basis function (of age)
	N23	Continuous	B-spline basis function (of age)
	N33	Continuous	B-spline basis function (of age)
	sex	Categorical	Distribution of the sexes
Socio-demographic	marsta	Categorical	Marital status
	maxedu	Continuous	Maximum education in household
	lninek	Continuous	Natural logarithm of: household income + 1
	lnwrth	Continuous	Natural logarithm of: household property/capital + 1
	ethnicnor	Categorical	Born in Norway (Yes/No)
Frailty	preintot	Continuous	Number of preadmissions regardless of cause
	preantdiagn	Continuous	Number of pertinent codiagnoses at previous admission
	dementia	Categorical	Yes / No
Severity	logdistance	Continuous	Natural logarithm of: distance from home to hospital + 1
	hemorrhage	Categorical	Yes / No
	moved in	Categorical	Transferred from another hospital. Yes / No

Table 6-34: Description of candidate models included in analysis.

Model name	Variables included in model	Interactions
M6	age/gender set, socio-demographic set, frailty set, severity set, second order terms in preadmiss and prenumbdiagn	(age/gender set, socio-demographic set, frailty set): hemorrhage, Yearfactor:hospital
M7	age/gender set, socio-demographic set, frailty set, severity set, yeardiff, second order terms in yeardiff, preadmiss and prenumbdiagn	(age/gender set, socio-demographic set, frailty set): hemorrhage, (Yeardiff, Yeardiff squared):hospital

M8	age/gender set, socio-demographic set, frailty set, severity set, yeardiff, second order terms in yeardiff, preadmiss and prenumbdiagn	(age/gender set, socio-demographic set, frailty set): hemorrhage, Yeardiff:hospital
M9	age/gender set, socio-demographic set, frailty set, severity set, yeardiff, second order terms in yeardiff and preadmiss	(age spline functions, prenumbdiagn): hemorrhage, Yeardiff:hospital
M17	age/gender set, socio-demographic set excluding ethnicnor, frailty set excluding dementia, severity set, yeardiff, second order terms in yeardiff and prenumbdiagn	(age spline functions, prenumbdiagn): hemorrhage
M31	age/gender set, frailty set excluding dementia, severity set, yeardiff, second order terms in yeardiff and prenumbdiagn	(age spline functions, prenumbdiagn): hemorrhage

Table 6-35: Analysis of deviance for stroke model selection.

Data set	Model name	Resid. Df	Resid. Deviance	Df	Deviance change	P-value	AIC change
subset	M6	19746	15351.96				
subset	M7	19766	15374.47	-20	22.52	0.313	-17.48
subset	M8	19775	15380.98	-9	6.51	0.688	-11.49
subset	M9	19784	15393.37	-9	12.39	0.192	-5.61
subset	M17	19787	15396.86	-3	3.49	0.322	-2.51
all	M7	49145	39601.50				
all	M8	49203	39658.04	-58	56.54	0.530	-59.46
all	M9	49212	39674.51	-9	16.47	0.058	-1.53
all	M17	49215	39675.37	-3	0.86	0.835	-5.14
all	M31	49279	39893.05	-64	217.68	<1e-4	89.68

The stepwise procedure selected as the final model M17, which includes interactions between hospital and year. Using the minimum AIC criterion leads to the same conclusion. All 49363 cases with complete (non-missing) data were used in fitting this model. Of these, 7 had weights exactly zero.

We do not know if a patient lived in a nursing home or not. However, it is reasonable to believe that dementia to some extent is a proxy for living in a nursing home. In this context, it is interesting to note that the variable dementia was not significant.

The order of the splines used for age effect was tested by comparing the AIC of the final model with the AIC using degree 4 splines (ie two internal knots, placed at the 33% and 67% percentiles 72 and 81). With an increase in AIC by 3.18, the degree 4 splines were rejected.

Two measures of the final model's prediction ability were computed: area under the receiver operating characteristic, or C statistic, and the optimal rate of correct classification.

Table 6-36: Stroke final model – measures of prediction ability.

C statistic	0.73
Classification rate	0.84

Hosmer-Lemeshow goodness-of-fit statistics were computed for the final model. The values are good, especially in light of the large sample size.

Table 6-37: Hosmer-Lemeshow goodness-of-fit statistics for final stroke model.

Data set	C statistic	Df	P-value
Subset	6.47	8	0.594
All	13.65	8	0.091

To check for influential observations, various influence measures were computed for the final model. The deviance residuals ranged from -1.69 to 2.98, which give no cause for concern. The largest leverage hat statistic was 0.11. 14 observations showed high or very high hat values, above 0.05. All these cases came from the Hsp5 hospital, which only had a total of 107 cases and a high hospital effect standard deviation (see Table 6-41 below). As noted earlier, it is reasonable to standardize hat values by the number of cases per hospital. Looking at cases with hat value greater than 0.01 and standardized hat value greater than 3, we find a set of 1450 cases with moderately high leverage. This set belongs to the hospitals Hsp5 and Hsp55, has a large proportion of transferred patients, and is otherwise unremarkable. Hsp55 accounts for 25% of the total number of the transferred cases, meaning that it is reasonable to expect cases from this hospital to have increased influence on the model, via the variable "moved in". The maximum Cook's distance was 0.0033, and 11 cases were above 0.001. These cases came from small hospitals, meaning that their leverage will be high, and had a very high mortality. This may be a manifestation of the relatively poor fit of the model for the very high mortalities, as noted earlier. The influential cases were judged to be acceptable and therefore retained in the model.

6.6.5 Covariate parameter estimates

The parameter estimates for risk adjustment covariates are displayed in the following table:

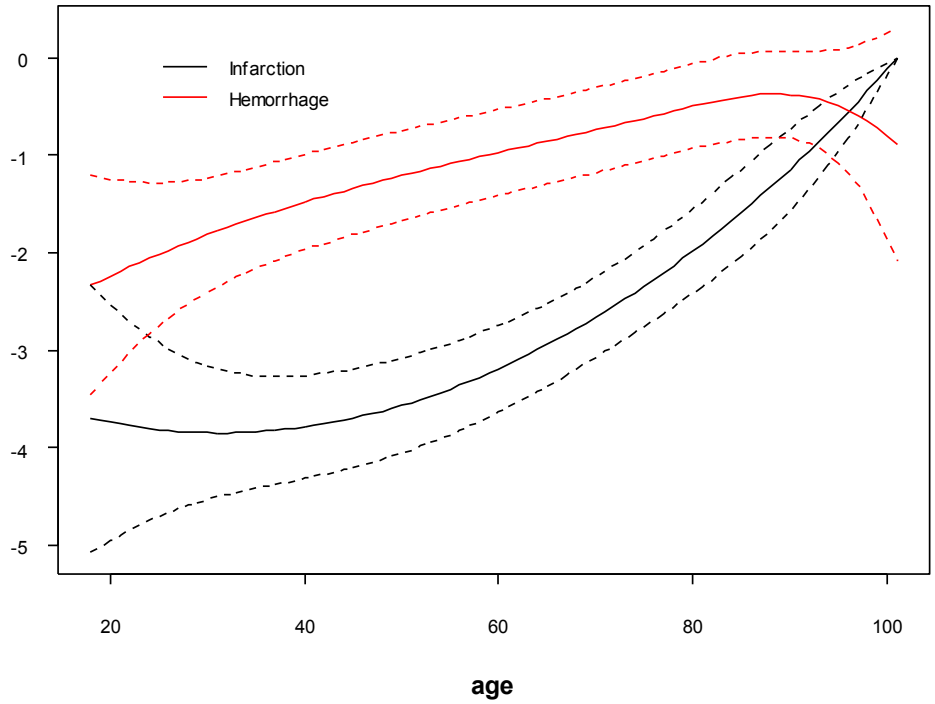
Table 6-38: Final stroke model – parameter estimates.

Model term		Estimate	Std. Error	Z value	Two-sided p-value.
(Intercept)		1.314	0.341	3.85	0.00012
N03	age function	-3.702	0.701	-5.28	<1e-4
N13	age function	-4.137	0.481	-8.59	<1e-4
N23	age function	-3.369	0.192	-17.55	<1e-4

Model term		Estimate	Std. Error	Z value	Two-sided p-value.
N33	age function	-0.935	0.287	-3.26	0.00113
hemorrhage		-0.891	0.614	-1.45	0.147
Yeardiff	=year-2001	-0.144	0.034	-4.25	<1e-4
Yeardiff squared		-0.026	0.0077	-3.38	0.00072
female sex		0.041	0.028	1.44	0.151
maxedu		-0.011	0.0053	-2.12	0.034
logincome		-0.073	0.025	-2.97	0.00300
logworth		-0.028	0.0070	-3.96	<1e-4
logdistance		0.000	0.018	-0.21	0.837
preadmiss		0.128	0.031	4.13	<1e-4
prenumbdiagn		1.430	0.131	10.89	<1e-4
prenumbdiagn squared		-0.571	0.134	-4.27	<1e-4
moved in		-0.565	0.129	-4.38	<1e-4
married/cohabitant		-0.022	0.025	-0.89	0.375
not married		0.198	0.034	5.87	<1e-4
divorced or separated		-0.124	0.042	-2.96	0.00306
widowed		-0.052	0.025	-2.07	0.038
N03:hemorrhage interaction	interaction terms between age function and type of stroke	2.256	0.994	2.27	0.023
N13:hemorrhage interaction		3.671	0.943	3.89	<1e-4
N23:hemorrhage interaction		3.312	0.523	6.33	<1e-4
N33:hemorrhage interaction		1.689	0.754	2.24	0.025
hemorrhage:prenumbdiagn interaction	interaction term between number of previous diagnoses and type of stroke	-0.410	0.157	-2.60	0.00924

The dependency of mortality on age, for the two types of stroke, is shown in Figure 6-16 below. The main effect of stroke type is incorporated in the plot. Note that the figure must be interpreted with some care because of the remaining interactions between type of stroke and number of previous diagnoses (Strictly interpreted, the figure applies to the case with no previous diagnoses).

Figure 6-16: Final stroke model – joint effect of age and type of stroke (with uncertainty bands).



The mortality increases strongly with age, less sharply so for hemorrhages than infarctions. From the figure, we see that having hemorrhage leads to a large increase in mortality (The estimated curves do not meet at the right end of the age interval, but as is apparent from the uncertainty band, this is due to estimation uncertainty. There are relatively few cases with very high age). The number of previous diagnoses also has a relatively strong effect on mortality. Relatively smaller increases in mortality follow from having previous hospital admissions and from never having been married or cohabitating.

Being transferred into the hospital has a fairly large effect in the direction of reduced mortality, and must be viewed as an indicator of reduced disease severity. The likely explanation is that on the balance, transfers tend to come after the initial period of high mortality. It is thus the less severely ill patients that tend to be transferred,

The socio-demographic variables have relatively small, but significant, effects. They act, as would be expected, to reduce mortality with increasing education, income or fortune. Distance from home to hospital, as well as gender, has no significant effect. These two variables are still retained in the model for the sake of consistency between the three diseases studied.

Distance from home to hospital is not significant. A priori, we may think of two mechanisms for an effect of distance on probability of death: One is the selection of patients that have survived a long transport and therefore are less severely ill. The other mechanism, widely believed to exist, that doctors do not always send stroke patients to hospital if the distance is very large. Our data do not confirm either of these mechanisms. However, it may be that they just happened to cancel each other out in our data.

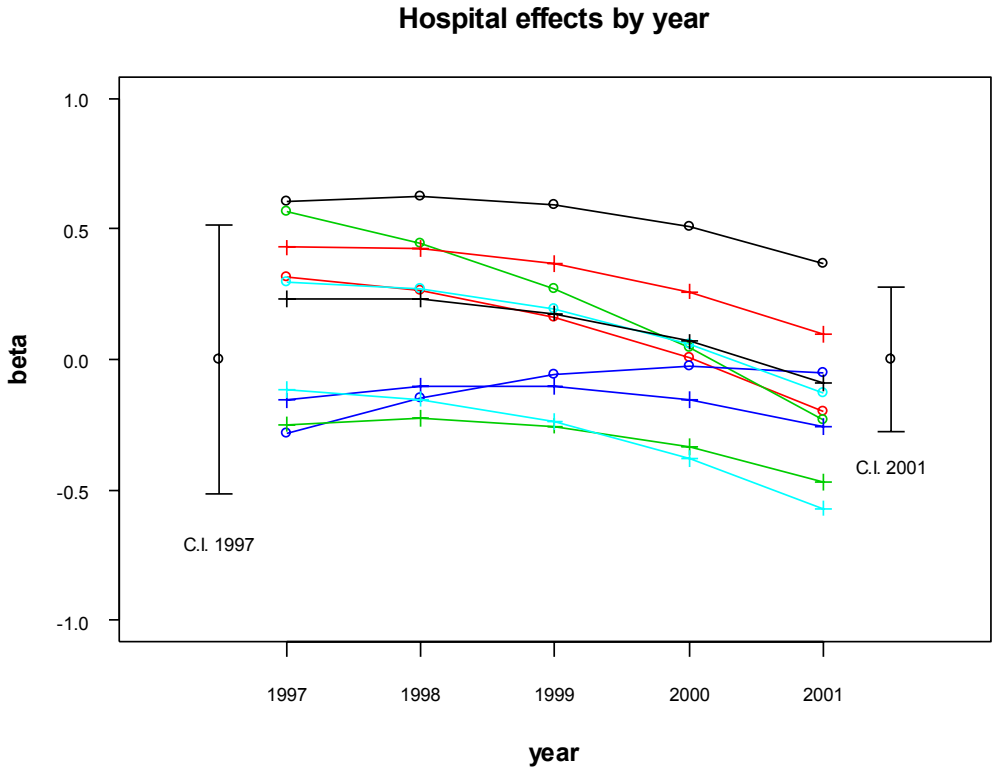
The number of previous admissions and previous diagnoses act as expected to increase the risk of dying. The reason for the second order term in the number of previous diagnoses is that this variable does not act linearly on mortality, rather has an effect that is linear at first but more or less stops to increase after a certain number of previous diagnoses are reached.

6.6.6 Variation of hospital performance over time

An important question in assessing the quality indicators is whether they show stability over time or have large temporal variability with no reasonable explanation. In the model, temporal variability not explained by risk adjustment covariates is either assigned to random variation or incorporated in the year/hospital part of the model.

The figure below illustrates the effects of calendar year and hospital. If all interactions between hospital and Yeardiff were zero, all curves would have been parallel. The individual interaction parameters ranged from -0.392 to 0.426, with 1st and 3rd quartiles of -0.074 and 0.093 respectively. For clarity, only the ten largest hospitals are shown.

Figure 6-17: Effect of hospital and year for the ten largest hospitals. Confidence intervals are indicated for the start and end years only. These must be superimposed on every curve to display their uncertainty.



The overall downward trend follows a second-order polynomial (in year-2001) with a first-order term significantly less than zero. The trend is therefore strictly decreasing towards the end of the period. Because of the presence of (statistical) interactions, it should be noted that the downward trend must be interpreted as an average trend (This average is hospital- and not case-weighted). The overall downward trend is more rapid in the latter part of the period.

The polynomial trend seems fairly consistent among the hospitals, with a few exceptions. We have not made a thorough, formal analysis of the variability of this pattern, except for making the observation that only a small number (5) of the individual interaction terms are significantly different from zero when judged by separate t-tests at the 1% level.

Also, note the relatively large uncertainty, shown only for the start and end of the observation period. One cannot therefore draw any conclusions about the apparent maxima, minima and reversals of time trend. Note that the way we have formulated our model is not symmetrical with respect to time. This introduces certain skewness in the uncertainty, with higher precision in the final year, as apparent from the figure.

From the calculations of residual autocorrelations reported in section 6.10, it follows that there is no indication of any monthly or weekly pattern of variation except random fluctuations.

It is inherent in the definition of the quality indicator that it be a single number, applying to a certain time-period. However, as in the present case, the data may reveal information about differences in trend, which may be an important part in the overall hospital performance picture.

6.6.7 Model robustness

The number of codiagnoses, (a sum of pertinent codiagnoses from previous admissions plus those under the current admission - but excluding those that may be considered an effect of treatment) was initially used as a proxy for patient frailty. When included in the model, this variable had a parameter estimate of -0.060, significant at the 0.1% level, i.e. in the direction of reduced risk, which obviously is contrary to expectation. This variable, which is discussed further in section 6.8.1 below, was excluded from the final model.

Hospital effects

Hospital effects are the log-odds of death within 30 days at the various hospitals, compared to the average hospital, controlling for risk adjustment covariates (see chapter 5.6).

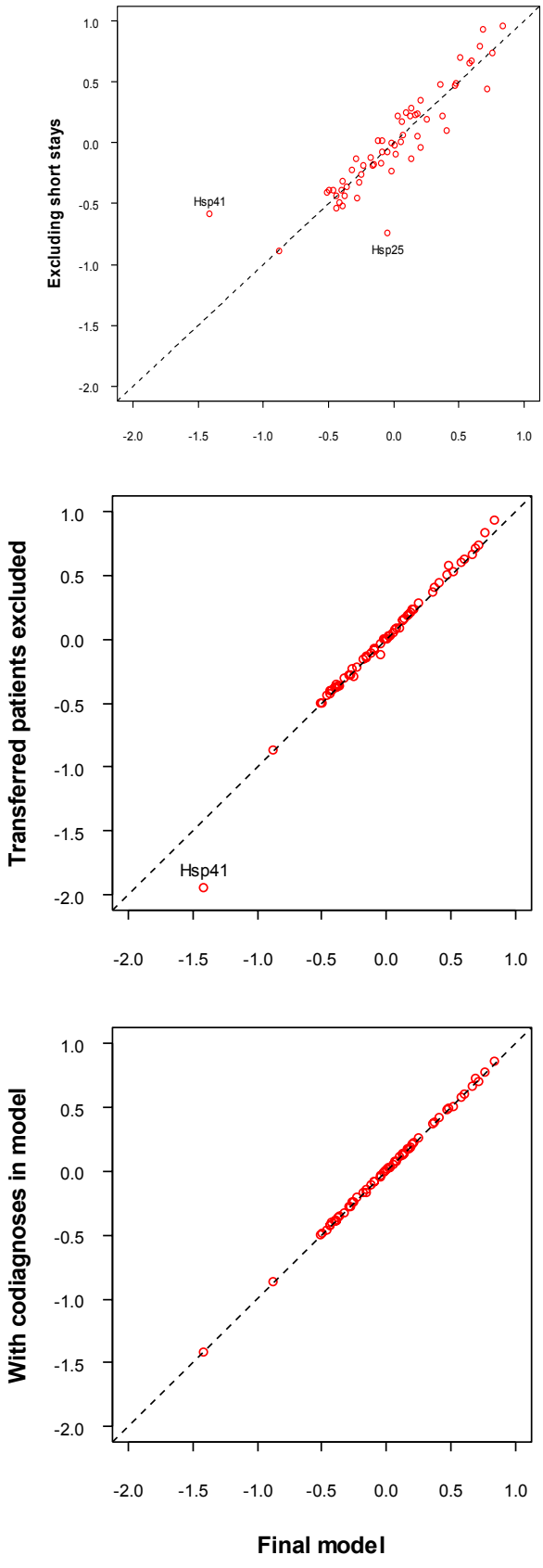
An effect of e.g. -0.182 means that this particular hospitals has 10% reduced log-odds for death, while an effect of e.g. 0.41 means that the log-odds of death are 50% greater than in the average hospital, given the levels of the risk adjustment covariates.

We were concerned about the robustness of our analysis with respect to two factors:

- inclusion/exclusion of transferred patients
- inclusion/exclusion of the risk adjustment covariate “number of codiagnoses excluding complications”
- inclusion/exclusion of short admissions (<2 days). Some of our concerns about data quality apply to short stays in particular, see 6.9.2 below.

To examine the robustness of our final model, we compared the hospital effect estimates of the final model with estimates obtained after varying the above factors. The figures below show the results. The hospitals with apparently large changes are labeled.

Figure 6-18: Robustness of final stroke model. Hospital effects from alternative models against effects from final model.



The changes in hospital effects are very small, both with respect to exclusion of transferred patients and inclusion of codiagnoses in the model.

The exception is Hsp41. Without transferred patients, the effect for Hsp41 decreased by 0.54. This is in fact smaller than the estimate's standard deviation under the final model. The proportion of transferred patients is very large in this hospital.

When excluding stays less than two days, the changes are somewhat larger. Again, the effect for Hsp41 is particularly sensitive. The performance of Hsp25 seems improved, although this is a small hospital with a correspondingly large standard deviation.

The table shows the statistical measures of the changes (see 5.6.5).

Table 6-39: Robustness of final stroke model. Statistical measures of changes in hospital effects.

Measure	Transferred patients excluded	Model without codiagnoses	Excluding short stays
AAD	0.026	0.007	0.120
RAAD	0.077	0.023	0.366
correlation	0.991	1.000	0.900
rank correlation	0.998	0.999	0.926

It had been suggested to exclude the patients that had died by day 5. The resulting average absolute change (AAD) in hospital effects was 0.163.

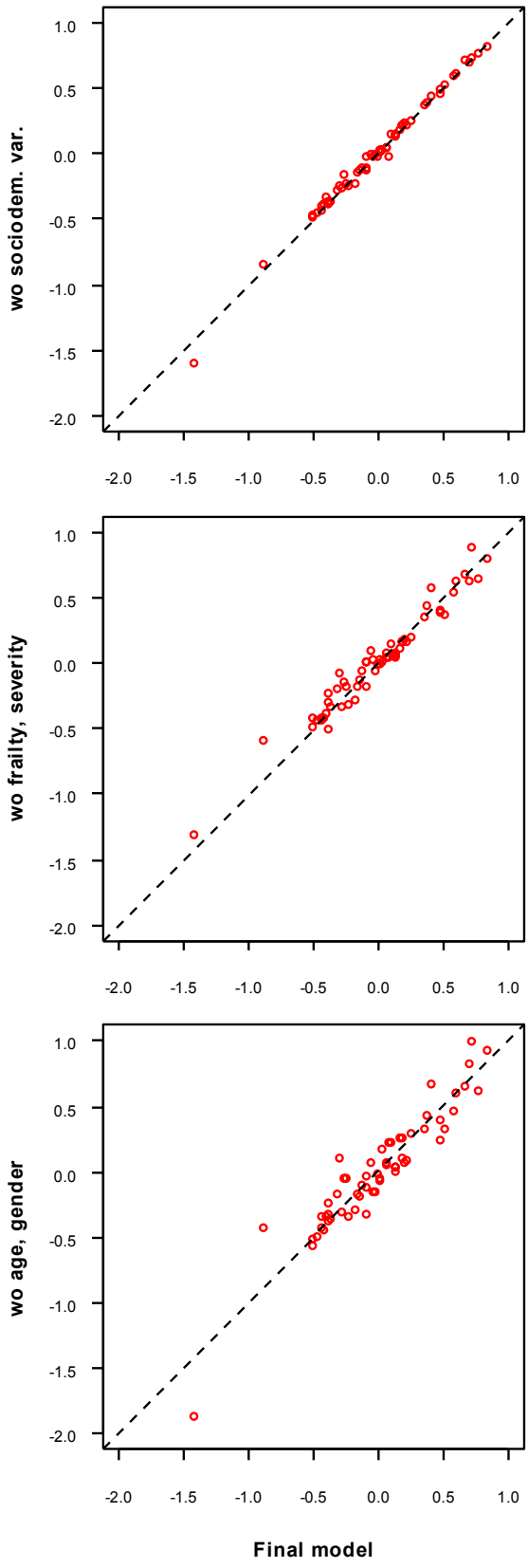
We concluded that the model robustness was acceptable, with the provision that for Hsp41, the estimated hospital effect was highly dependent on the inclusion/exclusion of transferred patients and thus must be regarded as unreliable.

6.6.8 Comparison of results with and without risk adjustment

To estimate the magnitude of the bias in the risk adjustment, we compute various bias indicators, following (1). We have investigated the change in hospital effect estimates when successively smaller sets of risk adjustment covariates are used. As explained in 6.9.1, we may regard these changes as plausible upper bounds for the true bias.

The figure below shows estimates from our final model plotted against those of the reduced models. First, we remove socio-demographic variables, then both socio-demographic and frailty/severity variables, and last, also age and gender, leaving no risk adjustment covariates in the model.

Figure 6-19: Comparison of results for stroke with and without risk adjustment. Hospital effects from reduced models against effects from final model.



The overall magnitude of the changes are very small when socio-demographic variables are excluded, fairly small when frailty and severity variables are also excluded, and moderate when age and gender are excluded in addition.

Quantitative measures of the change in effects are shown in the table below (see 5.6.5).

Table 6-40: Indicators for bias in stroke risk adjustment. Statistical measures of changes in hospital effects.

Measure	Socio-demographic set excluded	Frailty/severity set also excluded	Age/gender set also excluded
AAD	0.022	0.068	0.110
RAAD	0.069	0.220	0.343
correlation	0.996	0.979	0.941
rank correlation	0.994	0.979	0.941

The overall conclusion is that assessment of relative hospital performance is not very sensitive to the exclusion of socio-demographic variables, a finding that was also seen elsewhere (80). Excluding frailty and severity indicators, results in somewhat larger changes in hospital effect estimates. The additional changes due to exclusion of age and gender are of the same order.

6.6.9 Hospital effects

To judge the importance of the estimated hospital effects (defined in 5.4.2), the three decision procedures described earlier were applied:

- the single hospital rule: testing one hospital by the standard test. A one-sided p-value, if less than 10%, is reported
- the follow-up list rule: testing each hospital by the standard test, but with level according to the number of cases. A listed hospital is reported
- the multiple hypotheses testing rule: all effects are tested simultaneously by a multiple testing procedure. A simultaneous p-value, if less than 10%, is reported.

Hospital effects

Hospital effects are the log-odds of death within 30 days at the various hospitals, compared to the average hospital, controlling for risk adjustment covariates (see chapter 5.6).

An effect of e.g. -0.182 means that this particular hospitals has 10% reduced log-odds for death, while an effect of e.g. 0.41 means that the log-odds of death are 50% greater than in the average hospital, given the levels of the risk adjustment covariates.

The table below shows the shrinkage and raw estimates, the raw estimate standard deviation, the decision procedure results and number of cases per hospital.

Table 6-41: Stroke final model – hospital effect estimates (positive estimate indicates increased mortality).

Hospital	Shrinkage estimate	Raw estimate	Raw estimate Std.Error	P-value (one-sided)	Listed for follow-up	P-value (multiple testing)
Hsp1	0.477	0.605	0.141	<1e-4	yes	0.0010
Hsp3	0.103	0.134	0.151			
Hsp4	0.138	0.259	0.255			
Hsp5	0.123	0.840	0.658			
Hsp6	-0.032	-0.043	0.162			
Hsp7	-0.094	-0.155	0.219			
Hsp8	0.098	0.137	0.172			
Hsp9	-0.240	-0.392	0.217	0.035	yes	
Hsp10	0.112	0.138	0.131			
Hsp11	-0.391	-0.457	0.112	<1e-4	yes	0.0030
Hsp12	0.017	0.032	0.245			
Hsp14	-0.137	-0.245	0.242			
Hsp15	-0.182	-0.276	0.196	0.080		
Hsp16	-0.129	-0.172	0.157			
Hsp17	-0.075	-0.087	0.110			
Hsp18	0.044	0.060	0.163			
Hsp19	0.202	0.380	0.256	0.069		
Hsp20	-0.050	-0.092	0.250			
Hsp21	0.124	0.185	0.191			
Hsp23	0.289	0.585	0.275	0.017	yes	
Hsp24	0.152	0.190	0.136	0.082		
Hsp25	-0.020	-0.046	0.316			
Hsp26	-0.278	-0.372	0.158	0.00931	yes	
Hsp28	-0.007	-0.010	0.168			
Hsp29	-0.258	-0.434	0.225	0.027	yes	
Hsp30	-0.159	-0.287	0.245			
Hsp31	0.386	0.517	0.158	0.00054	yes	0.058
Hsp32	0.466	0.762	0.217	0.00022	yes	0.025
Hsp34	0.169	0.212	0.136	0.060		
Hsp35	-0.011	-0.013	0.133			
Hsp36	-0.157	-0.387	0.329			
Hsp37	0.367	0.475	0.147	0.00063	yes	0.067
Hsp38	-0.116	-0.257	0.300			
Hsp39	-0.126	-0.145	0.107	0.087		
Hsp40	-0.110	-0.228	0.284			
Hsp41[a)]	-0.185	-1.410	0.701	0.022	yes	
Hsp42	0.099	0.724	0.686			
Hsp43	0.056	0.064	0.107			

Hospital	Shrinkage estimate	Raw estimate	Raw estimate Std.Error	P-value (one-sided)	Listed for follow-up	P-value (multiple testing)
Hsp44	-0.326	-0.508	0.203	0.00630	yes	
Hsp46	-0.171	-0.385	0.304			
Hsp47	0.018	0.022	0.123			
Hsp48	-0.149	-0.318	0.290			
Hsp49	0.131	0.172	0.152			
Hsp50	0.006	0.008	0.143			
Hsp51	0.312	0.693	0.301	0.011	yes	
Hsp52	0.106	0.209	0.268			
Hsp53	-0.322	-0.492	0.198	0.00639	yes	
Hsp55	0.403	0.481	0.120	<1e-4	yes	0.0036
Hsp56	-0.311	-0.438	0.174	0.00604	yes	
Hsp57	-0.299	-0.357	0.120	0.00149	yes	
Hsp58	-0.233	-0.878	0.453	0.026	yes	
Hsp59	0.141	0.411	0.377			
Hsp60	0.275	0.363	0.154	0.00924	yes	
Hsp61	0.058	0.077	0.155			
Hsp62	-0.098	-0.120	0.130			
Hsp63	-0.258	-0.416	0.213	0.025	yes	
Hsp64	0.472	0.668	0.176	<1e-4	yes	0.0086
Hsp65	0.066	0.100	0.198			
Hsp66	-0.047	-0.086	0.249			

a) Results are sensitive to the exclusion of transferred patients

The estimated effects have estimation error standard deviations ranging from 0.11 to 0.70, with a median of 0.20. Comparison of the above results with the raw case fatality rates show that whether a hospital has effects significantly different from zero does not have a simple dependence on the case fatality rate nor the effect's standard deviation, as long as it is not very high. This is not surprising in view of the relatively large changes over the period.

The shrinkage adjusted hospital effects have a population standard deviation of 0.22, as compared to the raw estimates' population standard deviation of 0.42. Figure 6-20 below shows estimated probability densities for both the raw and the shrinkage estimates. We have indicated the indifference interval (i.e. the interval around zero where the odds of dying deviate no more than 10% from the average) as well as the alert limits (i.e. the points where the odds of dying deviate from the average by a factor of two), cf. section 5.6.2.1. In Figure 6-21, the same data have been translated to an absolute risk scale. The hospital-specific probability of death refers to a hypothetical average patient, and is not necessarily equal to the average mortality for that hospital.

Figure 6-20: Estimated probability densities for hospital effect estimates.

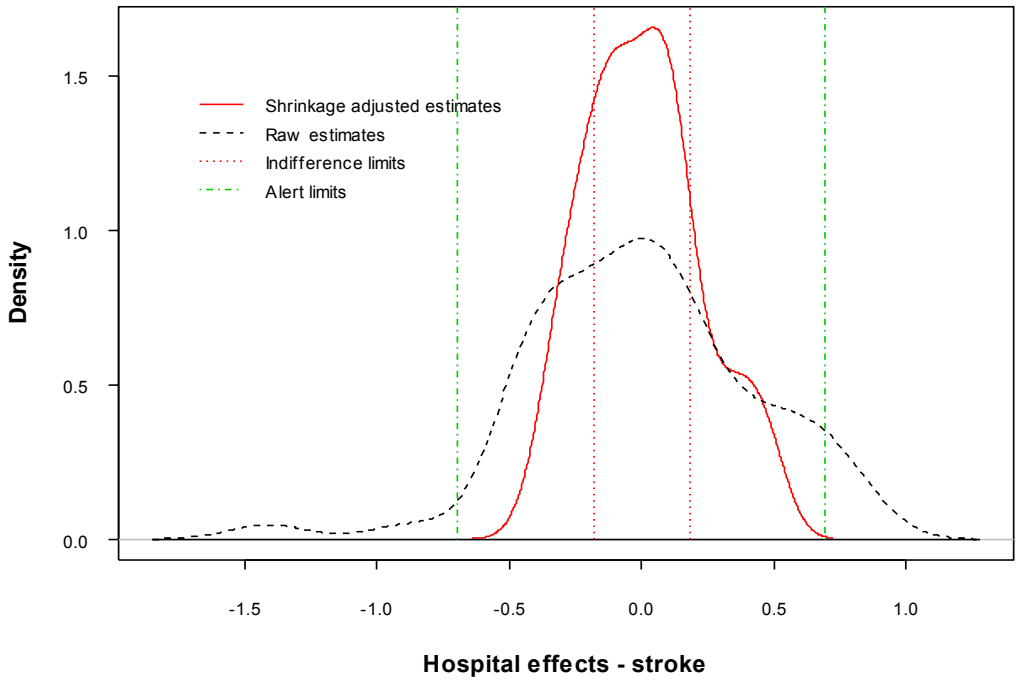
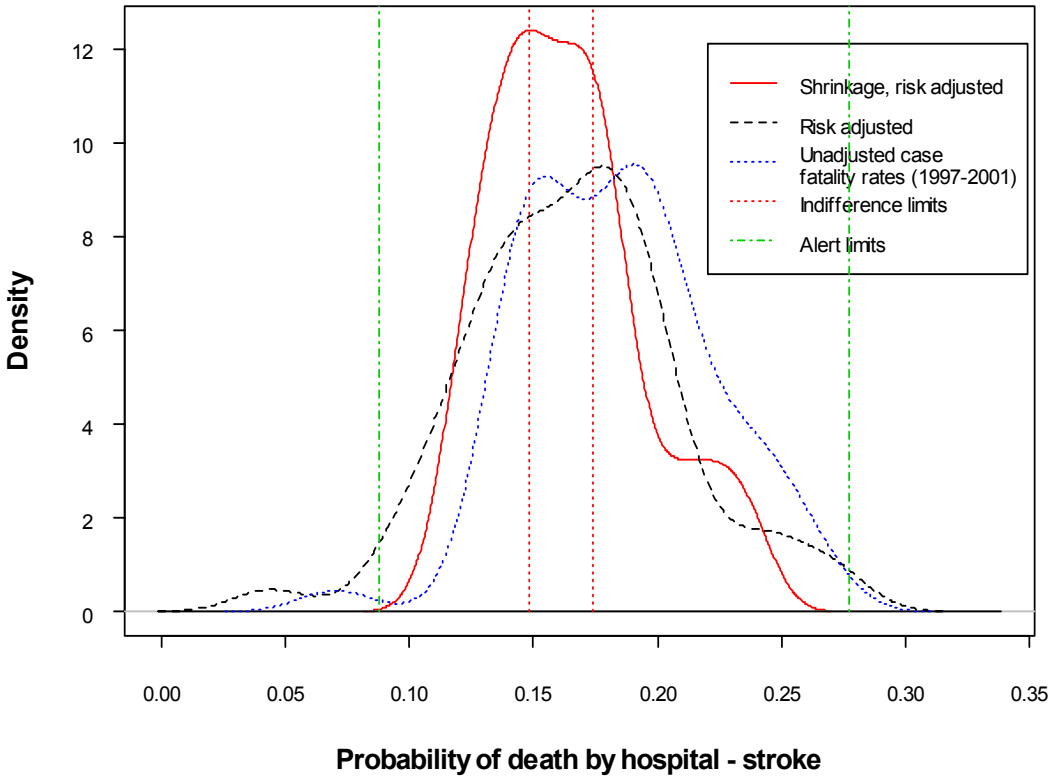


Figure 6-21: Estimated probability densities for hospital-specific probability of death.



