

Ole Håkon Kvelstad Hallem

Feil i innkomstjournaler etter innføring av talegjenkjenning

- Sekretærskrevet vs. legeprodusert tekst

Veiledere/medforfattere:

Eirik Nordengen, 5. års medisinstudent ved NTNU.

Vidar Rao, ass.lege, ph.d., Nevrokirurgisk avdeling/INM –
NTNU/St. Olavs Hospital.

Sasha Gulati, overlege, førsteamanuensis, Nevrokirurgisk
avdeling/INM – NTNU/St.Olavs Hospital.

Ole Solheim, hovedveileder, overlege, professor,
Nevrokirurgisk avdeling/INM – NTNU/St.Olavs Hospital.

Hovedoppgave i Medisin
Trondheim, Desember 2015

Norges teknisk-naturvitenskapelige universitet
Det medisinske fakultet
Institutt for Nevromedisin



Kunnskap for en bedre verden

SAMMENDRAG

Bakgrunn

Innføring av talegjenkjenning som verktøy for klinisk dokumentasjon og ledsagende nedbemanning av sekretærtjenesten har ført til at sykehusleger nå i større grad enn tidligere skriver egne notater, med eller uten hjelp av talegjenkjenning. Denne studien undersøkte om omleggingen fra sekretærskrevet tekst til legeprodusert tekst har hatt betydning for antall språklige feil i pasientjournaler.

Materiale og metode

Vi inkluderte 400 fortløpende signerte innkomstjournaler fra St. Olavs Hospital, hvorav 200 av disse var legeproduserte og 200 sekretærskrevne. En blindet gjennomgang av journalteksten ble gjort etter forhåndsdefinerte kriterier. Inter-observatør reliabilitet ble undersøkt for de første 20 journalene med Intraclass correlation coefficient. Primært endepunkt for studien var antall setninger som grunnet språklige feil ble utolkbare for leser.

Resultater

Vi fant at 21% av innkomstjournalene med legeprodusert tekst inneholdt én eller flere utolkbare feil, mot 11% av journalene som var diktert av lege og skrevet av sekretær ($p= 0.009$). Vi fant videre at blindingen i studien fungerte godt og Intraclass correlation coefficient viste høy grad av enighet mellom observatørene.

Fortolkning

Innkostjournaler skrevet av legen selv, med eller uten hjelp av talegjenkjenning, inneholder flere alvorlige feil enn innkomstjournaler diktert av lege og skrevet av helsesekretær. Dette kan ha betydning for pasientsikkerhet og dokumentasjonskravet i spesialisthelsetjenesten.

INTRODUKSJON

Det har vært store IKT-endringer i helsesektoren de senere år. En større nasjonal gjennomgripende endring er innføring av talegjenkjenning (TGK) som et verktøy for klinisk dokumentasjon i elektronisk pasientjournal (EPJ). Ved hjelp av TGK kan legen lese inn tekst direkte via et dikteringssystem til datamaskinen. Denne lydinformasjonen blir så prosessert og et tekstforslag blir presentert for legen omgående for korreksjon og signering. Dette har medført at leger ved mange norske sykehus nå utformer egen tekst gjennom TGK eller ved at de skriver selv, til forskjell fra at teksten blir diktert av lege og deretter skrevet av helsesekretær, som var vanligere før.

Innføring av TGK har vært debattert og det er delte meninger om hvor godt systemet fungerer (1,2,3). Noen er fornøyde og synes systemet fungerer tilfredsstillende, mens andre foretrekker å skrive teksten selv. Andre igjen fortsetter å diktere som før slik at teksten senere skrives av sekretær og sendes tilbake til dikterende lege for gjennomsyn og signering. Mange steder kan legene velge mellom tradisjonell diktering til sekretærtjenesten og produksjon av egen tekst med eller uten hjelp av TGK, mens man andre steder ikke lenger tillater diktering til sekretærtjeneste slik at journaltekst helt og holden blir legeprodusert.

