

CHALMERS



GÖTEBORGS UNIVERSITET

Matematiska modeller av läkemedelsprojekt

Känslighetsanalys som verktyg i beslutsfattning

Examensarbete för kandidatexamen i matematik vid Göteborgs universitet

Kandidatarbete inom civilingenjörsutbildningen vid Chalmers

Abraham Deniz

Elijah Ferreira

Erik Johansson

Hanna Johansson

Jacob Lindbäck

Linus Wiskman

Institutionen för matematiska vetenskaper

Chalmers tekniska högskola

Göteborgs universitet

Göteborg 2017

Matematiska modeller av läkemedelsprojekt

Känslighetsanalys som verktyg i beslutsfattning

Examensarbete för kandidatexamen i matematik vid Göteborgs universitet

Abraham Deniz Elijah Ferreira

Kandidatarbete i matematik inom civilingenjörsprogrammet Bioteknik vid Chalmers

Hanna Johansson

Kandidatarbete i matematik inom civilingenjörsprogrammet Teknisk fysik vid Chalmers

Linus Wiskman

Kandidatarbete i matematik inom civilingenjörsprogrammet Teknisk matematik vid Chalmers

Erik Johansson Jacob Lindbäck

Handledare: David Bolin Matematiska vetenskaper
 Magnus Ytterstad Captario

Examinator: Maria Roginskaya
 Marina Axelson-Fisk

Institutionen för matematiska vetenskaper
Chalmers tekniska högskola
Göteborgs universitet
Göteborg 2017

Populärvetenskaplig presentation

Kostnaden för att ta fram ett nytt läkemedel har tredubblats under de senaste 20 åren, samtidigt som endast en liten andel av alla upptäckta läkemedel når marknaden. Detta beror huvudsakligen på att många nya läkemedel inte klarar av de höga krav som ställs. Det är därför viktigt att veta hur läkemedelsutveckling kan effektiviseras. Vi har med hjälp av matematik undersökt om hur man kan åstadkomma just detta.

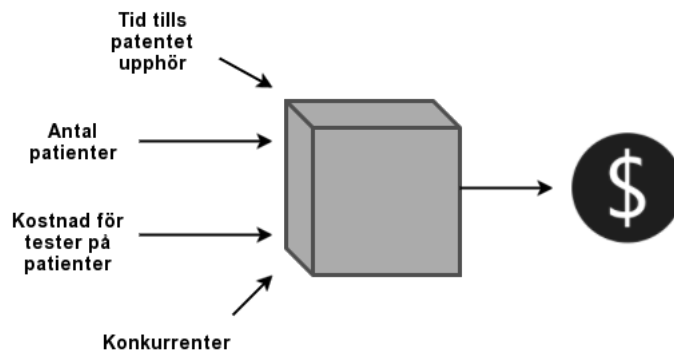
Läkemedelsutveckling

Läkemedelsutvecklingen innefattar hela processen från upptäckten av ett nytt läkemedel till att det börjar säljas på marknaden. I processen utför läkemedelsföretaget ett flertal tester för att undersöka om läkemedlet är säkert och effektivt. Det är dessa tester som tar lång tid och är kostsamma. Efter varje test måste läkemedelsutvecklaren bestämma om det är värt att fortsätta med kommande tester, eller om risken att projektet inte blir lönsamt är för stor. Även om ett läkemedel klarar alla tester så finns det andra anledningar som kan påverka lönsamheten av projektet. Exempelvis är lanseringstiden betydelsefull. Om någon konkurrent börjar sälja ett liknande läkemedel först så riskerar man att få väsentligt mycket lägre försäljningsvolym. Detta kan i värsta fall resultera i att man inte täcker upp kostnaderna för framtagningen av läkemedlet. Dessutom måste vinsterna för lyckade projekt täcka förlusterna av de misslyckade, vilket drar upp vinstkraven. Därför är det viktigt att identifiera vad som är viktigt för att få upp vinsten i läkemedelsprojekt. Det är här matematiken kommer in.