For reasons outlined earlier, the distribution of the shrinkage estimates must be regarded as the more accurate, overall representation of the set of Norwegian hospitals with respect to stroke mortality, despite their bias towards zero. A considerable proportion of

the hospitals falls outside the indifference interval (-0.182, 0.182). There is apparently a small subset with close and high values, somewhat below 0.5.

The estimated effect distribution extends to very low (≈-0.5) and very high (≈-0.6) values. It should be noted that this range corresponds to a relative odds ratio of 3, which is considerable. Based on the simulation experiment described in 5.6.4, we found that if the shrinkage estimates reported above were the actual, true effects, the chance of obtaining a sample maximum as small as we have observed here (0.477), is only 25%, and similarly for the sample range. We may thus conclude that the true spread in hospital effects is likely to be larger than in the diagram.

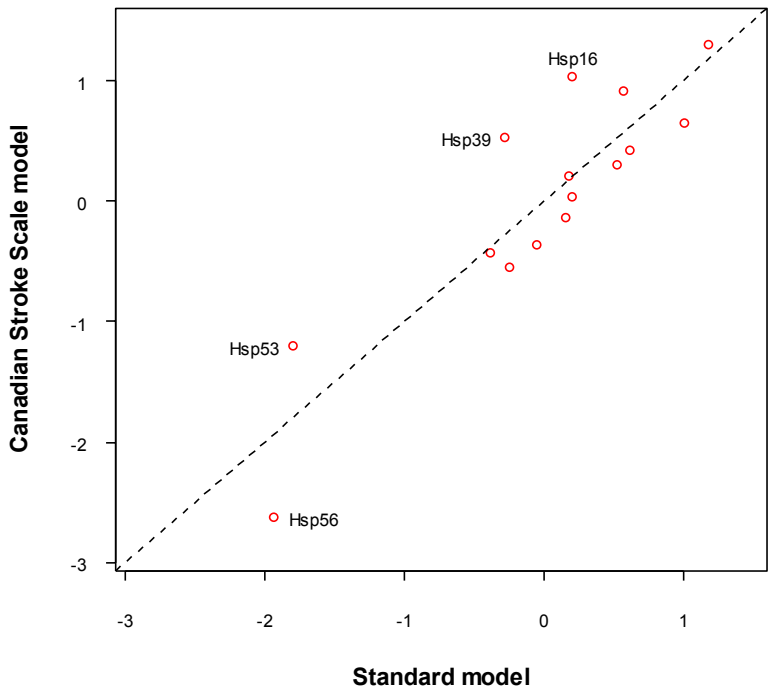
The raw estimates show more spread, due to estimation variance.

6.6.10 Comparison with the clinical data: Canadian Stroke Scale

In a sub-study of 15 selected hospitals, individual results of 30-day mortality were compared to values for each patient (50 per hospital) of the Canadian Stroke Scale (CSS) that assesses patient frailty on an individual basis (68;69). The analysis showed, not unexpectedly, a significant relationship between probability of death and CSS (higher probability of death in patients with higher CSS scores) on the individual level.

On the reduced data set, a comparison was made between estimated hospital effects with and without CSS as an additional explanatory variable in the model. In both cases, logistic regression models were fitted using stepwise regression. The resulting effects are plotted in the figure below.

Figure 6-22: Hospital effects with Canadian Stroke Scale added against effects from standard model.



Using clinical data for risk adjustment does not apparently lead to smaller variability among hospitals. The standard deviation of the shrinkage estimates did indeed increase from 0.68 to 0.75.

Bootstrapping was used to test whether any of the apparent changes were significantly different from zero. A large number of data sets were drawn from the original without replacement. For each replicate data set, the two models were re-fitted and differences in hospital effects recorded, resulting in a large set of bootstrap replicates. A difference was declared significant if it was more than two standard deviations (with respect to. the bootstrap samples) above or below zero.

Eventually, only two hospitals (Hsp16 and Hsp39) turned out having effect changes significantly greater than zero. For these hospitals, it thus appears that only using administrative data for risk adjustment leads to an underestimate of the true effect. However, during a review of findings, it was discovered that many of the admissions for stroke at Hsp 16 were actually for rehabilitation. However, coding used was imprecise and could not be used to segregate the admissions as true admissions or admissions for rehabilitation, and all admissions were thus used.

Due to the limited sample size of the clinical data set, the results are somewhat inconclusive. The correlation values are fairly good. The AAD value is high, but this is likely a result of the high random variability. One interpretation of this sub analysis is that the data quality for around 15% of the hospitals is unsatisfactory, leading to bias. It is difficult, however, to quantify the magnitude of this bias.

Table 6-42: Indicators for bias in stroke risk adjustment. Statistical measures of changes in hospital effects.

Measure	With CSS
AAD	0.36
correlation	0.89
rank correlation	0.81
fraction changed	13%

6.7 30-DAY MORTALITY AFTER HIP FRACTURE

6.7.1 Inclusion/exclusion criteria

The ”Index admission” is the first admission for hip fracture (ICD-9 codes 820 with all subgroups, ICD-10 codes S72.0, S72.1 and S72.3) in the calendar year. Table 9-5 shows the distribution of the various codes among the index cases. One patient may have more index admissions if she/he has had new hip fractures in different years, or if she/he has been transferred, since each hospital will then be counted as a separate index admission. 50205 cases were classified as index admissions for hip fracture from these criteria. Of these cases, 6.9% were followed by death within 30 days.

Criteria for exclusion were: 1) under 65 years of age, 2) admissions followed by accidental death (external cause of death, violence etc.), 3) neither dead on arrival nor emergency and 4) admissions to hospitals with very few cases in the period 1997-2001. Hospitals that dropped under 100 cases after using the first three exclusion criteria were dropped. The final sample size was 41862 cases.

In the final data set, a small number of cases had missing values for one or more variables. These cases were included or excluded from analyses according to whether the particular analysis depended on the variables in question. Missing values were

almost without exception associated with socio-demographic variables, mainly level of education.

In Appendix 9.1, we list the participating hospitals and the corresponding aliases used in tables and figures.

Table 6-43: Numbers of cases for hip fracture removed based on the exclusion criteria. Some exclusions relate to more than one criterion.

Exclusion criterion	Excluded hospitals	Excluded cases
Under 65 years of age	-	4708
Admissions followed by accidental death	-	372
Neither dead on arrival nor emergency	-	3525
Hospitals with less than 100 admissions	6	245
Hospitals falling under 100 admissions after criteria 1-3 are applied	3	61[a]

a) Remaining cases at these hospitals after applying criteria 1-3

Table 6-44: Number of index cases selected based on above selection criteria.

Not Selected	Selected	Selected and with complete socio-demographic variables
8343	41862	41511

6.7.2 Observed mortality

Table 6-45 shows the variation in proportions of dead at each hospital for each year. As can be seen in Figure 6-23, relative 30-day mortality has remained rather even from 1998 to 2001. However, in the first months of 2000 there is an unexplained drop in mortality which could be artifactual.

Table 6-45: Hip fracture. Variation in 30-day case fatality for each hospital per year of admission. Percentages not shown if number of admissions is below 50 per year (six hospitals with less than 50 admissions/year, each of the five years, are omitted).

Hospital	Year of admission				
	1997	1998	1999	2000	2001
Hsp01	6.2	7.4	6.5	13.3	7.8
Hsp03	8.6	4.4	9.7	5.4	5.0
Hsp04				11.3	
Hsp06	8.1	3.7	11.7	4.3	7.9
Hsp07	8.6	7.6		4.9	3.8
Hsp08	3.8	6.2	8.2	4.4	5.6
Hsp09	3.1	9.4	6.8	13.0	6.2
Hsp10	6.8	5.9	9.8	10.8	6.7

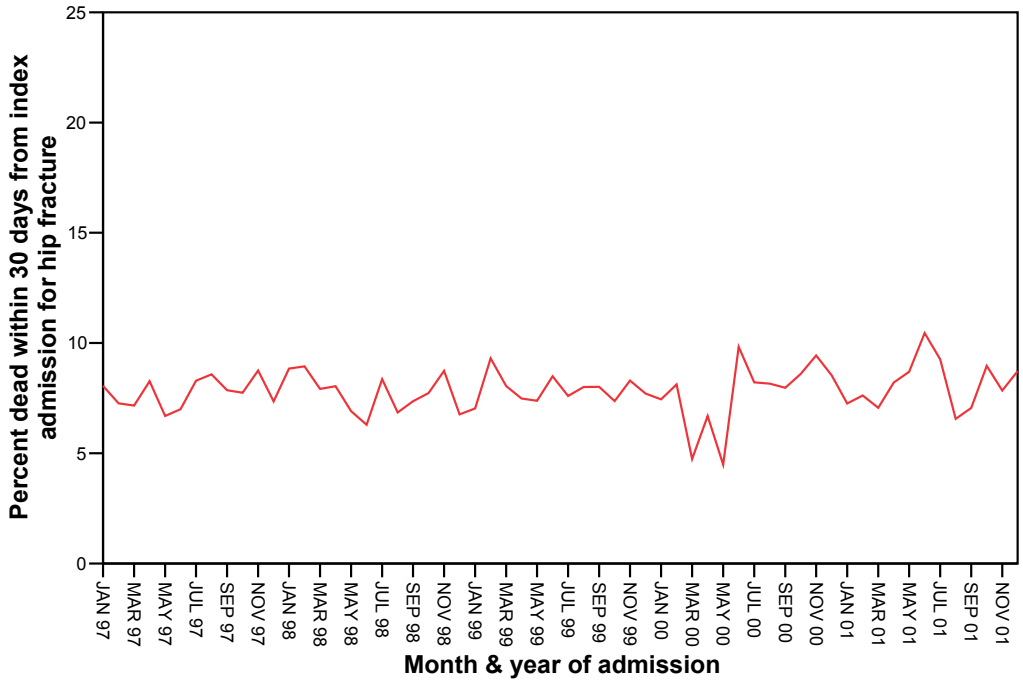
Hospital	Year of admission				
	1997	1998	1999	2000	2001
Hsp11	7.9	8.9	8.8	8.4	9.2
Hsp12	12.3	12.8	15.4	6.3	4.9
Hsp14	5.8	5.2	7.3	3.1	8.2
Hsp15			9.8	9.2	14.3
Hsp16	6.5	8.9	7.8	7.6	5.9
Hsp17	8.8	9.4	9.0	7.6	11.8
Hsp18	5.4	7.4	14.0	7.2	8.6
Hsp19				2.0	
Hsp20		8.0		8.1	5.4
Hsp21	9.2	10.8	7.4	12.0	11.1
Hsp24	9.3	8.1	8.9	8.8	11.2
Hsp26	12.7	6.4	10.6	10.7	8.8
Hsp28	10.3	3.9	7.5	8.9	7.1
Hsp29	6.1	10.7	7.3	8.2	8.8
Hsp30		3.3	6.7	15.1	14.3
Hsp31	4.7	4.9	1.2	7.9	3.8
Hsp32	12.7	14.0	6.7	5.5	8.5
Hsp34	7.6	8.6	6.1	10.6	7.3
Hsp35	10.7	10.5	7.7	8.6	11.5
Hsp36	12.3	4.4	10.3	4.5	8.6
Hsp37	10.7	8.1	5.0	6.7	10.7
Hsp38	9.8	15.1			
Hsp39	6.3	4.8	6.8	6.0	7.5
Hsp40	8.8	5.7	7.4	1.7	11.3
Hsp43	5.6	8.2	9.2	7.0	6.9
Hsp44	4.7	7.6	7.9	6.1	5.4
Hsp46	3.6				
Hsp47	6.5	8.2	8.0	7.5	5.4
Hsp48	15.4	6.6	10.7	6.8	8.6
Hsp49	4.8	9.2	5.5	5.9	8.9
Hsp50	7.8	7.5	9.1	9.9	7.2
Hsp51	9.9	11.0	4.2	4.4	8.5
Hsp52			8.0		
Hsp53	5.2	3.1	4.3	5.8	11.3
Hsp55	9.2	9.5	6.1	9.8	8.5
Hsp56	3.5	6.4	4.0	4.9	3.6
Hsp57	7.8	6.0	6.6	8.7	7.3
Hsp60	7.9	7.9	6.6		
Hsp61	11.4	3.9	9.3	7.4	6.7
Hsp62	9.0	11.0	7.3	7.8	9.1
Hsp63	5.6	8.0	3.3	2.1	6.9
Hsp65	2.9	9.4	17.0	3.6	11.2
Hsp66	13.0		5.7		6.6

Table 6-46: Hip Fracture: Number of cases dead at 30-days for each hospital, together with the total number of index cases per year, after applying exclusion criteria.

Hospital	1997		1998		1999		2000		2001	
	Number of hip fracture cases	Dead within 30 days	Number of hip fracture cases	Dead within 30 days	Number of hip fracture cases	Dead within 30 days	Number of hip fracture cases	Dead within 30 days	Number of hip fracture cases	Dead within 30 days
Hsp1	161	10	162	12	200	13	195	26	205	16
Hsp2	24	0	27	4	34	3	31	1	22	2
Hsp3	151	13	182	8	186	18	167	9	181	9
Hsp4	38	6	35	3	41	2	53	6	41	2
Hsp6	149	12	135	5	145	17	141	6	140	11
Hsp7	58	5	66	5	46	5	61	3	53	2
Hsp8	80	3	97	6	97	8	91	4	107	6
Hsp9	64	2	64	6	74	5	77	10	65	4
Hsp10	222	15	186	11	204	20	194	21	210	14
Hsp11	432	34	370	33	398	35	379	32	402	37
Hsp12	65	8	86	11	52	8	64	4	81	4
Hsp13	27	1	21	0	29	1	40	0	34	0
Hsp14	52	3	58	3	55	4	64	2	61	5
Hsp15	.	.	25	1	102	10	98	9	105	15
Hsp16	215	14	214	19	206	16	197	15	203	12
Hsp17	272	24	277	26	321	29	353	27	339	40
Hsp18	147	8	162	12	136	19	139	10	162	14
Hsp19	43	4	45	6	39	3	50	1	48	5
Hsp20	38	1	50	4	48	7	62	5	56	3
Hsp21	76	7	93	10	108	8	100	12	90	10
Hsp23	41	4	26	4	28	3	44	5	30	5
Hsp24	247	23	247	20	236	21	238	21	233	26
Hsp25	24	1	23	1	26	5	39	4	41	1
Hsp26	158	20	156	10	160	17	150	16	148	13
Hsp28	116	12	128	5	106	8	123	11	127	9
Hsp29	131	8	103	11	110	8	110	9	125	11
Hsp30	47	3	60	2	60	4	53	8	63	9
Hsp31	85	4	81	4	81	1	76	6	132	5
Hsp32	55	7	86	12	119	8	165	9	164	14
Hsp34	238	18	278	24	246	15	265	28	259	19
Hsp35	272	29	229	24	272	21	243	21	262	30
Hsp36	73	9	90	4	97	10	88	4	93	8
Hsp37	169	18	186	15	222	11	224	15	224	24
Hsp38	51	5	53	8	29	2	38	3	44	3
Hsp39	380	24	421	20	428	29	381	23	425	32
Hsp40	57	5	53	3	54	4	60	1	62	7
Hsp43	499	28	466	38	531	49	517	36	507	35
Hsp44	128	6	132	10	214	17	279	17	299	16
Hsp46	55	2	36	3	32	0	28	2	36	3
Hsp47	108	7	364	30	324	26	305	23	280	15
Hsp48	52	8	61	4	56	6	59	4	70	6

Hospital	1997		1998		1999		2000		2001	
	Number of hip fracture cases	Dead within 30 days	Number of hip fracture cases	Dead within 30 days	Number of hip fracture cases	Dead within 30 days	Number of hip fracture cases	Dead within 30 days	Number of hip fracture cases	Dead within 30 days
Hsp49	186	9	196	18	201	11	202	12	203	18
Hsp50	180	14	212	16	186	17	191	19	180	13
Hsp51	71	7	91	10	96	4	68	3	82	7
Hsp52	45	2	25	2	50	4	40	0	37	4
Hsp53	134	7	128	4	138	6	137	8	133	15
Hsp55	315	29	305	29	343	21	328	32	363	31
Hsp56	170	6	140	9	176	7	162	8	197	7
Hsp57	359	28	385	23	351	23	381	33	411	30
Hsp58	31	2	38	2	33	2	37	2	31	3
Hsp60	151	12	139	11	76	5	4	0	10	5
Hsp61	202	23	232	9	226	21	230	17	240	16
Hsp62	400	36	399	44	505	37	436	34	464	42
Hsp63	72	4	88	7	91	3	94	2	87	6
Hsp64	30	2	36	1	15	1	15	0	15	2
Hsp65	68	2	85	8	94	16	84	3	89	10
Hsp66	54	7	49	6	53	3	44	5	61	4

Figure 6-23: 30-day case fatality of hip fracture as a function of month and year of admission (percent).



6.7.3 Distribution of covariates

Summary measures of the distribution of risk adjustment covariates are shown in Table 6-47 below. The table shows minimum, maximum and standard deviation between

hospitals (i.e. the various statistics for the hospital means), as well as the overall mean and standard deviation, for each covariate. For categorical covariates, the indicator variables for each category (except the first (reference) category) are used. Apparently, most of the variability is on the individual rather than on the hospital level (a notable exception is distance between home and hospital).

Table 6-47: Distribution summaries for hip fracture covariates. Categorical variables in percentage units.

Variable	Between hospital min	Between hospital max	Between hospital std. dev	Overall std.dev	Overall mean
Age	80.24	82.99	0.73	7.13	81.59
Sex: female	66.20	81.70	3.02	43.72	74.26
Marital status: not married	5.00	18.02	2.81	30.92	10.71
Marital status: divorced or separated	0.47	14.41	2.41	22.43	5.31
Marital status: widowed	46.38	63.40	3.29	49.78	54.72
Maximum education in household (years)	9.20	13.18	0.69	2.75	10.48
Natural logarithm of: household income + 1	11.50	12.08	0.10	0.59	11.73
Natural logarithm of: household property/capital + 1	10.70	12.79	0.30	2.03	12.10
Natural logarithm of: distance from home to hospital + 1	0.24	2.47	0.44	0.82	0.98
Number of preadmissions regardless of cause, weighted	0.13	0.44	0.07	0.48	0.31
Number of pertinent codiagnoses at previous admission, weighted	0.05	0.20	0.03	0.23	0.13
CCDS stages 2.3 to 3.3	0.00	2.61	0.53	7.90	0.63
Transferred from another hospital	0.35	48.55	7.58	24.12	6.20

6.7.4 Model selection

Starting with a model having interactions between hospital and year, we performed a stepwise backward model selection. The largest model (M6 in the table below) was only fitted to the reduced data set consisting of the ten largest hospitals. With this exception, the final model selection was performed on the complete data set.

Note that we do not make any a priori assumptions on the yearly variation of performance. Both smooth and sudden changes are possible, and the hospitals may evolve independently over time.

In each step of the selection process, variables or (predefined) sets of variables were removed from the model when the standard log-likelihood (deviance) test, comparing the resulting model to the previous one, was non-significant at the 5% level. When the test was significant, the selection process stopped.

After inspection of the estimated coefficients for all three diseases, ethnicity appeared non-significant. The socio-demographic variables were then regrouped so that ethnicity could be tested separately.

Testing the selected model on the complete data set showed that a more flexible function should be chosen for age dependency. The selection process was then repeated using 4-interval splines.

The backward selection process for the extended a priori model is summarized in the tables below.

Table 6-48: Variable names and type.

Variable set	Variable name	Type of variable	Description
	yearfactor	Categorical	The years 1997-2001 as factor
	yeardiff	Continuous	Year – 2001
	hospital	Categorical	Hospital code 1…66
Age/gender	N03	Continuous	Two-interval B-spline basis function (of age)
	N13	Continuous	Two-interval B-spline basis function (of age)
	N23	Continuous	Two-interval B-spline basis function (of age)
	N33	Continuous	Two-interval B-spline basis function (of age)
	N05	Continuous	Four-interval B-spline basis function (of age)
	N15	Continuous	Four-interval B-spline basis function (of age)
	N25	Continuous	Four-interval B-spline basis function (of age)
	N35	Continuous	Four-interval B-spline basis function (of age)
	N45	Continuous	Four-interval B-spline basis function (of age)
	sex	Categorical	Distribution of the sexes
Socio-demographic	marital	Categorical	Marital status
	maxedu	Continuous	Maximum education in household
	logincome	Continuous	Natural logarithm of: household income + 1
	logworth	Continuous	Natural logarithm of: household property/capital + 1
	ethnicnor	Categorical	Born in Norway (Yes/No)

Variable set	Variable name	Type of variable	Description
Frailty	preadmiss	Continuous	Number of preadmissions regardless of cause
	prenumbdiagn	Continuous	Number of pertinent codiagnoses at previous admission
Severity	logdistance	Continuous	Natural logarithm of: distance from home to hospital + 1
	CCDS stage	Categorical	Dichotomized (1.1 to 2.2) or (2.3 to 3.3)
	moved in	Categorical	Transferred from another hospital. Yes / No

Table 6-49: Description of candidate models included in analysis.

Model name	Variables included in model	Interactions
M6	age/gender set, socio-demographic set, frailty set, severity set, yearfactor, prenumbdiagn squared	Yearfactor:hospital
M7	Age/gender set, socio-demographic set, frailty set, severity set, yeardiff, squares of yeardiff, preadmiss and prenumbdiagn	Yeardiff :hospital
M8	age/gender set, socio-demographic set, frailty set, severity set, yeardiff, squares of yeardiff, preadmiss and prenumbdiagn	Yeardiff:hospital
M9	age/gender set, socio-demographic set, frailty set, severity set, yeardiff, squares of yeardiff, preadmiss and prenumbdiagn	Yeardiff:hospital
M10	age/gender set, socio-demographic set, frailty set, severity set, yeardiff, squares of yeardiff, preadmiss and prenumbdiagn	
M11	age/gender set, socio-demographic set, frailty set, severity set, yeardiff, predadmiss squared	
M17	age/gender set, socio-demographic set, excluding ethnicnor, frailty set, severity set, yeardiff, squares of preadmiss and yeardiff	
M19	age/gender set, socio-demographic set, excluding ethnicnor, frailty set, severity set, preadmiss squared	
M21	age/gender set, frailty set, severity set, preadmiss squared	
M31	age/gender set, frailty set, severity set, preadmiss squared	

Table 6-50: Analysis of deviance for hip fracture model selection.

Data set	Spline order	Model name	Resid. Df	Resid. Deviance	Df	Deviance change	P-value	AIC change
subset	3	M6	17563	8576.66				
subset	3	M7	17583	8598.98	-20	22.32	0.324	-17.68
subset	3	M8	17592	8609.23	-9	10.26	0.330	-7.74
subset	3	M9	17593	8609.86	-1	0.62	0.429	-1.38
subset	3	M10	17602	8614.28	-9	4.42	0.881	-13.58
subset	3	M11	17603	8614.28	-1	0.00	0.982	-2.00
subset	3	M17	17604	8615.28	-1	1.00	0.317	-1.00
subset	3	M19	17605	8615.45	-1	0.16	0.685	-1.84
subset	3	M21	17611	8628.44	-6	12.99	0.043	0.99
all	5	M7	41317	19859.99				
all	5	M8	41373	19926.34	-56	66.35	0.162	-45.65
all	5	M9	41374	19926.60	-1	0.26	0.607	-1.74
all	5	M11	41431	19974.03	-57	47.43	0.813	-66.57
all	5	M17	41432	19974.81	-1	0.78	0.378	-1.22
all	5	M19	41433	19975.24	-1	0.44	0.509	-1.56
all	5	M31	41439	20016.00	-6	40.76	<1e-4	28.76

The stepwise procedure selected as the final model M19, which has no year dependency and 5-th order spline functions for age. Using the minimum AIC criterion would have led to the same conclusion. All 41511 cases with complete (non-missing) data were used in fitting this model. Of these, 2 had weights exactly zero.

The order of the splines used for age effect was tested by comparing the AIC of the final model with the AIC using degree 5 splines (ie three internal knots, placed at the median and quartiles of age, and four intervals). With a decrease in AIC by 1.68, the degree 3 splines were rejected.

Two measures of the final model's prediction ability were computed: area under the receiver operating characteristic, or C statistic, and the optimal rate of correct classification.

Table 6-51: Hip fracture final model – measures of prediction ability.

C statistic	0.73
Classification rate	0.92

Hosmer-Lemeshow goodness-of-fit statistics were computed for the final model. The values are good, especially in light of the large sample size.

Table 6-52: Hosmer-Lemeshow goodness-of-fit statistics for final hip fracture model.

Data set	C.statistic	df	P-value
subset	7.88	8	0.446
all	14.32	8	0.074

To check for influential observations, various influence measures were computed for the final model. The deviance residuals ranged from -1.99 to 3.05, which is no cause for concern. The largest leverage hat statistic was 0.14. As noted earlier, it is reasonable to standardize hat values by the number of cases per hospital. Looking at cases with hat value greater than 0.01 and standardized hat value greater than 3, we find a 254 cases with high leverage. All of these came from the Hsp55 hospital. Hsp55 accounts for 31% of the total number of the transferred cases, meaning that it is reasonable to expect cases from this hospital to have increased influence on the model, via the variable "moved in".

The maximum Cook's distance was 0.0069, and 62 cases were above 0.001. These cases mostly came from small hospitals, meaning that their leverage will be high, and had a very high mortality. The influential cases were judged to be acceptable and therefore retained in the model.

6.7.5 Covariate parameter estimates

The parameter estimates for risk adjustment covariates are displayed in the following table:

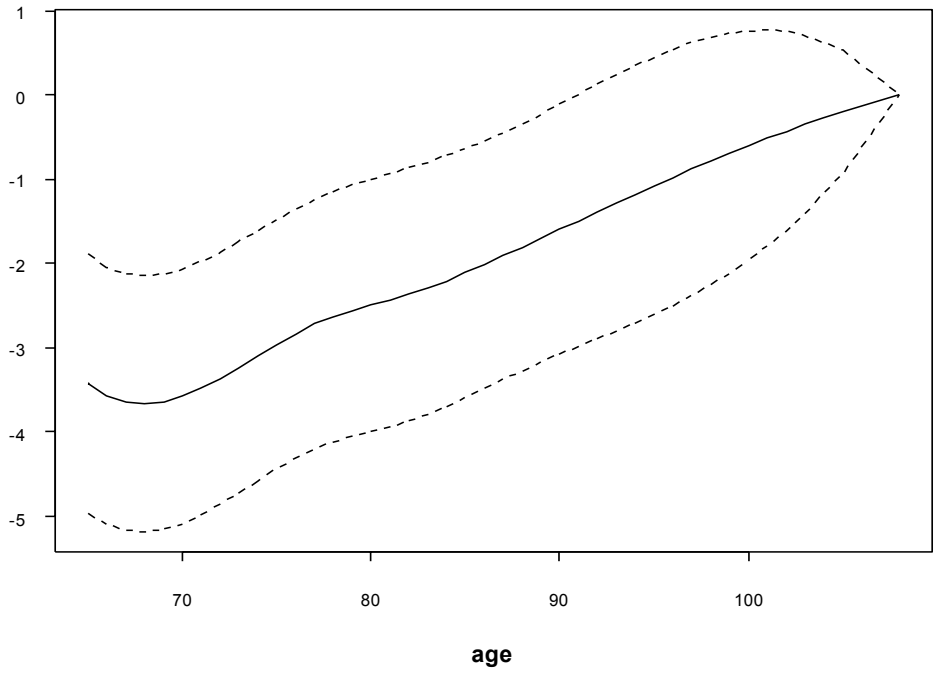
Table 6-53: Final hip fracture model – parameter estimates.

Model term		Estimate	Std.Error	Z value	Two-sided p-value
(Intercept)		1.076	0.950	1.132	0.258
N05	age function	-3.428	0.787	-4.355	<1e-4
N15	age function	-4.115	0.833	-4.939	<1e-4
N25	age function	-2.703	0.742	-3.642	0.00027
N35	age function	-2.419	0.785	-3.082	0.00205
N45	age function	-1.365	0.688	-1.985	0.047
N55	age function	-0.401	1.051	-0.381	0.703
female sex		-0.926	0.043	-21.322	<1e-4
maxedu		-0.023	0.0083	-2.704	0.00684
logincome		-0.025	0.055	-0.445	0.656
logworth		-0.040	0.0098	-4.116	<1e-4
logdistance		0.054	0.028	1.908	0.056
preadmiss		0.126	0.045	2.809	0.00498
prenumbdiagn		1.278	0.174	7.355	<1e-4
prenumbdiagn squared		-0.447	0.170	-2.633	0.00847
CCDS stages 2.3 to 3.3		1.919	0.143	13.467	<1e-4
moved in		-0.927	0.199	-4.666	<1e-4

Model term		Estimate	Std.Error	Z value	Two-sided p-value
married/cohabitant		-0.012	0.042	-0.290	0.772
not married		0.044	0.050	0.885	0.376
divorced or separated		0.00	0.072	0.045	0.964
widowed		-0.035	0.037	-0.966	0.334

The dependency of mortality on age is shown in the figure below.

Figure 6-24: Final hip fracture model – effect of age. With uncertainty band.



Mortality increases strongly with age. The CCDS stage and number of previous diagnoses also have strong effects on mortality. A relatively smaller increase in mortality follows from having previous hospital admissions. Being female strongly reduces mortality.

Being transferred into the hospital has a strong effect and must be viewed as an indicator of reduced disease severity. The likely explanation is that on the balance, transfers tend to come after the initial period of highest mortality. It is thus the less severely ill patients that tend to be transferred.

Among the socio-demographic variables, education and property/capital have relatively small, but significant, effects. They act, as would be expected, to reduce mortality with increasing education or fortune. Marital status has no significant effect, nor has distance from home to hospital. These variables are, however, retained in the model for the sake of consistency between the three diseases studied.

The number of previous admissions and previous diagnoses act as expected to increase the risk of dying. The reason for the second order term in the number of previous

diagnoses is that this variable does not act linearly on mortality, but has an effect that increases more and more slowly as the variable increases.

6.7.6 Variation of hospital performance over time

An important question in assessing the quality indicators is whether they show stability over time or have large temporal variability with no reasonable explanation. In the model, temporal variability not explained by risk adjustment covariates is either assigned to random variation or incorporated in the year/hospital part of the model.

In the final model, there is no effect of calendar year on mortality. Note that the model may still predict non-random year-to-year variation in mortality if the covariates show sufficient variation.

From the calculations of residual autocorrelations reported in section 6.10, it follows that there is no indication of any monthly or weekly pattern of variation except random fluctuations.

6.7.7 Model robustness

When included in the model, the number of codiagnoses excluding complications had a parameter estimate of 0.23, significant at the 0.1% level, and thus acting to increase risk as one would expect. However, since this variable was shown to be poorly recorded for the other two diseases (see section 6.8.1), it was considered too unreliable to include in the final model.

