

# Prosessuell generering av medisinske kasuistikker

*Et begrenset eksperiment*

Knut Sindre Åbjørsbråten

Veileder: Lars Aabakken



Prosjektoppgave ved Det medisinske fakultetet

UNIVERSITETET I OSLO

Høst 2015

# Prosessuell generering av medisinske kasuistikker

*Et begrenset eksperiment*

© Knut Sindre Åbjørsbråten

2015

Prosessuell generering av medisinske kasuistikker – et begrenset eksperiment

Knut Sindre Åbjørsbråten

<http://www.duo.uio.no/>

Trykk: Reprosentralen, Universitetet i Oslo

# Abstract

**BACKGROUND:** Procedural generation is a technique used in video game development where algorithms are used to create game assets such as landscapes and foliage. It enables game studios to create greater amounts of content than would be feasible using artists alone. For medical students, case books offer a way to apply and challenge their theoretical knowledge of diagnostics and therapy. Could procedural generation be used for generating medical cases? An unstructured search of medical literature reveals little mention of this method being explored. This project implemented and tested a limited experiment to see if the technique was applicable also in this field.

**METHODS:** Three software tools were written: (1) A modelling tool to describe the variety in a condition's presentation, findings, lab results etc. (2) A tool that picked elements from these models in a predictable way and outputted medical cases as datasets. (3) A browser based client that presented an excerpt of a random case and asked medical students to arrive at diagnosis and treatment by requesting additional information about the patient from the case dataset. To incentivize rational choices in what information to retrieve, each request came at a cost in terms of virtual money, time and patient discomfort. After solving each case a score was presented. Students were recruited through social media to test the client tool and answer a survey about their experience.

**RESULTS:** There were 30 respondents. Significant bias possibilities make the results unreliable. The students reported the system to be "interesting" (87%), "useful" (87%), "fun" (77%) and "exciting" (77%). 23% found the system difficult. Less than 14% found it "frustrating", "overwhelming", "confusing" or "boring". Most (> 75%) perceived the medical quality to be good. The same number responded they would use such a system again, and recommended it to others, if it had more content. A similar number believed it would make them better equipped to make real diagnostic considerations, and that it would improve retention of the material studied in books.

**CONCLUSION:** Procedural generation may be applied in medical education to generate dataset based medical cases, but significant challenges exist. Generating cases with multi-morbidity and polypharmacy appears especially difficult. Development may be costly and consume significant amounts of expert time.

# Interessekonflikter

Forfatter kan tjene økonomisk på programvaren og konseptet utviklet til denne oppgaven hvis den kommersialiseres. Muligheten for dette kan avhenge av de rapporterte resultatene.



# Innholdsfortegnelse

Abstract .....	IV
Interessekonflikter .....	V
Innholdsfortegnelse .....	VII
1 Bakgrunn .....	1
1.1 Motivasjon .....	1
1.2 Prosessuell generering .....	2
1.3 Tidligere studier og systemer .....	4
2 Metode og gjennomføring .....	6
2.1 Datastruktur for virtuelle pasienter .....	6
2.2 Prosessuell generering med flytskjemamodeller .....	7
2.3 Programvare og modeller utviklet .....	10
2.4 Utprøving på studenter .....	12
3 Resultater .....	15
4 Diskusjon .....	19
4.1 Aspekter ved bruk av flytskjemamodeller .....	19
4.2 Diagnostisk verdi av ulike symptomer og funn .....	21
4.3 Nivå av helsetjenesten .....	22
4.4 Feilkilder i statistikken .....	23
5 Konklusjon .....	24
6 Videreutvikling .....	25
Litteraturliste .....	26
Vedlegg – Skjermbilder .....	27





# 1 Bakgrunn

## 1.1 Motivasjon

Bøker benyttet på medisinstudiet heller erfaringsmessig mot å prioritere opplisting av tilstander, deres symptomer og relevante undersøkelser. Disse forsøker å gi en oversikt over hva som kan feile en pasient og danner viktige diagnostiske overveielser å ha med seg den dagen man møter en pasient med tilsvarende problem. Over mange år med studier utgjør dette en stor informasjonsmengde. Mange studenter opplever allikevel en bratt læringskurve i møte med den påfølgende klinikken knyttet til å danne seg en indre diagnostisk prosess som skiller mellom tilstander og forutsier mulige forhold man bør være forberedt på. Det er en vesensforskjell i å ha hørt om en sykdom og å gjenkjenne den blant andre. Kunnskap om sykdommer er grunnleggende, men i mange situasjoner vil også en god prosess for å skille de fra hverandre være minst like viktig. Erfaringsmessig er ikke den listepregede kunnskapen om sykdommer og patologi lett overførbar til en god evne til dette, samt at den store informasjonsmengden er vanskelig å huske.

Med den mer kliniske hverdagen mot slutten av studiet, og gjennom praksis og arbeid på studentlisens, skjer imidlertid denne læringen organisk og spontant fordi man utfordres. Antall pasienter man møter øker eksplosivt. Man lærer av feil og forglemmelser som forhåpentligvis plukkes opp av erfarne bakvakter, slår opp kunnskap aktivt og struktureres inn i rutine til den aktuelle avdelingen. Kunnskapen fra studiet danner et bakteppe som muliggjør denne type læring, hvor studiets innhold integreres gjennom gjentatt praktisk bruk. Det erverves indre prosedyrer som hjelper til å tenke på det som er viktig – i alle fall innenfor den populasjonen man måtte møte og dens prevalenser.

Kunne mer av denne læringsprosessen vært flyttet til tidligere i studiet ved at studentene aktivt møtte mange av de samme problemstillingene de senere ville treffe i praksis med forelesninger friskere i mente? Hva om møtet med de første pasientene skjedde etter å ha trent mye teoretisk på akkurat de problemstillingene som møter studenten i klinikken? Som snart ferdig utdannet lege skulle forfatter av denne oppgaven gjerne ønsket seg et slikt læringsverktøy. Kontinuerlig repetisjon av tidligere gjennomgått materiell fra studiet med en mer diagnostisk vinkling kunne vært gunstig. Dette kunne tenkes å gi økt trygghet på egen

kunnskap, redusere erosjonen av kunnskap det er lenge siden man har brukt og å styrke den senere praktiske læringen gjennom bredere kunnskapsgrunnlag.

Karakteristikkene for et slikt læringssystem kan være:

- A. Gi studenten konkrete teoretiske arbeidsoppgaver med overføringsverdi til reelle kliniske utfordringer
- B. Tilby mengdetrening gjennom variasjon og insentiver til regelmessig bruk
- C. Gradvis økende problemområde gjennom studiet i takt med det som er gjennomgått

Tradisjonelt ville dette være en formidabel utfordring, men informasjonsteknologi har en rekke teknikker som kanskje kan sikre akkurat disse karakteristikkene. Denne oppgaven tester en slik metode som kan tenkes å danne grunnlaget for mengdetrening på virtuelle pasienter, og i et lengre perspektiv innføre elementer av såkalt gamification – bruk av konkurranseinstinkt for å få brukere til å yte.

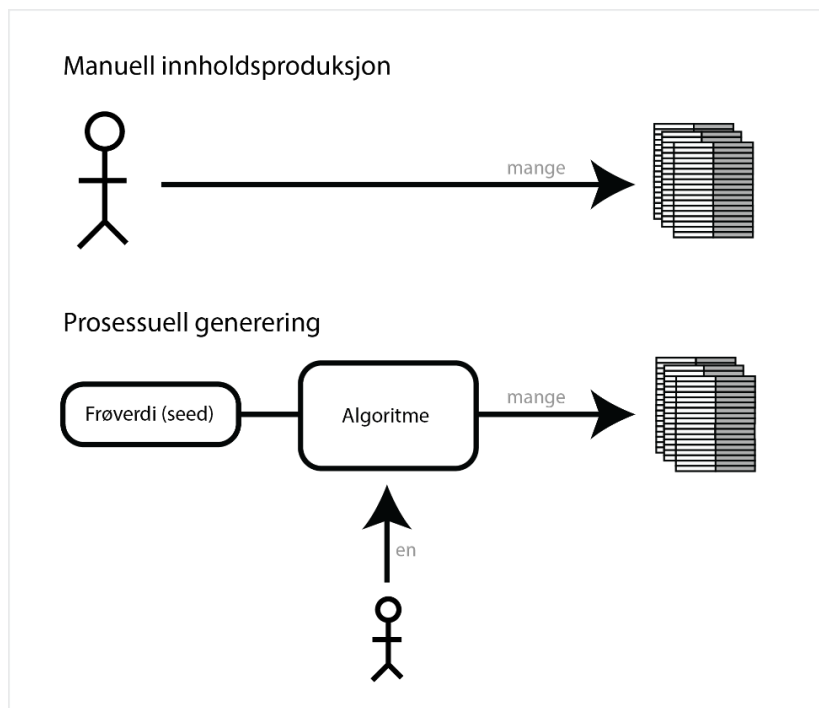
Problemstillingene for oppgaven har vært:

1. Å utvikle et system for automatisk generering av medisinske kasuistikker
2. Teste et antall genererte kasuistikker på medisinstudenter for å vurdere kvaliteten
3. Identifisere svakheter, styrker og eventuelt videreutviklingspotensiale ved denne tilnærmingen

Oppgaven er en studentoppgave ved Universitetet i Oslo, utført i tidsrommet 2013 – 2015.

## 1.2 Prosessuell generering

Teknikken som testes i denne oppgaven kalles *prosessuell generering*. Begrepet spenner bredt og benyttes for å beskrive data som produseres av en prosess. Typisk er denne prosessen en algoritme i en datamaskin. Eksempler er fraktaler, verdener og grafiske elementer i dataspill (Smith, 2008) og til og med musikk (Eno, 1996). Kraften ligger i at en enkel algoritme kan produsere komplekse og varierte resultater.