Matematiska modeller

Med matematikens hjälp kan man ställa upp *matematiska modeller* för utvecklingen av läkemedel. När man skapar en modell försöker man efterlikna verkligheten för att kunna förutspå vad som kan hända. Vidare kan en modell ses som en förenkling av någonting komplicerat, till någonting mer överskådligt. Rent matematiskt innebär detta att man försöker uttrycka det man är intresserad av med hjälp av en formel, som tar in värden som har att göra med det man försöker beskriva. Dessa värden kallar vi för variabler. En variabel skulle kunna vara hur många konkurrenter som utvecklar liknande läkemedel, eller hur mycket ett test kostar. Ett klassiskt sätt att beskriva denna formel är som en svart låda. Den svarta lådan tar in flera variabler, och ger ut exempelvis vinsten på ett projekt. Problemet är att man i allmänhet

inte vet i förväg om t.ex. hur många konkurrenter man har, eller hur höga testkostnaderna kommer bli. Därför måste man göra flera kvalificerade gissningar, och beräkna med hjälp av formeln vinsten för varje gissning. Matematiken förser oss med ett systematiskt sätt att göra detta. Med hjälp av så kallade *Monte Carlo*-metoder kan en förväntad vinst fås fram på ett effektivt sätt genom datorberäkningar.



En bild på en matematisk modell för ett läkemedelsprojekt. Den stora lådan är vår modell.

Captario och vårt arbete

Naturligtvis kan en matematisk modell för ett läkemedelsprojekt se ut på flera olika sätt. Företaget Captario har tagit fram ett nät-baserat verktyg där användaren själv kan skapa en modell och använda Monte Carlo-metoder för att uppskatta exempelvis förväntad vinst. Det är där vi kommer in i bilden. I vårt arbete har vi gjort en undersökning av vilka variabler som är viktigast för att driva upp vinsten på ett läkemedelsprojekt. De mest informativa metoderna kan vara väldigt tidskrävande att genomföra, även om snabba datorer finns tillgängliga. Dock sparas beräkningstid om modellen har färre variabler. Därför började vi med att sortera bort variabler som inte påverkade vinst tillräckligt mycket med hjälp enklare metoder. Efter att vi hade färre variabler gjorde vi en undersökning av vad som spelade störst roll i vad man får för vinst med hjälp av de mer sofistikerade metoderna. Ett tydligt resultat vi fick var att det allra viktigaste, ur lönsamhetsperspektiv är att se till att läkemedlet lanseras så tidigt som möjligt. En spontan tanke är kanske att det är kostnaden som kommer vara avgörande för att garantera vinst. Skälet till att tiderna är såpass viktiga är för att försäljningsvolymen sjunker om konkurrenterna lanserar läkemedlet först. Om man då inte hade haft hjälp av matematiken för att ta reda på detta, skulle man möjligtvis kommit fram till helt fel lösningar på sitt problem.

Sammanfattning

Läkemedelsutveckling styrs idag av att aktörer investerar i de läkemedelsprojekt som tros kunna ge stor avkastning. Hela processen från att ha hittat ett potentiellt läkemedel till eventuell försäljning på marknaden innefattar en rad olika studier och kontroller med höga kostnader. Dessutom svarar de allra flesta läkemedelsprojekt som genomgår denna process ej mot de höga kraven som ställs på nya läkemedel och läggs därmed ned. Detta resulterar sedermera i att läkemedelsföretaget inte får tillbaka det investerade kapitalet. Mot denna bakgrund har företaget Captario utvecklat ett nätbaserat verktyg, Captario SUM. I denna programvara kan användaren konstruera en matematisk modell över sitt läkemedelsprojekt och simulera utfallet för att erhålla en prognos över lönsamheten och andra intressanta aspekter av projektet.

Vi har undersökt vilken betydelse variationer i variablerna till en sådan modell har på prognosen som erhålls. Dessutom har vi begränsat oss till att enbart fokusera på prognosen över *nettonvärde*, som är ett mått på lönsamhet. Till vår hjälp använder vi oss utav metoder inom fältet *känslighetsanalys* som har implementerats på en representativ modell av ett läkemedelsprojekt. Ett sekundärt mål under projektets gång har varit att försöka hitta lämpliga metoder som Captario i framtiden ska kunna integrera med sitt verktyg. I förlängningen vill Captario kunna erbjuda sina kunder en känslighetsanalys av sina modeller.