We were concerned about the robustness of our final model with respect to two factors:

- inclusion/exclusion of transferred patients
- inclusion/exclusion of the risk adjustment covariate "number of codiagnoses excluding complications"

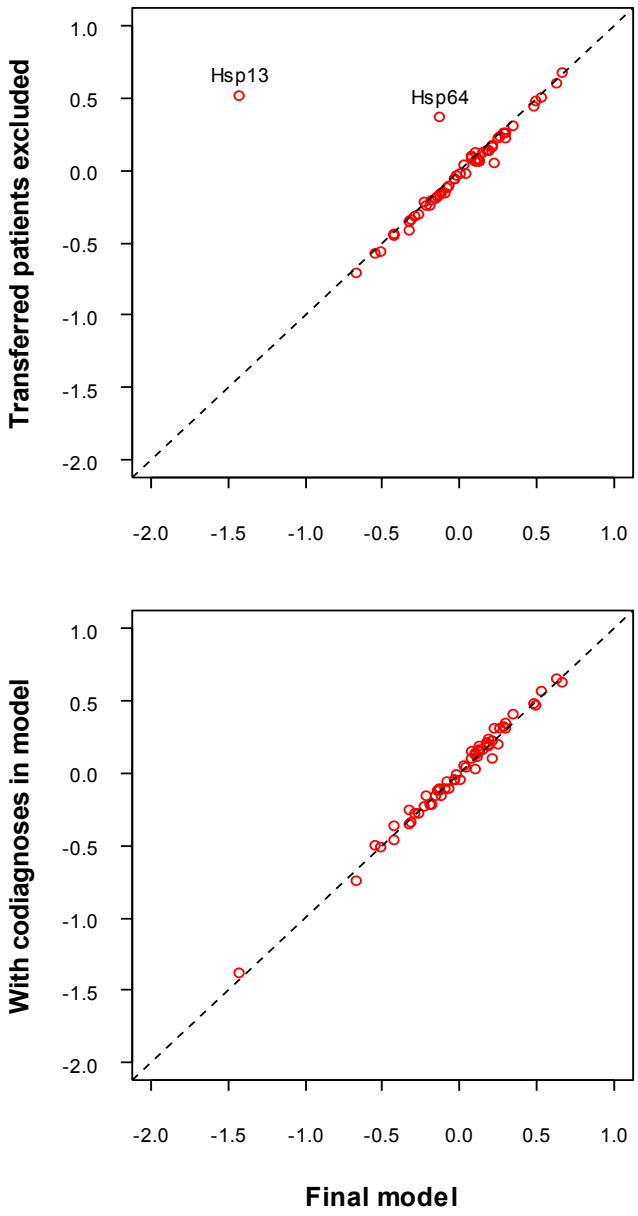
Hospital effects

Hospital effects are the log-odds of death within 30 days at the various hospitals, compared to the average hospital, controlling for risk adjustment covariates (see chapter 5.6).

An effect of e.g. -0.182 means that this particular hospitals has 10% reduced log-odds for death, while an effect of e.g. 0.41 means that the log-odds of death are 50% greater than in the average hospital, given the levels of the risk adjustment covariates.

To examine the robustness of our final model, we compared the hospital effect estimates of the final model with estimates obtained after varying the above factors. The figures below show the results. Hospitals with apparently large changes are labeled.

Figure 6-25: Robustness of final hip fracture model. Hospital effects from alternative models against effects from final model.



Without transferred patients, the effects for Hsp13 and Hsp64 increased by 1.94 and 0.49 respectively. For Hsp13, this is more than two standard deviations, for Hsp64, around one standard deviation. These hospitals are distinguished by very high rates of transferals to other hospitals. The uncertainty in effect estimates was very high for both hospitals due to their low number of cases. For the other hospitals, changes were very small.

Inclusion of codiagnoses in the model leads to very small changes in the effect estimates.

The table shows the statistical measures of the changes (see 5.6.5).

Table 6-54: Robustness of final hip fracture model. Statistical measures of changes in hospital effects.

Measure	Transferred patients excluded	Model without codiagnoses
AAD	0.088	0.032
RAAD	0.357	0.123
correlation	0.652	0.993
rank correlation	0.850	0.986

Including the number of codiagnoses had very little effect.

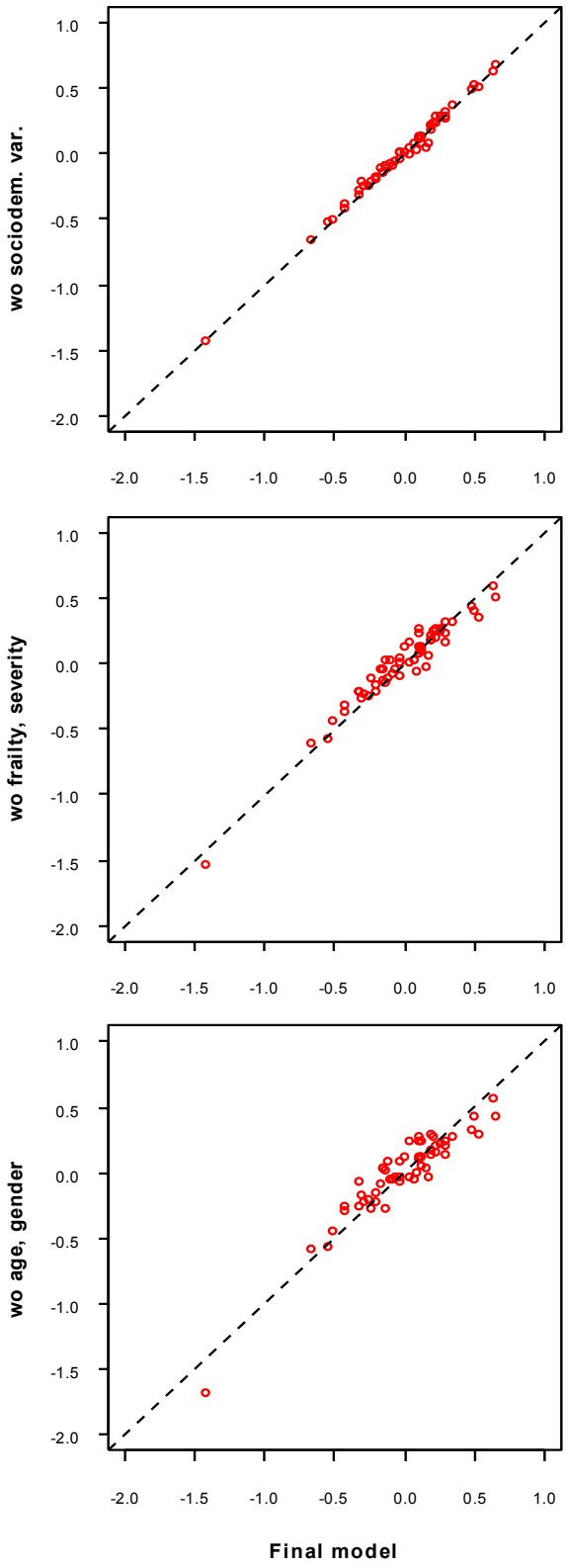
We concluded that the model robustness was acceptable, with the provision that for Hsp13 and Hsp64, results were highly dependent on inclusion/exclusion of transferred patients and thus must therefore be regarded as unreliable.

6.7.8 Comparison of results with and without risk adjustment

To estimate the magnitude of the bias in the risk adjustment, we compute various bias indicators, following (1). We have investigated the change in hospital effect estimates when successively smaller sets of risk adjustment covariates are used. As explained in 6.9.1, we may regard these changes as plausible upper bounds for the true bias.

The figure below shows estimates from our final model plotted against those of the reduced models. First, we remove socio-demographic variables, then both socio-demographic and frailty/severity variables, and last, also age and gender, leaving no risk adjustment covariates in the model.

Figure 6-26: Comparison of results for hip fracture with and without risk adjustment. Hospital effects from reduced models against effects from final model.



The overall magnitude of the changes are very small when socio-demographic variables are excluded, fairly small when frailty and severity variables are also excluded, and moderate when age and gender are excluded in addition.

Quantitative measures of the change in effects are shown in the table below (see 5.6.5).

Table 6-55: Indicators of bias in hip fracture risk adjustment. Statistical measures of changes in hospital effects.

Measure	Socio-demographic set excluded	Frailty/severity set also excluded	Age/gender set also excluded
AAD	0.021	0.065	0.099
RAAD	0.084	0.271	0.417
correlation	0.995	0.970	0.937
rank correlation	0.990	0.949	0.903

The overall conclusion is that assessment of relative hospital performance is not very sensitive to the exclusion of socio-demographic variables, a finding that was also seen elsewhere (80). Excluding frailty and severity indicators result in somewhat larger changes in estimated hospital effects. The additional changes due to exclusion of age and gender are of the same order.

6.7.9 Hospital effects

To judge the importance of the estimated hospital effects (defined in 5.4.2), the three decision procedures described earlier were applied:

- the single hospital rule: testing one hospital by the standard test. A one-sided p-value, if less than 10%, is reported
- the follow-up list rule: testing each hospital by the standard test, but with level according to the number of cases. A listed hospital is reported
- the multiple hypotheses testing rule: all effects are tested simultaneously by a multiple testing procedure. A simultaneous p-value, if less than 10%, is reported.

Hospital effects

Hospital effects are the log-odds of death within 30 days at the various hospitals, compared to the average hospital, controlling for risk adjustment covariates (see chapter 5.6).

An effect of e.g. -0.182 means that this particular hospitals has 10% reduced log-odds for death, while an effect of e.g. 0.41 means that the log-odds of death are 50% greater than in the average hospital, given the levels of the risk adjustment covariates.

The table below shows the shrinkage and raw estimates, the raw estimate standard deviation, the decision procedure results and number of cases per hospital.

Table 6-56: Hip fracture final model – hospital effect estimates (positive estimate indicates increased mortality).

Hospital	Shrinkage estimate	Raw estimate	Raw estimate Std.Error	P-value (one-sided)	Listed for follow-up	P-value (multiple testing)
Hsp1	0.083	0.107	0.126			
Hsp2	0.028	0.087	0.340			
Hsp3	-0.151	-0.207	0.143	0.074		
Hsp4	0.060	0.127	0.250			
Hsp6	-0.065	-0.092	0.153			
Hsp7	-0.159	-0.319	0.236	0.088		
Hsp8	-0.238	-0.415	0.203	0.021	yes	
Hsp9	-0.081	-0.142	0.205			
Hsp10	0.103	0.131	0.124			
Hsp11	0.164	0.187	0.087	0.016		
Hsp12	0.130	0.215	0.190			
Hsp13[a)]	-0.140	-1.426	0.715	0.023	yes	
Hsp14	-0.194	-0.422	0.255	0.049	yes	
Hsp15	0.306	0.500	0.187	0.00379	yes	
Hsp16	0.141	0.182	0.127	0.076		
Hsp17	0.421	0.492	0.097	<1e-4	yes	<1e-4
Hsp18	0.095	0.128	0.138			
Hsp19	-0.010	-0.021	0.251			
Hsp20	0.048	0.101	0.246			
Hsp21	0.127	0.189	0.165			
Hsp23	0.314	0.638	0.239	0.00382	yes	
Hsp24	0.223	0.269	0.106	0.00574	yes	
Hsp25	-0.045	-0.127	0.318			
Hsp26	0.084	0.108	0.128			
Hsp28	-0.047	-0.068	0.160			
Hsp29	-0.079	-0.116	0.161			
Hsp30	0.063	0.114	0.211			
Hsp31	-0.275	-0.541	0.231	0.00968	yes	
Hsp32	0.361	0.530	0.161	0.00050	yes	0.052
Hsp34	0.161	0.197	0.111	0.039		
Hsp35	0.296	0.352	0.102	0.00029	yes	0.031
Hsp36	-0.096	-0.155	0.183			
Hsp37	0.231	0.295	0.124	0.00878	yes	
Hsp38	0.023	0.048	0.247			
Hsp39	-0.219	-0.258	0.098	0.00441	yes	
Hsp40	-0.083	-0.170	0.242			
Hsp43	-0.022	-0.025	0.086			

Hospital	Shrinkage estimate	Raw estimate	Raw estimate Std.Error	P-value (one-sided)	Listed for follow-up	P-value (multiple testing)
Hsp44	-0.146	-0.193	0.134	0.075		
Hsp46	-0.077	-0.230	0.330			
Hsp47	0.026	0.032	0.110			
Hsp48	0.169	0.298	0.207	0.074		
Hsp49	-0.020	-0.026	0.135			
Hsp50	0.010	0.013	0.125			
Hsp51	0.043	0.076	0.205			
Hsp52	-0.109	-0.289	0.303			
Hsp53	-0.214	-0.323	0.168	0.027	yes	
Hsp55	0.553	0.663	0.106	<1e-4	yes	<1e-4
Hsp56	-0.436	-0.667	0.172	<1e-4	yes	0.0055
Hsp57	-0.071	-0.083	0.095			
Hsp58	-0.104	-0.304	0.328			
Hsp60	0.138	0.230	0.193			
Hsp61	0.127	0.161	0.122	0.094		
Hsp62	0.194	0.220	0.086	0.00539	yes	
Hsp63	-0.268	-0.510	0.224	0.011	yes	
Hsp64[a)]	-0.026	-0.124	0.461			
Hsp65	0.167	0.258	0.173	0.069		
Hsp66	0.163	0.303	0.218	0.082		

[a)] Results are sensitive to exclusion of transferred patients

The estimated effects have estimation error standard deviations ranging from 0.086 to 0.72, with a median of 0.17.

Comparison of the above results with the raw case fatality rates shows that the hospitals with effects significantly different from zero generally have both a low estimation standard deviation and a high (or low) raw case fatality rate.

The shrinkage adjusted hospital effects have a population standard deviation of 0.19, as compared to the raw estimates' population standard deviation of 0.35. The first figure below shows estimated probability densities for both the raw and the shrinkage estimates. We have indicated the indifference interval (i.e. the interval around zero where the odds of dying deviate no more than 10% from the average) as well as the alert limits (i.e. the points where the odds of dying deviate from the average by a factor of two), cf. section 5.6.2.1. In the second figure, the same data have been translated to an absolute risk scale. The hospital-specific probability of death refers to a hypothetical average patient, and is not necessarily equal to the average mortality for that hospital.

Figure 6-27: Estimated probability densities for hospital effect estimates.

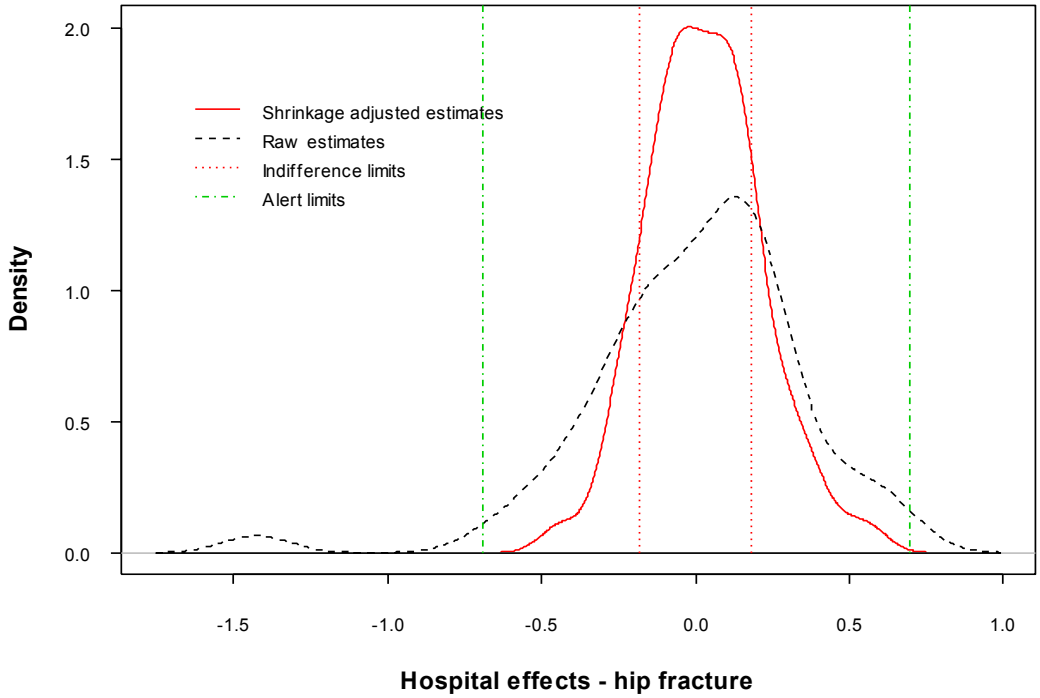
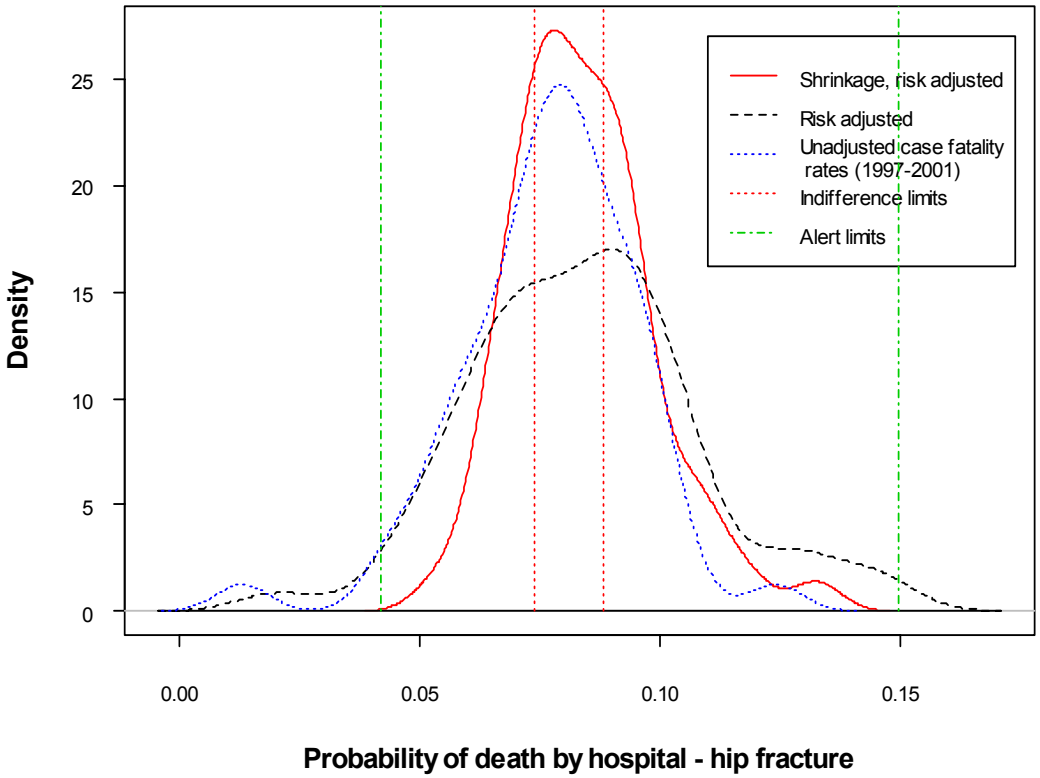


Figure 6-28: Estimated probability densities for hospital-specific probability of death.



For reasons outlined earlier, the distribution of the shrinkage estimates must be regarded as the more accurate, overall representation of the set of Norwegian hospitals with

respect to stroke mortality, despite their bias towards zero. A considerable proportion of the hospitals falls outside the indifference interval (-0.182, 0.182).

The estimated distribution extends to very low (≈-0.5) and very high (≈-0.6) values. It should be noted that this range corresponds to a relative odds ratio of 3, which is considerable. Based on the simulation experiment described in 5.6.4, we found that if the shrinkage estimates reported above were the actual, true effects, the chance of obtaining a sample maximum as small as we have observed here (0.553), is only 6%, and similarly for the sample range. We may thus conclude that the true spread in hospital effects is very likely to be larger than in the diagram above.

The raw estimates show more spread, due to estimation variance.

6.8 DERIVED RESULTS

6.8.1 Recording of codiagnoses

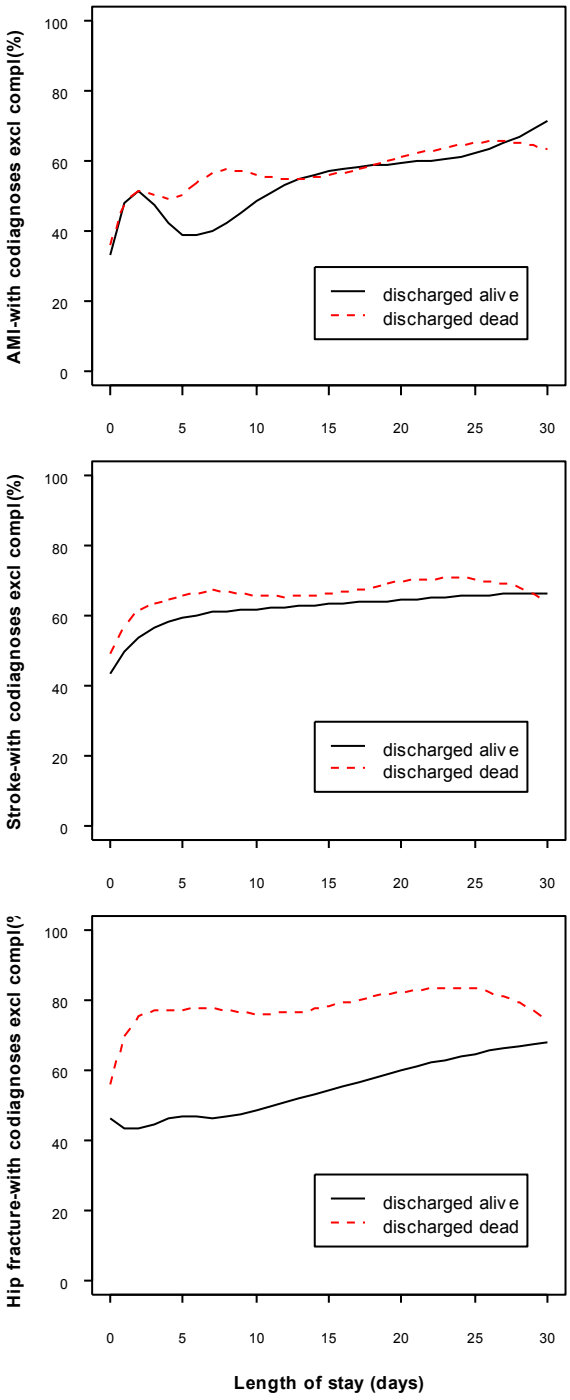
It has been noted that the potential risk adjustment covariate "number of pertinent codiagnoses" has a parameter estimate that is significantly different from zero for all three diseases, but with the wrong sign for AMI and stroke. This means that codiagnoses have, disturbingly, the effect of actually decreasing the mortality, albeit weakly. A similar phenomenon has been reported earlier for the effect of having a history of cancer on AMI mortality (81), while other studies report effects of codiagnoses with signs as expected (82).

We believe the reason is that codiagnoses are not recorded for the very ill or for those who die very soon after admission. In Figure 6-29, we compare the proportion of patients with codiagnoses as a function of time for those discharged alive with those discharged dead. At any given time, the patients who are discharged alive have the lowest number of codiagnoses (with allowance for some statistical fluctuations in the curves).

Our interpretation of the paradoxical behaviour of this variable is that it is not a data quality problem, but rather a consequence of a situation that must be rather similar for all hospitals. However, this variable causes problems for our statistical method, which does not take timing of events into account (other than death before or after 30 days). It was excluded from the model. Ideally, the total event history should be incorporated in the statistical model. Because of lack of information about the timing of diagnoses, onset of codiagnoses etc, this is not straightforward.

This interpretation is corroborated by the fact that this variable has the same, paradoxical effect for AMI and stroke, but has the expected effect for hip fracture, where short-term death is less frequent.

Figure 6-29: Proportion with codiagnoses excluding complications as function of length of stay.



6.8.2 Common structure of the quality indicators

Given several quality indicators for a hospital, it is tempting to try to combine these indicators into a composite hospital quality index. With only three diseases, this would obviously yield a poor overall picture of the hospital's general performance. Index construction is therefore not pursued in this report.

If a composite index were to be built at a later stage, a desirable property of the component quality indicators entering is that they measure the same underlying property, at least to some extent. This would be reflected in a correlation structure where, roughly speaking, the correlations are positive and not too different. Promisingly, the observed correlation matrix (see Table 6-57) of the three quality indicators studied has just this form.

Table 6-57: Correlation matrix of shrinkage estimators for hospital effects.

Indicator	AMI	stroke	hip fracture
AMI	1.000	0.250	0.179
Stroke	0.250	1.000	0.319
Hip fracture	0.179	0.319	1.000

6.8.3 Precision - power functions

The empirical power functions are determined by the way hospital effect standard deviations depend on total sample size. It was checked by graphing that assuming standard deviations are inversely proportional to the square root of sample size, is in reasonable accord with the actual historical data. We have estimated the logarithm of the proportionality factor σ_0 as the median of $\log(sdev)+\frac{1}{2}\log(n)$, resulting in the following table:

Table 6-58: Parameter of standard deviation function.

Indicator	Estimated σ_0
AMI	2.807
Stroke	4.993
Hip fracture	3.983

The power functions for testing single hypotheses can now be easily calculated, as explained previously. We assume one-sided testing throughout. For our purposes, it is most interesting to evaluate the power function for the indifference and alert limits for beta, of 0.182 and 0.693 respectively.

We have used the multiple contrasts testing procedure in this study. The power function of this procedure is not easily computed. In the figures below, we display the power of testing at 0.15%-level per hospital as a more readily computed approximation to the true power function. This method of multiple testing, via Bonferroni adjustment, is usually regarded as having less power than the multiple contrast procedure and thus provides an informal lower bound for the true multiple testing power function.

Figure 6-30: AMI – power function as function of number of cases per year. Parameters are level and alternative beta value.

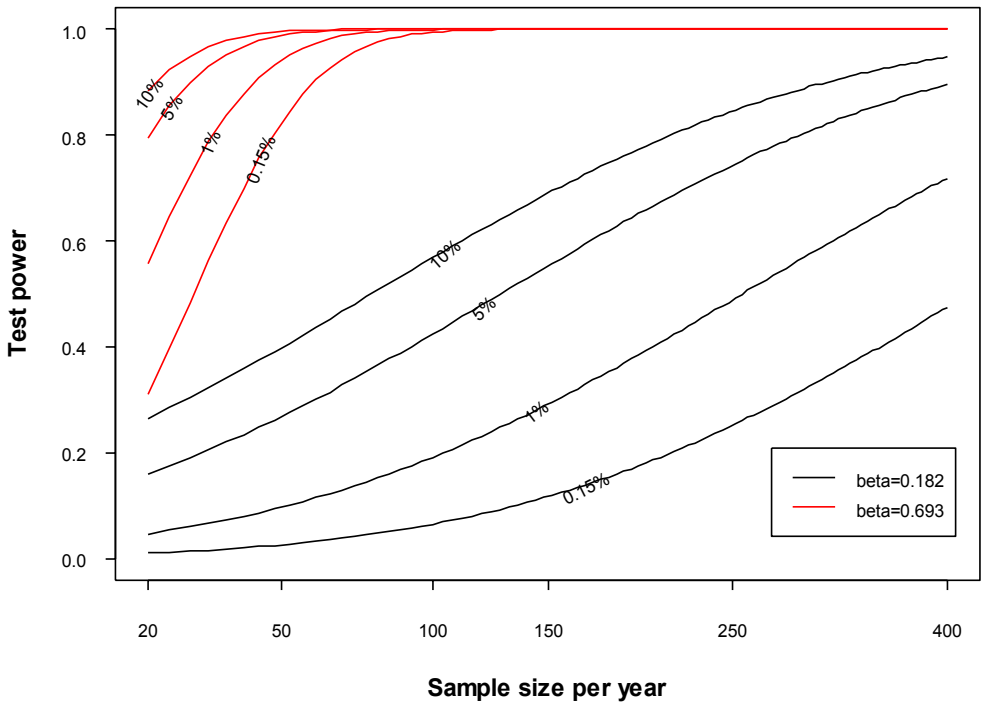


Figure 6-31: Stroke – power function as function of number of cases per year. Parameters are level and alternative beta value.

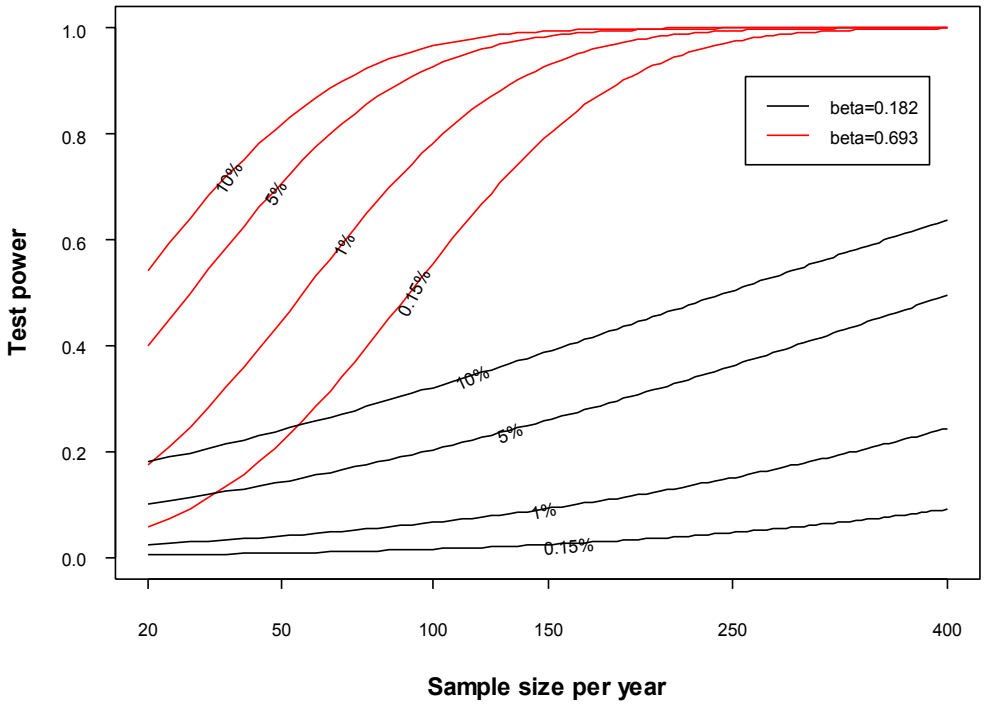
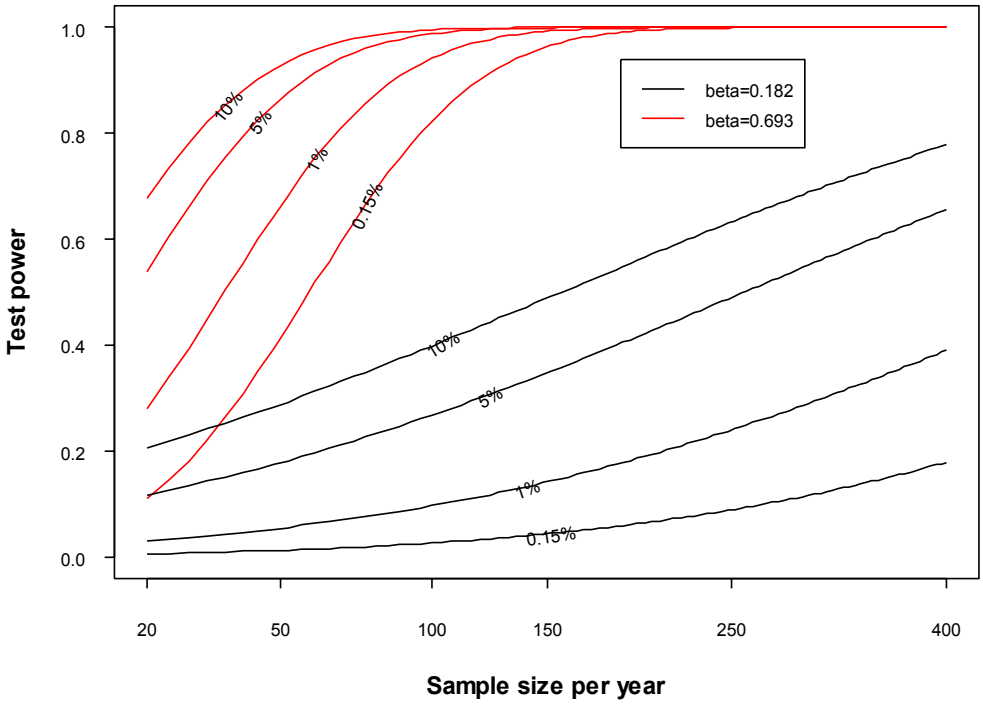


Figure 6-32: Hip fracture– power function as function of number of cases per year. Parameters are level and alternative beta value.



Based on the power functions, we may draw some conclusions about hypothesis testing versus confidence intervals. The difference is partly one of how a significance test is interpreted. When testing levels are chosen judiciously, there is only low probability of declaring significantly larger than zero a hospital effect that is within the indifference interval, except for AMI in the case of medium to large hospitals. Thus, when a hospital effect is declared statistically significantly different from zero, we may feel confident that the magnitude of the difference is significant in the ordinary sense of being of some importance.

6.9 ASSESSMENT OF BIAS MAGNITUDES

An important underlying question is whether we have accounted for all important systematic differences in case-mix between hospitals. Indeed, as in all observational studies, there is always the possibility of yet another, unknown risk factor. Based on our data alone, we can only provide indirect evidence and plausibility arguments to prove or disprove the existence of bias in the indicators.

Any bias must be the result of one or more risk factors that:

- are unknown or measured with error
- have non-negligible effects on mortality
- show enough variation between hospitals that their effect is comparable in magnitude to the hospital effects.

6.9.1 Empirical estimates

The approach implicit in the thorough and wide-encompassing AHRQ report (1), can be formulated along the following lines:

- Presumably, all major risk factors are accounted for. Any unknown or unmeasured and sufficiently unbalanced biasing factor is likely to have a smaller impact than those already included in the risk adjustment.
- The bias associated with the risk adjustment covariates can be measured by comparing results with and without risk adjustment.