Figur 1 - Prosessuell generering. Til forskjell fra manuell produksjon av innhold, eksempelvis ved å skrive en kasuistikkbok, kan en algoritme/prosess i noen tilfeller benyttes til å produsere det samme. Denne algoritmen kan lage langt mer innhold enn hva man manuelt ville klare.

Et aktuelt eksempel på slik generering er dataspillet *No Man's Sky* (Khatchadourian, 2015). Her utforsker spilleren et fiktivt univers og kan besøke  $10^{18}$  ulike plater i virkelighetsnær størrelse. Spilleren kan utforske planetene ned til bakkenivå med landskap, dyreliv, ulik atmosfære m.m. Denne datamengden er nesten ufattelig stor og kan ikke lagres på noe kjent medium. Spillet kan heller ikke benytte kunstnere til å lage alle disse elementene. For å danne sitt univers bruker spillet derfor prosessuell generering. Universet lages av en algoritme og kun den delen av universet spilleren befinner seg i lages til enhver tid. Algoritmene har komponenter laget av kunstnere, og disse muteres i ulike retninger. Det benyttes gjerne *seeds* – frøverdier – som påvirker resultatet av algoritmen. Samme frøverdi gir typisk samme resultat hver gang. I *No Man's Sky* er spillerens posisjon frøverdien, og bestemmer hva som vises på skjermen. Neste spiller som besøker samme posisjon vil derfor se det samme siden frøet er likt.

Prosessuell generering kan tenkes å generere medisinske kasuistikker ut fra en beskrivelse av hvordan ulike sykdommer varierer i sin presentasjon og mutere innholdet innenfor disse gitte rammene. Eksempelvis kan det defineres at hemoglobinverdien til en virtuell pasient vil ligge innenfor et intervall, eller at det av og til kan tilkomme et ekstra sett symptomer. En slik

metode benyttes i denne oppgaven. Så lenge en diagnose alltid følger et mønster kan man tillate alle variasjoner som ligger innenfor mønsteret.

Hvis denne teknikken kan anvendes innen medisinske problemstillinger, tilfredsstill den kriterium B nevnt innledningsvis – variasjon. Siden man elektronisk kan trykke opp mange ulike pasienter med samme diagnose muliggjøres mengdetrening fordi kasuistikkene sjelden er like. Også kriterium C – gradvis økende fag- og problemområde – muliggjøres hvis algoritmen justeres etter hvor en student er i sitt utdanningsløp. Det gjenværende kriteriet nevnt over er overføringsverdi til klinisk praksis. Dette forutsetter at løsning av kasuistikkene krever virkelighetsnær mønstergjenkjenning og kunnskap om undersøkelser. Hvis suksess i systemet krever at man må oppsøke og tolke den sammen informasjonen som for virkelig pasient, kan vi anta at systemet vil ha en overføringsverdi.

### 1.3 Tidligere studier og systemer

Et raskt usystematisk litteratursøk viser at pasientsimulering ikke er nytt, men at de fleste benytter beslutningstrær eller forhåndsskrevne sekvenser av tekster ikke ulik tradisjonelle bøker. Dette er ikke prosessuell generering, og disse resultatene utelukkes. Virtuelle pasienter brukes også innen kirurgi via forskjellige typer simuleringsutstyr som skopimaskiner, noe som også utelatt.

Tabell 1 Treff i litteratursøk ved bruk av søketerm: (((medical OR patient) AND (case OR record OR story)) OR patient) AND ((computer generated) OR (procedural generation) OR (procedural generation) OR (simulated)).

Database	Relevante innen 50 første treff
PubMed	0 av 22325
IEEE Xplore	0 av 537
Google Scholar	1 av mange (patent) <sup>1</sup>
Google nettsøk	1 av mange

Funnet fra Google Scholar inkluderer en amerikansk patent fra 2002. Metoden har likheter, men benytter i prinsippet en annen teknikk for generering. Interaksjonen med

<sup>1</sup> Beskrivelse av patentet finnes på <https://www.google.com/patents/US7024399>

kasuistikk løseren har også likheter. Det er ikke innenfor denne oppgavens rammer å utrede et eventuelt overlapp eller dets betydning.

I perioden denne oppgaven var under utvikling presenterte Duffy og Wang en svært begrenset versjon av prosessuelle generering av medisinske kasuistikker (Duffy & Wang, 2015). Utover dette er det, som tabellen over viser, tilsynelatende lite å finne om emnet. Det er grunn til å tro at prosessuell generering er lite utforsket innen medisinsk e-læring.

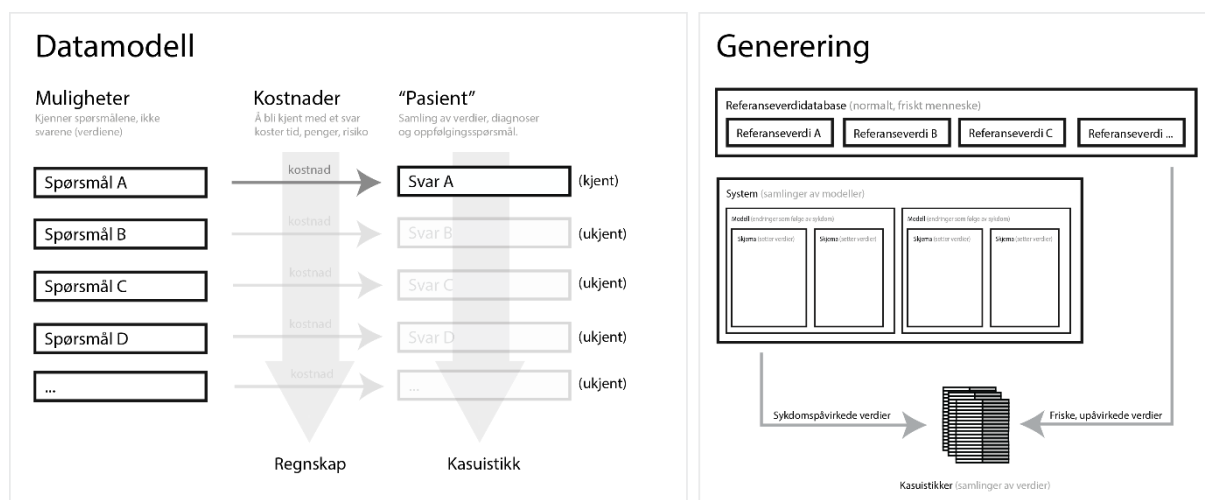
## 2 Metode og gjennomføring

### 2.1 Datastruktur for virtuelle pasienter

I denne oppgaven benyttes en enkel abstraksjon av klinisk virkelighet. Vi tenker oss en liste med de fleste viktige spørsmål en kliniker kan ønske å stille en pasient. Med spørsmål menes her den absolutt videste forstand og inkluderer anamnesticke spørsmål som alder, kjønn, plagenes varighet, klinisk inspeksjon, blodprøver, røntgenbildet m.m. *Pasienten* utgjøres av et sett svar på disse spørsmålene. *Kasuistikken* er kombinasjonen av en pasient og informasjon om dennes diagnoser, korrekt behandling og mer.

Innsamling av informasjon om en pasient har i virkeligheten en sammensatt kostnad for samfunnet, legen og pasienten. Bildeundersøkelser tar tid, mer invasive prosedyrer er ubehagelige og risikofylte. Enkelte prøver er kostbare rent økonomisk. Nær all innsamling av informasjon om pasienten tar tid. Når alt tolkes sammen, gir svarene ideelt sett diagnose og grunnlag for behandling. Slik er det også i dette systemet.

For å bruke modeller i en konsultasjonslignende situasjon kan vi la listen med mulige spørsmål være kjent i sin helhet, men skjule svarene. Studenten velger spørsmål hun eller han mener kan avklare diagnosen og slår opp svarene i kasuistikken. Innsamling av hvert svar assosieres med en takst i form av tid brukt, risiko for pasienten og økonomisk kostnad. Når studenten er ferdig og har oppgitt diagnose og behandling, sitter systemet i tillegg igjen med informasjon om hvilke spørsmål som er stilt og deres kostnad. Disse tre faktorene kan brukes i poengberegning og eventuelt som grunnlag for kommentarer. I denne oppgaven blir også oppfølgingsspørsmål og poeng fra disse tatt med i beregningen.



Figur 2 – Data- og genereringsmodell. I systemet er kasuistikker representert av mulige spørsmål en kliniker vil stille. Med spørsmål menes ikke bare anamnese, men også blodprøver, radiologi m.m. Under generering av en kasuistikk hentes normale svar fra en database på grunnlag av bl.a. alder og kjønn. Patologiske svar legges over dette. Følgelig er informasjon tilgjengelig for alle spørsmål, og studenten som løser kasuistikken kan få svar på et stort antall mulige spørsmål.

Datamodellen bruk i denne oppgaven tar ikke høyde for at verdiene kan endres over tid eller om utredning og behandling haster. Det er imidlertid ikke noe i veien for å lage et system hvor verdiene er en funksjon av tid hvor pasienter blir dårligere underveis eller symptomer endres.

## 2.2 Prosessuell generering med flytskjemamodeller

**Grunnprinsipp:** I dette forsøket benyttes generering organisert på fire nivåer: (1) referanseverdi, (2) flytskjema, (3) modell og (4) systemet som helhet. *Referanseverdiene* er en database med intervaller og standardbeskrivelser for alle spørsmål og brukes for å danne verdier som representerer det friske mennesket. Disse samvarierer kun med alder og kjønn. Avledning av en verdi fra en annen er mulig, men ikke utforsket i denne oppgaven utover teoretisk drøfting. *Flytskjemaene* består av *noder* som er plassert langs baner. Hver node inneholder informasjon som kan inngå i en kasuistikk, f.eks. tilstedeværelse av feber, en tidligere sykdom osv. Alle banene starter i samme node, og ender i en annen felles node. En akkumulator med en tom kasuistikk sendes gjennom banene i skjemaet. Den tar med seg informasjon fra nodene den møter langs banen. Når akkumulatoren kommer til siste node leveres den oppsamlede kasuistikken. *Modellene* er samlinger av flytskjemaer. Et flytskjema kan inngå i mer enn en modell, og kasuistikkene fra skjemaene slås da sammen. Dette kan brukes til å danne felles symptombilder og annet gjenbruk mellom kasuistikker. Hver modell representerer typisk en sykdomstilstand. *Systemet* er samlingen av modeller. Skal man se på

eksempelvis hvilke pre- og posttestsannsynligheter som møter studenten, må man vurdere systemet som helhet.