Hovedargumentene for innføringen av TGK var at journalnotater raskt ble tilgjengelig i EPJ, samtidig som at kvaliteten skulle øke blant annet ved at legen kunne ferdigstille og signere journaltekst med pasientdata ferskt i minne. Dette skulle også redusere epikrisetid, og føre til økonomiske besparelser gjennom nedbemanning av sekretærtjenesten. Hovedargumentene mot innføringen av TGK har vært at det kan føre til flere feil i journaler, det krever større årvåkenhet av legen ved korrekturlesing og at sekretæroppgaver pålegges legen på bekostning av tiden til pasientene. Mange mener også TGK eller egenprodusert tekst tar mer tid, samt at individuelle faktorer som språk, målform og dialekter i for stor grad påvirker kvaliteten.

I forbindelse med innføring av TKG er det mange steder gjennomført evalueringer i form av brukerfornøydhets (4). Hva denne omleggingen fra sekretærskrevet tekst til legeprodusert tekst har hatt å si for den språklige kvaliteten i journalene er imidlertid ikke forsøkt undersøkt objektivt.

I en blindet case-kontroll-studie undersøkte vi antall språklige feil i innkomstjournaler med legeprodusert tekst og sammenlignet med innkomstjournaler som ble diktert av lege og deretter skrevet av sekretær.

METODE OG MATERIALE

Denne studien er en hovedoppgave ved det medisinske fakultet, NTNU.

Regionale komiteer for medisinsk og helsefaglig forskningsetikk (REK) ble forelagt protokoll, men anså studien for å være et kvalitetssikringsprosjekt.

Studien er videre godkjent av Pasientvernombudet ved St. Olavs Hospital som et kvalitetssikringsprosjekt.

Studiepopulasjon

Vi søkte å inkludere 400 innkomstjournaler fra St. Olavs Hospital. En merkantilt ansatt ved fagavdelingen, St. Olavs Hospital inkluderte innkomstjournaler signert i tidsrommet 01.01.15, klokken 00.00 og fremover til 10.01.15 klokken 01.25, til datamaterialet bestod av 200 innkomstjournaler i hver av de to følgende gruppene:

- Legeprodusert tekst: 200 signerte fortløpende innkomstjournaler skrevet enten ved hjelp av TKG eller skrevet av legen selv.
- Sekretærskrevet tekst: 200 signerte fortløpende innkomstjournaler diktert av lege, men skrevet av sekretær.

De 400 inkluderte innkomstjournalene var signert av 85 ulike leger, der 38 leger hadde signert sekretærskrevet tekst, 59 hadde signert legeprodusert tekst, og 9 leger hadde signert journaler av begge typer tekst. Før en blindet 5. års legestudent uten erfaring med bruk av TKG (OHKH) fikk innkomstjournalene til gjennomlesing ble de aidentifisert og informasjon om hvilken gruppe de

tilhørte (legeprodusert eller sekretærskrevet) ble fjernet. For å validere studiens endepunkter med interobservatør-reliabilitet ble de 20 første journalene gjennomgått av en uavhengig observatør (EN)

Studievariabler

Journaltekstene ble gjennomgått og klassifisert etter følgende forhåndsdefinerte variabler:

- Primært endepunkt, heretter kalt "Utolkbare feil": Antall setninger som med beste vilje ikke gir mening for leser. Med det menes setninger som ut fra grammatiske, semantiske eller andre språklige feil eller mangler ikke er forståelige.
- Sekundære endepunkter:
 - "Tolkbare feil": Antall setninger som inneholder åpenbare grammatiske eller andre språklige feil, men som likevel ikke gjør setningen uforståelig.
 - Medikamentnavn stavet feil, for eksempel "Meftormin", i stedet for "Metformin".
 - Antatt dosefeil, for eksempel gram i stedet for milligram. Doseringer utenfor anbefalt maks- eller minimumsdose ut fra felleskatalogteksten.
 - Medikamentdose ikke oppgitt der det finnes flere mulige doseringer
 - Antall ord i den aktuelle innkomstjournalen.