Vi har efter denna undersökning hittat flera känslighetsanalysmetoder som lämpar sig väl för Captarios verktyg. En genomgående trend i samtliga metoder är att rekryterings-hastigheten för en av de större studierna var den variabel som hade störst betydelse för lönsamheten. Vidare kan denna rapport ses som ett första steg i analysen av läkemedelsmodeller, och det diskuteras hur man skulle kunna gå vidare med en djupare analys.

Abstract

Drug development is today controlled by investing in projects or drugs that are believed to provide a big return on investment. The entire process of finding a potential drug until it is possibly sold on the market involves a variety of tests and controls resulting in huge expenses. The vast majority of drug projects undergoing this process do not meet the high demands set on new drugs and is therefore discontinued. This subsequently results in a loss for the pharmaceutical company in that project. It is against this background that Captario has developed a cloud-based tool, Captario SUM. In Captario SUM, an user is able to construct a mathematical model over their pharmaceutical project and simulate the outcome to obtain a forecast of profitability and other interesting aspects of the project.

In this report we investigate the impact variations in the variables, of such a model, have on the forecast that is obtained. We have chosen to focus only on the forecast of *net present value* which is a measure of profitability. To our help, we use methods in the field of sensitivity analysis which have been implemented on a prototypical model of a pharmaceutical project. A secondary goal during the project has been to try to find suitable methods that Captario in the future will be able to integrate with the tool and further offer their customers a sensitivity analysis on their models.

After this investigation we found various sensitivity analysis methods that are well suitable for Captarios tool. All methods identified that the recruitment rate for one of the larger studies was the variable that had the greatest impact for profitability. Furthermore, this report can be seen as a first step in the analysis of drug development models and it is discussed how one could proceed with a deeper analysis.

Förord

Vi vill tacka Captario för att ha gett oss möjligheten att fördjupa oss i detta problem. Framförallt vill vi rikta ett stort tack till Magnus Ytterstad som har gett oss mycket insikter i arbetet, framförallt om hur det är att jobba i projekt. Vi vill även tacka vår handledare på Chalmers, David Bolin, som har gett oss mycket vägledning genom hela projektet och varit tillmötesgående när vi stött på problem.

Bidragsrapport

Under det här projektet har en dagbok och en individuell tidslogg förts. Båda dessa redovisar vad varje gruppmedlem bidragit med, men även vad vi har gjort varje vecka och de problem som har uppstått. Den större delen av projektet har bestått av individuellt arbete eller arbete i mindre grupper. Fördelning av arbetsuppgifter har skett löpande på veckovisa möten där vi även har diskuterat problem som har uppstått. Vi har i regel haft möten en gång i veckan med vår handledare på Captario för att diskutera projektets upplägg och mål.

MATLAB-funktionen som ligger till grund för implementeringen har skapats av Jacob och Linus. De har även implementerat Sobols index, FAST och Morris metod. Morris metod har implementerats med hjälp av Elijah. Den större delen av litteraturstudier och implementering av Monte Carlo-filtrering har också utförts av Elijah. Den populärvetenskapliga artikeln som följer med rapporten har skrivits av Erik och Abraham. De har även implementerat den grafiska känslighetsanalysen.

Varje kapitel i rapporten tillskrivs en eller flera författare. Alla i gruppen har under arbetet bidragit med korrekturläsning.