**Variasjon:** To mekanismer danner ulikheter mellom kasuistikkene laget gjennom samme modell: (1) byggebaner og (2) verdirandomisering. Noen noder har rolle som *delingsnoder* med mer enn en utgående bane. Delingsnodene har regler som definerer hvorvidt alle eller et tilfeldig utvalg av utgående baner skal følges når akkumulatoren skal videre. Via delingsnodene skapes derfor flere veier gjennom samme skjema. Hver kombinasjon av utgående baner skaper en kasuistikk hvor kun nodene nedstrøms får besøk av akkumulatoren. Hver av disse utfallene er en *byggebane*. Alle kasuistikker som er laget fra en byggebane vil ha verdier fra samme sett noder. Dette brukes f.eks. for å velge tilfeldig om en pasient skal ha en komplikasjon.

Noder kan også skape variasjon innenfor samme byggebane via *verdirandomisering*. Eksempelvis kan en CRP-verdi settes innenfor intervallet <30-50> i stedet for en spesifikk verdi. Tilsvarende kan tekst plukkes tilfeldig fra en liste mulige fraser eller ord og deler av en setning tilfeldig byttes ut. Det er et viktig prinsipp for forutsigbarheten og kvalitetssikringen til modellene at man kan analysere alle mulige byggebaner og spennet i verdiene deres noder lager.

**Sikring av samvariasjon:** Verdier er ofte avledet av andre eller må samsvare for å være meningsfulle. BMI avledes direkte av vekt og høyde. eGFR utregnes på grunnlag av alder, kjønn, vekt og S-kreatinin. S-kreatinin bør videre, ved nyresykdom, samsvare med den øvrige sykehistorien. Samtidig må også konsekvens av for eksempel høy BMI gjenfinnes i beskrivelsen av abdomens omfang ved klinisk undersøkelse, innsikt ved ultralyd etc. ad infinitum. Sikring av dette er komplisert. Det muliggjøres imidlertid ved å sette verdier som samvarierer langs samme byggebane, eventuelt legge inn konkrete utregninger for en verdi som baserer seg på andre verdier. Man kan alternativt innføre en skjult konfunderende faktor som ikke finnes i spørsmålslisten, men gir opphav til flere verdier. Dette utforskes ikke i oppgaven, men er relativt uproblematisk teknisk sett.

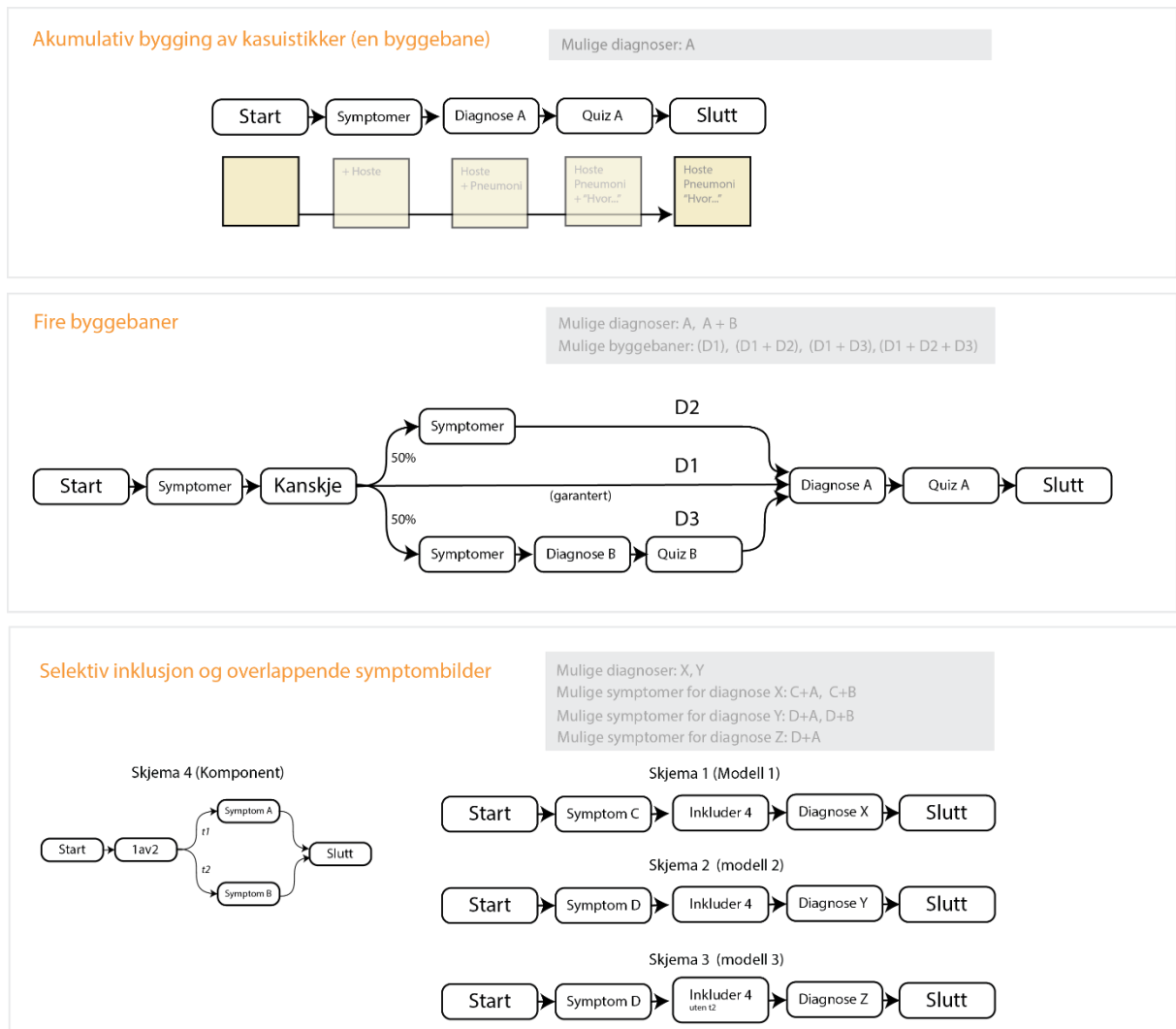
**Gjenbruk og overlapp:** Ved å si at et flytskjema kan inngå i mer enn en modell kan de samme verdiene – og deres variasjoner – havne i kasuistikker med ulike diagnoser. Når et skjema gjenbrukes, kan tiden investert i dette skjemaet redusere tiden brukt på andre modeller hvor det er naturlig å benytte nærliggende verdier. Et skjema kan eksempelvis beskrive



vanlige gjennomgåtte barnesykdommer og legge disse til sykdomshistorien. Dette skjemaet vil typisk være ønskelig å inkludere i de fleste modeller for å danne realistiske pasienter med normal helsehistorikk. Hvis man ønsker å modellere mer komplekse pasienter med tidligere sykdommer og behandlinger vil det være upraktisk å lage disse tilstandene på nytt for hver modell.

Vi kan tidvis ønske å ekskludere enkelte byggebaner fra de inkluderte skjemaene. Skal man lage kasuistikker på mononukleose må man ekskludere import av byggebaner fra skjemaet som setter inn mononukleose under tidligere sykdommer, da pasienten burde være immun. Tilsvarende kan man tenke seg å selektivt velge ut mulige elementer til tidligere sykdommer og behandlinger basert på epidemiologiske forhold der pasienten kommer fra. Gjenbruk av skjemaer har imidlertid noen konsekvenser som gir viktige begrensinger for bruk. Dette omtales i diskusjonskapittelet.

**Verdikollisjoner:** Hvis et skjema inkluderes i et annet kan det lett forekomme at det finnes mer enn en node som lager en verdi for samme spørsmål. Dette kan være fordelaktig eller negativt. Hvis senere spørsmålsverdier overskriver tidligere går informasjon tapt. Tidvis vil man dog ønske dette i gjenbruksøyemed. I neste avsnitt følger konkrete eksempler på dette. Eksempelvis kan det være nyttig at setninger legges gradvis til en røntgenbeskrivelse. Problemet med det siste er at hvis en node tidligere sier «Normale forhold i thorax», kan man ikke legge til «Stuvningsbilde». En node kan derfor bære en innstilling som sier om verdien skal legges til eller overskrives hvis tidligere verdier finnes.



Figur 3 – Flytskjemamodellen. Når kasuistikken skal lages, samles informasjon opp fra noder langs byggebaner, som i praksis er piler i flytskjemaene. Ved å variere hvilke byggebaner som tas med kan ulike verdier samles inn og variasjon skapes. Ytterligere variasjon kan også skapes ved at nodene stokastisk plukker sine verdier innenfor visse rammer. Hver gang datamaskinen kjører en modell lages slik en ny kasuistikk.