Eksempler på utolkbare feil:

- "Pasienten oppgir at han bukse med smerter"
- "Pas er opprinnelig fra i pannen"
- "Det ses to sår på venstre fotkrem"
- "Ved auskultasjon høres "....." over venstre lunge"

Eksempler på tolkbare feil:

- "Pasienten har streke smerter"
- "Våken og orientert pas"
- "Røykte i 5 år på 78-tallet"
- "Uøm ved rektal respirasjon"

For å evaluere kvaliteten på blindingen i studien gjettet gjennomleser under gjennomgang av materialet på om den aktuelle innkomstjournalen var legeprodusert eller sekretærskrevet.

Statistiske analyser

Styrkeberegning ble gjennomført med SamplePower versjon 3.0.1: Dersom man antar at andelen av journaler med feil som ikke gir mening for leser (utolkbare feil) er 10% vs. 20% krever det 200 journaler i hver gruppe for å ha 80% styrke til å påvise en forskjell med $p < 0,05$ og chi-square test. Resultatene er analysert ved hjelp av SPSS versjon 21. Statistisk signifikansnivå ble satt til $p < 0,05$. Q-Q-plot ble laget for å undersøke mulig normalfordeling av data. Andel innkomstjournaler med én eller flere feil ble sammenlignet med Pearson's Chi-Square test. Absolutt antall feil i de to gruppene ble sammenlignet med Mann-Whitney-U test. Interobservatør-reliabilitet for studiens endepunkter ble estimert ved intraclass correlation coefficient (ICC).

RESULTAT

Som vist i tabell 1 fant vi stor grad av enighet mellom de to observatørene med ICC-verdier godt over 0,75 (5) for både utolkbare- og tolkbare-feil.

Tabell 1: Inter-observatør-enighet vist med Intraclass correlation coefficient - ICC

Feilkategori	ICC*	95% KI
Utolkbare feil – primært endepunkt	0,89	0,73 – 0,96
Tolkbare feil	0,992	0,979 – 0,997
Medikamentnavn stavet feil	1	-
Antatt dosefeil	0,78	0,45 – 0,91
Medikamentdose ikke oppgitt	0,93	0,82 – 0,97

* To-veis-mixed consistency-definisjon, gjennomsnittsverdier.

Blindingen synes å ha fungert tilfredsstillende da gjennomleser (OHKH) kun gjettest korrekt gruppe (legeprodusert eller sekretærskrevet) i 54% av de 400 innkomstjournalene.

Som vist i tabell 2 fant vi en signifikant forskjell i andel feil i inntakstjournals med legeprodusert tekst og inntakstjournals med sekretærskrevet tekst. Det var nesten dobbelt så mange journals med én eller flere utolkbare feil i gruppen med legeprodusert tekst. Det var også signifikant forskjell mellom de to gruppene i totalt antall utolkbare feil, 62 vs. 26, $p=0,007$ (fig 1). For sekundære endepunkter fant vi også en signifikant forskjell i andelen journals med tolkbare feil (tab 2). Det var også signifikant forskjell mellom de to gruppene i totalt antall tolkbare feil, 265 vs. 161, $p=0,004$ (fig 1).