Syfte: Jacob

Bakgrund: Hanna / Elijah

Teori: Elijah / Jacob / Hanna / Abraham

Metod: Jacob / Linus / Hanna

Resultat: Samtliga i gruppen

Diskussion: Jacob / Hanna / Erik

Appendix: Abraham / Jacob

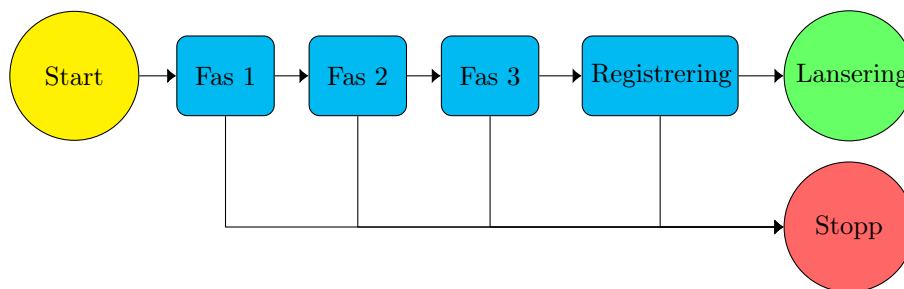
Innehåll

1	Bakgrund	1
1.1	Beslutsfattning inom läkemedelsutveckling	1
1.2	Captario och programvaran SUM	1
1.3	Känslighetsanalys	2
2	Syfte	2
3	Teori	3
3.1	Notation för allmänna modeller	3
3.2	Metoder inom känslighetsanalys	3
3.2.1	One-at-a-time-experiment	3
3.2.2	Morris metod	4
3.2.3	Monte Carlo-filtrering	5
3.2.4	Sobols index	5
3.2.5	Fourier Amplitude Sensitivity Test – FAST	7
4	Metod	10
4.1	Evaluering av NPV för modeller i SUM	10
4.1.1	Marknadsmodell	11
4.1.2	Val av responsfunktion	12
4.2	Val och implementering av känslighetsanalysmetoder	12
5	Resultat	14
5.1	Monte Carlo-filtrering	14
5.2	Morris metod	15
5.3	Sobols index och FAST	15
5.4	Grafisk känslighetsanalys	16
5.5	Relevans för Captario	17
6	Diskussion	18
6.1	Hantering av modeller	18
6.2	Applicering i SUM	19
6.3	Framtida frågeställningar	19
7	Referenser	21
	Bilagor	23
A	Morris metod på modell med Bernoullivariabler	23
B	Morris Implementeringsalgoritm	24
C	Modellbeskrivning av Accelerated	25
C.1	Variabelförteckning	25
C.2	Fasernas kostnader och tider	27
C.3	Marknadsmodellen	27
C.4	Andra projektvärdesstorheter	29

1 Bakgrund

1.1 Beslutsfattning inom läkemedelsutveckling

Läkemedelsutveckling innefattar hela processen från upptäckten av en läkemedelskandidat till lansering på marknaden. I processen sker ett flertal aktiviteter, bland annat toxiska studier, kliniska studier och rekrytering av försökspersoner. Processen från den första kliniska studien till lansering på marknaden brukar delas upp i fyra faser. Detta illustreras i figur 1. Faserna kan karaktäriseras av tester på helt friska individer, tester på ett fåtal patienter som har sjukdomen mot vilken läkemedelskandidaten verkar, långvariga tester på större grupper som bär på sjukdomen och till sist registrering av läkemedelskandidaten [1]. Utvecklingsprocessen är i regel lång och kostsam. Det finns exempel på fall då det tagit mer än ett decennium för ett läkemedel att ta sig igenom alla faser med kostnader på flera miljarder kronor som följd [2][3]. Dessutom är det endast några få procent av alla läkemedelskandidaterna som i slutändan lanseras på marknaden. Detta beror på att en övervägande del av alla substanser misslyckas i någon aktivitet.



Figur 1: Schematisk bild över de fyra faserna inom läkemedelsutveckling. Efter varje fas tas ett beslut om utvecklingen ska fortsätta eller om projektet ska läggas ner.

Mot den bakgrunden måste läkemedelsföretagen ständigt fatta beslut om de skall fortsätta med kommande aktiviteter eller om risken att projektet inte blir lönsamt är för stor. Även om ett läkemedel skulle ta sig igenom alla faser finns det fortfarande andra faktorer som skulle kunna göra projektet olönsamt, en sådan faktor är tiden. Vilken tidpunkt som lanseringen sker i förhållande till eventuella konkurrenter är avgörande för lönsamheten på ett projekt. Lanserar konkurrenterna ett liknande läkemedel först riskerar man att förlora en stor del av marknadsandelarna vilket resulterar i lägre vinst (Magnus Ytterstad, Analytics Director, Captario).