## 2.3 Programvare og modeller utviklet

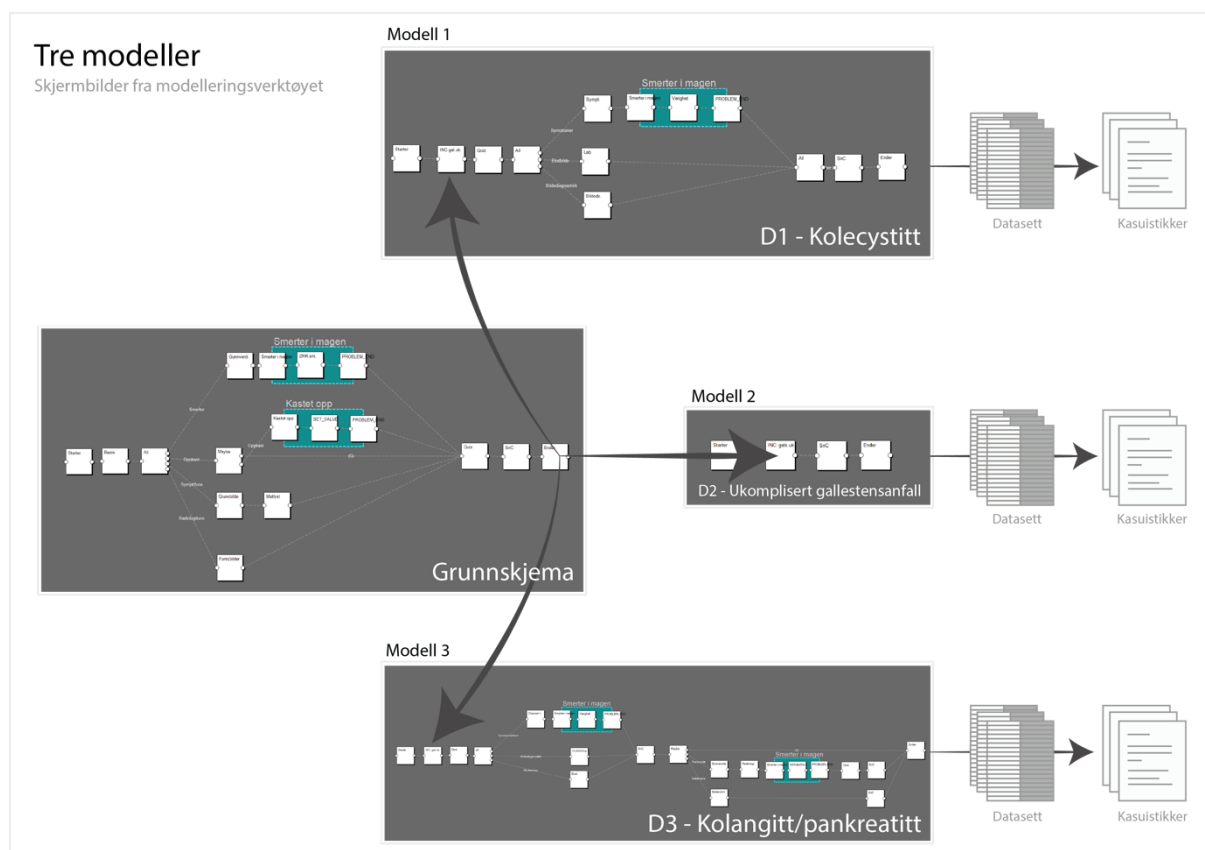
Det ble utviklet et modelleringsverktøy for konseptet beskrevet i avsnitt 2.2, en klient for å løse kasuistikker og diverse støtteprogrammer. Klienten ble laget for å la studentene løse kasuistikkene etter prinsippene beskrevet i avsnitt 2.1, og gi de en poengsum basert på bl.a. korrekt diagnose og terapi. Den ble bygget som en nettside for å senke terskelen for bruk. Det ble lagt inn ca. 370 mulig spørsmål basert på referanseverdier og forfatters egen kliniske erfaring. Alle diagnoser ble hentet fra ICD-10-katalogen. Behandlingsvalg basert bl.a. på legemiddelinformasjon fra Legemiddelverket (FEST-data) ble også importert. Klienten sammenlignet studenten med andre studenters bruk av tid, penger og risiko og viste dette etter

hver kasuistikk løst. Disse variablene ble noe tilfeldig satt i dette prosjektet da oppgaven primært var å teste prosessuell generering. Se vedlegg med skjermbilder for detaljer.

Med modelleringsverktøyet ble det laget seks modeller som inngikk i den endelige testen. Diagnosene var: akutt viral faryngitt, streptokokkfaryngitt/tonsillitt, mononukleose med EBV, ukomplisert gallestansanfall, kolecystitt og kolangitt. I tillegg var pankreatitt en mulig komplikasjon til kolangitt. Diagnoser med overlappende symptombilde ble valgt for å lage en diagnostisk utfordring ved gjentatt bruk.

Modellene med akutt abdomen illustrerer det grunnleggende modelleringsprinsippet:

*Grunnskjemaet* lager abdominale smerter i øvre høyre kvadrant, mulig med utstråling mot rygg og høyre skulder. Smertene har vart < 24 timer og er moderate. Det legges eventuelt til kvalme og, oppkast. Pasienten er utilpass og har smerter ved palpasjon i det aktuelle området. Radiologisvar legges også inn, og det vil finnes konkrementer i galleblære og galleganger på MR, CT, ultralyd, men ikke røntgen. *Modell 2* i illustrasjonen tar grunnskjemaet, legger på informasjon om at ukomplisert gallesteinsanfall er riktig diagnose. *Modell 1* angir riktig diagnose som kolecystitt og tar inn grunnskjemaet. Det forlenger tiden anfallet har vart, legger til veggfortykkelse i galleblæren til det tidligere radiologisvaret, leukocytose og feber. Noen ganger legges det på lett økning i bilirubin. *Modell 3* inkluderer grunnskjemaet, men legger nå til synlig eller usynlig ikterus, feber, sterkere allmennpåvirkning m.m. Radiologisvarene får nå gallestase fra ductus choledocus + konkrement i denne. Noen ganger legges det også til økt pankreasamylase og lipase samt betennelse i pankreas på radiologisvarene. Bevegelsestrang fjernes da fra smertebeskrivelsen samt at smerteintensiteten økes. For noen pasienter legges også til bakteriemi med gram negative tarmbakterier.



Figur 4 - Eksempelmodell fra eksperimentet m/skjernbilder fra modelleringsverktøyet. Et grunnskjema som beskriver gallestensanfall gjenbrukes i tre modeller. I disse modellene legges ytterligere symptomer og funn til, slik at kasuistikken utvikler seg til kolecystitt, kolangitt/pankreatitt og ukomplisert gallestensanfall. Hver av modellene danner grunnlag for prosessuell generering av kasuistikker som beskrevet i figur 3.

Modelleringen viser et viktig poeng: Systemet har som mål at det er selve mønsteret man kan finne i svarene som avgjør diagnosen og behandlingen. Siden grunnskjemaet inkluderes til alle vil de fleste av svarene være delvis like mellom pasientene, og man må orientere seg etter det som er ulikt mellom modellene. Merk at verdirandomisering brukes hele veien. I dette tilfellet er det radiologifunn, blodverdier som bilirubin, leukocytose, feber, grad av smertepåvirkning og smertens karakter som skal avgjøre diagnosen.

Merk at modellene kun ble kvalitetsvurdert av medisinstudenter med erfaring på akutt abdomen, lærebøker og nettbaserte oppslagsverk.

## 2.4 Utprøving på studenter

Det ble generert 20 kasuistikker for hver av de seks modellene som inngikk i eksperimentet. Studenter fra 6. – 12. semester ved medisinsk fakultet i Oslo ble invitert via Facebook til å løse kasuistikkene via klienten beskrevet over. Oslo-studentene har fra og med 5. semester

hospitert i akuttmottak med både kirurgiske og indremedisinske pasienter. På 10. semester vil de fleste komme i kontakt med halssmerter gjennom 6 ukers praksis på allmennlegekontor. Imidlertid hadde ikke dette kullet vært i praksis når oppgaven ble gjennomført, og 11. og 12. semesterstudentene var de eneste med denne erfaringen. Studentene ble bedt om å se en instruksjonsvideo og dernest løse et fritt antall kasuistikker før de fylte ut et tilbakemeldingsskjema. De kunne velge om de ønsket en helt tilfeldig valgt kasuistikk<sup>2</sup>, alltid en ny diagnose eller at nye diagnoser gradvis ble lagt til. Skjemaet ble laget for å få et inntrykk av hva studentene følte mens de brukte systemet, om de oppfattet den medisinske kvaliteten som god, om de ville bruke systemet igjen, hvor i utdanningen de befant seg, klinisk erfaring og teknisk interesse.

Tabell 2 – Tilbakemeldingsskjema. Denne tabellen viser spørsmål som ble gitt til studentene når de evaluerte kasuistikkene og systemet som presenterte disse. Se vedlegg for skjermbilder av selve kasuistikk-løsningsprosessen.

Spørsmål	Svarform
Ord som beskriver min opplevelse av systemet	Valg: Interessant, Overveldende, Kjedelig, Spennende, Frustrerende, Forvirrende, Nyttig, Brukervennlig, Gøy, Lett, Vanskelig
Jeg opplevde ingen tekniske problemer	Valg: Svært uenig, Noe uenig, Verken enig eller uenig, Noe enig, Svært enig
Jeg var god til å løse kasuistikkene om halssmerter	---
Jeg var god til å løse kasuistikkene om magesmerter	---
Jeg ble bedre til å løse kasuistikkene etterhvert	---
Systemet ga meg alle muligheter til prøver, undersøkelser osv. jeg trengte	---
Den medisinske kvaliteten på innholdet var god	---
I møte med en pasient med tilsvarende symptomer kan jeg ha nytte av å ha brukt dette systemet	---
Hvilke tekniske eller medisinske feil plaget deg eventuelt mest?	Fritekst

<sup>2</sup> Under testing ble det funnet at å få en helt vilkårlig kasuistikk var lite gunstig. Siden 50% av kasuistikkene var halssmerter og de øvrige 50% magesmerter samt en 16% sannsynlighet for å få samme diagnose to ganger på rad, ble det mer givende for studenten om systemet vekslet mer systematisk mellom mulighetene eller sikret ny diagnose hver gang.