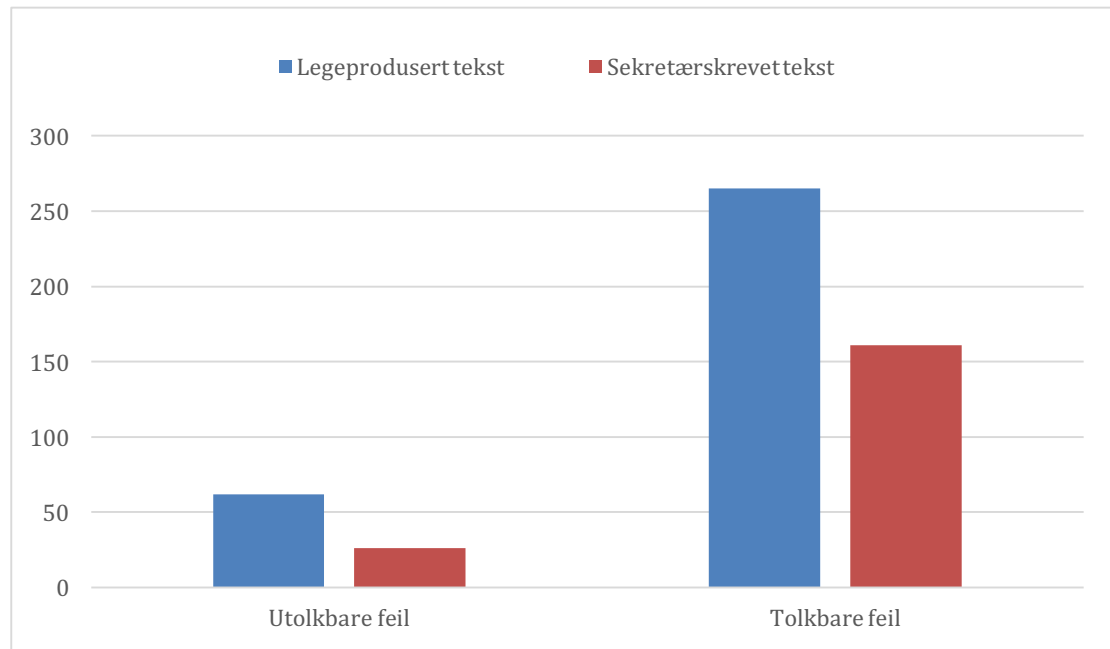
For de resterende sekundære endepunkter "medikamentnavn stavet feil", "antatt dosefeil i medikamentliste" og "medikamentdose ikke oppgitt", var det ikke signifikante forskjeller mellom gruppene (tab 2). Det var ikke signifikant forskjell i gjennomsnittlig antall ord i de to gruppene, med 450 ord i gruppen for legeprodusert tekst og 448 antall ord i gruppen for sekretærskrevet tekst ($p=0,462$). I en post hoc-analyse var det en signifikant positiv assosiasjon mellom utolkbare og tolkbare feil ($p = 0,012$).

Tabell 2: Andel journals av legeprodusert og sekretærskrevet tekst med minst én feil.

	Legeprodusert tekst N = 200	Sekretærskrevet tekst N = 200	p-verdi *
Kategori	Andel journals med minst én feil N(%)	Andel journals med minst én feil N(%)	
Primært endepunkt			
Utolkbare feil	21%	11%	0,009
Sekundære endepunkter			
Tolkbare feil	60%	48%	0,012
Med.navn stavet feil	11%	8%	0,306
Antatt dosefeil i med.-liste	10%	8%	0,485
Med.dose ikke oppgitt	6%	8%	0,433

* Sammenligner andel journals med minst én feil de to gruppene, Chi-square

Figur 1: Totalt antall språklige feil (utolkbare og tolkbare) i 200 journaler med legeprodusert tekst og 200 journaler med sekretærskrevet tekst.



DISKUSJON

I denne blindede case-kontroll-studien fant vi at innkomstjournaler skrevet av legen selv, med eller uten hjelp av TGK (legeprodusert tekst), er assosiert med signifikant flere feil enn innkomstjournaler som er skrevet av sekretær etter diktering. Forskjellen gjelder både for alvorlige språklige feil som gjør at setninger mister mening for leser, og mindre alvorlige feil som gjør at setninger fortsatt beholder meningen. Det var nært dobbelt så mange innkomstjournaler med én eller flere utolkbare setninger og over dobbelt så mange utolkbare setninger totalt i innkomstjournaler med legeprodusert tekst. Funnene gir grunn til bekymring for om innføring av TGK og overgang til mer legeprodusert tekst kan ha hatt klare negative konsekvenser i form av å øke antall språklige feil i EPJ.

Det var ingen signifikant forskjell mellom gruppene for øvrige studievariabler, herunder antatte dosefeil, manglende informasjon om dose i medikamentliste, stavefeil i medikamentliste samt antall ord. Vi fant en positiv assosiasjon mellom utolkbare feil og tolkbare feil og det synes som om én feil i journal øker sjansen

for at det er flere feil. Dette kan antyde at det kan være forskjell på hvor nøye den enkelte lege redigerer og retter opp tekst før signering.