We may thus compute a plausible upper bound for the bias in our quality indicators. In the present study, we may also assume that all risk factors that are unaccounted for must be associated with severity or frailty. Note that in the AHRQ, a different risk adjustment system (APR-DRG) was used.

Comparisons of hospital effects, with and without risk adjustment, are found in 6.5.8 (for AMI), 6.6.8 (for stroke) and 6.7.8 (for hip fracture).

6.9.2 Theoretical estimates

Another line of argument is to consider known, possible biasing mechanisms and estimate plausible, quantitative bounds on their effects. Firstly, we will assume that errors in the registration of age, gender, time of death etc are negligible. We will consider the following, hypothetical mechanisms that may be at work to introduce bias in the effects of a particular hospital:

1. Erroneous coding of patients that are dead on arrival, to the effect of including them as index admissions.
2. Selection mechanisms acting before admission (e.g. doctors sending seriously ill patients to one particular hospital).
3. Erroneous coding or variability in diagnosis, resulting in cases with very low severity being erroneously included (the opposite mechanism of (1)).
4. Erroneous or lacking coding, or variation or error in diagnosis, of codiagnoses and risk covariates.
5. Risk covariates not included in the model.
6. Epidemiological or geographic variation in severity and/or frailty.

Basically, (1) and (3) are the same mechanisms. In principle, one could consider erroneous inclusion of conditions with very high mortality in (1), other than dead of arrival. This would have been relevant if we had chosen less severe diagnoses for our indicators. Mechanism (3) will tend to produce short stays. By excluding short stays, one may get an indication of the magnitude of both mechanisms (1) and (3).

Selection mechanisms act to shift the balance between high and low severity cases. Hopefully, this should be caught by the risk adjustment. We will therefore consider (2) in conjunction with mechanism (4).

Having covariates that are not represented in the model will also resemble mechanism (4) and will be considered a special case of this. We have seen in the data indications that codiagnoses are not always being coded. This particular error acts to reduce the apparent probability of death.

It must be borne in mind that only effects that differ between hospitals can introduce any bias. The exception is measurement error in the covariates, coupled with significant differences across hospitals in the distribution of these covariates.

Eventually, we have chosen a simplified model for bias as follows: Data for one hospital are contaminated by data that are systematically in error. The errors manifest themselves by inducing a change in log-odds of dying. Errors are due to one of the causes below:

1. Coding and diagnosis variability or error. The strongest effects will be seen in the limiting cases where the erroneously included cases have probability of death either ≈0 or ≈1 (This is approximated by beta values of ±10).
2. Variability or error in coding and diagnosis of risk covariates. For the error effect magnitudes, we have looked informally at the estimated model parameters to find moderately large effects that could be regarded as representative of rather general risk determinants.
3. Epidemiological/geographic variation. The error effect magnitude has been chosen equal to the estimated effect of being born in Norway, which is one variable that relates to epidemiological differences between groups of patients (This effect is not included in the final model as it is non-significant).

The result of contaminating the data as described is to change the probability of death, away from the base value corresponding to the uncontaminated data. We have computed the ensuing increment in apparent probability of death on the logistic scale. This increment is the bias in the mortality indicator.

For the effect magnitudes associated with covariate error, we have, somewhat arbitrarily, taken as a point of departure the mean of the absolute values of (a) the effect of being transferred into the hospital, and (b) the effect of having one previous diagnosis. The covariate errors may act to increase or decrease risk. For the case of omitted codiagnoses, the error will be in the direction of increasing risk. For the case of clinical variables, direction of the error may be in both directions, depending on the actual case. We are looking at the extreme possibilities where the error is always in one direction for the hospital in question.

For the contamination rates, we have looked at two scenarios: one moderately high and one very high rate. We believe the moderately high value to be in the upper range of plausible values. This is partly supported by the literature, see 7.1.3. The actual numbers have been arrived at by considering the between-hospitals spread in covariates and using our prior judgment. For the dead-on-arrival rates, rough estimates can be inferred from the data quoted in the following sections.

Results are shown in Table 6-59 to Table 6-61 below. The entries in the "beta difference" column are the resulting change in hospital effect for a hospital with contaminated data. We assume that only one contaminating factor applies at a time and that uncontaminated data have probabilities of death equal to the mean case fatality rates for the three diseases.

6.9.3 30-day Mortality for Acute Myocardial Infarction

Exclusion of severity and frailty covariates resulted in average, absolute changes in hospital effects (AAD) of 0.048. That can be compared with the median of the standard

deviation of the hospital effect's estimation errors of 0.11, the adjusted effect standard deviation of 0.065 as well as the absolute standards of the indifference and alert limits.

We see that the upper bound for bias AAD is 40% of the median standard deviation and 26% of the indifference limit. In our opinion, this indicates that bias is not a serious problem.

The AAD derived from (1)-Appendix 7 is 0.43, which refers to the total effect of all risk adjustment covariates. This is still much higher than our corresponding number, 0.105. A possible explanation is that the Norwegian hospitals receive a more uniform mix of cases than US hospitals.

Theoretical bias estimates for systematic errors have been obtained as described in 6.9.2. We have been able to draw on results in (59) for the effects of clinical variables on probability of death. In this reference, systolic blood pressure on admission was shown to be a very important predictor for death. Based on the reported model parameters, we have very roughly estimated the actual error effect of not including blood pressure, within our framework.

Table 6-59: Effects of hypothetical biasing mechanisms – AMI.

Error cause	Contamination rate	Error effect (beta)	Change in apparent hospital effect (beta)
Dead on arrival	0.02	10	0.104
Dead on arrival*	0.05	10	0.248
Zero mortality cases	0.1	-10	-0.128
Zero mortality cases*	0.3	-10	-0.423
Covariate error, increasing risk	0.15	0.5	0.085
Covariate error, decreasing risk	0.15	-0.5	-0.065
Covariate error, increasing risk*	0.3	0.5	0.165
Covariate error, decreasing risk*	0.3	-0.5	-0.133
Epidemiological variation, increasing risk	0.25	0.12	0.031
Epidemiological variation, decreasing risk	0.25	-0.12	-0.029

*) High value for contamination rate.

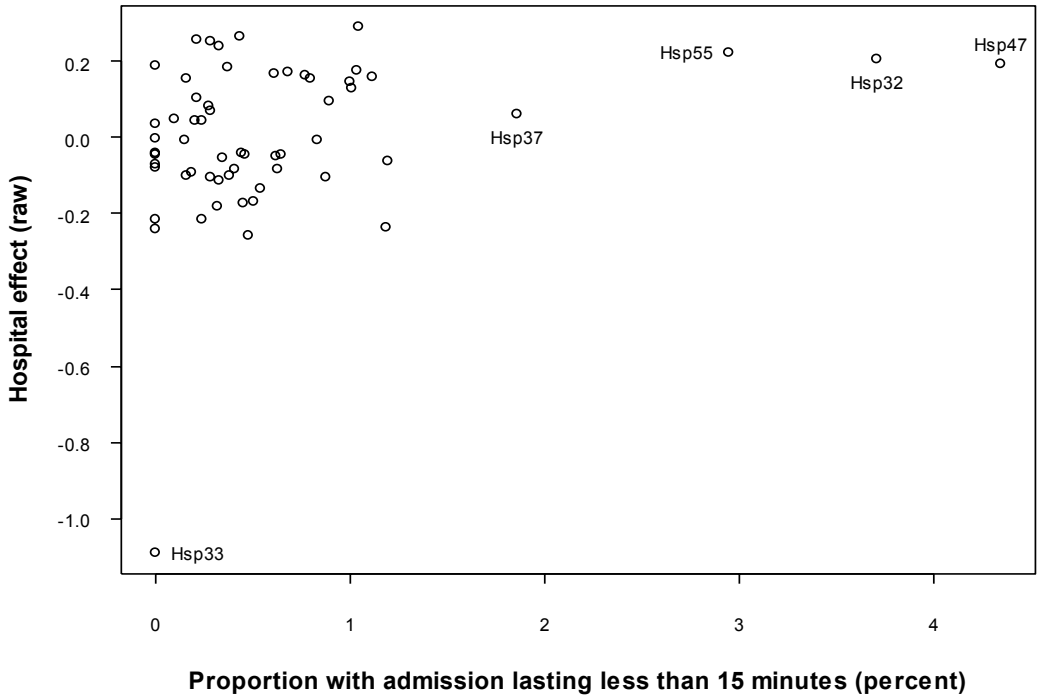
We see that a high rate of contamination with dead-on-arrival cases, if present, will have serious effects on bias, as well as a high rate of inclusion of cases with very low mortality. Moderate rates of low-mortality contamination have non-negligible effects, with magnitude similar to the AAD values from Table 6-26. All other effects are negligible, with the possible exception of high contamination rate of covariate error, which is inside the indifference interval but comparable to the actual estimated effects.

The data gives some indications as to the problem concerning the reporting of dead on arrival. We believe that the admission date and time are accurately recorded. The discharge date and time are probably not as reliable. However, we may use these times

to calculate an apparent duration for an admission in minutes. It turns out that the proportion of admissions with very short apparent durations varies considerable from hospital to hospital. The mortality among these cases is 0.96. Figure 6-33 shows the proportion of stays with duration not exceeding 15 minutes, plotted against the estimated (raw) hospital effects. As is readily apparent, we may suspect that the high effect estimate of the hospitals Hsp55, Hsp32 and Hsp47 are at least partly due to registration and coding practices (For Hsp55 at least, this is corroborated by the fact that the cases in question have not been assigned a proper ward code).

On the other hand, it has been estimated that the short-term mortality from AMI is around 30% - half of which occurs during the first two hours. One would therefore expect 1%-2% mortality per 15 minutes in the initial phase.

Figure 6-33: Proportion of very short admissions vs hospital effect – AMI.



6.9.4 30-Day Mortality for Stroke

Exclusion of severity and frailty covariates resulted in average, absolute changes in hospital effects (AAD) of 0.068. This can be compared with the median of the standard deviation of the hospital effects' estimation errors of 0.20, the adjusted effect standard deviation of 0.22 as well as the absolute standards of the indifference and alert limits.

We see that the bias upper bound AAD is around one third of the median standard deviation, and less than 40% of the indifference limit. In our opinion, this indicates that bias is not a serious problem.

The AAD derived from (1)-Appendix 7 is 0.17, which refers to the total effect of all risk adjustment covariates. This is still higher than our corresponding number, 0.11, by about 50%. A possible explanation is that the Norwegian hospitals receive a more uniform mix of cases than US hospitals.

Theoretical bias estimates for systematic errors have been obtained as described in 6.9.2.

Table 6-60: Effects of hypothetical biasing mechanisms – stroke.

Error cause	Contamination rate	Error effect (beta)	Change in apparent hospital effect (beta)
Dead on arrival	0.02	10	0.112
Dead on arrival*	0.05	10	0.267
Zero mortality cases	0.15	-10	-0.193
Zero mortality cases*	0.3	-10	-0.417
Covariate error, increasing risk	0.15	0.48	0.081
Covariate error, decreasing risk	0.15	-0.48	-0.062
Covariate error, increasing risk*	0.3	0.48	0.159
Covariate error, decreasing risk*	0.3	-0.48	-0.128
Epidemiological variation, increasing risk	0.25	0.088	0.022
Epidemiological variation, decreasing risk	0.25	-0.088	-0.022

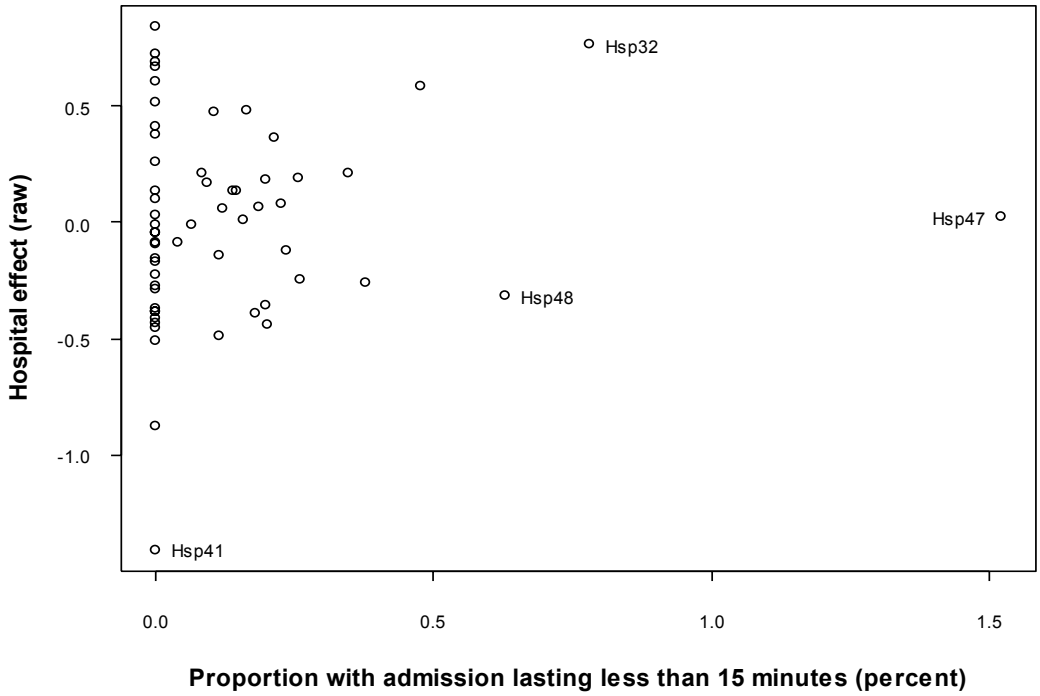
*) High value for contamination rate.

We see that high rate of contamination with dead-on-arrival cases or zero-mortality cases, if present, will have serious effects on bias. The lower rate of inclusion of cases with very low mortality has a moderate effect. Errors in risk covariates have relatively small effects, with magnitude similar to the AAD values from Table 6-40, and within the indifference interval. Epidemiological effects are negligible.

The proportion of apparently very short stays was computed in the same way as we did for AMI (see 6.9.3). In Figure 6-34, the proportion of stays apparently lasting less than 15 minutes are plotted against estimated hospital effects. For stroke, there is also variability in proportions of very short stays. Though obviously showing discrepancies between hospitals in important aspects of recording and coding, this observation does not otherwise seem to invalidate 30-day mortality for stroke as a quality indicator, as there is no clear association with high hospital effect. Compared to AMI, the proportions were much smaller and the cases in question had lower mortality (0.75).

In view of these bias magnitude estimates, it is not readily apparent how the large spread in observed hospital effects could be explained by bias alone.

Figure 6-34: Proportion of very short admissions vs hospital effect – stroke.



6.9.5 30-Day Mortality for Hip Fracture

Exclusion of severity and frailty covariates resulted in average, absolute changes in hospital effects (AAD) of 0.065. That can be compared with the median of the standard deviation of the hospital effect's estimation errors of 0.17, the adjusted effect population standard deviation of 0.19 as well as the absolute standards of the indifference and alert limits.

We see that the bias upper bound AAD is less than half the median standard deviation, and less than half the indifference limit. In our opinion, this indicates that bias is not a serious problem.

The AAD derived from (1)-Appendix 7 is 0.26, which refers to the total effect of all risk adjustment covariates. This is still much higher than our corresponding number 0.099. A possible explanation is that the Norwegian hospitals receive a more uniform mix of cases than US hospitals.

Theoretical bias estimates for systematic errors have been obtained as described in 6.9.2.

Table 6-61: Effects of hypothetical biasing mechanisms – hip fracture.

Error cause	Contamination rate	Error effect (beta)	Change in apparent hospital effect (beta)
Dead on arrival	0.01	10	0.137
Dead on arrival*	0.02	10	0.259
Zero mortality cases	0.15	-10	-0.174
Zero mortality cases*	0.3	-10	-0.379
Covariate error, increasing risk	0.15	0.65	0.123
Covariate error, decreasing risk	0.15	-0.65	-0.077
Covariate error, increasing risk*	0.3	0.65	0.234
Covariate error, decreasing risk*	0.3	-0.65	-0.159
Epidemiological variation, increasing risk	0.25	0.22	0.059
Epidemiological variation, decreasing risk	0.25	-0.22	-0.051

*) High value for contamination rate.

We see that a high rate of contamination with dead-on-arrival or zero mortality cases, if all present, will have serious effects on bias. Covariate error acting to increase apparent risk will have small to moderate effect. All other effects are of small or negligible magnitude, within the indifference interval.

As we did for AMI and stroke, the proportion of apparently very short stays was computed. The proportion of stays not exceeding 15 minutes is very small (0.00031) and these cases do not have very high probability of death (0.18), indicating that improper or variable recording of dead on arrival is not a problem for hip fracture mortality as a quality indicator.

In view of these bias magnitude estimates, it is not readily apparent how the large spread in observed hospital effects could be explained by bias alone.

6.10 COMPARISON WITH MULTILEVEL METHODS

In the report (2) one of the main objections to the statistical analysis was that it did not include multilevel versions of the logistic regression analysis.

With our data, multilevel analysis could conceivably be applied to several levels:

1. Hospital. The main effects (i.e. level) have been taken care of in the logistic regression. Another modelling possibility would be to use random, hospital-specific coefficients for the covariates.
2. Department or ward. We have collected data on the first department and ward a patients received treatment. It was, however, decided early on not to present results pertaining to individual departments or wards. However, if the variability between these units is high, a correlation could be introduced

between cases. The consequence could be an apparent precision in our estimates that is under- or overestimated.
3. Doctor or doctor team. We do not have the relevant data.
4. Units of time, such as month or week. This would capture temporal correlation between outcomes for patients treated in the same unit at closely spaced points in time (In practice, one would probably work in a time series framework instead).

With random coefficient models, the relative performance of hospitals is assumed to depend on the level of one or more covariates. To get a single summary performance indicator, some sort of comparison on a suitably chosen standard population must be carried out, as in (83). In our view, this is detracts from the applicability of this approach.

In addition, we carried out statistical analysis inside a framework that allowed parameter differences between hospitals. We estimated for each covariate under scrutiny a parameter τ that measures the underlying degree of spread of regression parameters b across hospitals; when $\tau=0$, there is no spread, and the hospitals are equal in performance with respect to the covariate in question. We then observed that the τ parameters were estimated as either zero or negligibly small. The modeling and methodology is laid out in detail in a technical report (84).

A difficulty with using random effects for levels nested within hospitals (e.g. ward) is that the random parameters will be confounded with the hospital effects. In a linear framework, this need not cause any problems, but in a loglinear model, there remains the question of definition of hospital effect. For instance, a hospital may choose to treat the most severe cases in one ward. This will lead to large between-ward effects, but should probably not be included in a model for risk adjustment between hospitals.

We tried out the random effects approach on the hospital level for comparison with the standard logistic regression analysis, for selected models. The multilevel analysis was performed in MLWin version 1.10 and in SAS version 8.2.

In the multilevel setting, we added hospitals as level 2 units, and the individual patients as level 1 units. We specified individual α coefficients (constants) for the different hospitals. The other exploratory variables were included as fixed, using exactly the same variables as in the standard logistic regression analysis (85).

Reassuringly, the multilevel parameter estimates agree with those from the logistic regression to the second decimal digit.

There remains the question as to whether between-patient correlations stemming from department/ward and/or time closeness are strong enough to seriously distort our findings regarding the indicators' precision. An informal model check was performed on the deviance residuals for the models using data only from the ten largest hospitals.

The autocorrelation between residuals from cases from the same hospital with admission time differences rounded to a fixed number of days (ranging from 1 to 50) was computed. The correlations were below 0.1, with varying sign. A resampling estimate of variance indicated that none of the autocorrelations was significant.

An analysis of variance of the residuals within each hospital showed between-ward variability with an R^2 (or intraclass correlation coefficient) generally below 0.05. This is sufficiently low that we expect no important variance inflation (or deflation) to occur. The exceptions were three hospitals with R^2 values for AMI between 0.13 and 0.21. It

turned out that these hospitals all had one ward code (possibly indicating missing data for ward) with very high probability of death and very short admission times. When these wards were excluded, the R^2 dropped to the level of the other hospitals. We interpret this as a selection effect or coding artifact that does not indicate any correlation between the outcomes for individual patients.

6.11 CONTROL OF DATA QUALITY

6.11.1 Missing Data

The number of missing values in each variable derived from the PASs and Statistics Norway, ranged from less than 1% to 3%.

6.11.2 Quality control of data

There were two forms of quality control of data. In the first, each hospital was to do a manual check using hospital journals of 50 patients per disease category. In the second form, one doctor from the study group, controlled 50 patients (the same patients as in the first control) per disease at each of 15 hospitals selected as representative of size, type and geographic location.

21 hospitals did not conduct the planned manual check of the collected data and another five did it only partially. The hospital management gave two main reasons:

1. Some argued that, regarding the information requested, the quality of the data in the PAS is identical to the electronic medical records.
2. Some argued that due to major organizational change the hospitals could not give priority to this task because it would be too expensive.

The data collected were checked against date and time of admission, main diagnosis and index diagnosis. Detailed data can be found in the appendix (9.6). The results are summarized in Table 6-62.

Table 6-62: Reported error rates as reported from the hospitals' doctors.

	Average no. of records controlled	Date and time of admission	Main diagnosis	Index diagnosis
		% errors	% errors	% errors
AMI -Average	48.2	0.8%	4.7%	1.7%
Stroke - Average	48.4	0.9%	4.9%	2.9%
Hip Fracture - Average	48.7	1.6%	4.2%	2.2%

Between records scrutinized in the hospitals and the data gathered electronically there are minor discrepancies, 1-2% for the variables date and time of admission and index diagnosis, while the percentage of errors found was larger for main diagnosis (4-5%). As can be seen from the detailed results, the variation between hospitals is large. The error rates are very high for a few hospitals and very low or zero for many hospitals.

The discrepancies were often due to one or more of the following:

- When scrutinizing the journals it is likely that the local doctors have detected some diagnoses that have not been recorded in the patient administration system of the hospital, and have added these;
- For patients transferred between departments, the patient administration system of some of the hospitals allow separate sets of diagnoses;
- There were different coding practices for coding diagnoses during the period 1997-2001. Some patient administration systems and some hospitals coded up to two main diagnoses per department stay. Our system allows the recording of only one main diagnosis;
- We deliberately selected the first hospital stay in each year as index stay. For patients with more than one hospital stay within a calendar year coding errors can occur because the hospital doctors that controlled the data could compare data with wrong admission.

In order to map the cause of the reported errors we selected the three hospitals with the highest reported number of errors within each disease group. We found that 80% of the reported errors were on a level of detail not relevant for this study. The doctor had added the fourth dimension to the diagnosis code (8200 instead of 820) and thereby refined the diagnoses but the patient still suffered the same disease: hip fracture. For the purpose of the analyses in this study, both diagnoses were aggregated to the same level, an index admission for hip fracture. Excluding these errors gives what we may say is the relevant table of errors (for this study) detected in the quality control of the data (Table 6-63).

Table 6-63: Estimated relevant error rates.

	Average no. of records controlled	Date and time of admission	Main diagnosis	Index diagnosis
		% errors	% errors	% errors
AMI -Average	48.2	0.8%	0.9%	0.3%
Stroke - Average	48.4	0.9%	1.0%	0.6%
Hip Fracture - Average	48.7	0.8%	0.8%	0.4%

The error rates found by the doctor from the study group are summarized in Table 6-64 below, and compared to errors found by the hospital doctors at the same selected hospitals. Detailed data can be found in the appendix (9.6).

Table 6-64: Average error rates in some important variables, for the hospitals examined by the study group doctor. Error rates reported by hospital (Hsp) or study group doctor.

	Average no. of records controlled	Date of admission		Main diagnosis		Index diagnosis	
		Hsp	Study group	Hsp	Study group	Hsp	Study group
AMI	47.8	0.79%	0.41%	6.2%	2.6%	3.3%	2.3%
Stroke	47.2	1.1%	0.15%	5.0%	3.5%	3.7%	3.5%
Hip Fracture	48.3	2.5%	1.4%	3.6%	2.9%	1.7%	2.9%

We see that on the average, the study group doctor reported fewer errors than the hospitals themselves did.

Error rates for admission date, main diagnosis and index diagnosis, both at reported by the hospital themselves and reported by the study group doctor are provided in the appendix. As can be seen there, although error rates were usually 0, there were exceptions with some hospitals having large error rates.

7. DISCUSSION

This study aimed at evaluating the use of probability of death 30-day post admission for acute myocardial infarction, stroke and hip fracture as an indicator for quality of service in hospitals. Such an evaluation requires a clear concept of what the indicator measures, for whom it is intended and how the evaluation should take place.

7.1 AVAILABILITY AND QUALITY OF DATA

7.1.1 Completeness of population

We do not know whether there is a discrepancy between the actual number of patients admitted to hospital, and the number recorded in the patient administration systems. Such a discrepancy is unlikely for several reasons, though mainly because the recording of diagnoses in the patient administration system is an important prerequisite for correct estimation of DRG-points and income. We have also been informed from the Norwegian Patient Registry (NPR, Unn H Kvam, personal communication) that lack of registrations on individual level in the PASs hardly is a problem. However, there are indications that coding practices relating to patients arriving dead, under resuscitation or dying very shortly after admission are not uniform between hospitals. This is discussed further below. In addition, patients having symptoms of a severe disease can be admitted for observation, to be released after a short time, when it is documented that the patient is not in fact severely ill.

However, a complete sample based on data given at discharge is not necessarily the correct sample to analyze. The data set does not distinguish between diagnoses present at admission and acquired during the hospital stay. It is not unlikely for these reasons that our sample is “more than complete”, that is it includes patients that should have been excluded as their cases are not relevant to the study. This problem is most likely to occur for AMI, since we included admissions with AMI as secondary diagnosis (we did so with hip fracture, too, but this is a less likely complication to hospital treatment).

7.1.2 Completeness of information

Incompleteness of information is due to two problems: 1) lack of codes in the PAS that reflect important information, and 2) existing coding that is improperly or incompletely reported.

The most critical coding problem is that the routines regarding the registration of patients dead on arrival evidently differ between hospitals. They are recorded as dead on arrival in some hospitals, and included among admissions, while other hospitals do not record such cases at all. For improvement of comparability between hospitals, we strongly advise to establish consistent procedures as to the registration of persons arriving dead or under resuscitation, including the time of death.

In general, the patient administration system does not contain all the information we would have liked. This is because this system is not primarily designed for the purpose of generating quality indicators.

The numbers of missing values from the PASs were very few and not sufficient to have any substantial effects on the results. However, some of the results point to one problem connected to studies based on registries; difficulty of spotting lack of information. When e.g. codiagnoses are absent, we cannot know if the reason is that the patient did not suffer any other conditions, or that such conditions have not been recorded. The latter is known to occur sometimes.

Fewer diagnoses were observed among patients that died in hospital. This may be related to a greater utility of recording codiagnoses in patients that are healthy enough to be saved and/or rehabilitated. The average number of codiagnoses per admission is also lower than one would expect in the age groups most often represented in this study. However, controlling for length of stay, the proportion with codiagnoses among those who were discharged alive was lower than among those who died in hospital.

A final point is that recording of codiagnoses did not use all available codes. For example, it is possible to register whether the fracture was in the right or left hip. Existence of this information would have facilitated identifying a new occurrence of hip fractures as opposed to readmission for the same fracture. However, the code was not found. In addition, coding was often recorded using only three digits and not the final digit after the decimal point. This was a major problem in the use of staging (CCDS) especially for hip fracture.

The main criticism of the previous study on in-hospital mortality was that risk adjustment only was made for age and sex. It was claimed that certain hospitals receive more high risk patients than others. This led to intensive efforts in the present study to attempt to account for these differences. Many available methods are used in the literature to account for disease severity, such as Acute Physiological and Chronic Health Evaluation (APACHE II), Computerized Severity Index (CSI), Patient Management Categories (PMC), Medisgroups (MDGRP), Simplified Acute Physiology Score (SAPS), Coded Disease Staging (CDS), and Clinical Criteria Disease Staging System (CCDSS – the one we chose to use). APACHE II, CSI, MDGRP are based on physiological variables whereas PMC, CDS, and CCDSS are based on journal discharge abstracts. There have been indications that the predictive abilities of the journal based indexes are as good as those based on physiological parameters (86;87). Regardless, physiological measures were not available and could not be tested. CCDSS seemed to be useful for acute myocardial infarction, could be useful for hip fracture if available coding had been recorded completely but was not useful for stroke. For stroke, a categorization into two subdiagnoses (hemorrhage and infarction) was used as covariate.

Codiagnoses (from previous hospital admissions, if any) were used as a proxy for patient frailty. Ideally, more standardized procedures for adjustment for comorbidity should be developed. We recommend investigating the feasibility of substituting our ad hoc list of codiagnoses by the Charlson comorbidity index or another internationally widely used instrument. It was outside the scope of the present methodological study to undertake the relatively large task of constructing such indicators. The ASA score should have been available electronically for all patients having undergone an operation. The research shows that it was not so in many of the hospitals. Some hospitals did not have IT systems that register this variable and some hospitals used more than one IT

system to register ASA score for different patients. It was not cost effective to attempt to collect this information for all hospitals at this stage of the research.

We were not able to collect complete information about important characteristics of the organization of the hospitals. Many hospitals did not respond to our questionnaires or they did not have the required information. We wanted to collect information about the use of pertinent diagnostic and therapeutic procedures at an individual level from all hospitals. Such data are available, but not readily and not always in the PAS, which made us choose not to use such data in our primary analyses. However, see the discussion in 7.2.4 below.

We aimed to collect data from the laboratory and X-ray/CT/MR data systems. This is possible, and we did it to some extent. The complexity of the task, the fact that about 30% of the hospitals did not have quality data available, and data for 1997-99 were not recorded, made us decide not to apply these data in the present study.

Finally, smoking is an important explanatory factor in the development and in part in the outcome of disease and should ideally have been included as a covariate. This parameter was not available to us. However, we are not sure that it would have substantially changed results. This is because rather precise and complete information was included on socio-demographic parameters, in themselves closely correlated to smoking. In addition, inclusion of the socio-demographic variables had relatively little effect on the hospital estimates.

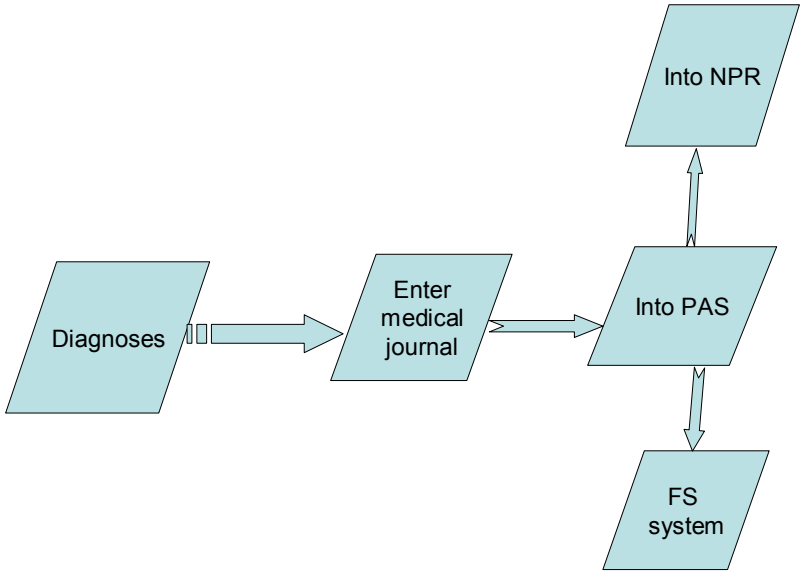
7.1.3 Quality of data

The quality control revealed that more than 99% of the data in our data set for the following variables: date and time of admission, main diagnosis and index diagnosis, was correct as compared to the patients' medical records.

During the quality control of data, the hospital doctors reported coding discrepancies in the main diagnosis in less than 5% of the sample. We found that a large part of these discrepancies were explained by different precision levels in coding, and not relevant for our study. There were different coding practices during the period 1997-2001. Some patient administration system and some hospitals coded up to two main diagnoses per department stay. Our system allows the recording of only one main diagnosis. The discrepancy between actual PAS coding and our data were minimal.

It should be noted that diagnoses to some degree have inherent and unavoidable variability. Usually, coding audits explicitly do not question the original diagnosis. Diagnostic errors may therefore be underreported. In our study, however, no such reservation was made, and some of the errors reported were probably due to medical reconsideration of diagnosis. As shown in the figure below, information passes through many steps where discrepancies may arise.

Figure 7-1: Information chain from diagnosis to database.



To investigate whether the data errors had any correlation with the estimated hospital effects, regression analysis were performed, with hospital effects as response variables and diagnosis error rates as explanatory variables. Separate analyses were carried out for the hospitals' own control and the study group doctor's, for each disease group.

One statistically significant effect was observed. For AMI, there is a significantly negative coefficient of hospital effect against the diagnosis error rates. The magnitude of the estimated regression coefficients implies a total decrease in hospital effect of 0.25-0.35 when the error rate goes from 0% to the maximum observed, which is much larger than average. The coefficients are non-significant when the hospital with largest error rates (Hsp43) is excluded. We may conclude that there is little evidence that the error mechanisms detected in this study have caused significant bias.