Under de senaste 20 åren har den totala kostnaden för att framställa och utveckla ett läkemedel tredubblats [3] och sannolikheten att nå marknaden är fortsatt låg med något negativ utveckling [2][3]. Med de fortsatta stora riskerna och den kostnadsmissigt negativa utvecklingen i branschen är det av stor vikt för läkemedelsföretag att kunna bygga trovärdiga modeller över läkemedelsprojekt och göra pålitliga riskanalyser av utvecklingsprocessen. Utan effektiv läkemedelsutveckling blir det ännu svårare och mer kostsamt för nya läkemedel att nå marknaden och sedermera patienter. Samtidigt tyder mycket på att behovet av läkemedel för kroniska sjukdomar kommer öka [3]. Dessa läkemedel är förhållandevis dyra, och därmed är frågan om effektiva läkemedelsprocesser inte bara viktiga för läkemedelsföretagen utan även samhället i stort.

1.2 Captario och programvaran SUM

Captario är ett Göteborgsbaserat företag som grundades 2012. De hjälper läkemedelsutvecklare att fatta strategiska beslut genom att modellera risker och uppskatta projektvärdesstorheter för ett läkemedelsprojekt. Captario har utvecklat ett nätbaserat verktyg, Captario SUM, där användaren kan bygga en modell över utvecklingsprocessen av en läkeme-

delskandidat. I verktyget kan användaren specificera antaganden kring tid och kostnad för olika aktiviteter samt konstruera en marknadsmodell. Under dessa antaganden uppskattar SUM projektvärdesstorheter, exempelvis nettonuvärde och tid tills projektet är avslutat. Nettonuvärde, eller NPV från det engelska ordet *net present value*, är kanske den viktigaste projektvärdesstorheten och därför kommer vår analys av en modell från SUM i huvudsak kretsa kring denna. NPV definieras som nuvärdet av alla framtida intäkter och kostnader[4] vilket fungerar som ett mått på lönsamhet. Resterande projektvärdesstorheter som SUM beräknar samt förklaringar till dessa finns i bilaga C. Sammanfattningsvis kan användaren få en prognos på hur lönsamt ett projekt kan bli, vilket sedermera kan användas som underlag vid beslutsfattning. Förutom en prognos över projektvärdesstorheter är det av intresse för användaren att få information om vad i modellen och ens antaganden som ger ett lönsamt projekt. Mer specifikt, eftersom det mesta i modellen varierar efter antaganden från användaren, är det intressant att ta reda på vilka osäkerheter från dessa antaganden som har stor påverkan på responsen från en modell.

1.3 Känslighetsanalys

Matematiska modeller används, bland annat inom läkemedelsbranschen, i allt större utsträckning för att simulera utfall när traditionella experiment är för dyra eller omöjliga att genomföra [5]. Modeller kan vara väldigt komplexa och bero på många variabler där fördelningarna oftast är okända och måste därför uppskattas. För att förstå hur dessa antaganden påverkar ens resultat, krävs att en så kallad *känslighetsanalys* utförs. Känslighetsanalys bygger på att analysera hur robusta modeller är under störning i indata. Metoderna inom detta fält delas in i två kategorier; lokal samt global känslighetsanalys, vilka förklaras ingående i kapitel 3. En definition av känslighetsanalys given av Andrea Saltelli, en av de ledande matematikerna inom området, lyder:

”The study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input.” [6]

Den metod som Captario använder idag för att analysera vilka variabler som bidragit till osäkerhet i utdata är att göra ett så kallat tornadodiagram. Dessa har emellertid visat sig kräva många simuleringar för att ge tillförlitliga resultat. Dessutom kan denna metod inte påvisa interaktiva effekter mellan variabler. Vidare är det intressant för Captario att testa mer sofistikerade metoder inom känslighetsanalys, för att på så sätt kunna ge kunden bättre verktyg för att förstå sin konstruerade modell på ett bättre sätt.