Et tilsvarende system med mer innhold vil gjøre det mye lettere å huske relevant kunnskap lest i lærebøkene	Valg: Svært uenig, Noe uenig, Verken enig eller uenig, Noe enig, Svært enig
Et tilsvarende system med mer innhold ville jeg benyttet igjen	---
Et tilsvarende system med mer innhold ville gjøre meg bedre rustet til å gjøre diagnostiske overveielser i klinikken	---
Et tilsvarende system med mer innhold ville jeg anbefalt til medstudenter	---
Et tilsvarende system med mer innhold burde benyttes i medisinstudiet	---
Hvilket semester ved medisinstudiet UiO er du på? *	Valg: 1. semester, 2. semester, .. 12. semester, Studerer ikke medisin ved UiO
Jeg benytter tradisjonelle kasuistikkbøker mye	Valg: Svært uenig, Noe uenig, Verken enig eller uenig, Noe enig, Svært enig
Jeg er tidlig ute med å ta i bruk nye digitale verktøy som kan hjelpe meg i utdanningen	---
Jeg har klinisk erfaring med pasienter med halssmerter	---
Jeg har klinisk erfaring med pasienter med akutt abdomen	---
Jeg har mye klinisk og praktisk erfaring med pasienter generelt	---
Jeg mestrer medisinstudiet godt	---

Systemet lå tilgjengelig på nett i 3 uker. Det ble gitt en påminnelse til studentene før den siste uken.

### 3 Resultater

Merk at store feilkilder trolig eksisterer i datagrunnlaget, noe som diskuteres senere.

Deltagelsen var lav med totalt 30 besvarelser fra minst 400 eksponerte. Blant disse dominerte svar fra 11. semester (14 av 30 besvarelser, 110 eksponerte). Rundt halvparten løste færre enn 4 kasuistikker (17 av 30). Se Figur 5- Grafisk fremstilling av . Rådata og analysesteg er tilgjengelig på nett<sup>3</sup>.

Blant ord som beskriver opplevelsen, dominerer positive adjektiver med «interessant», «nyttig», «gøy» og «spennende» hos over 75%. 63% av besvarelsene valgte «brukervennlig», mens kun 13% fant systemet «overveldende». Kun 17% svarte at systemet var «lett», og kun 7% mente det var «kjedelig».

Tabell 3 - Fordeling av adjektiver (fargekodet, grønn = positiv, gul = nøytral, rød = negativt). Studentene ble bedt om å krysse på ord som passet med deres opplevelse av systemet. Tabellen angir prosentvis hvor mange av besvarelsene som hadde adjektivet.

Adjektiv	% av besvarelsene
Interessant	87
Nyttig	87
Gøy	77
Spennende	77
Brukervennlig	63
Vanskelig	23
Lett	17
Frustrerende	13
Overveldende	13
Forvirrende	10
Kjedelig	7

For skalerte spørsmål med svar fra «Svært uenig» til «svært enig» er kategoriene her omkodet slik at 1 tilsvarer minst enig og 5 mest enig. Vi kan dele inn spørsmålene i tre grupper: (1)

<sup>3</sup> Innsamlet datasett og script benyttet til statistisk analyse er tilgjengelig på <http://folk.uio.no/knutsaa/permanent/data/cbox2015/>

vurdering av systemet, (2) vurdering av konseptet og (3) studentens egenvurdering. Tabell 4 - Svar oppsummert oppsummerer svarene og deres fordeling.

Tabell 4 - Svar oppsummert hvor 1 = svært uenig. 5 = svært enig.

Spørsmål n=30	$\mu$	$\sigma_x$	Min	25 %	50 %	75 %	Maks
<b>Vurdering av systemet</b>							
Jeg opplevde ingen tekniske problemer	4.2	1.2	2.0	4.0	5.0	5.0	5.0
Jeg var god til å løse kasuistikkene om halssmerter	3.4	1.2	1.0	2.0	3.5	4.0	5.0
Jeg var god til å løse kasuistikkene om magesmerter	3.0	1.0	1.0	2.0	3.0	4.0	5.0
Jeg ble bedre til å løse kasuistikkene etterhvert	3.9	0.9	2.0	3.0	4.0	4.8	5.0
Systemet ga meg alle muligheter til prøver, undersøkelser osv. jeg trengte	4.4	1.1	2.0	4.0	5.0	5.0	5.0
Den medisinske kvaliteten på innholdet var god	4.4	0.7	2.0	4.0	5.0	5.0	5.0
I møte med en pasient med tilsvarende symptomer kan jeg ha nytte av å ha brukt dette systemet	4.4	0.8	2.0	4.0	5.0	5.0	5.0
<b>Vurdering av konseptet</b>							
Et tilsvarende system med mer innhold vil gjøre det mye lettere å huske relevant kunnskap lest i lærebøkene	4.7	0.6	3.0	4.3	5.0	5.0	5.0
[...] med mer innhold ville jeg benyttet igjen	4.9	0.3	4.0	5.0	5.0	5.0	5.0
[...] med mer innhold ville gjøre meg bedre rustet til å gjøre diagnostiske overveielser i klinikken	4.6	0.6	3.0	4.0	5.0	5.0	5.0
[...] med mer innhold ville jeg anbefalt til medstudenter	4.8	0.5	3.0	5.0	5.0	5.0	5.0
[...] med mer innhold burde benyttes i medisinstudiet	4.7	0.6	3.0	4.0	5.0	5.0	5.0
<b>Studentens egenvurdering</b>							
Jeg er tidlig ute med å ta i bruk nye digitale verktøy som kan hjelpe meg i utdanningen	3.1	1.1	1.0	2.0	3.0	4.0	5.0
Jeg har klinisk erfaring med pasienter med halssmerter	3.5	1.6	1.0	2.0	4.0	5.0	5.0
Jeg har klinisk erfaring med pasienter med akutt abdomen	3.7	1.3	1.0	3.0	4.0	5.0	5.0
Jeg har mye klinisk og praktisk erfaring med pasienter generelt	3.7	1.1	1.0	3.0	4.0	4.0	5.0
Jeg mestrer medisinstudiet godt	4.1	0.9	1.0	4.0	4.0	5.0	5.0

De fleste (>50% svært enig, > 75% noe enig) angir at kvaliteten på det medisinske innholdet var god, at systemet ville ha overføringsverdi til klinikk og at de fikk et tilstrekkelig utvalg



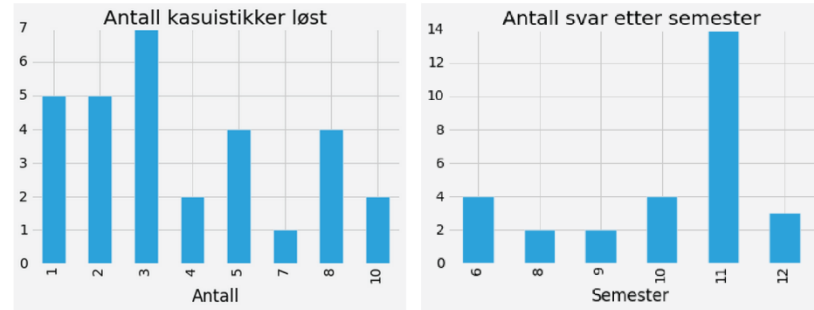
muligheter. Ingen svarte at de var noe eller meget uenige på disse spørsmålene. Det er større spredning i om studenten klarte å løse kasuistikkene. 50% mener de ble bedre til å løse de etterhvert. Tekniske problemer ser ikke ut til å ha vært en betydelig faktor.

Konseptet kommer tilsynelatende godt ut. Over 75% er svært enig i at et tilsvarende system med mer innhold ville de benyttet igjen og anbefalt til andre. Over 75% er noe enige og 50% svært enige i at et system som dette, med mer innhold, burde brukes i medisinsk utdanning, at det vil gjøre dem i bedre stand til å foreta diagnostiske overveielser i klinikken og at det vil gjøre det lettere å huske kunnskap fra lærebøker.

Studentene føler de mestrer medisinstudiet godt (> 75% noe eller meget enig). Ca. halvparten av studentene svarer de er noe enige i at de har mye klinisk og praktisk erfaring generelt, og under 25% sier seg noe uenig i dette.

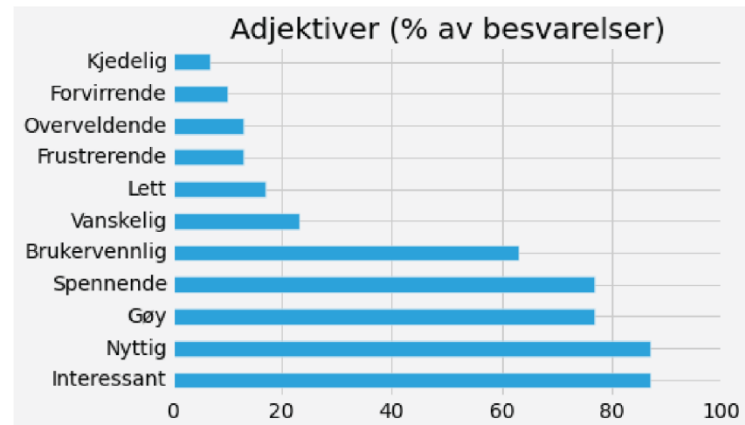
## Semesterfordeling og antall kasuistikker