Other reports that have quality controlled coding practices in Norwegian hospitals, have found a greater percent of discrepancies than this study reports. Two studies address the accuracy of coding in Norwegian administrative databases (as reported to the NPR database which is the basis for hospital funding), one based on a coding audit of 2001-data from 5 randomly selected hospitals (88) and one based on a coding audit of 2003-data from 14 randomly selected hospitals (89). Both studies found large error rates in coding of main diagnosis (38% and 41%) and large between-hospital variations in error rate (overall range 26% - 51%). Also, an overcoding of codiagnoses was found. A substantial amount of the main diagnoses were even coded to the wrong ICD-10 chapter (17% and 16%). Of these gross errors, 18% and 28% resp. were due to unwarranted switching of main versus codiagnosis. Chapter error rates were reported in (88) to vary considerably between ICD-10 chapters: for ch. IX (e.g. AMI and stroke) the error rate was 9% and for ch. XIX (e.g. hip fracture) it was 19%. Though one may suspect that less serious episodes to some extent will end up coded as AMI or stroke, the present authors question the plausibility of very large rates of gross errors for emergency admissions for the important disease groups considered. More disease- and error-specific data are needed before any strong conclusion can be drawn.

Both studies were aimed at examining if coding was used deliberately with the intention of increasing income through increased DRG value. The main conclusion was that coding practice which is specific and documented is little known and followed. Since the quality control in our study was performed by doctors and not coding specialists, they would not necessarily find irregularities in coding. Typical mistakes were: including more codiagnoses than allowed by current rules, or incorrect selecting of main and codiagnoses (often as a result of hospital transfer, where the original coding was kept, when the situation has changed after treatment). This type of error had little effect on this study since the index diagnoses were chosen from both main diagnoses and codiagnoses (except for stroke) and diagnoses codes were grouped into larger categories where many of these errors would have not effect. In addition, the codiagnoses were simply counted and grouped as an indication of the general health of the patient. It is important that validation against medical records adhere to a strict protocol specifying the kind of discrepancies to look for: diagnostic variability inside or outside the "normal" range, coding that is not documented in records but still may or may not be correct, variability in translating diagnosis into ICD coding, and lastly violations of the coding procedures that may or may not be sound from a medical point of view and may or may not lead to misleading comparisons between hospitals. Differences in validation protocol probably explain some of the differences between our results and others.

In the case of acute myocardial infarction, there is a considerable literature concerning coding and risk adjustment. The effects of undercoding of severity and comorbidities in administrative databases is studied in (90) by a simulation experiment. The conclusion is that undercoding in itself is very unlikely to account for outlier status of most hospitals. Risk-adjusted hospital-specific mortality rates based on administrative data and data reabstracted from journals were compared in (91). The correlation coefficient between the two mortality rates was in the range 0.83-0.85 (depending on the choice of variables for risk adjustment). The correspondence with respect to outlier status was fair (κ=0.38). Several risk-adjustment methods, some using clinical data, were compared in (92). Of the pairs of correlation coefficients between z-scores of probability of death for each hospital, using different risk adjustment method, ¾ were above 0.80 and ¼ above 0.9, The agreement between outlier status across pairs was fair to good. The sensitivity of AMI diagnosis to diagnostic criteria has already been mentioned. It should be noted that in Norway, the large increase in admissions, due to new troponine-based criteria, seemed to occur from 2000 to 2001 (93).

We initially tried to use the codes distinguishing between acute stroke treatment and rehabilitation. There were problems with one hospital, but since the data sets eventually used are restricted to emergency admissions, we do not regard this as a significant problem for the analysis. Several studies address the accuracy of ICD-9 coding for stroke in administrative databases. A European study (94) reports a positive predictive value (PPV) for ischemic stroke of 71%. A US literature review reports a PPV of 79% for stroke (95), while a Norwegian study (96) finds a PPV of 89% (when not discerning between first-ever and recurrent stroke) for the stroke codes relevant here. Several risk-adjustment methods, some using clinical data, were compared in (97). It was found that some pairs of methods ranked a large proportion of patients very differently by predicted probability of death (No hospital-level results were reported).

Concerning hip fracture, the Norwegian study (98) found a positive predictive value (PPV) of 85% for hip fracture. For one of the three hospitals studied, sensitivity was as low as 46%. However, the study period was before the introduction of the current DRG

based funding system, and we find it unlikely that underreporting on this scale is present in our data. The other hospitals had sensitivities of 98% and 99%. The only published coding audit (99) in Norway that gives data specifically for hip fracture, reports a PPV of 92% and a sensitivity of 88%, based on a small data set from 2000.

In our sensitivity study, we have used a low value of 70% and a high value of 85% (90% for AMI) for PPV, excluding contamination with "dead on arrival". In light of published results for stroke and hip fracture, these values seem reasonable and slightly conservative in the sense of not being too optimistic.

7.2 EVALUATION OF VALIDITY OF 30-DAY MORTALITY AS QUALITY INDICATOR AT HOSPITAL LEVEL

7.2.1 Face validity

Acute myocardial infarction (AMI) is one of the major causes of death in Norway. At least 25% of those who die of AMI, die during the acute phase of the attack. Treatment of AMI is closely related to survival and standard procedures of care have been developed. The timing of care is essential to its effectiveness. Stroke is also a major cause of death in Norway. Norway has one of the highest frequencies of hip fracture in the world. Survival depends on proper treatment, organizational aspects of hospital function and patient characteristics. Therefore, these indicators are associated with high face validity.

It is difficult to make international comparisons of results of this kind, mainly because exact information about data availability, quality, and collection is not readily available. Looking at Sweden, reports on mortality after myocardial infarctions are available (100). The report does not give the opportunity to make direct comparisons, but it seems as our data for admissions and deaths are similar to those reported in Sweden. The figures for Sweden have been validated and diagnoses in patient administration systems were found to be good approximations of actual diagnoses given (101). Data from the Swedish Stroke Registry and Swedish Hip Registry do not give case fatality figures that are comparable (102;103). We draw no conclusions as to the comparability of our results to other countries. The literature behind the use of mortality as a quality indicator is thoroughly reviewed in the HTA report published by AHRQ (1).

Another form of validity is comparability over time. Our data is from 1997-2001, and cannot be extrapolated to 2004 for several reasons. For AMI, diagnostic criteria and treatment procedures are radically changed the last few years. For stroke, changes in the organization of treatment, particularly the implementation of stroke units, have occurred continuously. A new coding standard, ICD-10, was introduced in the middle of the observation period. A large health care reform affecting all hospitals was put into power after 2001. We emphasize that the results are of historical interest, and should not be used as a description of contemporary hospital quality.

7.2.2 Precision

For a quality indicator to be usable, the statistical uncertainty cannot be too high, or in the terms of Davies et al. (1), the precision must be sufficiently high. We know that this is a very real issue in the context of hospital quality. This calls for a rather delicate discussion of precision. As already noted, we must consider carefully the decision

problem at hand. The precision may well be sufficient for some decision-makers but not for others.

At the core of all decision-making, however, lies some form of statistical hypothesis testing. In the language of decision theory, we can make two different decisions: to state that a certain hospital has performance different from the rest, or to abstain from drawing any conclusions about the hospital. In the first case, there is a certain probability of making an error. This is a so-called type I error, the probability of which is controlled by the testing level. In the second case, we may fail to identify a hospital with poor performance. This is the so-called type II error. The power function is the probability of not making type II errors. Thus, we are led to the consideration of power functions, and to rephrasing the question regarding reliability as follows: do we have a low type II error probability, while keeping the relevant type I error probability under control? As explained in 5.6.2.1, we will judge power functions by their value in the alternative points beta = ±0.683, the alert limits. These limits correspond to hospitals having odds of death within 30 days either one half or twice the average. ,In this context, one should also keep in mind the distribution of actual estimated beta effects, which will tell us how often, on the average, we will conclude that a hospital is different.

7.2.2.1 30-day Mortality for acute myocardial infarction

This indicator is considered in the literature to have good precision for at least the larger hospitals. The shrinkage adjusted hospital effects (see 5.6.4) have a standard deviation of 0.065, substantially lower than the corresponding figure derived from (1) (Appendix 7), of 0.18.

From Figure 6-30, we may draw the following conclusions about the indicator AMI mortality:

- For medium-sized and large hospitals, with more than 100 admissions per year, the chance of detecting an effect at the alert limit is close to 100% with a test level down to 0.15%. This includes multiple testing procedures.
- For small hospitals, with less than 100 admissions per year, the chance of detecting an effect at the alert limit ranges is above about 85% at the 5% level.
- For large hospitals (more than 240 cases), power stays above 75% down to the indifference limit when testing at 5% level.
- If the distribution of hospital effects in 2001 is taken as representative for later studies, one would not expect many hospitals to have effects large enough to be detected.

AMI is the most precise of the indicators studied. When care is taken to use the three different decision rules appropriately, this mortality index has good precision for use as a quality indicator in Norwegian hospitals.

7.2.2.2 30-day Mortality for stroke

The shrinkage adjusted hospital effects (see 5.6.4) have a standard deviation of 0.22, somewhat smaller than the corresponding figure derived from (1) (Appendix 7), of 0.32.

From Figure 6-31, we may draw the following conclusions about the indicator stroke mortality:

- For medium-sized and large hospitals, with more than 100 admissions per year, the chance of detecting an effect at the alert limit is better than 90% with a test level of 5%.
- For small hospitals, with less than 100 admissions per year, the chance of detecting an effect at the alert limit ranges from about 40% to 90% at the 5% level. At the 10% level, power exceeds 50% for all hospitals.
- With a multiple testing procedure, the power ranges from 50% to almost 100% for medium and large hospitals, but down to less than 10% for the smallest. This means that we cannot give any guarantee about the overall error probability without sacrificing the chances of detecting large effects small hospitals.
- If the distribution of hospital effects in 2001 is taken as representative for later studies, one would expect several hospital effects to be large enough to be detected with reasonable probability.

The stroke indicator performs reasonably well from the precision viewpoint. When care is taken to use the three different decision rules appropriately, we feel that this mortality index has sufficient precision for use as a quality indicator in Norwegian hospitals.

7.2.2.3 30-day Mortality from hip fracture

The shrinkage adjusted hospital effects (see 5.6.4) have a standard deviation of 0.19, much smaller than the corresponding figure derived from (1) (Appendix 7), of 0.63.

From Figure 6-32, we may draw the following conclusions about the indicator hip fracture mortality:

- For medium-sized and large hospitals, with more than 100 admissions per year, the chance of detecting an effect at the alert limit is better than about 95% at the 1% level.
- For small hospitals, with less than 100 admissions per year, the chance of detecting an effect at the alert limit is more than 50% at the 5% level.
- With a rigorous multiple decision procedure, the power exceeds 80% except for the small hospitals, but ranges down to near 10% for the very smallest. This means that we cannot give any guarantee about the overall error probability without seriously sacrificing the chances of detecting large effects for some of the hospitals.
- For large hospitals, there is more than about 50% chance of detecting deviations down to the indifference value, using the 5% level.
- If the distribution of hospital effects in 2001 is taken as representative for later studies, one would expect several hospital effects to be large enough to be detected with reasonable probability.

The hip fracture indicator performs fairly well with respect to precision. When the appropriate decision rule is used, this mortality index has sufficient precision for use as a quality indicator in Norwegian hospitals.

7.2.2.4 Decision rule recommendations

For all the three indicators studied in this report, it is necessary to distinguish between the different classes of decision-makers and have available a decision rule that is a compromise between single and multiple hypotheses testing.

For use in one hospital only, using single hypothesis testing at 1% to 5% level, serious performance deviations will usually be detected in large or medium-sized hospitals. In small hospitals, the chances of detection are smaller, so one should test at the 10% level.

It is outside the scope of the present study to determine the decision rule in detail. For illustration of the degree of statistical uncertainty involved, the following pragmatic rule for follow-up listing is suggested:

Table 7-1: Decision rule for follow-up listing – definition.

Number of cases per year		(One-sided) testing level per hospital
Lower bound	Upper bound	
-	40	10%
40	160	5%
160	300	2%
300	-	1%

For the period 1997-2001, the estimated error probabilities of this rule are shown in the following table:

Table 7-2: Decision rule for observation listing – error probabilities.

Number of cases per year	Testing level (type I error probability)	Type II error probability		
		AMI	Stroke	Hip fracture
30	0.100	0.041	0.338	0.118
60	0.050	0.004	0.264	0.085
120	0.020	<0.001	0.089	0.014
240	0.010	<0.001	0.007	<0.001
400	0.100	<0.001	<0.001	<0.001

It should be noted that the usefulness of a follow-up list rests on the assumption that it can be followed up with moderate costs and in such a way that the consequences of an erroneous listing of an adequately performing hospital has acceptable consequences.

7.2.3 Minimum bias

In the literature, there is much discussion of the bias resulting from using in-hospital mortality and length of stay as outcome variables. In the present study, 30-day mortality is used that removes this source of bias.

We have a fairly complete set of data for Norwegian patients and there is no reason to expect serious bias in the material. The observed admission rates are at a national level approximately as expected and support this opinion. On the other hand, regional differences in admission rates may possibly reflect differences in admission procedures and pre-hospital handling of patients in Norway. Such differences could influence risk adjustment and thereby hospital performance.

For evaluation of empirical results for precision and bias, the AHRQ report uses performance categories that unfortunately for the most part do not apply to our modeling framework. The exception is the bias measure rank correlation of hospital effects before and after risk adjustment, where the categories are as in Table 7-3.

Table 7-3: Quality Indicator performance categories for bias: rank correlation before and after risk adjustment.

QI bias performance	Fair	Good	Very good
Rank correlation statistic	Less than 0.750	0.750 to 0.949	0.950 or greater

There are several sources of information bias, differing and negligent coding practices are probably the most important ones. Coding practices may be part of a culture of a hospital and therefore may differ on the provider level. Uniform and consistent handling of cases that are dead on arrival is critical.

We have computed empirical and theoretical estimates for the bias magnitudes. The rank correlation statistics for the three diseases resulted in their being classified as in the "Good" category defined in Table 7-3, for stroke almost in "Very good". For all three diseases, plausible upper bound for the average, absolute bias in the hospital effects is in the range 0.09-0.11. This is well within the indifference interval. True bias of this size should not cause substantial error in the assessment of a hospital's performance.

A sensitivity analysis yielded theoretical upper bounds for the bias associated with coding error, diagnosis variability, insufficient risk adjustment and patient variability between hospitals. Our theoretical estimates seem to indicate that bias associated with insufficient or imprecise risk adjustment is even smaller, in the range 0-0.05. The theoretical estimates indicate that diagnostic or coding error leading to inclusion of cases with very low probability of death may lead to somewhat higher bias, but still within or slightly outside the indifference interval. However, inclusion of cases dead on arrival will seriously bias the hospital effects. From our data, it seems that this may be a potential problem mostly for the AMI indicator.

The sub-study using the Canadian Stroke Scale (CSS) indicated that 2/15 (13%) of the hospitals had biased hospital effects for stroke. The results from this sub analysis are somewhat inconclusive, but were not felt to invalidate our conclusions concerning bias.

A fair comparison between hospitals must correct for case-mix at admission. When using administrative data, it is necessary to use information on comorbidity to adjust for patient frailty and codiagnoses (i.e. staging) to adjust for severity. The ideal data for risk adjustment would contain a complete history of the admission – when did the various symptoms appear, when were diagnoses made, when was treatment given and how did the patient respond. A basic problem related to risk adjustment for comorbidity and disease severity using PAS data, is that these data are created at discharge. Codiagnoses may thus be acquired during the stay. Some of these represent complications and should

ideally have been removed prior to assessing differences in hospitals since they may reflect poor quality of treatment. Also short stays, including the most severe cases leading to early death, tend to be associated with few recorded codiagnoses. The method of analysis we have used, where the precise time until death is disregarded, is not suited for such time-dependent covariates.

The most important risk adjustment variables were age and sex. Proxy variables for comorbidity, while leading to meaningful and understandable mortality models and being important on the individual level, had only slight influence on the hospital estimates. The probable reason is that in Norway, choice of hospital in emergency cases is based on geographical criteria. For stroke in particular, this study indicates that disease severity is important to control for on the patient level. Other disease severity indices should be investigated.

Another important variable used in risk adjustment was whether the patient was transferred from another hospital. Being transferred is associated with a large decrease in probability of death, contrary to the widely held assumption that it is the most severely ill patients that are transferred most often, which indeed has been shown to hold in other countries (104). Excluding all transferred patients strongly influenced the results for the few hospitals with large proportions of transferred patients. Distance between home and hospital was used as proxy for time from onset of symptoms to admission and thus part of disease severity. Such timing data would make it possible to model transfers between hospitals in a more satisfactory way (104).

If our risk adjustment were based on a model that predicts outcome with very little uncertainty, we would be justified in concluding that there are no important risk factors we have not accounted for (assuming that coding practices are uniform and that risk adjustment is based on admission characteristics alone). A commonly used measure for degree of certainty in prediction is the C statistic, or area under the receiver-operating curve. It is interesting to compare our models with those of the more well-known clinical prediction models. For AMI, seven prediction models for 30-day mortality were validated on a large data set (59). The C values ranged from 0.74 to 0.78 for models using administrative information or both administrative and clinical information, compared to the value 0.74 for our model. A similar study for stroke reported C values from 0.60 to 0.77 for models based on administrative data (105), compared to our value of 0.73.

The risk adjustment models of the present study thus compares favorably with the more elaborate clinical models. This is an indication that our risk adjustment is adequate. Note however, that models based on administrative records may get some of their apparent predictive ability as a result of using diagnostic codes for high-severity states that arise late in the hospital stay (e.g. coma) and should properly be treated as complications, see (106;107). Also, even sophisticated models based on extensive clinical data have been shown to adjust inadequately for risk differences between patients from different admission sources, at least for intensive care units (104). In that study, the hypothesis is put forward that a history of failed standard treatment is a stronger prognostic factor than biophysiological measurements.

Distance from home to hospital was a significant covariate for AMI – longer distance was associated with lower risk of death. The interpretation is not straight forward. Is the explanation that the farther you are away from home, the more likely it is that you die before you get to hospital? Since the association was seen for the most acutely life-

threatening diagnosis only, this is probably the best explanation. For hip fracture, there is a marginally significant (p=0.056), positive association between distance and probability of death. For future use of the quality indicator, we strongly advise to establish procedures for having the exact time of onset of symptoms recorded at arrival.

In our study, clinical data were only collected for the sub analysis using the Canadian Stroke Scale, see 6.6.10. Unfortunately, a larger, clinical data sample would have been needed to draw any definite conclusions about coding accuracy in general.

As a control for risk adjustment, this study advises for use: age (via spline functions), sex, patient frailty as measured using the proxies number of previous admissions and number of pertinent codiagnoses from previous admissions; and disease severity, using proxies distance from home, simplified CCDSS staging and being transferred from another hospital. Not all variables are equally important, and parameters such as availability and time-frame may be taken into account when deciding on the final variable list. Socio-demographic variables and marital status may be important in the future and should ideally also be included.

7.2.4 Construct validity

The expert groups were asked to evaluate the findings according to their perceived knowledge of the quality of care at Norwegian hospitals. The results differed.

The AMI group was divided but did not consider that the results necessarily reflected their perception of quality of cardiological care in Norway. They also raised serious doubt about the results generated in this difficult time-period of change (1997-2001), since so many changes have occurred particularly in cardiology. On the other hand, the variability was not large for AMI mortality. Moreover, the group was generally not in opposition to the crude observations of case fatality or the staging procedure used. They recognized that the data were good enough for use by the hospitals as a source for quality improvement, but not for the construction of an external quality indicator.

The stroke group disapproved of the results of the initial staging analyses, and suggested the crude division into brain infarctions and hemorrhages. The group also determined that the results showed that hospitals led by distinguished leaders dedicated to the improvement of stroke treatment in Norway had significantly lower case fatality. They accepted the large variability for stroke case fatality, compared to AMI and hip fracture case fatality, and stated that it was in accordance with their expectations that modern stroke treatment probably is not implemented throughout the country.

The hip fracture group was not able to state clearly whether the results were in accordance with expectations at the units observed. They accepted 30-day mortality as an interesting evaluation of hip fracture treatment, but emphasized that it was probably of limited value as to the quality of the surgical treatment.

We have some data about the hospitals that could be related to the quality of care. The most specific data are from the questionnaire sent to each hospital. Because of the less than satisfactory response rate for the questionnaire, we have only made a rather limited analysis of these data, reported elsewhere (108). In particular, few of the outlying hospitals were represented.

In this sub analysis, regression analyses with hospital effects as dependent variable, were performed in two stages. In the first stage, including all hospitals, the following

explanatory variables were used: hospital averages of risk analysis variables, proportion of patients older than the third quartile (per disease), teaching status and volume (log-transformed).

In the second stage, the number of specialists and beds in specialist units was entered in the model. For AMI, troponine use was also included.

Few clear and unequivocally interpretable conclusions emerged from the analysis. Many of the explanatory variables are strongly correlated and may act as proxies for regional differences. The regression analyses are weighted to account for the hospital effect's standard deviation, meaning that the relatively few major hospitals will have a large influence on the results.

One overall conclusion is that risk analysis variables that are important in the patient level, on the hospital level have negative influences on hospital effects. This means that the hospitals with more severely ill or frail patients seem to perform better. However, the relation is not very strong, with very moderate values of R^2. The exception is the model for hip fracture including questionnaire variables.

We cannot determine conclusively the reason for the negative correlation between severity and frailty on one hand and apparent mortality on the other hand. One obvious possibility is undercoding of comorbidities. Another factor may be the correlation between early death and scarcity of codiagnoses noted earlier. However, it might simply be that good hospitals take greater care to identify, and hence record, comorbid conditions. The same phenomenon was reported in (90), where the authors concluded that undercoding of comorbidities was unlikely to account for outlier status of hospitals.

For stroke, there was a statistically significant association between case volume and mortality. As one would expect, high case volume acts in the direction of decreased mortality.

The last conclusion is that for hip fracture, the number of specialist beds per admission per year was significant and acted in the direction of decreased mortality. The model had an appreciable R^2 of 0.82. Changing the weighted number of previous diagnoses from the first to the third quartile would decrease the hospital effect by 0.20.

7.2.5 Fosters Real Quality Improvement.

Use of the quality indicator may add a further stimulus to the suspected overcoding of disease seriousness motivated by the DRG based financing system. Otherwise, there are no indications that release of this indicator would create incentives or rewards that would lead providers to improve results without improving quality of care.

7.2.6 Application

The indicator is used widely, and is well documented in the HTA report (1). A literature survey was performed to examine the comparability of the results measured in this study with special emphasis on the variability and range of differences between individual hospitals with those found in international studies.

Our most extreme (shrinkage) estimates of hospital effects for stroke and hip fracture have absolute magnitudes around 0.5. These estimates are subject to random error which is likely to be in the direction of too small effect magnitudes. Apparently, our results show that for these two disease groups, there is a substantial range in mortality between hospitals. The question arises whether this can be reconciled with what is known on this subject and in particular with what is known about effects of proven medical interventions.

Directly comparable figures come from studies comparing either individual hospitals, or small groups of hospitals. Studies of hospital types or interventions pool a large number of hospitals and thus tend to average out any hospital-specific differences.

Differences in decisions about curtailing care could possibly explain some of the observed variability in mortality, and should be subject of future investigations. In particular, this may be important in stroke and hip fracture with many elderly patients. In the predominantly US literature, differences have been reported for several population subgroups.

Acute Myocardial Infarction

In (109), based on administrative data from 180 hospitals, a standard deviation for the true hospital effects of 0.14 is reported for acute myocardial infarction.

Some studies examine various kinds of quality ratings for hospitals in relation to 30D mortality for AMI (110-112), using clinical data for risk adjustment. A difference in log-odds of 0.50, between the the highest and lowest rated groups, is found in (112). In (111), a range in log-odds of 0.20 is found between hospital peer groups, among the top rated hospitals (The corresponding range for the lower rated hospital category was 0.13). In (110), the log-odds difference between top rated hospitals and the rest was 0.26, and the difference between top-rated and similarly equipped hospitals 0.24.

Several studies report on the relation between various hospital characteristics and AMI mortality. The effect of teaching status was studied in (113) using administrative data. A difference in log-odds of in-hospital mortality of 0.16, between minor and major teaching hospitals, was reported. Based on a very large set of administrative data, a risk adjusted 30D effect range between hospital types of 0.37 was found (43). In (114), an effect of 0.64 was found between metropolitan and non-metropolitan hospitals, using clinical data. A significant association between hospital volume and mortality was reported in (115). Being admitted by a physician treating less than 5 cases per year was found in (116) to increase log-odds of 30D mortality by 0.30 compared to those treating more than 24 cases per year. After controlling for physician specialty, the effect was generally larger, from 0.28 to 0.91. The study controlled for both patient and hospital characteristics. Based on estimates for treatment effects from the literature, the net, hypothetical effect of proven interventions on actual patient populations is put at 0.46 in (117). For an empirical approach, see (118), where it is noted that 22% of the patients

were treated neither with aspirin nor thrombolytics, making a significant contribution to the overall mortality.

For AMI, the variability in hospital effects we find are somewhat in the lower end of the spectrum that can be inferred from the literature, as one would expect in view of the bias towards zero of our shrinkage estimates. It may also be due to greater variability in treatment in the time periods and hospital populations studied. It seems reasonable to conclude that our results do not run counter to what is known from the literature.

Stroke

In the case of stroke, several studies (119-121) compare hospital effects across countries or hospitals using clinical data for case-mix correction. The range in hospital effects seen (i.e. log-odds ratios) vary from 0.3 to 1.0. In one study, the effect range increased from 0.77 to 1.07 after service indicators were controlled for in the case-mix model. No explanation for the large differences was found (121). In the above figures, no correction (eg shrinkage estimation) has been made for the sampling variability of the reported individual effects. However, the standard errors were in the range 0.07-0.13 and thus not large enough to be of great influence, except for (121), where it is 0.63. One study of 30D in-hospital mortality (122), based on administrative data from 180 hospitals, does make a shrinkage correction, and reports a standard deviation for the true hospital effects of 0.006.

Comparisons between types of hospitals report moderate effects, from 0.13 to 0.37, relating to teaching status (43) or resource use (123).

A large, systematic review (124) estimates an effect of 0.20 from having a stroke unit. The median follow-up time for the reviewed experiments was one year.

For stroke, the variability in hospital effects we find are within the range of reported results from other comparisons of hospitals. In view of the bias towards zero of shrinkage estimates, our results are somewhat in the upper end. There is a discrepancy between observed effect ranges in these studies and what is easily explained on the basis of known effects of interventions (121), which is found in our study as well. It seems reasonable to conclude that our results do not run counter to what is known from the literature.

Hip Fracture

Concerning hip fracture, the study (125), using age and sex standardization as case-mix adjustment, reports a hospital effects range of 0.77 for 30D mortality in patients with fractured neck of femur (ICD-9 codes 820, 821.0-1). The sampling standard deviations of the hospital effects are small (0.09-0.17). In (126), using clinical data for case-mix adjustment, one hospital is reported to have a significant effect of -1.97 (-3.22 to -0.73) in 1992 (In a later (1997) follow-up study (127), this hospital is no longer found to be significantly different from the others).

The report (128) discusses upper bounds to the theoretically possible improvement in perioperative mortality for a specialized hip fracture unit. The report is based on clinical data. The limit for improvement in log-odds is found to lie in the range 0.7 – 1.5. Interestingly, the authors find that the effect of actively curtailing care is 0.47.

Based on a very large set of administrative data, a risk adjusted 30D effect range between hospital types of 0.23 was found for lower extremity fracture repair (43).

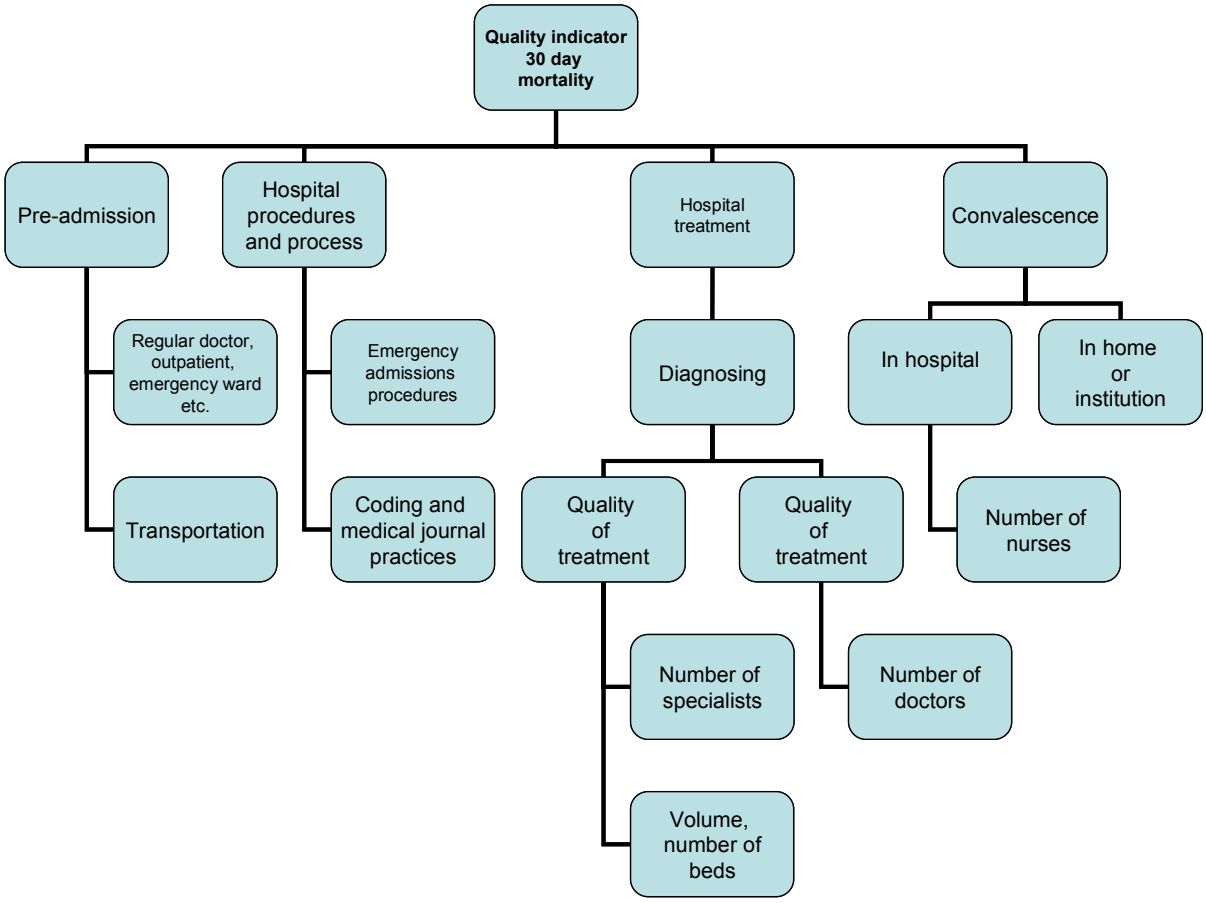
In a prospective study (129) with case-mix adjustment based on clinical data, delay until surgery is found to have an effect (on the log-odds scale) of 2.08 (0.77 to 3.38) per square-root transformed day on 1-year mortality. A systematic review (130) finds an effect of regional versus general anesthesia of 0.42 (0.19) on 1 month mortality.

For hip fracture, the variability in hospital effects in the present study seems compatible with the limited amount of results in the literature. In view of the bias towards zero of shrinkage estimates, our variability results may be somewhat large. There is a discrepancy between observed effect ranges in these studies and what is easily explained on the basis of known effects of interventions, apparent in our study as well.

7.3 CONCLUSIONS

Evaluating hospital health care quality includes not only evaluating treatment of diseases, but also evaluating administrative routines and processes, such as number of days at hospital, delay to operation, not enough capacity for patients, etc, as seen in the figure below.

Figure 7-2: The web of decisions and processes determining outcome and quality of care.



Differences in 30-day mortality between hospitals may depend on any decision or process in the figure. Also, there may well be perfectly legitimate reasons for two

hospitals to differ in procedures, processes and resource allocations. Factors outside the control of the hospital may also be important.

7.3.1 Are there substantial performance differences between hospitals?

Our most extreme (shrinkage) estimates of hospital effects for stroke and hip fracture have absolute magnitudes around 0.5. Roughly, this means that for some hospitals, the probability of death within 30 days is 65% greater than average. These estimates are subject to random error which is likely to be in the direction of too small effect magnitudes. Apparently, our results show that for these two disease groups, there is a substantial range in mortality between hospitals. The question arises whether this can be reconciled with what is known on this subject and in particular with what is known about effects of proven medical interventions.

Hospital effects

Hospital effects are the log-odds of death within 30 days at the various hospitals, compared to the average hospital, controlling for risk adjustment covariates (see chapter 5.6).

An effect of e.g. -0.182 means that this particular hospitals has 10% reduced log-odds for death, while an effect of e.g. 0.41 means that the log-odds of death are 50% greater than in the average hospital, given the levels of the risk adjustment covariates.

For AMI, the variability in hospital effects we find is fairly small and well within the limits reported in the literature.

For stroke, the variability in hospital effects we find are within the range of reported results from other comparisons of hospitals, though somewhat in the upper end. There is a discrepancy between observed effect ranges in these studies and what is easily explained on the basis of known effects of interventions (see the discussion in (121)), which is found in our study as well. It seems reasonable to conclude that our results do not run counter to what is known from the literature.

For hip fracture, the variability in hospital effects in the present study seems compatible with the limited amount of results in the literature. In view of the bias towards zero of shrinkage estimates, our variability results may be somewhat large. There is a discrepancy between observed effect ranges in these studies and what is easily explained on the basis of known effects of interventions, apparent in our study as well.

Especially for stroke and hip fracture, differences in the way hospitals make decisions about curtailing care could possibly explain some of the observed variability. This should be the subject of future investigations.