Plikten til å føre journal og kravene til journalens innhold er regulert i helsepersonelloven og i forskrift om pasientjournal. Journalen skal være legens arbeidsverktøy, men journalen skal også kunne benyttes til kontrollformål, for eksempel i forbindelse med tilsynssaker. Både Helsetilsynet og domstolene legger stor vekt på legens journalføring i tilsynssaker og i rettssaker. Helsetilsynet vil ved vurdering av om det foreligger brudd på forsvarlighetskravet i helsepersonelloven § 4 ofte se dette i nær sammenheng med legens journalføring. Dersom legen har skrevet en god og utfyllende journal som viser hvilke vurderinger som er foretatt med hensyn til pasienten, vil legen lettere kunne dokumentere at han eller hun har handlet i tråd med forsvarlighetskravet. Data fra pasientjournaler er også viktige i kvalitetsarbeid, kvalitetssikringsprosjekter og mange forskningsprosjekter. Innføringen av TGK i norske sykehus kan ha bidratt til økt antall uleselige journalopplysninger og kan dermed være en potensiell trussel for pasientsikkerhet og i siste instans også forsvarlighetsvurderinger av tilsynsmyndigheter.

Innføring av TGK ved norske sykehus skulle gi en kvalitetsgevinst gjennom en mer oppdatert journal, redusert epikrisetid, og gi en økonomisk gevinst gjennom reduksjon i behov av helsesekretærer. Denne studien er den første i Norge som ser på objektive språklige kriterier i journaler etter innføring av TGK. Vi har heller ikke funnet utenlandske publikasjoner som ser på det samme. Det er imidlertid et tankekors at ingen til nå har dokumentert faktiske følger av en så stor endring av rutiner som det har vært å gå over til legeprodusert tekst. Det kan argumenteres for at studier som denne skulle vært gjort før TGK og andre gjennomgripende IKT-system ble innført i stor skala. Det kan videre sees på som et paradoks at mens innføring av ny diagnostikk og behandling er underlagt dels strenge krav til utprøving, kontroll og kost-nytte-analyser, kan gjennomgripende endringer i IKT-løsninger i helsesektoren gjøres uten kvantitative data eller objektive pasientsikkerhetsvurderinger.

Studien gir ikke noe svar på mekanismene bak resultatet. Studien er ikke randomisert, men legene valgte selv å bruke TGK, skrive selv eller diktere til sekretær, så vi kan ikke med sikkerhet vite at feilene ville vært unngått ved bruk av sekretær i gruppen med legeprodusert tekst. En svakhet er at vi ikke i ettertid kunne finne ut om journalene er skrevet med TGK eller skrevet av legen selv. Av journalene med legeprodusert tekst er det derfor vanskelig å vite om de fleste feilene kommer fra de som bruker TGK eller de som skriver selv. I mange tilfeller brukes dessuten kombinasjoner med større og mindre grad av redigering fra legens side. Det burde imidlertid være fullt mulig å lagre lydfiler etter digital diktering og la sekretærer skrive samme tekst for å sammenligne TGK og sekretærprodusert tekst mer direkte. En oversikt fra TGK-gruppa ved St. Olavs (data fra mai 2014-sept 2014) tyder på at omkring 45% bruker TKG, ca. 45% skriver selv og 10% dikterer dikteres av lege og skrives av sekretær. Den høye andelen som skriver selv uten bruk av TGK skyldes nok i stor grad likevel innføring av TGK med samtidig nedbemanning av sekretærtjenesten. Som mulig følge av tidspress er nok prosentandelen som skrives av sekretær høyere for inntakstjournaler enn annen legedokumentasjon i EPJ.