n = 30



## Adjektiver som beskriver brukeropplevelsen

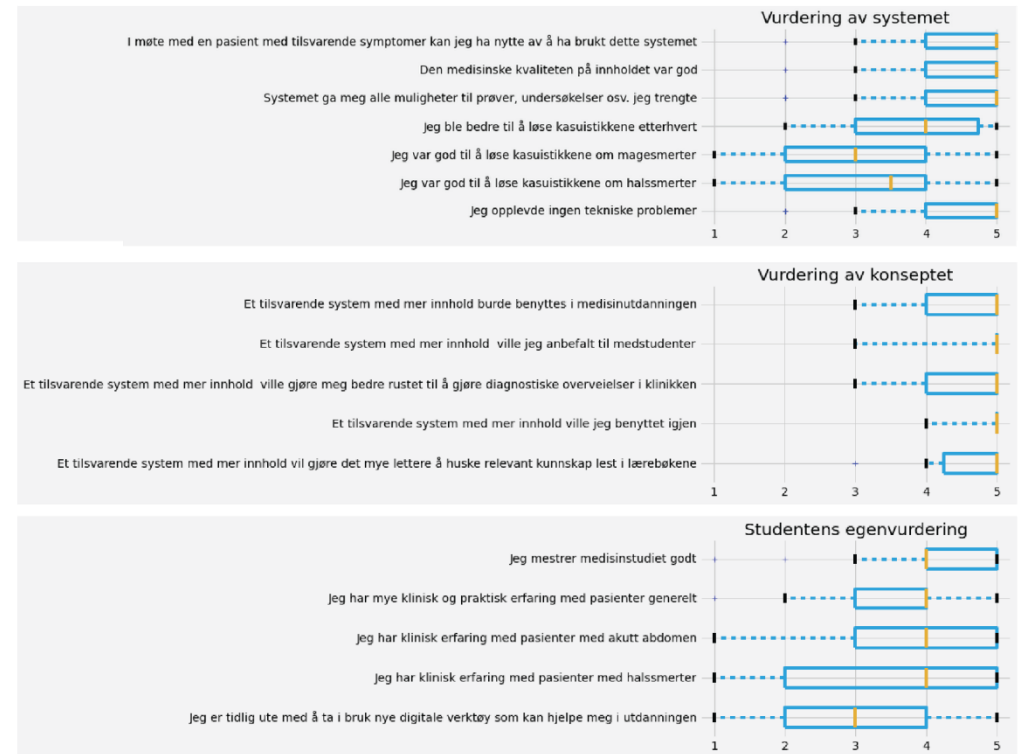
n = 30



## Svarfordeling

n = 30

Stiplede linjer angir maksimum og minimum  
Boksene representerer 25- og 75-percentilen  
Delelinjen i boksen er median



Figur 5- Grafisk fremstilling av resultater.

## 4 Diskusjon

Under utviklingen av systemet har det blitt klart at flytskjemamodellen har mange aspekter som kompliserer bruk innen medisinsk utdanning. Disse er viktige å ta i betraktning ved vurdering av bruk på større skala. Statistikken innsamlet har også endel svakheter den må tolkes i lys av.

### 4.1 Aspekter ved bruk av flytskjemamodeller

**Tidsbruk og læringskurve:** Å lage de tre modellene som ble beskrevet i metodekapittelet tok over en dag for oppgaveforfatteren som allerede kjente systemet godt. For en profesjonell utgave med bedre kvalitetssikring vil tiden trolig øke. Modelleringen krever naturlig nok også grundig innsikt i det som modelleres. Realistiske verdiintervaller for patologiske prøvesvar kan eksempelvis være vanskelige å finne. Et ubesvart spørsmål er om tidsbruket vil rettferdiggjøres av læringsutbyttet. Imidlertid vil gjenbruk gjøre at senere modeller kan utvikles raskere. Vedlikehold og korreksjoner over tid vil også kreve tid. Det er realistisk å tro at å modellere en medisinsk spesialitets mest sentrale diagnoser vil kreve måneders arbeid av en fageksperts tid. Årlig vedlikehold må også påregnes.

**«Røde hatter»:** Det er lett å innføre utilsiktede signaler i kasuistikkene som kan brukes til å gjenkjenne diagnosen. Hvis alle pasienter med hjertesvikt, og ingen andre, har på seg en rød hatt er diagnosen enkel. Utilsiktet introduksjon av slike «røde hatter» kan skje via skrivefeil i beskrivelser som er unike for en modell, eller måter å ordlegge seg på som ikke går igjen mellom kasuistikkene m.m. Dette gjør at prosessen med å lage modeller enten krever innsikt i hvordan man ordlegger seg i de andre, eller utstrakt gjenbruk av skjemaer. Eventuell utvidelse av systemet til også å inkludere bilde- og lydmateriell i kaustikker vil også vanskeligjøres av samme grunn hvis ikke kildemateriellet er stort nok til at samme bilde sjeldent møtes igjen. Verdirandomisering svekker denne effekten for tekst og tall da den samme røde hatten ikke oppstår hver gang.

Røde hatter vil også florere i et system med få modeller. Hvis man kun modellerte et svært begrenset utvalg tilstander, for eksempel kun en enkelt modell som lager følelse av instabilitet i kneet, vil dette også umiddelbart avgjøre diagnosen for en student som har løst kasuistikker fra modellen noen ganger tidligere. All annen variasjon innført innad i samme modell vil

derfor bli irrelevant og sløsing med tid å legge inn hvis det er diagnosen som er poenggivende.

**Gjenbruk krever innsikt:** Gjenbruk krever at den som modellerer har kjennskap til skjemaet som inkluderes. I mononukleoseeksempelet fra metodekapittelet må man vite at skjemaet for barnesykdommer inkluderer gjennomgått mononukleose med Epstein-Barr-virus og sørge for å ekskludere de relevante byggebanene slik at pasienten ikke beskrives å ha fått sykdommen på nytt, og gir selvmotsigende blodverdier for IgG/IgM antistoff. Modelleringsverktøyet kan trolig gi støtte i dette ved å informere bruker om det som inkluderes og eventuelle kollisjoner.

**Gjenbruk svekkes av medisinsk kompleksitet:** Gjenbruk danner en liten grad av *nettverkseffekt* – jo flere gjenbrukbare skjemaer i systemet, jo større verdi får de andre da de i teorien kan brukes sammen for å lage mer sammensatte og realistiske kasuistikker. Samtidig vil kompleksiteten innen det medisinske fagfeltet raskt dempe denne goden. Når flere medisinske tilstander finnes samtidig hos en pasient kan resultatet være uforutsigbart. Å forutsi hvordan tilstandene i felleskap påvirker ulike verdier er vanskelig. Dette vanskeliggjør tilfeldig valgte kombinasjoner av sykdommer hos pasienter da interaksjonene mellom sykdommene også må modelleres. Dette begrenser mulighetene for denne typen sammensatte pasienter betydelig da antall kombinasjoner som må tas høyde for raskt blir uoverkommelig. På samme måte legger dette demper på mulighet for polyfarmasi.

Sammensatte pasienter kan alternativt lages ved å legge inn relevante tidligere sykdommer som komplikasjoner i modellene. Eksempelvis kan inklusjon av kjent magesår hos smertepåvirket pasient være relevant med tanke på hvilke smertestillende medikamenter studentene bør gi. Her må man imidlertid passe seg for røde hatter. Hvis kun en eller to modeller har magesår i listen over tidligere sykdommer avslører man modellens diagnose billig ved at dette blir en diagnostisk markør.

Et alternativ man kan forsøke, er å utvide programvaren til å ta hånd om kollisjoner mellom sykdommer via regler for hva som skjer hos en pasient som har de ulike kombinasjonene. Dette utforskes ikke i oppgaven, men kan tenkes å være et alternativ fordi mange sykdommer trolig har få interaksjoner. Hvis man modellerer de 10 vanligste folkesykdommene og begrenser pasientene til å kun ha maksimalt to får vi  $\binom{10}{2} + \binom{10}{1} = 55$  kombinasjoner som må vurderes for hver modell man lager, og en gang for interaksjoner i kombinasjonene selv. Hvis antallet regler som må lages blant disse 55 er lavt fordi få faktisk interaksjoner finnes, kan

dette kanskje være mulig, men krevende. Antallet kombinasjoner øker imidlertid raskt, som for ved 2 eller 1 tilfeldige fra 50 sykdommer, som gir  $\binom{50}{2} + \binom{50}{1} = 1225$ . Å skulle ta stilling til interaksjoner med 1225 kombinasjoner av sykdommer hver gang man lager en modell er helt urealistisk.

En bedre løsning vil trolig være en liste tidligere sykdommer pasientene kan ha og for hver kasuistikk velge ut maksimalt en. Man kan definere, for eksempel 50 vanlige tilstander og deres behandling (hypertensjon, nyresvikt, leversvikt osv.) og si at alle modeller må ta disse i betraktning og justere for hver enkelt, men ikke kombinasjoner av dem. Dette er kanskje ikke uoverkommelig, men vil medføre endel ekstra arbeid pr. modell.

**Poenggivende kriterier:** Kasuistikkene bør kunne løses for å sikre mestringsfølelse, spesielt hvis man skal dra nytte av studentens konkurranseinstinkt for å gi motivasjon. I dette eksperimentet ble det gitt mindre poeng for uspesifikk diagnose, og full pott for spesifikk. Eksempelvis 5 poeng for K85.9 – uspesifisert akutt pankreatitt og 10 poeng for K85.1 – biliær akutt pankreatitt. I praksis står man fritt til å velge om man vil legge seg på en modell hvor kun spesifikk diagnose får poeng, hvor uspesifikk diagnose gir like mye eller hvor man for eksempel kun gir poeng for korrekt behandling. Man tester imidlertid ulike ting. Ved spesifikk diagnose kreves også mulighet for å komme frem til denne, og det kan innebære at man presser frem det som ville representere et dramatisk overforbruk av diagnostiske verktøy hos en virkelig pasient. Dette er kanskje akseptabelt hvis man ønsker å teste studentens forståelse av behandling og diagnostikk på tvers av spesialiteter og spesialistnivåer.