7.3.2 Can and should 30-day mortality based on administrative data be used as a quality indicator?

The main issue is whether mortality measures based on administrative data are valid indicators of true, hospital-specific mortality, while accounting for presumed bias, resulting from inaccurate coding, diagnostic variability and less than ideal case-mix adjustment. On the one hand, there remains a possibility that the bias of the indicator is large enough to influence the comparison between hospitals in a significant way. On the

other hand, the results indicate that there are unacceptably high differences between hospitals. A review of the literature indicates that these differences seem to agree with those reported internationally. It is the role of the public health authorities to weigh the risk of incorrectly exposing hospitals as having poor quality, against the possibility that large apparent discrepancies in mortality reflect a true situation.

The purpose of this study was to evaluate the use of 30-day mortality as a quality indicator in Norway by examining the precision, and the robustness of the indicator to different methods of risk adjustment that examine the known and measurable sources of bias.

In the international literature, the issue is vigorously debated, and different authors make very different conclusions. Unfortunately, with some exceptions, the literature does not apply with sufficient precision to Norwegian hospitals and their present practices. It is noteworthy, however, that 30-day mortality was evaluated in the HTA report by AHRQ (1) as one of 200 quality indicators to be evaluated, and was accepted (among a group of 45) as an internal quality indicator for the disease groups acute myocardial infarction, stroke and hip fracture. In addition, the indicator is used in several countries.

Besides bias, the most important criterion is precision. We have shown that the mortality indicator can be used to identify, with good statistical precision, hospitals where the probability of dying is appreciably different from the average. We have identified three perspectives of decision-making: A) the individual hospital, B) public authorities and policy makers, C) the public. The statistical methods and their associated parameters differ among these perspectives.

Mortality indicators can aid hospitals in their efforts for excellence. In addition, hospital administrations are presumably in a better position to judge the bias in their own quality indicator numbers resulting from inaccurate coding practices.

Since the results of 30-day mortality reflect disease states in emergency admissions, and that patients cannot in reality select the site of emergency admission, this would indicate that this indicator is less relevant and useful for decision-making for the general public.

The present study is limited by the lack of clinical data and independent validation of diagnoses and codes. Within these limits, we have performed a study of plausible bias magnitudes indicating that unacceptable bias is probably avoided. Still, we feel that the issue is not settled in a satisfactory way. Further study, geared towards resolving the bias question, is recommended.

We have identified some less fundamental issues that need to be addressed: the need for more reliable registration of very early deaths, or choosing a strategy to reduce the sensitivity of 30D mortality to these cases (particularly for acute myocardial infarction), as well as the use of correct decision rules to identify hospitals as performance outliers. It is necessary to finalise the decision rules and their parameters, based on discussion with the various users of the indicators.

Further studies should include

- evaluation of the results using clinical and laboratory data in addition to information from journals and direct communication with hospitals.
- further development of risk adjustment methods such as Charlson Comorbidity Index and DRG index.

- expansion of the data base to 2005 such that problems involving the transition from ICD-9 to ICD-10 can be eliminated and that present-day coding practices can be validated.
- in particular, the hospitals scoring exceptionally high or low in probability of death must be covered adequately in the validation, which should attempt to identify the causes of any performance deviations, e.g. by interviews with hospital administration.

On a somewhat longer term, we note that our results are obtained from a rather homogenous population of hospitals. If trends towards greater specialization continue, patients may be selected and transferred to a greater degree than now, and special methods may be needed to handle this.

The criteria suggested by the HTA (Health Technology Assessment) report (1) were used as a conceptual framework for the evaluation. The results are summarized in Table 7-4 below.

Table 7-4: Evaluation of quality indicators.

<table>
<tr><th>Evaluation criterion</th><th colspan="3">Conclusion</th></tr>
<tr><td>Face validity</td><td colspan="3">The disease categories are major causes of death. It is possible to provide results on a year-by-year basis.</td></tr>
<tr><td rowspan="3">Precision</td><td colspan="3">We have judged precision (reliability) as the ability to have low type II error probability, while keeping the relevant type I error probability under control. Proper decision rules, based on the user's decision perspective, are to be applied. Error probabilities are low for AMI and hip fracture, and acceptable for stroke.
The study group's assessment of precision based on type of quality indicator and disease category[a)]:</td></tr>
<tr><td>AMI</td><td>Stroke</td><td>Hip fracture</td></tr>
<tr><td>Good</td><td>Good</td><td>Good</td></tr>
<tr><td rowspan="3">Minimum bias</td><td colspan="3">Without good coverage of clinical data, there will necessarily be some uncertainty whether data quality and risk adjustment is adequate to exclude any case-mix bias in hospital comparisons. However, there were few indications that systematic differences in case-mix did in fact exist between hospitals. Robustness testing resulted in few differences between models.
Theoretical sensitivity studies seem to indicate that most kinds of bias are of small to moderate magnitude. It is, however, necessary to investigate further the coding practices for dead on arrival.
The study group's assessment of minimum bias based on type of quality indicator and disease category[a)]:</td></tr>
<tr><td>AMI</td><td>Stroke</td><td>Hip fracture</td></tr>
<tr><td>Acceptable [b)]</td><td>Acceptable</td><td>Acceptable</td></tr>
<tr><td>Construct validity</td><td colspan="3">There was no clear indication that outlier status for an individual hospital could be explained by hospital characteristics.</td></tr>
<tr><td>Fosters Real Quality Improvement</td><td colspan="3">The indicator may provide further stimulus to incorrect coding. Otherwise, there are no indications that using this indicator would create incentives that would lead providers to improve performance without improving quality of care.</td></tr>
</table>

Evaluation criterion	Conclusion
Application	The indicator is widely used, and is well documented in the HTA report published by AHRQ. For stroke and hip fracture, there are strong indications that there are substantial differences between hospitals in probability of death after 30 days. A review of the literature resulted in the conclusion that the substantial performance differences found in this study do not run counter to what is known from the literature for AMI or stroke and to a lesser degree hip fracture.

a) Criteria for evaluation of quality indicators are based on those found in the HTA report published by AHRQ (Agency for Health Care Research and Quality)(1).

b) On the condition that uncertainties concerning coding of dead on arrival is satisfactorily resolved.

Limited to PAS data and national statistics, this study recommends the following list of risk adjustment variables:

- age (via spline functions),
- sex,
- patient frailty as measured using the proxies number of previous admissions and number of pertinent codiagnoses from previous admissions,
- disease severity, using the proxies distance from home, simplified CCDSS (Clinical Criteria Disease Staging System) staging and being transferred from another hospital,
- distance from home and socio-demographic data.

The predictive value of distance from home must be weighed against the fact that it is currently not available in the same time-frame as the PAS data. The socio-demographic variables had predictive value on the individual level, but not on the hospital level. In the future, the socio-demographic case-mix may well become less uniform between hospitals. It is therefore desirable to retain these variables in the model.

7.4 FUTURE RECOMMENDATIONS

The indicator that has been mostly accessible historically has been in-hospital case fatality. In-hospital case fatality was the indicator used in the earlier study on mortality indicators in Norway (2). It is also the indicator used in the AHRQ survey (1). However, there is universal agreement that in-hospital case fatality includes many biases and should be replaced if possible with probability of death a fixed number of days post-admission, or based on total survival (Cox analysis). The most commonly used is 30-day. However, 28-days have also been used. In this report we are focused on 30-day as a commonly accepted indicator. However, 30 days is not an even number of weeks and can contain a bias if there exists differences in probability of death per weekday or based on admission days. Since this was suggested for stroke (see section 6.3.4), we suggest that in the future, also 28 days should be considered.

A closer exploration of which point in time that provides the best point for measuring probability of death for different diseases using Cox Proportional Hazard models will follow. There are indications that 30 days after admission not to be the optimal point for any of the three diseases we have studied, except possibly for stroke.

Although this study focused on acute myocardial infarction, stroke and hip fracture, other disease categories should be considered in the future for investigation. These may include mortality from congestive heart failure, GI hemorrhage, chronic obstructive lung disease and pneumonia. In addition, post-procedural probability of death should also be investigated.

With a sufficiently large number of quality indicators, a composite index for overall quality of hospitals could be constructed, and this possibility should then be investigated. Because patients with different diseases are admitted to different organizational units, it is not clear a priori that such an index would yield a meaningful indicator for the whole hospital. Short of constructing an index, an analysis of the correlations between the various indicators may still yield important information and should be undertaken.

7.4.1.1 Data availability and quality

While much information about treatment quality must rest on prospectively recorded data in disease specific registries, we think that further refinement of 30-day mortality based on PAS data is possible and advisable. Since in the future, the hospitals will know which data are required, they will be more ready to record them in order to obtain fair comparisons with other hospitals. But extra efforts needed for recording procedures have to be balanced against the gain in information. Researchers always want more information, but this should not lead to exhaustion on the clinical side. We must remember that the refinement we seek primarily is needed to make fair comparisons, the crude data are by and large indicative of what goes on. The following summarizes the needs in data availability identified in this study:

- There is a need for more accurate diagnostic coding. In particular, a more detailed description of the patient's state at admission and time from onset of symptoms to admission should be implemented. In addition, information concerning whether or not the admission is related to rehabilitation should be more precisely noted.
- We need to develop better ways of recording important organizational data from the hospitals, including costs, treatment units and procedures, and relevant data from laboratories and radiology units.

7.4.1.2 Research and development needs

There exist further issues that might profitably be investigated studied before concluding on the use of 30-day mortality as a quality indicator. These include:

- Survival models with added severity states should be investigated to take care of the timing of codiagnoses, and to utilize the information in the time to death variable. An observation period of 30 days after admission may not be optimal. Ideally, times of important events such as treatment, diagnoses etc should be available.
- A more sophisticated study of how bias is produced by various error mechanisms should be undertaken, with the aim of providing more precise bias estimates.
- Correct and uniform coding of dead on arrival or being under resuscitation is critical. The journals for patients suspected to be in these groups should be examined and recoded if necessary. Interviews must be carried out to

ascertain the routines and practices at each hospital. Consistent coding of dead on arrival or being under resuscitation, as well as date and time of death, must be established.

- The study should be reanalyzed using a later period (up to 2005) so that the analyses can be restricted to using only ICD-10 coding. Shifting coding practices during the study period added uncertainties.
- A further development of other methods to define patient frailty, such as the Charlson comorbidity index and/or the ASA score (or other internationally widely used instrument or the use of DRG codes.
- A further development of other methods to define disease severity. Many available models are used in the literature, including Acute Physiological and Chronic Health Evaluation (APACHE II), Computerized Severity Index (CSI), Patient Management Categories (PMC), Medisgroups (MDGRP), Simplified Acute Physiology Score (SAPS), Coded Disease Staging (CDS), and Clinical Criteria Disease Staging System (CCDSS – the one we chose to use). APACHE II, CSI, MDGRP are based on physiological variables whereas PMC, CDS, and CCDSS are based on journal discharge abstracts. There have been indications that the predictive abilities of the journal based indexes are as good as those based on physiological parameters (86;87).
- An in-depth study should be made within the hospitals that are placed on the follow-up list (both those positively or negatively deviant from the mean), to find explanations for the differences found. This would serve two purposes: as a check for overlooked flaws in risk adjustment, and provide a test of face validity.
- A more systematic analysis of available data and questionnaire information to further test the question of face validity.
- A more in-depth study of the underlying patterns in transferral of patients between hospitals. This includes an analysis of type of hospitals that patients move between, socio-demographic characteristics of patients that are transferred and disease severity and patient frailty issues involved in the transfer process.
- Improve proxy of distance from home to hospital of admission by including a new set of parameters, distance from home to all hospitals in catchment area. This information is to be used to studying possible selection effects.
- Development of more sophisticated decision rules for drawing conclusions about hospital performance.
- Investigate the possibility and feasibility of correlating other currently available indicators such as patient satisfaction, queue size and time from entering queue to treatment, number of patients in corridors, etc. to results in this study.

8. REFERENCES

(1) Davies SM, Geppart J, McClellan M, McDonald KM, Romano PS, Shojania K.G. Refinement of the HCUP Quality Indicators. AHRQ Publication No. 01-0035, 1-427. 5-1-2001. Rockville, MD, Agency for Healthcare Research and Quality.

(2) Guldvog B, Kopjar B. In-hospital mortality rates in Norway 1994-97. 2 - 1999, 1-90. 1999. Lørenskog, Norway, Foundation for Health Services Research.

(3) Allison JJ, Kiefe CI, Weissman NW, Person SD, Rousculp M, Canto JG et al. Relationship of hospital teaching status with quality of care and mortality for Medicare patients with acute MI. JAMA 2000; 284(10):1256-1262.

(4) Curtin LL. An integrated analysis of nurse staffing and related variables: effects on patient outcomes. Online J Issues Nurs 2003; 8(3):5.

(5) Person SD, Allison JJ, Kiefe CI, Weaver MT, Williams OD, Centor RM et al. Nurse staffing and mortality for Medicare patients with acute myocardial infarction. Med Care 2004; 42(1):4-12.

(6) Mark BA, Harless DW, McCue M, Xu Y. A longitudinal examination of hospital registered nurse staffing and quality of care. Health Serv Res 2004; 39(2):279-300.

(7) Aiken LH, Clarke SP, Sloane DM, Sochalski J, Silber JH. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. JAMA 2002 Oct 23-30 288;1987-1993.

(8) Pronovost PJ, Angus DC, Dorman T, Robinson KA, Dremsizov TT, Young TL. Physician staffing patterns and clinical outcomes in critically ill patients: a systematic review. JAMA 2002; 288(17):2151-2162.

(9) Sosial- og Helsedepartementet. Rapport fra Arbeidsgruppen for Utvikling av Kvalitetsindikatorer for Behandlingstilbudet i Somatiske Sykehus. Sosial- og Helsedepartementet, editor. 1-88. 27-4-2001.

(10) Garnick DW, DeLong ER, Luft HS. Measuring hospital mortality rates: are 30-day data enough? Ischemic Heart Disease Patient Outcomes Research Team. Health Serv Res 1995 Feb 29:679-695.

(11) Iezzoni LIed. Risk adjustment for measuring healthcare outcomes. Chicago: Health Administration Press, 1997.

(12) Brinkley J. US releasing lists of hospitals with abnormal mortality rates. New York Times , 1. 3-12-1986.

(13) Keeler EB, Kahn KL, Draper D, Sherwood MJ, Rubenstein LV, Reinisch EJ et al. Changes in sickness at admission following the introduction of the prospective payment system. JAMA 1990 Oct 17 264:1962-1968.

(14) Knaus WA, Harrell FE, Jr., Lynn J, Goldman L, Phillips RS, Connors AF, Jr. et al. The SUPPORT prognostic model. Objective estimates of survival for seriously ill hospitalized adults. Study to understand prognoses and preferences for outcomes and risks of treatments. Ann Intern Med 1995; 122(3):191-203.

(15) Montague TJ, Ikuta RM, Wong RY, Bay KS, Teo KK, Davies NJ. Comparison of risk and patterns of practice in patients older and younger than 70 years with acute myocardial infarction in a two-year period (1987-1989). Am J Cardiol 1991; 68(9):843-847.

(16) Weintraub WS, Craver JM, Cohen CL, Jones EL, Guyton RA. Influence of age on results of coronary artery surgery. Circulation 1991; 84(5 Suppl):III226-III235.

(17) Daley J, Forbes MG, Young GJ, Charns MP, Gibbs JO, Hur K et al. Validating risk-adjusted surgical outcomes: site visit assessment of process and structure. National VA Surgical Risk Study. J Am Coll Surg 1997; 185(4):341-351.

(18) Njolstad I, Arnesen E. Preinfarction blood pressure and smoking are determinants for a fatal outcome of myocardial infarction: a prospective analysis from the Finnmark Study. Arch Intern Med 1998 Jun 22 158:1326-1332.

(19) Ayanian JZ, Epstein AM. Differences in the use of procedures between women and men hospitalized for coronary heart disease. N Engl J Med 1991; 325(4):221-225.

(20) Gozum ME, Gonnella JS, Louis DZ. Disease Staging Clinical Criteria. 4 ed. Ann Arbor Michigan: MEDSTAT group, 1994.

(21) Selmer R. Blood pressure and twenty-year mortality in the city of Bergen, Norway. Am J Epidemiol 1992 Aug 15 136:428-440.

(22) Selmer R, Tverdal A. Mortality from stroke, coronary heart disease and all causes related to blood pressure and length of follow-up. Scand J Soc Med 1994 Dec 22:273-282.

(23) [Ischemic stroke associated with atrial fibrillation: the demographic and clinical characteristics and 30-day mortality in a hospital stroke registry. The European Community Stroke Project, Florence Unit]. Ann Ital Med Int 1996 Jan -Mar 11:20-26.

(24) Ellekjaer EF, Wyller TB, Sverre JM, Holmen J. Lifestyle factors and risk of cerebral infarction. Stroke 1992 Jun 23:829-834.

(25) Greenfield S, Sullivan L, Silliman RA, Dukes K, Kaplan SH. Principles and practice of case mix adjustment: applications to end-stage renal disease. Am J Kidney Dis 1994; 24(2):298-307.

(26) Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987; 40(5):373-383.

(27) Poses RM, McClish DK, Smith WR, Bekes C, Scott WE. Prediction of survival of critically ill patients by admission comorbidity. J Clin Epidemiol 1996; 49(7):743-747.

(28) Dahl E. Sosial ulikhet i helse: Artifakter eller seleksjon? [Social inequality in health: artifacts or selection?]. FAFO 170. 1994. Oslo, FAFO.

(29) Dahl E, Elstad JI. Recent changes in social structure and health inequalities in Norway. Scand J Public Health 2001; Suppl 55:7-17.:7-17.

(30) Elstad JI. Social inequalities in health and their explanations. 9/2000. 2000. Oslo, NOVA/Norwegian Social Research.

(31) Valkonen T, Martelin T, Rimpelä A, Notkola V, Savela S. Socio-economic mortaltiy differentials in Finland 1981-90. Population, 1993:1. 1993. Helsinki, Statistics Finland.

(32) Fox Je. Health inequalities in Euroepan countries. Aldershot: Gower, 1989.

(33) Lynch JW, Smith GD, Kaplan GA, House JS. Income inequality and mortality: importance to health of individual income, psychosocial environment, or material conditions. BMJ 2000; 320(7243):1200-1204.

(34) Townsend P, Davidson N. Inequalities in Health. The Black Report. Harmondsworth: Penguin Books, 1982.

(35) Wilkinson RG. Unhealthy societies. London: Routledge, 1996.

(36) Wilkinson RG, Kawachi I, Kennedy BP. Mortality, the social environment, crime and violence. Sociology of Health & Illness 1998; 20:578-597.

(37) Wilkinson RG. Inequality and the social environment: a reply to Lynch et al. J Epidemiol Community Health 2000; 54(6):411-413.

(38) American Cancer Society. Cancer and the poor: A report to the nation. Atlanta: The Society, 1989.

(39) Sherbourne CD, Stewart AL. The MOS social support survey. Soc Sci Med 1991; 32(6):705-714.

(40) Sherbourne CD, Meredith LS, Rogers W, Ware JE, Jr. Social support and stressful life events: age differences in their effects on health-related quality of life among the chronically ill. Qual Life Res 1992; 1(4):235-246.

(41) Case RB, Moss AJ, Case N, McDermott M, Eberly S. Living alone after myocardial infarction. Impact on prognosis. JAMA 1992; 267(4):515-519.

(42) Williams RB, Barefoot JC, Califf RM, Haney TL, Saunders WB, Pryor DB et al. Prognostic importance of social and economic resources among medically treated patients with angiographically documented coronary artery disease. JAMA 1992; 267(4):520-524.

(43) Yuan Z, Cooper GS, Einstadter D, Cebul RD, Rimm AA. The association between hospital type and mortality and length of stay: a study of 16.9 million hospitalized Medicare beneficiaries. Med Care 2000 Feb 38:231-245.

(44) Block BM, Sirio CA, Cooper GS, DiGiuseppe DL, Rosenthal GE. Use of intensive care-specific interventions in major teaching and other hospitals: a regional comparison. Crit Care Med 2000 Apr 28:1204-1207.

(45) Fisher ES, Wennberg JE, Stukel TA, Skinner JS, Sharp SM, Freeman JL et al. Associations among hospital capacity, utilization, and mortality of US Medicare

beneficiaries, controlling for sociodemographic factors. Health Serv Res 2000 Feb 34:1351-1362.

(46) Glenn LL, Jijon CR. Risk-adjusted in-hospital death rates for peer hospitals in rural and urban regions. J Rural Health 1999 Winter 15:94-107.

(47) Dudley RA, Johansen KL, Brand R, Rennie DJ, Milstein A. Selective referral to high-volume hospitals: estimating potentially avoidable deaths. JAMA 2000 Mar 1 283:1159-1166.

(48) McGrath PD, Wennberg DE, Dickens JDJ, Siewers AE, Lucas FL, Malenka DJ et al. Relation between operator and hospital volume and outcomes following percutaneous coronary interventions in the era of the coronary stent. JAMA 2000 Dec 27 284:3139-3144.

(49) Ornato JP, Peberdy MA, Chandra NC, Bush DE. Seasonal pattern of acute myocardial infarction in the National Registry of Myocardial Infarction. J Am Coll Cardiol 1996 Dec 28:1684-1688.

(50) Green J, Wintfeld N. How accurate are hospital discharge data for evaluating effectiveness of care? Med Care 1993 Aug 31:719-731.

(51) Ciccone G, Bertero D, Bruno A, Canavese C, Ciccarelli E, Ivaldi C et al. [Quality of data or quality of care? Comparison of diverse standarization methods by clinical severity, based on the discharge form, in the analysis of hospital mortality]. Epidemiol Prev 1999 Oct -Dec 23:286-293.

(52) Thomas JW, Hofer TP. Accuracy of risk-adjusted mortality rate as a measure of hospital quality of care. Med Care 1999 Jan 37:83-92.

(53) Hofer TP, Hayward RA. Identifying poor-quality hospitals. Can hospital mortality rates detect quality problems for medical diagnoses? Med Care 1996 Aug 34:737-753.

(54) Gowrisankaran G, Town RJ. Estimating the quality of care in hospitals using instrumental variables. J Health Econ 1999 Dec 18:747-767.

(55) Green J, Passman LJ, Wintfeld N. Analyzing hospital mortality. The consequences of diversity in patient mix. JAMA 1991 Apr 10 265:1849-1853.

(56) Merlo J, Ostergren PO, Broms K, Bjorck-Linne A, Liedholm H. Survival after initial hospitalisation for heart failure: a multilevel analysis of patients in Swedish acute care hospitals. J Epidemiol Community Health 2001 May 55:323-329.

(57) Polanczyk CA, Rohde LE, Dec GW, DiSalvo T. Ten-year trends in hospital care for congestive heart failure: improved outcomes and increased use of resources. Arch Intern Med 2000 Feb 14 160:325-332.

(58) Lied TR, Kazandjian VA, Hohman SF. Impact of risk adjusted clinical outcomes methodology--quality measures on hospital mortality data: a statistical and case study approach. Am J Med Qual 1999 Nov -Dec 14:255-261.

(59) Krumholz HM, Chen J, Wang Y, Radford MJ, Chen YT, Marciniak TA. Comparing AMI mortality among hospitals in patients 65 years of age and older: evaluating methods of risk adjustment. Circulation 1999 Jun 15 99:2986-2992.

(60) Van Ruiswyk J, Hartz A, Kuhn E, Krakauer H, Young M, Rimm A. A measure of mortality risk for elderly patients with acute myocardial infarction. Med Decis Making 1993 Apr -Jun 13:152-160.

(61) Rosenthal GE, Baker DW, Norris DG, Way LE, Harper DL, Snow RJ. Relationships between in-hospital and 30-day standardized hospital mortality: implications for profiling hospitals. Health Serv Res 2000 Mar 34:1449-1468.

(62) Kuhn EM, Hartz AJ, Krakauer H, Bailey RC, Rimm AA. The relationship of hospital ownership and teaching status to 30- and 180-day adjusted mortality rates. Med Care 1994 Nov 32:1098-1108.

(63) Rolstad OJ, Stromme JH, Mangschau A. [New cardiac markers--clinical benefits in early diagnosis of acute heart disease]. Tidsskr Nor Laegeforen 2001; 121(4):415-420.

(64) Stromme JH, Rolstad OJ, Mangschau A. [Troponins and other biochemical cardiac markers--time for a change]. Tidsskr Nor Laegeforen 2000; 120(16):1863-1869.

(65) Myocardial infarction redefined--a consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction. Eur Heart J 2000 Sep 21:1502-1513.

(66) Vik-Mo H. Intravascular ultrasound improves the quality of coronary interventions, but is it worth the extra time and money invested? Scand Cardiovasc J 2001; 35(2):67-69.

(67) Stromme JH, Halvorsen S, Frederichsen P. [Diagnoses and increased levels of troponin T among discharged patients]. Tidsskr Nor Laegeforen 2001; 121(26):3041-3045.

(68) Hantson L, De Weerdt W, De Keyser J, Diener HC, Franke C, Palm R et al. The European Stroke Scale. Stroke 1994 Nov 1925;2215-2219.

(69) Goldstein LB, Chilukuri V. Retrospective assessment of initial stroke severity with the Canadian Neurological Scale. Stroke 1997 Jun 1928;1181-1184.

(70) R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2004.

(71) Chambers JM, Hastie TJ. Statistical Models in S. Chapman & Hall/CRC, 1991.

(72) Greenland S. Introduction to Regression Models. In: Rothman K, Greenland S, editors. Modern Epidemiology, Second Edition. Lippincott Williams & Wilkins, 1998: 359-399.

(73) McCullagh P, Nelder JA. Generalized linear models. 2 ed. Chapman & Hall/CRC, 1999.

(74) Bretz F, Genz A, Hothorn LA. On the numerical availability of multiple comparison procedures. Biometrical Journal 2001; 43(5):645-656.

(75) Tukey JW. Conclusions vs. decisions. Technometrics 1960; 2:423-433.

(76) Austin PC, Alter DA, Tu JV. The use of fixed- and random-effects models for classifying hospitals as mortality outliers: a Monte Carlo assessment. Med Decis Making 2003; 23(6):526-539.

(77) McCulloch CE, Searle SR. Generalized, Linear, and Mixed Models. Wiley, 2001.

(78) Stukenborg GJ, Wagner DP, Harrell FE, Jr., Oliver MN, Kilbridge KL, Lyman J et al. Hospital discharge abstract data on comorbidity improved the prediction of death among patients hospitalized with aspiration pneumonia. J Clin Epidemiol 2004; 57(5):522-532.

(79) Piccirillo JF, Tierney RM, Costas I, Grove L, Spitznagel EL, Jr. Prognostic importance of comorbidity in a hospital-based cancer registry. JAMA 2004; 291(20):2441-2447.

(80) Alter DA, Austin PC, Naylor CD, Tu JV. Factoring socioeconomic status into cardiac performance profiling for hospitals: does it matter? Med Care 2002; 40(1):60-67.

(81) Normand ST, Glickman ME, Sharma RG, McNeil BJ. Using admission characteristics to predict short-term mortality from myocardial infarction in elderly patients. Results from the Cooperative Cardiovascular Project. JAMA 1996 May 1 275:1322-1328.

(82) Tu JV, Austin PC, Walld R, Roos L, Agras J, McDonald KM. Development and validation of the Ontario acute myocardial infarction mortality prediction rules. J Am Coll Cardiol 2001; 37(4):992-997.

(83) Normand S-LT, Glickman ME, Gatsonis CA. Statistical Methods for Profiling Providers of Medical Care: Issues and Applications. Journal of the American Statistical Association 1997; 92(439):803-814.

(84) Hjort N, Mowinckel P. Survival after hospitalisation in Norwegian hospitals: Aspects of statistical modelling. 2004. Dept. of Mathematics, University of Oslo. Statistical Report.

(85) Goldstein H. Multilevel Statistical Models. 1995. London, New York, Edward Arnold, Wiley.

(86) Alemi F, Rice J, Hankins R. Predicting In-Hospital Survival of Myocardial-Infarction - A Comparative-Study of Various Severity Measures. Medical Care 1990; 28(9):762-775.

(87) Thomas JW, Ashcraft MLF. Measuring Severity of Illness - 6 Severity Systems and Their Ability to Explain Cost Variations. Inquiry-the Journal of Health Care Organization Provision and Financing 1991; 28(1):39-55.

(88) Midttun L, Sverrbo E, Thorsen G, Steinum O. Er det sammenfall mellom journalopplysninger og innrapporterte data? En studie av 500 pasientopphold ved norske somatiske sykehus i 2001. STF78 A035504, 1-74. 6-5-2003. Trondheim, Norway, SINTEF.

(89) Jørgenvåg R, Hope ØB. Kvalitet på medisinsk koding og ISF-refusjoner. I hvilken grad er journalgjennomgang et nyttig verktøy? NPR, editor. STF78 A055501, 1-64. 22-2-2005. Oslo, SINTEF.

(90) Austin PC, Tu JV, Alter DA, Naylor CD. The impact of under coding of cardiac severity and comorbid diseases on the accuracy of hospital report cards. Med Care 2005 Aug 1943;801-809.

(91) Romano PS, Chan BK. Risk-adjusting acute myocardial infarction mortality: are APR-DRGs the right tool? Health Serv Res 2000 Mar 1934;1469-1489.

(92) Jayes RL, Zimmerman JE, Wagner DP, Draper EA, Knaus WA. Do-not-resuscitate orders in intensive care units. Current practices and recent changes. JAMA 1993; 270(18):2213-2217.

(93) Hagen TP, Reikvam A. [Marked increase of the number of myocardial infarctions following introduction of the new diagnostic criteria]. Tidsskr Nor Laegeforen 2003 Nov 6 123;3041-3043.

(94) Rinaldi R, Vignatelli L, Galeotti M, Azzimondi G, de Carolis P. Accuracy of ICD-9 codes in identifying ischemic stroke in the General Hospital of Lugo di Romagna (Italy). Neurol Sci 2003 Jun 1924;65-69.

(95) Williams GR, Jiang JG, Matchar DB, Samsa GP. Incidence and occurrence of total (first-ever and recurrent) stroke. Stroke 1999 Dec 30:2523-2528.

(96) Ellekjaer H, Holmen J, Kruger O, Terent A. Identification of incident stroke in Norway: hospital discharge data compared with a population-based stroke register. Stroke 1999 Jan 30:56-60.

(97) Iezzoni LI, Shwartz M, Ash AS, Mackiernan YD. Predicting in-hospital mortality for stroke patients: results differ across severity-measurement methods. Med Decis Making 1996 Oct -Dec 1916;348-356.

(98) Lofthus CM, Cappelen I, Osnes EK, Falch JA, Kristiansen IS, Medhus AW et al. Local and national electronic databases in Norway demonstrate a varying degree of validity. J Clin Epidemiol 2005 Mar 1958;280-285.

(99) Fylkesrevisjonen i Sør-Trøndelag. Innsatsstyrt finansiering - kodepraksis. Forvaltningsrevisjonsrapport 2000, 1-41. 2000. Fylkesrevisjonen i Sør-Trøndelag.

(100) Myocardial infarctions in Sweden 1987-2000. 2003. Stockholm, Socialstyrelsen.

(101) Värdering av diagnoskvaliteten för akut hjärtinfarkt i patientregistret 1987 och 1995. 2000. Stockholm, Socialstyrelsen.

(102) Analyserande rapport från Riks-Stroke för heåren 1999 og 2000. 2002. Umeå, Västerbotten Läns Landsting.

(103) Thorngren K-G. Rikshöft-SAHFE. 2004. Stockholm, Socialstyrelsen.

(104) Rosenberg AL, Hofer TP, Strachan C, Watts CM, Hayward RA. Accepting critically ill transfer patients: adverse effect on a referral center's outcome and benchmark measures. Ann Intern Med 2003; 138(11):882-890.

(105) Iezzoni LI, Shwartz M, Ash AS, Hughes JS, Daley J, Mackiernan YD. Using severity-adjusted stroke mortality rates to judge hospitals. Int J Qual Health Care 1995; 7(2):81-94.

(106) Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernan YD. Judging hospitals by severity-adjusted mortality rates: the influence of the severity-adjustment method. Am J Public Health 1996; 86(10):1379-1387.

(107) Romano PS, Chan BK. Risk-adjusting acute myocardial infarction mortality: are APR-DRGs the right tool? Health Serv Res 2000 Mar 34:1469-1489.

(108) Helgeland J. 30D - forklaringsmodeller på sykehusnivå (in Norwegian). 2005. Oslo, Norway, Norwegian Knowledge Centre for the Health Services. Working paper.

(109) Every NR, Maynard C, Schulman K, Ritchie JL. The association between institutional primary angioplasty procedure volume and outcome in elderly Americans. J Invasive Cardiol 2000 Jun 12:303-308.

(110) Cleves MA, Golden WE. Assessment of HCFA's 1992 Medicare hospital information report of mortality following admission for hip arthroplasty. Health Serv Res 1996 Apr 31:39-48.

(111) Roos LL, Walld RK, Romano PS, Roberecki S. Short-term mortality after repair of hip fracture. Do Manitoba elderly do worse? Med Care 1996 Apr 34:310-326.

(112) Krumholz HM, Rathore SS, Chen J, Wang Y, Radford MJ. Evaluation of a consumer-oriented internet health care report card: the risk of quality ratings based on mortality data. JAMA 2002; 287(10):1277-1287.

(113) Polanczyk CA, Lane A, Coburn M, Philbin EF, Dec GW, DiSalvo TG. Hospital outcomes in major teaching, minor teaching, and nonteaching hospitals in New York state. Am J Med 2002 Mar 112;255-261.

(114) Vu HD, Heller RF, Lim LL, D'Este C, O'Connell RL. Mortality after acute myocardial infarction is lower in metropolitan regions than in non-metropolitan regions. J Epidemiol Community Health 2000 Aug 1954;590-595.