2 Syfte

Syftet med detta projekt är att hjälpa Captario att få en djupare förståelse över hur variationer i indata påverkar de projektvärdesstorheter SUM returnerar. För att åstadkomma detta använder vi metodik inom känslighetsanalys för att identifiera de variabler som bidrar med mest känslighet i modellen. Detta kan i förlängningen användas som ett välgrundat underlag för Captarios kunder om vilka variabler som bidrar mest till variation i utvariablerna. Detta är dels intressant ur modelleringssynpunkt för att förstå hur osäkerheter i antagandena påverkar osäkerheten i utvariabler, men också vilka variabler som behöver kontrolleras för att öka sannolikheten att projekt blir lönsamma. Vårt fokus ligger på att analysera känsligheten i projektvärdesstorheten NPV, och vår analys kommer begränsas till en specifik modell som speglar hur läkemedelsframtagning ser ut generellt. På så sätt testar vi vilka metoder som lämpar sig bäst för Captarios produkt.

3 Teori

Känslighetsanalys är ett brett område och innefattar många olika metoder. De metoder som används i det här projektet, och presenteras i avsnitt 3.2, är bland de vanligaste inom området. I avsnitt 3.1 beskrivs hur vi definierar en modell som känslighetsanalyser kan utföras på. I styckena nedan förklaras skillnaden mellan två typer av känslighetsanalyser; lokal och global.

Lokal känslighetsanalys är den enklaste typen av känslighetsanalys. Det bygger på att en variabel X_i varieras med en förbestämmd störning Δ kring ett nominellt värde. Hur denna störning påverkar responsen undersöks genom att olika partiella derivator uppskattas i det nominella värdet. Metoderna ger information om vilka variabler som har en linjär effekt på responsen kring det nominella värdet [7]. Om modellen består av ett större antal variabler så är fördelarna med en lokal känslighetsanalys dess tidseffektivitet och kan därmed behöva relativt få beräkningar. Nackdelarna är däremot att endast förändringar kring ett nominellt värde utforskas och att eventuella interaktionseffekter mellan variabler förbises.

För att kunna upptäcka interaktionseffekter kan man använda global känslighetsanalys. Gemensamt för dessa metoder är att de undersöker störningar i hela indatarummet och ger på så sätt information kring hur känsligheten i responsen kan förklaras i form av osäkerheten i modellens ingående variabler [8]. I förlängningen ger detta mer detaljerad information över vilka variabler som bidrar mest till känsligheten i responsen. Begränsningarna med global känslighetsanalys ligger i att metoderna kan vara väldigt beräkningsmässigt tidskrävande, framförallt om antalet variabler är stort [8].

3.1 Notation för allmänna modeller

Metoder för att utföra en känslighetsanalys beskrivs i avsnitt 3.2 och där används en allmän modell som exempel. I det här avsnittet beskrivs den notation som har valts för att beskriva en sådan modell. När en modell analyseras antas det att en deterministisk funktion existerar och är tillgänglig, vilken evaluerar responsen av intresse Y för en given vektor \mathbf{X} av variabler. Alltså gäller det att

$$Y = h(\mathbf{X}) \tag{3.1}$$

där $\mathbf{X} = (X_1, X_2, \dots, X_p) \in \mathbb{R}^p$ är vektorn innehållande alla variabler. Denna vektor kommer vara stokastisk med någon viss fördelning. För modeller i SUM kan det utan inskränkning antas att alla variabler är oberoende. Av bekvämlighetsskäl begränsas analysen till det fall då utfallsrummet av projektets variabler är hyperkuben $K^p = [0, 1]^p$ och slumpvariablerna är likformigt fördelade över denna mängd. Detta är ingen inskränkning, ty Y kan omskrivas till

$$Y = (h \circ g)(\mathbf{Z}) \tag{3.2}$$

där $\mathbf{Z} \sim \text{unif}([0, 1]^p)$ och är oberoende samt $g_i(\mathbf{Z}) = F_{X_i}^{-1}(Z_i)$ ¹. I fortsättningen kommer vi låta $f = h \circ g$.

3.2 Metoder inom känslighetsanalys

De metoder som presenteras i det här avsnittet är uteslutande globala känslighetsanalysmetoder. Varför just dessa metoder valts ut motiveras i avsnitt 4.2.