## 4.2 Diagnostisk verdi av ulike symptomer og funn

Med de seks modellene brukt under utprøving på studenter blir pretestsansynlighet for kolangitt 16.6% (1 av 6 modeller). Ved magesmerter økte dette til 33% (1 av 3 modeller) og ved magesmerter + feber til 50% (1 av 2 modeller). På samme vis var synlig ikterus patognomonisk for kolangitt siden ingen andre kasuistikker inneholder dette symptomet. Prediktive verdier og sannsynlighetsratioer svarer altså ikke til virkeligheten. Sensitivitet og spesifisitet var heller ikke satt opp til å være realistisk. Spørsmålet er om denne skjevheten er viktig eller ikke, og i hvilken grad. I virkeligheten varierer pretestsansynlighet og prediktive verdier mye etter hvor i helsesystemet man praktiserer. I et e-læringsystem gir det mening å variere prevalens av tilstander i læringsøyemed, eksempelvis ved å gradvis innføre nye

diagnoser til studenten, eller kun plukke diagnoser fra emnet som studeres på et gitt tidspunkt. Det kan derimot være et mål å etterstrebe realistisk sensitivitet og spesifisitet for å sikre at systemet beveger seg mot forhold som ligner på virkeligheten. Gitt at en test for en sykdom er  $SP =$  sann positiv,  $FN =$  falsk negativ,  $SN =$  sann negativ eller  $FP =$  falsk positiv<sup>4</sup>:

$$\text{sensitivitet} = \frac{SP}{SP + FN} = \frac{P(\text{positiv})}{P(\text{syk})}$$

$$\text{spesifisitet} = \frac{SN}{SN + FP} = \frac{P(\text{negativ})}{P(\text{ikke syk})}$$

Hvis en student ber om å få en serie kasuistikker med kun en spesifikk diagnose er prevalens av denne sykdommen 1.0, FN 0, FP 0. For at variasjonen i kasuistikkene innen dette grensetilfellet sammen skal danne et virkelighetsnært bilde av pasientpopulasjonen de etterligner må derfor sensitivitet og spesifisitet også være realistiske. Hvis man ønsker å oppnå korrekte verdier for disse må dette justeres for i byggebanene til en modell, et aspekt ikke utforsket i denne oppgaven grunnet tidsbegrensinger. Imidlertid kan man tenke seg at det kan finnes læringsverdi i å øke antallet i en syk populasjon som har for eksempel et spesifikt men sjeldent symptom. Hvis symptomet opptrer for sjeldent vil ikke studenten i stor grad møte det selv om det kan være viktig å vite om det. Det kan være klokt å tillate en viss grad av slik karikering av et sykdomsbilde, men studenten bør gjøres klar over dette. Man kan altså velge både å øke prevalensen av sykdommer selektivt, og styrke eller svekke prevalensen av et symptom innenfor pasienter med samme sykdom.

Faktiske verdier for disse målene vil være tidkrevende å oppdrive, og man bør trolig basere seg på erfarne klinikers inntrykk av kasuistikkene som kommer ut av modellene. Et analyseverktøy med mulighet for å visualisere og direkte justere all variasjon en modell lager kan trolig hjelpe.

### 4.3 Nivå av helsetjenesten

Hvilket nivå av helsetjenesten man forsøker å etterligne er fleksibelt. For oppgaven ble det valgt å ikke ta hensyn til dette. Alle muligheter er derfor umiddelbart tilgjengelige. Dette kan

---

<sup>4</sup> Formler etter (Bjørndal, 2007)

være en god og enkel strategi som tester kunnskap fra flere nivåer av helsetjenesten. Et system kan imidlertid også lett konstrueres for å simulere et spesifikt nivå ved å redusere antall spørsmål tilgjengelig. En variant vil være et system hvor mulige spørsmål avhenger av hvilken spesialitet man ”henviser” pasienten til, og hvor man starter med spørsmål tilgjengelig i allmennpraksis osv. Poenggivning kan også baseres på hvor man valgte å legge inn pasienten hvis man ønsker å fokusere på dette.

## 4.4 Feilkilder i statistikken

Det var lav svarprosent i undersøkelsen, og resultatene er sterk forskjøvet mot 11.semester. Det er grunn til å tro at mange deltok fordi de kjente forfatteren. Det er også rimelig å anta at de som svarte kan være de som er tidligst ute med nye tekniske løsninger. Populasjonen undersøkt er heller ikke spesielt erfaren med reelle pasienter. Utprøving på erfarne klinikere kunne gitt en helt annen vurdering av den medisinske kvaliteten på innholdet og utvalget av muligheter. Det siste er spesielt relevant fordi modellene ikke ble kvalitetsvurdert eller laget av klinikere med erfaring fra det aktuelle medisinske området.

Undersøkelsen danner kun en analyse av førsteinntrykk og har sannsynligvis liten prediktiv verdi for å vite om deltageren ville fortsette å benytte systemet. Usikkerheten styrkes av at mange svarte med adjektivet «spennende» og «interessant» i beskrivelsesspørsmålet hvilket kan tolkes som at det hele farges av å være en ny erfaring. I en undersøkelse etter en periode med jevn bruk vil neppe «gøy» og «spennende» ha samme utslag. I sum er derfor trolig svarene forskjøvet i positiv retning, og kan vanskelige generaliseres til hele studentpopulasjonen eller langsiktig bruksmønster. Den lave svarprosenten gjør det også vanskelig å si noe om variasjon mellom semester, klinisk erfaring og antall kasuistikker løst. Imidlertid ville eksperimentet sannsynligvis plukket opp eventuelle grove grunnleggende feil og mangler i innhold og brukeropplevelse.

Spørsmålene i skjemaet var noe tvetydige. «Klinisk erfaring med pasienter» er åpen for tolkning og skiller eksempelvis ikke mellom pleie- og behandlingserfaringer.

Resultatene indikerer at å ha svært mange tilgjengelige valg med kjent ordlyd ikke var overveldende. Det er lite sannsynlig at dette ville forverre seg over tid da trenden er at erfaring med et system gjør det lettere å bruke.

## 5 Konklusjon

Prosessuell generering med flytskjemamodell som beskrevet i metodekapittelet fungerer og kan benyttes til å lage medisinske kasuistikker for mange diagnoser. Et uavklart element er grad av variasjon mellom disse kasuistikkene, og følgelig grad av mengdetrening og trening i mønstergjenkjenning et slikt system muliggjør. Multimorbiditet er utfordrende grunnet interaksjoner mellom sykdommer og pasienter med mange tidligere sykdommer er trolig urealistisk å generere av ressurshensyn.

Konseptet kommer godt ut ved testing på studenter, men det er stor risiko for seleksjonsskjevhet og andre betydelige feilkilder i vår undersøkelse. Studentenes opplevelse ved bruk var trolig preget av at systemet var nytt og spennende. De oppga at de ville benytte et tilsvarende system med mer innhold igjen og mente det ville gi overføringsverdi til klinisk hverdag. De mente også at et system som dette kan forsterke effekten av tradisjonell læring. De oppfattet kun i liten grad mange valg med kjent ordlyd som overveldende eller forvirrende.



## 6 Videreutvikling

Å utvikle et komplett system for prosessuell generering og befolke det med modeller vil være en resurskrevende oppgave både initielt og over tid. Imidlertid kan læringsutbyttet av et slikt system være stort. Flere nyttige elementer for studenter kan legges til, som eksempelvis dosering av medisiner, bilde- og lydolkning, pasienter som blir dårligere underveis m.m. Ved siden av det rent diagnostiske ble studentene som testet dette systemet eksponert for handelsnavn og navn på virkestoff i ulike medisiner (FEST-data), samt ICD-10 diagnoser. Dette, spesielt i kombinasjon med doseringer, kan være nyttig forberedelse til nybegynnerstillingene og klinisk arbeid med studentlisens. Et system som beskrevet her kan øke eksponeringen for disse elementene som det erfaringsmessig er vanskelig å pugge seg til. Målet er slik sett både å gi økt trygghet på egen kunnskap og kanskje redusere behovet for oppslag av basale ting.

Systemet vil gi best utbytte av tiden investert i utvikling og bruk hvis det inneholder flere mulige diagnoser med like symptombilder. Man kan tenke seg at man rullet ut diagnoser over noe tid, men læringskurven knyttet til modelleringsverktøyet og behovet for å ha kjennskap til hva annet som er modellert i systemet vil kreve en viss kontinuitet. Skal den konkrete programvaren skrevet for denne oppgaven brukes vil dette kreve at den videreutvikles noe, og eventuelt utvides med et analyseverktøy for å justere realismen i modellene. Den største utfordringen resursmessig vil trolig være dette, samt å finne, lære opp og lønne fagpersoner som kan legge ned tilstrekkelig arbeid i å utvikle en god katalog av modeller og vedlikeholde denne på sikt.

Videre utredning av prosessuell genererings styrker og svakheter på det medisinske e-læringsfeltet er å anbefale. Det kan tenkes at alternative informasjonstekniske metoder kan løse utfordringene beskrevet i denne oppgaven, men de vil neppe aldri helt unnslippe den store kostnaden knyttet til utvikling av modeller og den grunnleggende kompleksiteten det medisinske fagfeltet innehar.

# Litteraturliste

- Bjørndal, A. (2007). Hvordan avgjøre hva pasienten feiler. I A. Bjørndal, *Tenke- og arbeidsmåter i medisinen* (ss. 91-120). Oslo: Gyldendal Norsk Forlag AS.
- Duffy, J., & Wang, Z. (2015). Application of Procedural Generation as a Medical Training Tool. *International Conference on Health Informatics and Medical Systems*, (ss. 223-228).
- Eno, B. (1996). Evolving metaphors, in my opinion, is what artists do. *Imagination Conference*. San Fransico.
- Khatchadourian, R. (2015, Mai 18). *World Without End*. Retrieved from The New Yorker: <http://www.newyorker.com/magazine/2015/05/18/world-without-end-raffi-khatchadourian>
- Smith, H. (2008). *Procedural Trees and Procedural Fire in a Virtual World*. Hentet fra Intel: <https://software.intel.com/sites/default/files/Procedural-Trees-and-Procedural-Fire-Wp.pdf>

# Vedlegg – Skjermbilder

## Skjermbilder fra kasuistikkløsning (1 av 2)

0:00:30 | PAS, KOMFORTABEL, 0.00% RISIKO | 0-, NOK

**Gutt, 18, hoste, vondt i halsen.**

Anamnese

**Halsmerter:** Smerteretninger: VAS 2. Lokalisert til bakre del av svelget. Sår og øm.

**Klinisk undersøkelse**

Generell inspeksjon/inntrykk: Våken, klar. Huden er varm og lett klam. Ikke generalisert glandersvulst. Ikke bleik. Orienter for tid og sted. Alimemtilstand noe redusert. Ikke generaliserte ødemer. Samarbeider godt. **Glasgow coma scale:** GCS 15.