(115) Thiemann DR, Coresh J, Oetgen WJ, Powe NR. The association between hospital volume and survival after acute myocardial infarction in elderly patients. N Engl J Med 1999 May 27 340;1640-1648.

(116) Tu JV, Austin PC, Chan BT. Relationship between annual volume of patients treated by admitting physician and mortality after acute myocardial infarction. JAMA 2001; 285(24):3116-3122.

(117) Mant J, Hicks N. Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction. BMJ 1995 Sep 23 311;793-796.

(118) Reikvam A, Aursnes I. Hospital mortality from acute myocardial infarction has been modestly reduced after introduction of thrombolytics and aspirin: results from a new analytical approach. European Secondary Prevention Study Group. J Clin Epidemiol 1999 Jul 52:609-613.

(119) Weir NU, Sandercock PA, Lewis SC, Signorini DF, Warlow CP. Variations between countries in outcome after stroke in the International Stroke Trial (IST). Stroke 2001 Jun 32:1370-1377.

(120) Wolfe CD, Tilling K, Beech R, Rudd AG. Variations in case fatality and dependency from stroke in western and central Europe. The European BIOMED Study of Stroke Care Group. Stroke 1999 Feb 1930;350-356.

(121) Gray LJ, Sprigg N, Bath PM, Sorensen P, Lindenstrom E, Boysen G et al. Significant variation in mortality and functional outcome after acute ischaemic stroke between western countries: data from the 'Tinzaparin in Acute Ischaemic Stroke Trial'(TAIST). J Neurol Neurosurg Psychiatry 2005 Jul 26.

(122) Scott I, Youlden D, Coory M. Are diagnosis specific outcome indicators based on administrative data useful in assessing quality of hospital care? Qual Saf Health Care 2004 Feb 1913;32-39.

(123) Mukamel DB, Zwanziger J, Bamezai A. Hospital competition, resource allocation and quality of care. BMC Health Serv Res 2002 May 27 2002;10.

(124) Stroke Unit Trialists' Collaboration. Organised inpatient (stroke unit) care for stroke. 2001. John Wiley & Sons, Ltd.

(125) Goldacre MJ, Roberts SE, Yeates D. Mortality after admission to hospital with fractured neck of femur: database study. BMJ 2002; 325(7369):868-869.

(126) Todd CJ, Palmer C, Camilleri-Ferrante C, Freeman CJ, Laxton CE, Parker MJ et al. Differences in mortality after fracture of hip. BMJ 1995 Oct 14 311;1025.

(127) Freeman C, Todd C, Camilleri-Ferrante C, Laxton C, Murrell P, Palmer CR et al. Quality improvement for patients with hip fracture: experience from a multi-site audit. Qual Saf Health Care 2002 Sep 1911;239-245.

(128) Foss NB, Kehlet H. Mortality analysis in hip fracture patients: implications for design of future outcome trials. Br J Anaesth 2005 Jan 1994;24-29.

(129) Elliott J, Beringer T, Kee F, Marsh D, Willis C, Stevenson M. Predicting survival after treatment for fracture of the proximal femur and the effect of delays to surgery. J Clin Epidemiol 2003; 56(8):788-795.

(130) Parker MJ, Handoll HH, Griffiths R. Anaesthesia for hip fracture surgery in adults. Cochrane Database Syst Rev 2004 2004;CD000521.

(131) Gonnella JS, Louis DZ, Gozum MV.E, Callahan CA, Barnes CA. Disease staging: clinical criteria. 2003. Thomson, Medstat.

(132) Gonnella JS, Louis DZ, McCord JJ. The staging concept--an approach to the assessment of outcome of ambulatory care. Med Care 1976; 14(1):13-21.

9. APPENDICES

9.1 APPENDIX 1 – LIST OF HOSPITALS, AND IDENTIFICATION OF ALIASES

Table 9-1: Names and aliases of hospitals used in this study, sorted by alias number and hospital name.

Hospital-Full name	Short name	Alias
Akershus universitetssykehus HF	Ahus	Hsp17
Aker universitetssykehus HF	Aker	Hsp62
Sørlandet sykehus Arendal	Arendal	Hsp1
Sykehuset Østfold - Askim	Askim	Hsp42
Sykehuset Buskerud HF	Buskerud	Hsp35
Sykehuset Asker og Bærum HF	Bærum	Hsp16
Diakonhjemmets sykehus	Diakonhjemmet	Hsp61
Sykehuset Innlandet Elverum	Elverum	Hsp44
Feiringklinikken	Feiring	Hsp33
Førde sentralsjukehus - Florø	Florø	Hsp59
Sykehuset Østfold HF	Fredrikstad	Hsp55
Førde sentralsjukehus	Førde	Hsp29
Sykehuset Innlandet Gjøvik	Gjøvik	Hsp24
Sykehuset Østfold - Halden	Halden	Hsp54
Sykehuset Innlandet Hamar	Hamar	Hsp60
Helse Finnmark Hammerfest	Hammerfest	Hsp20
Haraldsplass Diakonale sykehus	Haraldsplass	Hsp49
Hålogalandssykehuset Harstad	Harstad	Hsp8
Haugesund sjukehus	Haugesund	Hsp3
Haukeland universitetssykehus HF	Haukeland	Hsp39
Helse Finnmark	Kirkenes	Hsp46

Alias	Short name
Hsp1	Arendal
Hsp2	Mandal
Hsp3	Haugesund
Hsp4	Tynset
Hsp5	Sarpsborg
Hsp6	UNN
Hsp7	Stokmarknes
Hsp8	Harstad
Hsp9	Volda
Hsp10	Kristiansand
Hsp11	St Olav
Hsp12	Notodden
Hsp13	Kragerø
Hsp14	Stord
Hsp15	Orkdal
Hsp16	Bærum
Hsp17	Ahus
Hsp18	Ringerike
Hsp19	Sandnessjøen
Hsp20	Hammerfest
Hsp21	Kongsberg

Hospital-Full name	Short name	Alias
Kirkenes		
Blefjell sykehus Kongsberg	Kongsberg	Hsp21
Sykehuset Innlandet Kongsvinger	Kongsvinger	Hsp31
Sykehuset Telemark Kragerø	Kragerø	Hsp13
Sørlandet sykehus Kristiansand	Kristiansand	Hsp10
Kristiansund sykehus	Kristiansund	Hsp65
Sykehuset i Vestfold - Larvik	Larvik	Hsp32
Sykehuset Levanger	Levanger	Hsp26
Sykehuset Innlandet Lillehammer	Lillehammer	Hsp50
Sørlandet sykehus Lister	Lister	Hsp48
Nordlandssykehuset Lofoten	Lofoten	Hsp52
Lovisenberg Diakonale sykehus	Lovisenberg	Hsp64
Lærdal sjukehus	Lærdal	Hsp38
Mandal sykehus	Mandal	Hsp2
Molde sjukehus	Molde	Hsp28
Helgelandssykehuset Mosjøen	Mosjøen	Hsp25
Sykehuset Østfold - Moss	Moss	Hsp37
Sykehuset Namsos	Namsos	Hsp63
Hålogalandssykehuset Narvik	Narvik	Hsp40
Nordfjord sjukehus	Nordfjord	Hsp30
Nordlandssykehuset Bodø	Nordland	Hsp53
Blefjell sykehus Notodden	Notodden	Hsp12
Odda sjukehus	Odda	Hsp58
Orkdal Sanitetsforenings sjukehus	Orkdal	Hsp15
Helgelandssykehuset Mo i Rana	Rana	Hsp66
Rikshospitalet HF	Rikshospitalet	Hsp41
Ringerike sykehus HF	Ringerike	Hsp18
Blefjell sykehus Rjukan	Rjukan	Hsp23
Sentralsjukehuset i Rogaland	Rogaland	Hsp57
	Røros	Hsp27
Helgelandssykehuset Sandnessjøen	Sandnessjøen	Hsp19

Alias	Short name
Hsp22	Ski
Hsp23	Rjukan
Hsp24	Gjøvik
Hsp25	Mosjøen
Hsp26	Levanger
Hsp27	Røros
Hsp28	Molde
Hsp29	Førde
Hsp30	Nordfjord
Hsp31	Kongsvinger
Hsp32	Larvik
Hsp33	Feiring
Hsp34	Telemark
Hsp35	Buskerud
Hsp36	Voss
Hsp37	Moss
Hsp38	Lærdal
Hsp39	Haukeland
Hsp40	Narvik
Hsp41	Rikshospitalet
Hsp42	Askim
Hsp43	Ullevål
Hsp44	Elverum
Hsp45	St Elisabeth
Hsp46	Kirkenes
Hsp47	Vestfold/Tbg
Hsp48	Lister
Hsp49	Haraldsplass
Hsp50	Lillehammer
Hsp51	Stensby

Hospital-Full name	Short name	Alias
Sykehuset Østfold - Sarpsborg	Sarpsborg	Hsp5
Aker Universitetssykehus - Ski sykehus	Ski	Hsp22
St. Elisabeths Hospital	St Elisabeth	Hsp45
St. Olavs Hospital	St Olav	Hsp11
Stensby sykehus	Stensby	Hsp51
Hålogalandssykehuset Stokmarknes	Stokmarknes	Hsp7
Stord sjukehus	Stord	Hsp14
Sykehuset Telemark	Telemark	Hsp34
Sykehuset Innlandet Tynset	Tynset	Hsp4
Ullevål universitetssykehus HF	Ullevål	Hsp43
Universitetssykehuset i Nord-Norge HF	UNN	Hsp6
Sykehuset i Vestfold - Tønsberg	Vestfold/Tbg	Hsp47
Volda sjukehus	Volda	Hsp9
Voss sjukehus	Voss	Hsp36
Ålesund sjukehus	Ålesund	Hsp56

Alias	Short name
Hsp52	Lofoten
Hsp53	Nordland
Hsp54	Halden
Hsp55	Fredrikstad
Hsp56	Ålesund
Hsp57	Rogaland
Hsp58	Odda
Hsp59	Florø
Hsp60	Hamar
Hsp61	Diakonhjemmet
Hsp62	Aker
Hsp63	Namsos
Hsp64	Lovisenberg
Hsp65	Kristiansund
Hsp66	Rana

9.2 APPENDIX 2 –THE CLINICAL CRITERIA DISEASE STAGING SYSTEM FOR ACUTE MYOCARDIAL INFARCTION, STROKE AND HIP FRACTURE

The system, most recently described by Gonnella et al. (131;132) is a classification system that uses diagnostic findings to produce comparable clusters of patients. Its core idea is that the three fundamental questions that must be answered before cases can be clustered into groups that can be meaningfully compared, are "Where?", "Why?" and "How serious?" Diagnostic labels do not regularly provide that information, and the Clinical Criteria Disease Staging system is designed to rectify that problem.

Their classification depicts the severity of the pathophysiological manifestations of the disease. Stage 1 is diseases with no complications, Stage 2 is the diseases with local complications, Stage 3 is diseases involving multiples sites or with systemic complications, Stage 4 is death.

Most diseases begin at Stage 1 and may develop through all later stages. There are, however, exceptions. Some diseases are self-limiting and do not include a stage 3 or 4. Some diseases begin at Stage 2 or 3. Examples are myocardial infarction, which may be viewed as a later stage of angina, or meningitis, which can be a complication of sinusitis, otitis media or bacterial pneumonia. And some diseases may exist before they can be discovered at Stage 1, like cancer in patients with a family history of carcinoma or a baby born to a mother having an infection at the time of delivery. Although no pathology is present, important risk factors may exist. Also, "the same conditions" are not always classified as the same stage: a patient hospitalized for "pneumonia" is normally a Stage 1 case. But if pneumonia occurred secondary to other problems, it will be classified as Stage 2. And in some cases it may score a 3: e.g. when it reflects the systemic nature of a problem (like botulism), and not just the involvement of the respiratory system.

It should be understood that stages are ordinal: Stage 2 is not twice as severe as Stage 1. Also, stages are not equivalent across diseases: Stage 1 for disease A may be more (or less) severe than Stage 1 for disease B – hyperglycemia (Stage 1 for diabetes mellitus) is not equivalent to HIV positivity (Stage 1 for AIDS). And even if they were equally severe on some common scale, the comparison would be affected by the fact that some conditions may be reversible (e.g. strokes or pulmonary embolisms) while others are temporary stations in a degenerative process (e.g. multiple sclerosis).

This raises the very important question of against which criteria staging scales are calibrated. As far as we have been able to ascertain through personal communication with the authors of the Clinical Criteria Disease Staging, the staging is driven by the natural history of the disease in question – in which the risk of in-hospital death is an important consideration. Other bases for staging might be construed, which may produce different classifications of cases by severity (e.g. mortality at 30-day after hospitalization, or one year (or other points in time). Also, treatment (e.g. medical or surgical) is not used for staging, nor has the patient's level of functioning or quality of life been taken into consideration.

The clusters produced by the Clinical Criteria Disease Staging system can be used for several purposes of clinical research, like for (as in our case) inter-hospital comparison of patient outcomes. This is probably its most important trait: it tries to ensure fair comparisons across service providers while accounting for risk adjustment differences involving disease severity.

Table 9-2: Stage for index admission for hip fracture.

Stage	ICD-10	ICD-9
Stage1.1	S72.0-S72.2	820.0, 820.2, 820.8 820.00-820.09, 820.20-820.21, 820.80)
Stage 2.2	S72.0-S72.2 and Stage 1.1	820.1, 820.3, 820.9 (820.10-820.19, 820.30- 820.31, 820.90) and Stage 1.1
Stage 2.3	M87.0, M87.8-M87.9 and Stage1.1-Stage2.2	733.4 (733.42)and Stage1.1-Stage2.2
Stage 2.4	I80.1- I80.2, I82,8 I86.8 Stage1.1-Stage2.2	453.8 (453.80) Stage1.1-Stage2.2
Stage 3.1	I26.9, T79.1 and Stage1.1-Stage2.2	415.1, 958.1 (415.10, 415.19, 958.10) and Stage1.1-Stage2.2
Stage 3.2	J80, J95.1-J95.3 and Stage1.1-Stage2.2	518.5 (518.50) and Stage1.1-Stage2.2
Stage 3.3	T79.4, R57.0-R57.1, R57.8-, R57.9 and Stage1.1-Stage 2.2	958.4, 785.5 (958.40, 785.50-785.59) and Stage1.1-Stage2.2
Stage 4	3.1-3.3 and dead on arrival	3.1-3.3 and dead on arrival

Table 9-3: Staging for index admission for stroke.

Stage	ICD-10	ICD-9
Stage3.1	I60.0-I60.9, I61.0-I61.6, I61.8, I61.9, I62.0, I62.1, I62.9, I63.0-I63.6, I63.8, I63.9, I67.5, I67.8, I69.0, I69.4, I69.8	430, 431, 432.0, 432.1, 432.9, 433.0-433.3, 433.9, 434.0, 434.1, 434.9, 436, 437.5, 437.6, 438, 437.1
Stage 3.2	H47.0, H49.0-H49.3, G51.0, G51.1, G51.9, G52.0-G52.3, G52.7-G52.9, R43.0, R43.2, R43.8, H93.3, H90.5, H90.6, H90.8, R13, H54.0-H54.7, I69.3, I69.8 and Stage3.1	377.4, 378.5, 350.8-353.6, 352.8,781.1, 388.5, 389.1, 389.8, 389.9, 787.2, 369.0-369.3, 369.6-369.9, 438.8 and Stage3.1
Stage 3.3	G11.9, H55, G24.9, R25.1, R25.8, G26, R27.0, R27.8, I69.3, G81.0, G81.1, G81.9, I69.3, I63.8 and Stage 3.1	334.3, 379.5, 781.0, 781.3, 438 and Stage 3.1
Stage 3.4	G81.0, G81.1, G81.9, I69.3, I63.8 and Stage 3.1-3.3	332.0, 332.1, 342.9, 438 and Stage 3.1-3.3
Stage 3.5	G82.0-G82.5, G83.0-G83.4, G83.8, G83.9, I69.3, I63.8 and Stage 3.1-3.4	344.0-344.6, 344.8, 344.9 and Stage 3.1-3.4
Stage 3.6	R40.2 and Stage 3.1-3.5	780.0 and Stage 3.1-3.5

Table 9-4: Staging for index admission for acute myocardial infarction.

Stage	ICD-10	ICD-9
Stage 3.1	I21.0-I21.9	410
Stage 3.2	I44.0-I44.7, I45.0, I45.1, I45.4-I45.6, I45.8, I45.9, I46.0, I46.9, I47.1-I47.2, I47.9, I48, I49.1- I49.5, I49.4, I49.8, I49.9, I30.0, I30.1, I30.8, I30.9, I20.0, I20.1, I20.8, I20.9 and Stage 3.1	426, 427.0-427.6, 427.8, 427.9, 420.9, 411.0
Stage 3.3	I50.0, I50.1, I50.9 , J81and Stage 3.1-3.2	428.0, 428.1, 428.9 and Stage 3.1-3.2
Stage 3.4	I23.6 (429.7 does not exist in ICD-10)	429.7 and Stage 3.1-3.3
Stage 3.5	I25.3 and Stage 3.1-3.4	414.1 and Stage 3.1-3.4
Stage 3.6	I26.9, I63.3, I63.5, I63.8, I63.9, I64, I66.0-I66.4, I66.8-I66.9, G46.0A-G46.8A, and Stage 3.1-3.5	415.1, 434.0, 434.1, 434.9, 436.0 and Stage 3.1-3.5
Stage 3.7	I23.2, I23.4, I23.5 and Stage 3.1-3.6	429.6 and Stage 3.1-3.6>
Stage 3.8	I47.2, I49.0, R57.0 and Stage 3.1-3.7	427.1, 427.4, 785.5 and Stage 3.1-3.7
Stage 3.9	I46.0, I46.9 and Stage 3.1-3.8	427.5 and Stage 3.1-3.8

9.3 APPENDIX 3 – THE FS-SYSTEM

The FS-system is a modular system of MS Access databases that was designed and developed by Tomislav Dimoski in close cooperation with the developers of Patient Administrative Systems (PAS) or IT specialists within the hospitals.

9.3.1 At the hospital

The administrative and medical patient data used by this study are extracted from the hospitals' PAS (See Figure 9-1). For each PAS a program that extracts data in accordance with a standard specification was developed within the original technological platform. This specification has been developing incrementally in the period 1995-2003 in order to satisfy specific needs of new research projects.

The module of the FS-system that is installed at the hospital imports the data extracted from PAS and LAB-systems and constructs a dataset that describes a hospital stay at ward, department and hospital level. The hospital module of the FS-system generates encrypted personal identifier and encrypted stay identifier. The FS-system programmatically selects the index admission upon criteria specified in the project.

The data from the FS-system extracted from the PAS were quality controlled at almost all Norwegian hospitals during the period 1995-2003. A random sample of data collected by the FS-system was compared on site with the data presented in the original schema from the PAS. Aggregated data from the FS-system were compared against data presented by standard reports developed within the PAS. All known system discrepancies that were identified during the period 1995-2003 were analyzed and corrected.

For this study data were also extracted from the hospitals' Clinical chemical laboratory system (LAB-system). For each LAB-system a program that extracts data within the original technological platform was developed. The LAB programs are not quality controlled yet, and the data imported from the LAB-system are not used in the analyses in this study so far.

9.3.2 Combination of data at Statistics Norway

The FS-system at the hospital exports a limited set of data from the hospital to Statistics Norway (SSB). SSB generates an encrypted national personal identifier for each patient. SSB exports this encrypted national personal identifier and a limited set of data from several registries to the module of the FS-system at the Norwegian Knowledge Centre for the Health Services.

SSB never receives medical information about the patient stay.

9.3.3 Treatment of data at the Norwegian Knowledge Centre for the Health Services

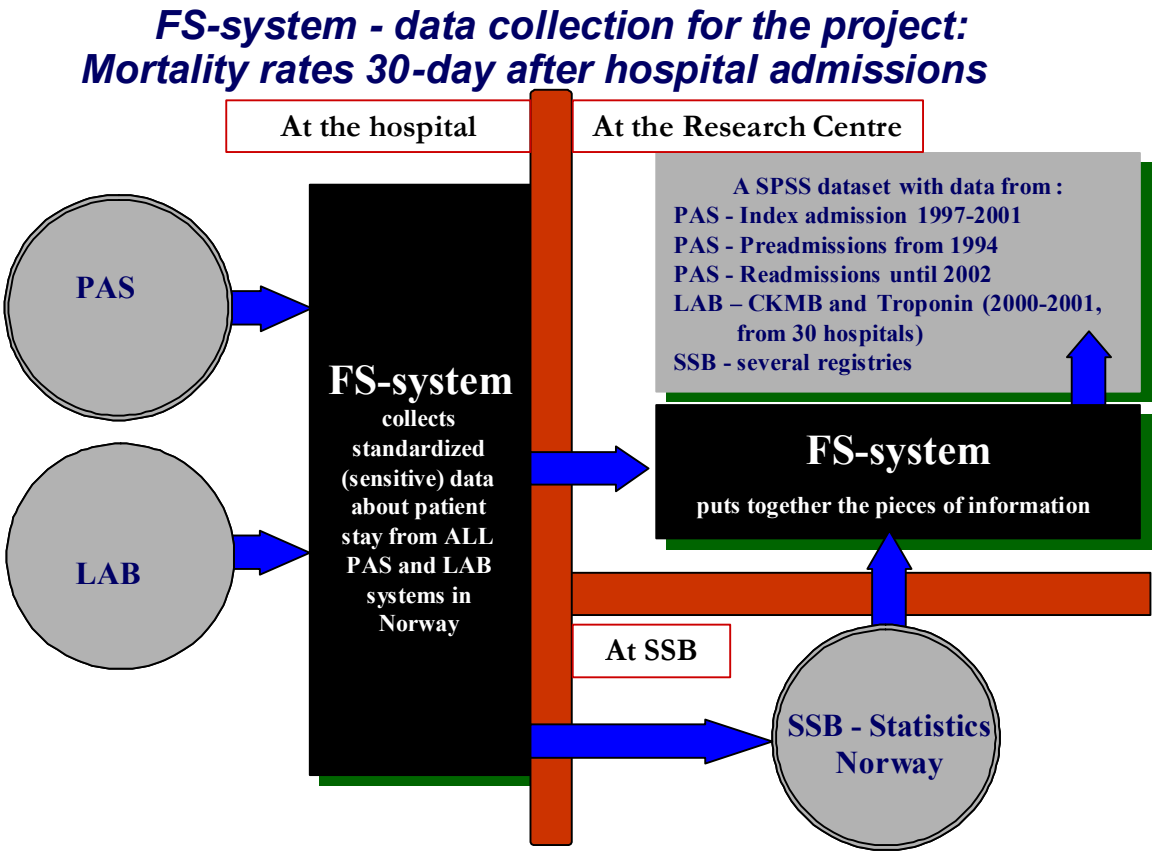
The module of the FS-system at the Norwegian Knowledge Centre for the Health Services reconstructs the pieces of information coming from PAS, LAB and SSB registries into a nationally uniquely identifiable set of data that describes a patient index admission within one or several hospitals. If the patient was transferred from one

hospital to another, the FS-system identifies the transfer and connects the data into an aggregated index admission dataset.

Data about all preadmissions back to 1994 and all readmissions until 2002 of the index patients within the index hospital have been collected.

The FS-system prepares the data in a standard format in order to be imported by the statistical program SPSS.

Figure 9-1: The fundamental principles of the FS-system.



9.4 APPENDIX 4 - INDEX CASES BY DIAGNOSTIC CODE

The table below shows how the three categories of index admission split into diagnosis codes (ICD-9 and ICD-10). Note that an index admission for e.g. AMI may also have a stroke code. The table also shows these cases.

Table 9-5: Index admissions by diagnostic code.

		Index admissions		
		Myocardial infarction	**Stroke**	**Hip fracture**
Codes for AMI at index admission			52341	49633
	410	21763	287	214
	I210	6255	57	31
	I211	5324	28	24
	I212	837	5	6
	I213	1958	17	22
	I214	7948	74	50
	I219	10010	263	225
Code for stroke at index admission		52318		49070
	431	45	2420	41
	4310		14	
	434	21	991	22
	4340	20	1399	24
	4341	42	951	6
	4349	153	6595	87
	436	393	8732	261
	4360		2	
	43690		1	
	I61			2
	I610	22	974	9
	I611	8	425	3
	I612	12	555	4
	I613	3	150	4
	I614	11	297	7
	I615	7	261	1
	I616	7	222	7
	I618	5	333	4
	I619	31	1192	30
	I63	13	7	9
	I630	26	499	14
	I631	21	290	3
	I632	3	323	7
	I633	59	3317	39

		Index admissions		
		Myocardial infarction	**Stroke**	**Hip fracture**
	I634	153	2419	40
	I635	48	2388	31
	I636	3	100	
	I638	44	1317	16
	I639	514	14663	302
	I64	113	2235	162
Code for hip fracture at index admission		53568	52805	
	820	46	30	5193
	8200	67	38	7714
	8201	3	3	185
	8202	65	42	5690
	8203	2		99
	8208	3	4	213
	8209	2	1	59
	S720	182	102	17831
	S7200	9	6	1026
	S7201			8
	S721	119	34	9349
	S7210	3	4	667
	S7211			3
	S722	26	3	2046
	S7220			118
	S7221			4

9.5 APPENDIX 5 – COVARIATES BY INDIVIDUAL HOSPITAL

Table 9-6: Patient volume variables – AMI.

Hospital	Total number of admissions 1997-2001*	Percent of all admissions are first time ami**	Total number of emergency admissions 1994-2001*	Percent of all emergency admissions are first time ami**	Percent of ami patients moved from another hospital	Percent of ami patients moved to another hospital
Halden	9094	1.7	1744	8.8	31.8	6.5
Sarpsborg	Data not available***	Data not available	Data not available	Data not available	13.8	5.4
Fredrikstad	153134	1.4	99003	2.1	1.7	10.3
Moss	40566	2.9	34391	3.4	1.9	7.8
Askim	10970	2.9	4932	6.4	4.8	4.1
Ski	10481	.7	1532	4.5	43.5	5.8
Feiring	17081	1.9	1854	17.6	68.2	6.4
Stensby	15930	2.9	12446	3.7	3.0	9.3
Ahus	150174	1.4	113247	1.9	2.7	8.0
Aker	133339	1.2	82114	1.9	5.0	4.1
Ullevål	232425	1.2	121203	2.3	11.6	6.9
Lovisenberg	53095	1.2	17327	3.7	6.1	6.6
Diakonhjemmet	36283	1.5	24982	2.2	7.6	4.2
Rikshospitalet	206428	.4	40737	2.2	57.6	14.0
Bærum	67376	1.6	50783	2.1	2.4	3.1
Kongsvinger	30797	2.0	20733	3.0	2.2	2.4
Elverum	88206	.7	55848	1.2	2.8	2.6
Tynset	11933	2.5	8303	3.6	1.0	7.3
Hamar	19956	5.1	14370	7.0	1.7	5.0
Lillehammer	89408	1.4	52356	2.5	1.0	9.3
Gjøvik	65487	2.0	43502	3.0	1.5	4.3
Ringerike	38786	1.8	27386	2.6	1.0	5.7
Buskerud	121245	1.2	74908	2.0	2.6	5.8
Kongsberg	27616	1.9	18049	2.9	.8	3.5
Notodden	15857	2.7	12072	3.6	1.8	7.1
Rjukan	10584	2.2	5626	4.1	.9	1.7
Vestfold/Tbg	112124	1.5	58263	2.9	2.8	3.4
Larvik	22029	2.6	14445	4.0	2.6	4.2
Telemark	112487	1.4	64938	2.3	2.0	4.3
Kragerø	6448	2.4	2984	5.2	13.6	5.8
Arendal	80142	1.4	51405	2.2	.8	2.6
Kristiansand	115122	1.4	73350	2.3	1.2	3.7
Mandal	5694	.2	2013	.5	45.5	18.2
Lister	19900	1.8	11129	3.3	.8	1.4
Rogaland	196330	1.4	133224	2.1	.7	1.7
Haugesund	79417	1.3	52562	2.0	.7	4.5
Stord	31289	1.4	23879	1.9	.2	2.9
Odda	12763	1.8	8740	2.6	.0	3.0
Haukeland	286672	.9	168242	1.6	10.3	1.0
Haraldsplass	42414	2.9	34525	3.6	.5	8.9
Voss	19803	1.7	13226	2.6	1.5	9.3
Lærdal	14219	1.9	10868	2.5	1.1	5.5
Førde	65340	.7	36212	1.3	2.1	6.0

Hospital	Total number of admissions 1997-2001*	Percent of all admissions are first time ami**	Total number of emergency admissions 1994-2001*	Percent of all emergency admissions are first time ami**	Percent of ami patients moved from another hospital	Percent of ami patients moved to another hospital
Florø	Data not available	Data not available	Data not available	Data not available	1.9	11.0
Nordfjord	15314	2.1	11935	2.7	.6	7.2
Volda	24233	1.9	16703	2.7	.7	5.6
Ålesund	96017	1.2	50488	2.2	.8	4.5
Molde	50628	1.5	29147	2.7	.8	1.4
Kristiansund	39323	1.2	24174	2.0	.6	3.1
Orkdal	32167	1.8	26586	2.2	2.5	6.3
St Olav	217802	1.3	121682	2.3	6.3	1.4
St Elisabeth	6084	.7	1093	3.8	19.5	14.6
Røros	5882	.1	812	.9	.0	14.3
Levanger	65022	1.4	44049	2.1	.9	3.6
Namsos	39080	1.6	23841	2.5	.8	5.4
Mosjøen	13018	1.7	9738	2.2	1.4	2.8
Sandnessjøen	21024	1.8	13732	2.7	1.1	7.3
Rana	27242	1.4	17309	2.2	.3	5.5
Nordland	90733	1.2	42372	2.5	1.3	5.4
Lofoten	16891	2.1	12775	2.7	.3	5.5
Stokmarknes	25056	1.5	18515	2.0	.8	4.1
Narvik	24157	1.8	17313	2.5	.2	6.3
Harstad	39564	1.7	23669	2.9	.4	5.1
UNN	149533	.9	57794	2.5	14.1	1.1
Hammerfest	31697	1.7	22863	2.4	.7	7.6
Kirkenes	19992	1.8	13474	2.7	.6	5.8

* Total number of admissions and emergency admissions are the numbers of all admissions to each hospital. Data from the central Norwegian Patient Register (NPR).

** Percentage of all (emergency) admissions are AMI, is the proportion of index admissions in this study as part of the total number of admissions to each hospital in the period.

*** Data not available refers to hospitals which are considered departments of another hospital. The data we received from NPR did in a few cases not distinguish between these departments and their “mother hospital”. This mean that data for Sarpsborg is included in the information given on Fredrikstad, and Florø is included in Førde.

Table 9-7: Patient volume variables – stroke.

Hospital	Total number of admissions 1997-2001	Percent of all admissions are stroke	Total number of emergency admissions 1994-2001	Percent of all emergency admissions are stroke	Percent of stroke patients moved from another hospital	Percent of stroke patients moved to another hospital
Halden	9094	1.6	1744	8.5	11.4	4.7
Sarpsborg	Data not available	Data not available	Data not available	Data not available	18.1	1.2
Fredrikstad	153134	1.2	99003	1.9	.4	31.1
Moss	40566	2.4	34391	2.8	.2	13.8
Askim	10970	2.6	4932	5.8	4.9	1.8
Ski	10481	.3	1532	2.3	44.4	8.3
Feiring	17081	.0	1854	.0	.	.
Stensby	15930	1.1	12446	1.4	4.1	5.9
Ahus	150174	1.8	113247	2.3	.5	8.8
Aker	133339	1.3	82114	2.1	.4	1.8
Ullevål	232425	1.2	121203	2.3	.5	2.5
Lovisenberg	53095	1.3	17327	4.1	.8	2.1
Diakonhjemmet	36283	2.5	24982	3.6	.1	.8
Rikshospitalet	206428	.2	40737	.9	12.9	33.7
Bærum	67376	1.9	50783	2.5	.4	2.7
Kongsvinger	30797	2.5	20733	3.8	.3	2.8
Elverum	88206	.9	55848	1.4	.4	3.4
Tynset	11933	2.4	8303	3.4	.0	3.8
Hamar	19956	4.8	14370	6.6	.2	3.2
Lillehammer	89408	1.5	52356	2.5	.2	4.9
Gjøvik	65487	1.9	43502	2.9	1.2	2.0
Ringerike	38786	2.2	27386	3.1	.2	3.1
Buskerud	121245	1.3	74908	2.1	.8	2.8
Kongsberg	27616	1.9	18049	2.9	.0	1.5
Notodden	15857	2.5	12072	3.3	.0	3.0
Rjukan	10584	2.0	5626	3.8	.0	3.7
Vestfold/Tbg	112124	1.6	58263	3.0	.5	2.6
Larvik	22029	1.8	14445	2.7	2.0	9.3
Telemark	112487	1.1	64938	1.9	.6	3.7
Kragerø	6448	3.8	2984	8.1	1.2	2.9
Arendal	80142	1.4	51405	2.2	.1	1.9
Kristiansand	115122	1.2	73350	1.9	.1	3.4
Mandal	5694	.4	2013	1.2	20.0	20.0
Lister	19900	1.7	11129	3.1	.0	1.7
Rogaland	196330	1.1	133224	1.6	.2	2.1
Haugesund	79417	1.3	52562	2.0	.3	2.7
Stord	31289	1.3	23879	1.7	.8	3.0
Odda	12763	1.3	8740	1.8	.0	5.0
Haukeland	286672	1.0	168242	1.6	1.0	2.1
Haraldsplass	42414	2.6	34525	3.2	.2	1.8
Voss	19803	1.5	13226	2.2	.7	6.4
Lærdal	14219	1.9	10868	2.5	1.1	2.6
Førde	65340	1.0	36212	1.7	1.1	4.6
Florø	Data not available	Data not available	Data not available	Data not available	.8	4.7
Nordfjord	15314	2.3	11935	3.0	.0	2.8

Hospital	Total number of admissions 1997-2001	Percent of all admissions are stroke	Total number of emergency admissions 1994-2001	Percent of all emergency admissions are stroke	Percent of stroke patients moved from another hospital	Percent of stroke patients moved to another hospital
Volda	24233	2.3	16703	3.4	.2	1.2
Ålesund	96017	1.1	50488	2.1	.3	2.4
Molde	50628	1.9	29147	3.2	.6	2.3
Kristiansund	39323	1.5	24174	2.4	.2	2.7
Orkdal	32167	1.9	26586	2.3	.2	2.5
St Olav	217802	1.3	121682	2.4	1.0	2.0
St Elisabeth	6084	Data not available	1093	Data not available	.	.
Røros	5882	.2	812	1.1	.0	11.1
Levanger	65022	1.8	44049	2.6	.2	3.4
Namsos	39080	1.8	23841	2.9	.3	1.3
Mosjøen	13018	2.0	9738	2.7	.4	3.9
Sandnessjøen	21024	1.6	13732	2.4	.6	4.6
Rana	27242	1.5	17309	2.4	.7	3.3
Nordland	90733	1.1	42372	2.3	1.6	5.1
Lofoten	16891	1.7	12775	2.3	1.0	4.2
Stokmarknes	25056	2.1	18515	2.8	.6	3.4
Narvik	24157	1.4	17313	2.0	.0	4.1
Harstad	39564	1.9	23669	3.1	.4	3.6
UNN	149533	.7	57794	1.9	2.2	3.8
Hammerfest	31697	1.4	22863	1.9	.5	6.7
Kirkenes	19992	1.6	13474	2.4	.0	3.4

* Total number of admissions and emergency admissions are the numbers of all admissions to each hospital. Data from the central Norwegian Patient Register (NPR).