3.2.1 One-at-a-time-experiment

One-at-a-time-experiment, eller OAT-experiment, är en kategori av metoder inom känslighetsanalys. I dessa metoder undersöker man vilken effekt man får på responsen Y när man varierar någon variabel X_i kring ett värde medan man håller de resterande variablerna fixerade [9]. Detta görs i tur och ordning för alla variabler där praxis i OAT-experiment är att

¹I det diskreta fallet ska $F_{X_i}^{-1}$ tolkas som den generaliserade inversen.

variera X_i i ett intervall kring dess nominella- eller standardvärde [10]. Den låga beräkningskostnaden är den stora fördelen med OAT-experiment men de tar dessvärre inte hänsyn till eventuella interaktionseffekter [9].

3.2.2 Morris metod

Morris metod [11] är en typ av OAT-experiment, vilken kan betraktas som en generalisering av de lokala känslighetsmetoderna med partiella derivator till en metod som besvarar vilka variabler som bidrar till den globala känsligheten. Metoden är förhållandevis beräkningsmässigt effektiv, samtidigt som den identifierar de variabler som bidrar till den globala känsligheten. Dessutom ger metoden en indikation av i vilka variabler responsfunktionen är icke-linjär och/eller vilka variabler som interagerar med andra. I och med att Morris metod tillhör gruppen OAT-experiment, kan inte interaktiva effekter mellan variabler hittas med hjälp av denna metod. Däremot säger metoden vilka variabler som inte interagerar, eller interagerar svagt med andra.

Metoden bygger på att storheter associerade till fördelningen av den *elementära effekten* för den i :te variabeln beräknas. Den elementära effekten defineras – för något utfall \mathbf{Z} enligt (3.3)

$$E_i(\mathbf{Z}) = \frac{f(\mathbf{Z} + \Delta \mathbf{e}_i) - f(\mathbf{Z})}{\Delta} \quad (3.3)$$

där Δ är en förbestämd fixerad störning och \mathbf{e}_i är den i :te kanoniska basvektorn. Vidare är (3.3) en numerisk uppskattning av $\frac{\partial f}{\partial Z_i}$. Fördelningen av E_i ger information om den globala känsligheten av den i :te variabeln. Om m stycken simuleringar av \mathbf{Z} genomförts ($\mathbf{Z}_1, \mathbf{Z}_2 \dots \mathbf{Z}_m$) kan följande storheter, (3.4) och (3.5), beräknas

$$\mu_i^* = \frac{1}{m} \sum_{j=1}^m |E_i(\mathbf{Z}_j)| \quad (3.4)$$

$$\sigma_i = \sqrt{\frac{1}{m} \sum_{j=1}^m (E_i(\mathbf{Z}_j) - \bar{E}_i)^2} \quad (3.5)$$

där \bar{E}_i är medelvärdet av de elementära effekterna för variabel i . Morris [11] föreslog att $\mu_i = \bar{E}_i$ ska betraktas istället för μ_i^* . Dock har μ_i^* blivit allt mer vanligt förekommande i vetenskapliga sammanhang under senare år. Bland annat introduceras denna av Compolongo [12]. Vi har mot den bakgrunden valt att studera μ_i^* istället för μ_i . Om μ_i^* är relativt stor innebär det att en störning i den i :te variabeln fortplantar sig, ty μ_i^* uppskattar den genomsnittliga absoluta avvikelserna under en störning av storlek Δ . σ_i , standardavvikelsen av den i :te elementära effekten, ger en indikation om den i :te variabeln interagerar med andra, eller om responsfunktionen f är icke-linjär i just denna variabel. Vidare bör denna metod användas för att identifiera variabler som är okänsliga och inte interagerar med andra. Ofta illustreras dessa index i en graf med μ^* på ena axeln och σ på andra. På så sätt kan användaren på ett kvalitativt sätt identifiera vilka variabler som är mer känsliga än andra, genom att analysera vilka punkter (μ_i^*, σ_i) som är längst bort ifrån origo.