UTFØR PLAN | FIERN VALGT

**Muligheter**

ANAMNESE | PSYKIATRISK

KLINISK UNDERSØKELSE

RADIOLOGISKOPI | PROSEDYRER

MIKROBIOLOGI | KLINISK-BIOKEMISK

MEDIKAMENTTERUSMIDLER | SØK

DIAGNOSE OG BEHANDLING

Todo

Ingen handlinger er valgt. Velg fra skjemaene nedenfor, eller søk.

Steg 1: Systemet presenterer en kort kasuistikk. Denne er datagenerert basert på modellen, og er i prinsippet noe ulik hver gang. Målet er at studenten må basere seg på mønstergjenkjenning.

0:01:24 | PAS, KOMFORTABEL, 0.00% RISIKO | 0-, NOK

**Anamnese**

Generell inspeksjon/inntrykk: Våken, klar. Huden er varm og lett klam. Ikke generalisert glandersvulst. Ikke bleik. Orienter for tid og sted. Alimemtilstand noe redusert. Ikke generaliserte ødemer. Samarbeider godt. **Glasgow coma scale:** GCS 15.

UTFØR PLAN | FIERN VALGT

**Muligheter**

ANAMNESE | PSYKIATRISK

KLINISK UNDERSØKELSE

RADIOLOGISKOPI | PROSEDYRER

MIKROBIOLOGI | KLINISK-BIOKEMISK

MEDIKAMENTTERUSMIDLER | SØK

DIAGNOSE OG BEHANDLING

Todo

Ingen handlinger er valgt. Velg fra skjemaene nedenfor, eller søk.

Steg 2: Studenten velger anamnestiske spørsmål, undersøkelser og tester han eller hun mener å trenge for å komme nærmere diagnosen.

0:02:59 | PAS, KOMFORTABEL, 0.00% RISIKO | 88-, NOK

**Gutt, 18, hoste, vondt i halsen.**

Anamnese

**Halsmerter:** Smerteretninger: VAS 2. Lokalisert til bakre del av svelget. Sår, var ghet, utvikling: VAF 5 dager.

**Klinisk undersøkelse**

Generell inspeksjon/inntrykk: Ikke generalisert glandersvulst. Alimemtilstand noe redusert. Ikke bleik. Samarbeider godt. Huden er varm og lett klam. Orienter for tid og sted. Våken. Klar. Ikke generaliserte ødemer. **Glasgow coma scale:** GCS 15.

UTFØR PLAN | FIERN VALGT

**Muligheter**

ANAMNESE | PSYKIATRISK

KLINISK UNDERSØKELSE

RADIOLOGISKOPI | PROSEDYRER

MIKROBIOLOGI | KLINISK-BIOKEMISK

MEDIKAMENTTERUSMIDLER | SØK

DIAGNOSE OG BEHANDLING

Todo

Ingen handlinger er valgt. Velg fra skjemaene nedenfor, eller søk.

Steg 3: Systemet gir umiddelbart svar. Studenten har nå lært noe mer om pasienten. Tid og penger brukt er oppdatert. Ny informasjon merkes med gult.

0:03:31 | PAS, KOMFORTABEL, 0.00% RISIKO | 88-, NOK

**Klinisk undersøkelse**

Generell inspeksjon/inntrykk: Ikke generalisert glandersvulst. Alimemtilstand noe redusert. Ikke bleik. Samarbeider godt. Huden er varm og lett klam. Orienter for tid og sted. Våken. Klar. Ikke generaliserte ødemer. **Glasgow coma scale:** GCS 15.

UTFØR PLAN | FIERN VALGT

**Muligheter**

ANAMNESE | PSYKIATRISK

KLINISK UNDERSØKELSE

RADIOLOGISKOPI | PROSEDYRER

MIKROBIOLOGI | KLINISK-BIOKEMISK

MEDIKAMENTTERUSMIDLER | SØK

DIAGNOSE OG BEHANDLING

Todo

Ingen handlinger er valgt. Velg fra skjemaene nedenfor, eller søk.

Steg 4: Studenten kan igjen velge å innhente ny informasjon. Dette kan gjøres så mange ganger det oppfattes nødvendig.

## Skjermbilder fra kasuistikk-løsning (2 av 2)

00:08:42 PAS, KOMFORTABEL, 0.00% RISIKO | 236,- NOK

**Gutt, 18, hoste, vondt i halsen.**

**Historie**  
 Refert seg til syk, vondt i hals, ikke sveigeværsket, egermitt feber/feberfølelse. Vært hjemme til 33.12. Opprager, ingen hoste. Ikke anod, lukt, klart sputum, kommet tilbake dagen før. Tidligere sykdommer: Allergisk nitt som skyldes pollen, Uspesifisert mononukleose tidligere barnealder. Helsestiller Medisinerer, ikke tidligere brukt medisin. Bruker ingen medisiner nå. (i halsområdet, Smermedisiner). Lokalisert til bakre del av svelget. Sår og am, VAS 2. Sært, var ghet, utvikling, Vart 5 dager.

**Relevante funn**  
 Generell inspeksjonsinntrykk: Ikke lukt, Våken klar. Smerterfølelse gjer. Halsen er varm og lett å røre. Ikke generalisert ganbeul. Ikke generalisert ødemer. Allmenntilstand noe redusert. Orienter for tid og sted. Glasgow coma scale: G:5 V:5 R:5. Puls: regelmessig, 66 /min. Respiration: ubesværet, 12/min. Temperatur: 36.6 °C. Collum u.s. Ingen stenosyber over karotidene, eller ausing over thyroidea. Normale konturer. Ingen palpable oppfyllinger eller tumorer. Hållrumm inspisjon insipatelt. Ikke inflammerte tonsiller. Bakre sveigeværsk etymearas. Egne tenner. Normalt faget ålthinne. Fullstg ålthinne. Lunge u.s. Vesikulær respirasjonstyd. Sp. vete grove pipeyder høyre side. Smerer perikardialt uten dekkning. Normale lungegrenser.

**Relevante**  
 StreptA-tungstest bakt: **Negativ**, Monospot test: **Negativ**.

**Todo**  
 Ingen handlinger er valgt. Velg fra sjernemne-netanfor, eller søk.

UTPØR PLAN | FERN VALGT

**Muligheter**

ANAMNESE | PSYKIATRISK  
 KLINISK UNNBSØKELSE  
 RADIOLOGISKE | PROSEDYRER  
 MIKROBIOLOGI | KLINISKE OBJEKTIV  
 MEDIKAMENTER, RUSMIDLER | SØK

DIAGNOSE OG BEHANDLING

Steg 5: Når studenten har innhentet informasjon over flere runder og mener diagnosen kan stilles, velger han eller hun "Diagnose og behandling".

0:09:11 PAS, KOMFORTABEL, 0.00% RISIKO | 236,- NOK

**TILBAKE**

**Hva er det pasienten feiler?**

Søk...

Uspesifisert akutt faryngitt (002.9)

HEMMEBILDE

**Hvilke tiltak vil du vurdere?**

Søk...

Når ikke gjenestetilstand, dannes satsingen på poeng, helning, utredning fra

Avventer/ingen behandling

FERN MERKEDE

FULLFØR OG GÅ VIDERE

Steg 7: Systemet spør om mulige tiltak. Her velger studenten å avvente. Data hentes fra Legemiddelverkets FEST-database m.m.

00:09:02 PAS, KOMFORTABEL, 0.00% RISIKO | 236,- NOK

**OK**

**Finn diagnoser, diagnosekoder**

Utskrift av diagnosekoder fra Legemiddelverket (FEST)

**faryngitt**

Genetokkalfaryngitt (A54.5)

Vesikulær faryngitt som skyldes enterovirus (B08.5)

Akutt nasale faryngitt (000)

Streptokokkfaryngitt (002.0)

Akutt faryngitt som skyldes annen spesifisert mikroorganisme (J2.8)

Uspesifisert akutt faryngitt (002.9)

Akutt faryngitt (005.0)

Kronisk nasalfaryngitt (J11.1)

Kronisk faryngitt (J11.2)

Steg 6: Systemet spør om diagnose. Dette søkes opp via ICD-10.

**Resultat**

**83 % korrekt**

- Du fikk 20 av maksimalt 24
- Feilregning for denne kasuistikk:
  - 2 poeng: "Når en eller flere av (spesifisert akutt infeksjon i nese, luftveier)
  - 10 poeng: "Hvis en eller flere av (spesifisert akutt faryngitt eller akutt nasalfaryngitt)
  - 10 poeng: "Hvis en eller flere av (avventende behandling)
  - Bløtge skal svar være: 3-5 dager"

**Kommentarer**

- Kunne fått 2 eksampelpoeng for å ha utvalgt utspesifisert med 2 kriterier (falsk) (fått på fargene)

**Deg vs. andre**

Meg: **80%**

Tid: 00:03:17 605:47

Periode: 26-09-2020

Risiko: 0

© 2020 Legemiddelverket. Alle rettigheter reservert. Dette er en del av Legemiddelverkets informasjonssystemer. For mer informasjon, se legemiddelverket.no

NY KASUISTIKK | EVALUERINGSHEMMA

Steg 8: Studenten mottar en poengsum basert på handlinger utført, diagnoser stilt og de tiltakene som ble foreslått. Resursene brukt sammenlignes med andre som har løst tilsvarende kasuistikker.