** Percentage of all (emergency) admissions are stroke, is the proportion of index admissions in this study as part of the total number of admissions to each hospital in the period.

*** Data not available refers to hospitals which are considered departments of another hospital. The data we received from NPR did in a few cases not distinguish between these departments and their "mother hospital". This mean that data for Sarpsborg is included in the information given on Fredrikstad, and Florø is included in Førde.

Table 9-8: Patient volume variables - hip fracture.

Hospital	Total number of admissions 1994-2001	Percent of all admissions are hip fracture	Total number of emergency admissions 1994-2001	Percent of all emergency admissions are hip fracture	Percent of hip fracture patients moved from another hospital	Percent of hip fracture patients moved to another hospital
Halden	9094	8.9	1744	46.4	97.3	1.2
Sarpsborg	Data not available	Data not available	Data not available	Data not available	86.8	13.2
Fredrikstad	153134	1.2	99003	1.9	1.8	47.1
Moss	40566	2.9	34391	3.4	4.0	16.4
Askim	10970	2.0	4932	4.5	93.7	1.8
Ski	10481	2.6	1532	18.1	79.4	11.2
Feiring	17081	.0	1854	.1	.0	100.0
Stensby	15930	2.9	12446	3.7	17.8	5.7
Ahus	150174	1.2	113247	1.6	4.8	16.0
Aker	133339	2.0	82114	3.2	3.5	8.8
Ullevål	232425	1.3	121203	2.4	3.0	4.6
Lovisenberg	53095	.6	17327	1.7	73.7	2.7
Diakonhjemmet	36283	3.5	24982	5.1	8.2	1.3
Rikshospitalet	206428	.0	40737	.2	16.0	23.5
Bærum	67376	1.8	50783	2.4	5.2	.9
Kongsvinger	30797	1.7	20733	2.5	2.3	2.7
Elverum	88206	1.4	55848	2.2	3.7	3.1
Tynset	11933	2.0	8303	2.9	4.6	5.9
Hamar	19956	2.2	14370	3.1	4.1	5.7
Lillehammer	89408	1.2	52356	2.1	1.5	4.3
Gjøvik	65487	2.1	43502	3.1	.8	3.5
Ringerike	38786	2.2	27386	3.1	.8	2.6
Buskerud	121245	1.2	74908	2.0	3.0	1.6
Kongsberg	27616	2.0	18049	3.0	2.4	4.4
Notodden	15857	2.5	12072	3.3	4.2	5.4
Rjukan	10584	1.9	5626	3.7	1.5	2.4
Vestfold/Tbg	112124	1.5	58263	2.8	11.9	1.6
Larvik	22029	3.1	14445	4.7	4.7	24.1
Telemark	112487	1.3	64938	2.3	3.0	12.9
Kragerø	6448	2.9	2984	6.2	84.9	5.9
Arendal	80142	1.4	51405	2.1	.9	2.6
Kristiansand	115122	1.0	73350	1.6	1.6	4.5
Mandal	5694	3.0	2013	8.6	8.7	5.8
Lister	19900	1.9	11129	3.4	7.4	1.9
Rogaland	196330	1.1	133224	1.6	1.0	.4
Haugesund	79417	1.2	52562	1.8	.9	.6
Stord	31289	1.1	23879	1.4	1.8	3.0
Odda	12763	1.5	8740	2.1	1.1	1.6
Haukeland	286672	.8	168242	1.4	5.0	1.4
Haraldsplass	42414	2.5	34525	3.1	2.0	8.0
Voss	19803	2.4	13226	3.6	1.5	1.9
Lærdal	14219	1.8	10868	2.3	4.0	3.2
Førde	65340	1.0	36212	1.8	5.8	4.3
Florø	Data not available	Data not available	Data not available	Data not available	44.4	37.0
Nordfjord	15314	2.1	11935	2.7	.9	3.4
Volda	24233	1.6	16703	2.3	.8	.8

Hospital	Total number of admissions 1994-2001	Percent of all admissions are hip fracture	Total number of emergency admissions 1994-2001	Percent of all emergency admissions are hip fracture	Percent of hip fracture patients moved from another hospital	Percent of hip fracture patients moved to another hospital
Ålesund	96017	1.0	50488	1.9	1.8	1.1
Molde	50628	1.4	29147	2.4	1.6	1.2
Kristiansund	39323	1.2	24174	2.0	1.4	1.0
Orkdal	32167	1.1	26586	1.4	3.0	9.6
St Olav	217802	1.1	121682	2.0	2.4	1.1
St Elisabeth	6084	0	1093	0	.	.
Røros	5882	1.7	812	12.1	1.0	3.1
Levanger	65022	1.4	44049	2.0	1.2	.8
Namsos	39080	1.2	23841	2.0	1.1	1.1
Mosjøen	13018	1.4	9738	1.9	4.9	4.4
Sandnessjøen	21024	1.2	13732	1.9	3.8	4.6
Rana	27242	1.2	17309	2.0	4.7	2.7
Nordland	90733	.9	42372	1.8	1.7	2.1
Lofoten	16891	1.3	12775	1.7	2.7	1.4
Stokmarknes	25056	1.3	18515	1.8	3.9	1.8
Narvik	24157	1.3	17313	1.8	1.3	.9
Harstad	39564	1.4	23669	2.4	1.8	1.4
UNN	149533	.6	57794	1.5	3.2	2.8
Hammerfest	31697	1.0	22863	1.3	11.5	6.6
Kirkenes	19992	1.2	13474	1.7	4.7	13.3

* Total number of admissions and emergency admissions are the numbers of all admissions to each hospital. Data from the central Norwegian Patient Register (NPR).

** Percentage of all (emergency) admissions are hip fracture, is the proportion of index admissions in this study as part of the total number of admissions to each hospital in the period.

*** Data not available refers to hospitals which are considered departments of another hospital. The data we received from NPR did in a few cases not distinguish between these departments and their "mother hospital". This mean that data for Sarpsborg is included in the information given on Fredrikstad and Florø is included in Førde.

Table 9-9: Socio-economic and demographic data for AMI patients.

Hospital	Patients admitted	Mean age	% aged 80 or older	% men	% born outside of Norway	Mean distance from home	Mean years of highest education/	Mean gross income (NOK)	Mean gross worth (NOK)	% married/ \cohabitant	% not married	% divorced	%t widowed
Halden	154	76	42	57	6	13	10	123351	243525	46	17	6	31
Sarpsborg	167	73	34	57	2	13	11	143181	307838	46	10	11	34
Fredrikstad	2077	70	27	63	3	19	11	164003	334996	57	6	9	27
Moss	1165	70	28	60	4	25	11	171255	359337	55	5	10	29
Askim	314	71	32	60	3	21	11	151736	371895	55	6	7	32
Ski	69	74	38	49	1	11	11	164199	317494	48	9	4	39
Feiring	327	61	4	72	3	193	12	259777	449805	68	6	15	11
Stensby	464	72	30	60	2	26	10	161035	342033	55	6	7	31
Ahus	2163	67	20	68	4	30	11	205077	394418	62	7	10	21
Aker	1565	71	29	56	7	8	11	177061	322336	47	8	14	30
Ullevål	2826	69	26	62	10	27	12	207748	406558	51	9	15	25
Lovisenberg	640	71	37	51	14	8	11	158976	262828	33	14	18	35
Diakonhjemmet	543	73	40	55	8	6	14	282708	812930	48	11	13	27
Rikshospitalet	912	60	5	73	6	151	12	251921	451118	65	8	14	12
Bærum	1082	70	26	62	7	18	13	263220	635754	57	7	12	24
Kongsvinger	630	70	24	64	2	31	10	154184	341746	53	10	10	27
Elverum	649	69	24	66	2	51	10	154430	342510	58	9	10	23
Tynset	301	72	35	63	1	69	11	145487	319980	57	10	5	28
Hamar	1010	70	25	61	1	18	11	162935	361009	57	7	8	28
Lillehammer	1285	71	27	63	2	72	11	153479	501668	55	12	7	25
Gjøvik	1293	70	27	62	1	45	10	149802	331806	55	8	10	27
Ringerike	702	71	29	64	2	55	10	167190	392010	57	9	8	26
Buskerud	1476	69	25	66	6	18	11	191464	374313	56	6	13	25
Kongsberg	518	70	29	64	2	31	11	172731	378181	57	8	9	26
Notodden	435	72	32	64	1	31	11	159972	337684	57	11	7	25
Rjukan	231	70	28	63	5	48	11	161759	279496	57	10	6	27
Vestfold/Tbg	1666	70	27	62	4	29	11	183992	404383	57	6	12	25
Larvik	572	71	28	66	3	18	11	186257	520472	60	5	8	26
Telemark	1521	70	27	63	3	27	11	170937	315626	57	7	10	26
Kragerø	154	72	33	58	1	14	11	149510	313386	47	12	10	31
Arendal	1116	71	30	60	3	33	11	160383	346523	55	9	8	28
Kristiansand	1656	70	26	63	5	36	11	178718	433156	56	7	8	27
Mandal	11	82	73	45	0	3	11	115209	453673	45	9	0	45
Lister	366	72	34	58	2	36	11	172478	436392	55	10	7	29
Rogaland	2787	70	29	62	3	31	11	196551	435211	58	8	8	25
Haugesund	1066	70	25	64	3	32	11	175307	410800	58	8	7	27
Stord	446	72	30	61	1	34	11	160983	325931	55	9	4	32
Odda	230	72	33	60	2	24	11	155815	324632	54	10	7	28
Haukeland	2696	69	26	65	3	41	11	198348	402586	60	8	9	22
Haraldsplass	1230	72	32	57	2	30	11	165553	300566	52	9	10	29
Voss	344	73	34	57	2	42	11	152209	330999	55	13	5	27
Lærdal	275	74	39	54	1	58	11	154331	380384	52	10	4	34
Førde	481	73	38	62	1	63	11	154356	353871	52	14	5	30
Florø	154	70	26	65	3	27	11	174260	351475	55	8	9	27
Nordfjord	318	72	31	60	3	51	11	176356	577284	56	11	5	29

Hospital	Patients admitted	Mean age	% aged 80 or older	% men	% born outside of Norway	Mean distance from home	Mean years of highest education/	Mean gross income (NOK)	Mean gross worth (NOK)	% married/ \cohabitant	% not married	% divorced	%t widowed
Volda	450	73	36	62	2	44	11	153568	340808	58	13	3	27
Ålesund	1107	72	34	63	1	30	11	161493	383341	56	9	6	28
Molde	774	72	31	63	2	43	11	160128	366811	59	8	6	27
Kristiansund	487	73	33	61	1	32	10	136529	278636	51	11	7	31
Orkdal	592	72	33	57	1	73	10	150449	326593	53	11	5	31
St Olav	2782	70	26	62	2	49	11	170192	332465	54	10	9	27
St Elisabeth	41	65	0	73	3	137	11	184080	316048	66	2	10	20
Røros	7	76	43	43	0	120	9	101243	148386	71	0	0	29
Levanger	941	70	27	62	1	45	11	152122	302409	57	7	7	29
Namsos	606	72	33	63	1	63	10	134540	266600	57	9	6	28
Mosjøen	218	71	28	61	0	26	10	156271	346762	61	6	7	26
Sandnessjøen	368	71	30	61	1	71	10	138195	263411	54	9	8	28
Rana	381	71	29	62	1	27	10	148099	300003	53	9	9	30
Nordland	1065	70	28	59	2	90	10	159414	314894	55	8	8	30
Lofoten	347	71	29	64	1	69	10	139773	285852	54	8	7	31
Stokmarknes	366	72	30	59	1	76	10	131493	261568	56	7	8	29
Narvik	430	70	30	62	2	63	10	154949	316867	54	10	8	27
Harstad	687	71	30	58	1	62	10	145038	251823	54	10	7	28
UNN	1419	67	19	65	1	192	11	172692	327813	57	12	11	21
Hammerfest	541	68	21	65	1	171	10	160749	303279	55	11	10	24
Kirkenes	362	69	19	64	1	196	10	156482	266054	48	17	9	26

Table 9-10: Socio-economic and demographic variables for stroke patients.

Hospital	patients admitted	Mean age	% aged 80 or older	% men	Pct born outside of Norway	Mean distance from home	Mean years of highest education/	Mean gross income (NOK)	Mean gross worth (NOK)	% married/ \cohabitant	% not married	% divorced	%t widowed
Halden	149	78	48	43	3	10	11	111817	270476	44	8	3	46
Sarpsborg	171	78	47	38	5	9	10	103353	215206	39	8	9	44
Fredrikstad	1854	73	36	50	5	20	11	148775	330461	52	8	7	33
Moss	978	75	40	51	3	22	11	153123	362799	50	5	9	36
Askim	285	75	37	48	2	16	10	129173	378381	41	9	9	40
Ski	36	77	36	39	6	10	11	144883	331780	56	3	11	28
Feiring	0	.	.	.	.	.	.	.	.	.	.	.	.
Stensby	169	79	53	48	1	27	10	122820	264871	44	7	7	43
Ahus	2638	72	31	53	4	31	11	179910	401905	53	7	9	32
Aker	1698	75	38	45	5	7	11	179519	345473	40	9	13	38
Ullevål	2810	73	38	45	7	11	12	184365	444194	40	11	15	34
Lovisenberg	709	76	46	44	8	4	11	148528	272253	27	16	16	41
Diakonhjemmet	910	78	52	45	7	4	14	266776	1369712	41	10	9	40
Rikshospitalet	365	55	7	58	5	154	13	220252	484835	54	19	12	12
Bærum	1275	75	36	51	5	17	13	210602	595966	51	5	10	33
Kongsvinger	782	75	37	53	1	28	10	128998	302802	47	9	6	39
Elverum	795	74	35	51	2	42	10	134907	346635	51	9	8	32
Tynset	286	78	54	47	1	56	10	121414	301793	48	12	4	36
Hamar	948	75	37	49	2	15	11	134767	336991	48	11	7	35
Lillehammer	1315	74	34	52	2	68	11	134025	348192	52	12	6	31
Gjøvik	1242	75	36	50	2	41	10	135304	302461	50	10	7	34
Ringerike	842	76	40	53	2	49	10	137668	364108	48	9	7	35
Buskerud	1594	73	35	56	4	19	11	161645	400823	49	6	10	34
Kongsberg	524	77	44	51	1	29	10	136118	357715	42	10	9	38
Notodden	395	77	48	52	1	30	11	132137	313115	47	14	6	33
Rjukan	216	76	44	38	1	42	10	126531	296034	45	11	6	38
Vestfold/Tbg	1760	74	38	51	4	21	11	158512	414241	50	6	10	34
Larvik	396	76	40	49	3	12	11	149349	403947	47	8	7	38
Telemark	1260	74	36	51	3	25	11	152048	343380	46	9	9	36
Kragerø	242	78	50	50	2	20	10	124785	390371	42	10	6	42
Arendal	1128	75	38	51	3	40	11	146572	373726	49	10	7	34
Kristiansand	1418	74	37	49	5	31	11	146966	388629	48	10	7	35
Mandal	25	81	68	36	8	26	10	110084	357628	28	12	12	48
Lister	347	77	47	51	3	36	11	127817	368001	47	12	4	37
Rogaland	2092	74	41	52	3	29	11	161602	417012	48	9	8	34
Haugesund	1046	75	38	49	2	26	11	137632	340300	46	10	6	38
Stord	395	76	47	52	2	27	10	137366	299301	52	8	4	35
Odda	161	77	47	53	1	29	11	147863	321755	45	14	3	39
Haukeland	2753	72	34	52	4	29	11	164246	382666	50	10	10	30
Haraldsplass	1091	79	54	45	3	22	11	140561	306371	40	11	6	42
Voss	296	77	47	49	1	40	11	133521	332634	45	17	3	36
Lærdal	269	75	45	48	2	53	11	134151	342760	51	13	4	32
Førde	624	74	39	55	1	67	11	139383	320626	48	16	5	31
Florø	127	77	42	50	2	34	10	128661	291087	46	10	4	39
Nordfjord	354	77	49	53	1	42	11	131554	384202	45	19	4	32

Hospital	patients admitted	Mean age	% aged 80 or older	% men	Pct born outside of Norway	Mean distance from home	Mean years of highest education/	Mean gross income (NOK)	Mean gross worth (NOK)	% married/ \cohabitant	% not married	% divorced	%t widowed
Volda	568	75	42	52	1	33	11	142932	397783	51	11	4	34
Ålesund	1039	75	40	53	2	29	11	150408	423529	48	13	6	32
Molde	944	75	40	51	1	44	11	138029	322070	50	11	5	34
Kristiansund	589	77	45	50	1	28	10	120267	256693	43	13	7	38
Orkdal	612	77	44	51	0	68	10	143257	294570	47	11	6	37
St Olav	2939	73	32	51	2	36	11	158967	369822	51	10	8	31
St Elisabeth	0	.	.	.	.	.	.	.	.	.	.	.	.
Røros	9	82	67	11	0	10	10	103533	265933	33	22	0	44
Levanger	1161	75	37	50	1	42	11	135144	295895	48	11	5	36
Namsos	701	76	43	49	0	61	10	119554	244193	51	9	5	36
Mosjøen	259	75	42	44	1	30	10	126908	258725	56	8	3	32
Sandnessjøen	328	75	39	49	0	67	10	122424	265278	43	12	7	38
Rana	421	75	40	50	1	36	10	125462	270348	49	11	3	36
Nordland	967	72	33	52	1	90	10	138667	301710	50	10	8	32
Lofoten	289	75	36	51	1	66	10	118194	229517	46	7	9	39
Stokmarknes	524	75	42	51	0	63	10	120797	237316	47	9	9	35
Narvik	345	75	38	45	2	57	10	151738	330898	41	12	8	39
Harstad	744	74	36	48	1	63	10	131010	263517	46	10	8	36
UNN	1109	70	26	54	1	126	10	152435	304377	50	15	9	26
Hammerfest	433	71	29	52	2	179	10	133548	258300	42	11	12	35
Kirkenes	322	73	32	52	2	187	10	141358	301542	42	15	9	34

Table 9-11: Socio-economic and demographic variables for hip fracture patients.

Hospital	patients admitted	Mean age	% aged 80 or older	% men	Pct born outside of Norway	Mean distance from home	Mean years of highest education/	Mean gross income (NOK)	Mean gross worth (NOK)	% married/ \cohabitant	% not married	% divorced	%t widowed
Halden	809	79	57	22	3	24	10	114822	254406	31	8	5	56
Sarpsborg	38	78	47	47	0	10	11	113968	200942	50	5	3	42
Fredrikstad	1838	79	56	27	3	20	10	117148	258862	33	8	6	53
Moss	1182	79	56	27	2	29	11	122032	301491	32	10	7	51
Askim	221	79	58	23	1	17	10	108957	276673	35	8	7	50
Ski	277	76	48	30	3	22	12	151527	319996	43	7	10	40
Feiring	1	74	0	100	0	140	8	134700	74200	100	0	0	0
Stensby	460	79	55	25	2	35	10	119413	297180	29	10	6	55
Ahus	1839	77	49	29	2	29	11	135236	311921	37	7	8	48
Aker	2617	80	62	24	4	5	11	134669	308749	23	14	10	53
Ullevål	2908	78	57	27	4	13	11	147640	359280	27	15	11	46
Lovisenberg	293	81	64	23	4	3	11	129038	263576	18	17	14	51
Diakonhjemmet	1275	81	65	25	5	5	13	222204	773715	29	13	10	48
Rikshospitalet	81	57	16	41	1	289	12	147300	272706	43	21	12	21
Bærum	1235	78	50	28	4	14	13	219872	1969414	36	8	9	47
Kongsvinger	527	79	53	28	1	30	10	110697	258327	33	10	6	51
Elverum	1210	79	55	29	1	44	10	118766	309176	31	11	6	52
Tynset	237	80	62	30	0	49	10	104797	227259	31	15	3	51
Hamar	441	77	50	26	1	22	10	107966	280206	33	15	4	49
Lillehammer	1100	77	50	33	1	66	10	118362	312361	35	16	5	43
Gjøvik	1348	79	54	30	1	44	10	113077	267933	33	12	4	50
Ringerike	845	79	56	30	2	59	10	138894	365788	31	13	6	49
Buskerud	1487	77	53	27	2	18	10	124117	289408	30	11	7	51
Kongsberg	541	79	60	32	1	25	10	125049	319564	29	11	6	55
Notodden	404	79	57	26	1	31	10	116576	265276	34	16	3	47
Rjukan	206	78	53	30	0	49	10	121224	240950	29	13	6	52
Vestfold/Tbg	1660	78	54	29	2	20	11	134463	348591	32	8	9	50
Larvik	684	79	55	27	3	13	11	128073	338139	32	8	7	52
Telemark	1471	79	55	28	1	26	10	124408	279402	32	9	7	52
Kragerø	185	79	52	24	1	25	11	120988	265697	25	12	8	55
Arendal	1099	79	57	29	3	39	11	122303	285418	31	13	6	50
Kristiansand	1168	78	54	27	3	26	11	124729	314700	34	13	6	47
Mandal	173	79	58	31	6	16	11	115878	308287	38	13	9	40
Lister	378	79	55	26	4	35	10	117792	333187	27	17	5	51
Rogaland	2138	79	60	26	2	24	11	127973	330185	32	12	6	50
Haugesund	965	80	61	27	1	28	10	113754	259809	31	13	4	52
Stord	333	81	61	24	2	32	10	107521	251944	32	11	3	54
Odda	186	80	60	27	1	33	11	114432	253544	28	15	2	55
Haukeland	2361	79	58	29	1	20	11	125300	251813	29	14	6	50
Haraldsplass	1071	80	60	26	1	15	11	125178	261685	32	12	6	50
Voss	473	81	64	29	0	42	10	111093	274567	32	16	3	49
Lærdal	252	79	59	28	1	51	10	112962	280774	31	18	2	50
Førde	652	80	62	31	0	66	10	106857	269453	31	18	2	49
Florø	27	80	44	22	0	27	10	93681	231615	44	19	0	37
Nordfjord	324	80	61	32	1	40	11	112151	252903	34	13	2	51
Volda	387	80	60	28	1	36	10	109578	249514	34	14	4	48
Ålesund	965	79	56	28	1	27	11	117037	273170	30	14	5	51

Hospital	patients admitted	Mean age	% aged 80 or older	% men	Pct born outside of Norway	Mean distance from home	Mean years of highest education/	Mean gross income (NOK)	Mean gross worth (NOK)	% married/ \cohabitant	% not married	% divorced	%t widowed
Molde	687	79	56	30	1	39	10	113386	278346	33	13	5	49
Kristiansund	484	78	55	29	1	30	10	109194	218558	32	11	6	51
Orkdal	365	80	61	27	0	60	10	108993	224886	28	11	6	55
St Olav	2419	78	55	30	1	35	11	124026	277267	33	12	7	48
St Elisabeth	0	.	.	.	.	.	.	.	.	.	.	.	.
Røros	98	78	57	34	1	37	10	111058	251892	41	17	5	37
Levanger	883	79	59	31	1	41	10	110617	266067	31	14	5	51
Namsos	476	79	59	27	1	64	10	103653	206666	33	11	4	51
Mosjøen	182	80	60	22	1	30	10	111064	215416	27	8	7	57
Sandnessjøen	261	80	63	31	0	58	9	101665	220672	30	9	5	56
Rana	338	76	48	34	1	37	10	115625	242288	38	9	7	46
Nordland	779	78	52	29	1	83	10	119177	250352	31	11	6	52
Lofoten	222	79	55	25	0	51	10	100518	150030	35	10	6	49
Stokmarknes	334	80	63	36	1	69	9	100334	173745	30	12	5	52
Narvik	320	79	53	29	0	45	10	119179	223678	31	13	3	52
Harstad	566	79	57	32	1	54	10	107828	201152	29	14	6	51
UNN	884	75	47	32	1	126	10	119639	256761	32	14	7	46
Hammerfest	304	76	47	29	2	165	10	110854	196108	26	13	8	53
Kirkenes	233	77	51	27	1	217	10	110140	176421	24	12	11	52

9.6 APPENDIX 6 – DATA QUALITY DETAILS

The tables below show the results of the data quality control for all participating hospitals, where information collected from PAS for each disease group, admission data, main diagnosis and index diagnosis were checked against the patient journals. This control was done by the hospitals themselves or, for a sample of hospitals, by a study group doctor.

The instructions for checking were not specific as to the definition or classification of errors, or the scope and level of detail. It is possible that different hospitals made different interpretation of the task, which may be the reason for the relatively large variability in reported error rates. We have indications that some hospitals made a thorough reevaluation of the data based on reading of journals. The study group doctor however,was told to check that the information in the journal was correctly transcribed into the PAS system, but not to check if the diagnosis was accurately made.

From the tables it is apparent that the error rates were zero for most hospitals. Some hospitals had very large error rates, however.

Table 9-12: Acute myocardial infarction - percentage data item errors per hospital. Data checked by hospital (Hsp) or by the study group. Checks not performed are denoted by a dash -.

Hospital	Admission date		Main diagnosis		Index diagnosis	
	Hsp	Study group	Hsp	Study group	Hsp	Study group
Hsp1	0	-	13	-	0	-
Hsp4	0	-	8	-	4	-
Hsp6	0	-	0	-	0	-
Hsp9	2	-	0	-	0	-
Hsp10	0	-	2	-	0	-
Hsp11	2	0	2	0	2	0
Hsp12	0	0	0	0	0	0
Hsp15	0	-	0	-	0	-
Hsp16	0	0	0	4	0	4
Hsp17	0	0	2	2	0	0
Hsp18	-	0	-	0	-	0
Hsp20	-	0	-	2	-	2
Hsp22	0	-	2	-	0	-
Hsp23	0	-	2	-	0	-
Hsp24	0	0	0	0	0	0
Hsp25	2	-	0	-	0	-
Hsp26	0	0	0	0	0	0
Hsp28	0	-	8	-	2	-
Hsp29	2	-	4	-	0	-
Hsp30	2	-	12	-	6	-
Hsp31	0	0	0	0	0	0
Hsp33	0	-	2	-	2	-
Hsp34	0	0	10	0	4	0
Hsp35	2	2	14	0	0	0

Hospital	Admission date		Main diagnosis		Index diagnosis	
	Hsp	Study group	Hsp	Study group	Hsp	Study group
Hsp36	2	-	2	-	2	-
Hsp38	0	-	12	-	0	-
Hsp39	2	0	10	2	14	2
Hsp40	2	-	0	-	0	-
Hsp41	2	-	8	-	0	-
Hsp43	4	2	29	22	19	20
Hsp46	2	-	2	-	0	-
Hsp49	0	-	0	-	0	-
Hsp50	0	-	12	-	0	-
Hsp51	2	-	2	-	2	-
Hsp52	0	-	2	-	0	-
Hsp53	0	2	2	0	0	0
Hsp56	0	-	6	-	0	-
Hsp57	0	-	0	-	0	-
Hsp58	0	-	12	-	10	-
Hsp59	0	-	0	-	0	-
Hsp62	0	0	10	7	4	7
Hsp63	2	-	0	-	0	-
Hsp64	0	-	0	-	0	-
Hsp65	2	-	8	-	0	-
max	4	2	29	22	19	20
ave	0.77	0.41	4.74	2.60	1.70	2.29

Table 9-13: Stroke - percentage data item errors per hospital. Data checked by hospital (Hsp) or by study group. Checks not performed are denoted by a dash -.

Hospital	Admission date		Main diagnosis		Index diagnosis	
	Hsp	Study group	Hsp	Study group	Hsp	Study group
Hsp1	0	-	10	-	0	-
Hsp4	2	-	20	-	20	-
Hsp6	0	-	4	-	0	-
Hsp9	0	-	0	-	0	-
Hsp10	0	-	6	-	0	-
Hsp11	2	2	0	4	0	4
Hsp12	0	0	2	0	0	0
Hsp15	0	-	4	-	0	-
Hsp16	4	0	2	0	0	0
Hsp17	0	0	2	2	2	2
Hsp18	-	0	-	0	-	0
Hsp20	-	0	-	0	-	0
Hsp22	0	-	6	-	0	-
Hsp23	2	-	0	-	0	-
Hsp24	0	0	2	0	0	0
Hsp25	0	-	4	-	4	-

Hospital	Admission date		Main diagnosis		Index diagnosis	
	Hsp	Study group	Hsp	Study group	Hsp	Study group
Hsp26	2	0	0	0	0	0
Hsp29	0	-	2	-	0	-
Hsp30	0	-	34	-	34	-
Hsp31	0	0	0	0	0	0
Hsp34	0	0	9	0	9	0
Hsp35	2	0	6	2	0	2
Hsp36	2	-	0	-	0	-
Hsp38	0	-	0	-	0	-
Hsp39	0	0	2	0	2	0
Hsp40	4	-	2	-	2	-
Hsp41	6	-	0	-	0	-
Hsp43	4	0	27	33	27	33
Hsp46	2	-	4	-	2	-
Hsp49	0	-	0	-	0	-
Hsp51	0	-	0	-	0	-
Hsp52	0	-	8	-	0	-
Hsp53	0	0	4	0	0	0
Hsp56	0	-	0	-	0	-
Hsp57	0	-	6	-	0	-
Hsp58	0	-	2	-	2	-
Hsp59	0	-	6	-	0	-
Hsp62	0	0	8	11	8	11
Hsp63	0	-	2	-	0	-
Hsp64	0	-	0	-	0	-
Hsp65	2	-	6	-	0	-
max	6	2	34	33	34	33
ave	0.89	0.15	4.88	3.47	2.87	3.47

Table 9-14: Hip fracture - percentage data item errors per hospital. Data checked by hospital (Hsp) or by study group. Checks not performed are denoted by a dash -.

Hospital	Admission date		Main diagnosis		Index diagnosis	
	Hsp	Study group	Hsp	Study group	Hsp	Study group
Hsp1	0	-	0	-	0	-
Hsp4	2	-	14	-	16	-
Hsp6	0	-	0	-	3	-
Hsp9	0	-	6	-	4	-
Hsp10	0	-	6	-	2	-
Hsp11	12	4	6	4	4	4
Hsp12	0	2	0	2	0	2
Hsp14	2	-	6	-	6	-
Hsp15	0	-	2	-	0	-
Hsp16	0	0	0	2	0	2

Hospital	Admission date		Main diagnosis		Index diagnosis	
	Hsp	Study group	Hsp	Study group	Hsp	Study group
Hsp17	0	0	2	2	2	2
Hsp18	-	0	-	0	-	0
Hsp20	-	0	-	4	-	4
Hsp22	0	-	2	-	0	-
Hsp23	0	-	0	-	0	-
Hsp24	2	0	8	8	0	8
Hsp25	2	-	2	-	2	-
Hsp26	2	0	2	0	0	0
Hsp28	0	-	10	-	8	-
Hsp29	0	-	4	-	0	-
Hsp30	0	-	14	-	16	-
Hsp31	0	0	2	2	0	2
Hsp34	0	2	0	2	0	2
Hsp35	0	0	2	0	0	0
Hsp36	6	-	0	-	0	-
Hsp38	0	-	15	-	0	-
Hsp39	10	11	8	0	6	0
Hsp40	0	-	12	-	8	-
Hsp41	6	-	0	-	0	-
Hsp43	4	0	10	14	10	14
Hsp46	2	-	6	-	0	-
Hsp51	0	-	2	-	0	-
Hsp52	0	-	12	-	0	-
Hsp53	0	0	6	0	0	0
Hsp56	0	-	4	-	0	-
Hsp57	0	-	2	-	0	-
Hsp58	6	-	0	-	0	-
Hsp59	0	-	0	-	0	-
Hsp62	2	2	0	2	0	2
Hsp63	0	-	0	-	0	-
Hsp65	4	-	0	-	0	-
max	12	11	15	14	16	14
ave	1.6	1.4	4.2	2.9	2.2	2.9