Inntakstjournalerne er hentet fra akuttmottaket ved St. Olavs Hospital, og generaliserbarheten til andre norske sykehus er avhengig av flere tenkelige faktorer, som f. eks andelen av legene som produserer egen tekst, holdningene til TGK, tid siden innføring av TGK, krav om bruk av TGK og antall sekretærer til rådighet. TGK ble innført ved St. Olavs Hospital i 2009, mer enn 5 år før studien ble utført. Inntakstjournalerne i studien var signert av hele 85 ulike leger slik at materialet nok i liten grad er sårbart for ekstremverdier til enkeltleger på vakt i det aktuelle tidsrommet journalene ble hentet fra. At journalene ble hentet fortløpende og ikke er et tilfeldig utvalg over en lengre tidsperiode skyldes praktiske forhold, men har neppe hatt større betydning for resultatene. Da mange ferske leger jobber i akuttmottaket kan det tenkes at mange enda ikke har nådd toppen av læringskurven når det gjelder journalføring og bruk av TGK. Hvorvidt funnene er generaliserbare til andre typer journalnotater vites ikke, men vi valgte å se på inntakstjournaler da disse er nokså standardiserte og da informasjonen i inntakstjournalerne ofte reflekterer pasientens første møte med

helsevesenet ved et akuttopphold. Feilaktige opplysninger kan derfor potensielt ha stor negativ konsekvens for logistikk, pasientsikkerhet og i eventuelle tilsynssaker.

Alle journaler i studien ble gjennomlest av én enkelt gjennomleser (OHKH), og hvilke feil som ble oppdaget og i hvilken feilkategori de ble plassert i, vil til en viss grad kunne være subjektivt. Undersøkelsen av interobservatør-reliabilitet i en del av materialet viser imidlertid svært stor grad av enighet mellom observatørene. Blindingen av observatøren var også god, noe som nok kan skyldes at gjennomleser selv ikke har erfaring med bruk av TKG.

Studien fra St. Olavs Hospital viser at det er betydelig flere språklige feil i innkomstjournaler med legeprodusert tekst sammenlignet med tekst skrevet av sekretær.

Hovedbudskap:

- Innkomstjournaler skrevet av legen selv, med eller uten bruk av talegjenkjenning, inneholder flere alvorlige språklige feil enn innkomstjournaler som er diktert og skrevet av sekretær.
- Innføring av talegjenkjenning skulle gi en kvalitetsgevinst for medisinsk dokumentasjon. Det bør nå vurderes om overgang til legeprodusert tekst ved norske sykehus i stedet har gitt en kvalitetsforringelse og dermed kan representere et pasientsikkerhetsproblem. Studien eksemplifiserer at det er fullt mulig, med relativt lite ressurser, å gjennomføre evalueringer av nye IKT-tjenester med objektive effektmål.

REFERANSER

1 Talegjenkjenning gir trøbbel. Dagens medisin, 2010:

<http://www.dagensmedisin.no/artikler/2010/04/29/talegjenkjenning-gir-trobbel/> (22.11.15)

2. Han får smerter i albuen når han spiser på folk. Dagens medisin, 2010:

<http://www.dagensmedisin.no/artikler/2010/04/30/han-far-smerter-i-albuen-nar-han-spiser-pa-folk> (22.11.15)

3. Fornøyd med talegjenkjenning. Dagens medisin, 2010

<http://www.dagensmedisin.no/artikler/2010/05/11/fornoyd-med-talegjenkjenning-i-allmennpraksis/> (25.11.15)

4. Fossberg AB, Gjerde A, Graver C, Universitetet i A. Talegjenkjenning: hvilke erfaringer har leger med bruk av talegjenkjenning? Grimstad: Forfatterne; 2012.

5. Portney LGa, Watkins LGa. Foundations of Clinical Research: applications to practice. Upper Saddle River, NJ: Prentice Hall; 2015, 3rd revised ed. Side 594-595.

TAKKSIGELSER

- Jostein Dale (avdelingssjef ved Mottaksavdelingen, St.Olavs Hospital) og Lars Eide Næss-Pleym (IKT-rådgiver ved Mottaksavdelingen, St. Olavs Hospital), for hjelp med datauttrekk.
- Stor takk til Rut Naversen, seksjonsleder i fagavdelingen, seksjon for kliniske IKT-funksjoner, St. Olavs Hospital og Anne Margrethe Hoston, systemkonsulent, St. Olavs Hospital for anonymisering og klassifisering av journaler.