En implementeringsmetod som föreslogs av Morris bygger på *slumpmässiga banor* i hyperkuben $[0, 1]^p$. En närmare beskrivning av denna finns beskriven i appendix B. Denna metod har emellertid under senare år förfinats av bland andra Compolongo [12], och bör vidare endast betraktas som en skiss över hur Morris metod implementeras på ett effektivt sätt. Morris implementeringsmetod [11] bygger på att hyperkuben diskretiseras likformigt med n steg i varje dimension. I detta diskretiseringsnät evalueras f i varje nod, vilka sedermera används i metoden. På så sätt kommer responsen inte beräknas fler gånger än nödvändigt, vilket är beräkningsmässigt fördelaktigt om f är tidskrävande att evaluera.

3.2.3 Monte Carlo-filtrering

Monte Carlo-filtrering (MCF) är en metod för att kartlägga hur variabler är fördelade i olika områden av responsen Y . Metoden går ut på att man delar upp simuleringsdatan i två disjunkta delmängder: acceptabla, G , och icke-acceptabla, B , efter förutbestämda villkor satt på utfallet av responsen. Exempelvis kan vissa utfall klassas som bra om responsen överstiger ett visst tröskelvärde ξ , vilket sedermera utgör villkoret vid indelning. MCF används framförallt inom diagnostisering av modeller för att ta reda på vilka variabler som bidrar till ett visst utfall för responsen [9].

För att implementera metoden genereras N stycken realiseringar av variabler $\mathbf{X}_1, \mathbf{X}_2 \dots \mathbf{X}_N$, vilka sedan används för att beräkna N värden av responsen $Y_i = h(\mathbf{X}_i)$, $i = 1, 2, \dots, N$. Därefter delas observationerna upp i respektive delmängd utifrån det förutbestämda värdet på ξ .

$$\begin{aligned} G &= \{(\mathbf{X}_i, Y_i); i = 1, \dots, N, Y_i > \xi\} \\ B &= \{(\mathbf{X}_i, Y_i); i = 1, \dots, N, Y_i \leq \xi\} \end{aligned} \quad (3.6)$$

För varje variabel i modellen kan nu två stycken empiriska fördelningar fås, $\hat{F}_i(X_i|G)$ och $\hat{F}_i(X_i|B)$, där $X_i|G \in G$ och $X_i|B \in B$. Ifall de två empiriska fördelningarna avviker mycket från varandra kan slutsatsen dras att variabeln X_i har stor påverkan på Y [8]. För att mäta denna avvikelse kan ett *Kolmogorov-Smirnovtest* (KS-test) tillämpas som testar ifall två stickprov är dragna från samma underliggande fördelning. Teststatistikan är det största avståndet mellan två fördelningar och definieras enligt

$$d_i = \sup_{X_i} |\hat{F}_i(X_i|G) - \hat{F}_i(X_i|B)|. \quad (3.7)$$

Hypotesen att $X_i|G$ har samma fördelning som $X_i|B$ förkastas med signifikansnivå α om

$$d_i > \sqrt{-\frac{1}{2} \ln\left(\frac{\alpha}{2}\right)} \cdot \sqrt{\frac{n+m}{nm}} \quad (3.8)$$

där n och m är lika med antalet X_i i mängden G respektive B .

MCF är ett bra sätt att kvantifiera vilka variabler i modellen som har störst påverkan på Y . Metoden ihop har många egenskaper gemensamt med global känslighetsanalys däribland att hela indatarummet tas i beaktning samt att alla variabler varieras samtidigt. Dock så tas ingen hänsyn till interaktionseffekter överhuvudtaget [13]. Vilka variabler som kommer ge utslag i ett KS-test kommer att vara högst beroende på vad man sätter tröskelvärdet till. Det bör även poängteras att om en variabel X_i har *någon* inverkan på Y , kommer ett KS-test att påfinna denna om stickprovsstorleken är tillräckligt stor. Därav bör man studera graferna av de empiriska fördelningarna för att undersöka om det finns en praktisk signifikant avvikelse.

3.2.4 Sobols index

Den mest centrala känslighetsanalysmetoden som vi valt att betrakta i denna rapport bygger på *variansdekomposition*. Detta innebär att variansen av responsvariabeln additivt delas upp i funktioner som är beroende av färre variabler. Dessa termer utnyttjas sedan för att uppskatta det bidrag variabler och interaktioner mellan variabler ger till den totala variansen i responsen.

Låt f vara responsfunktionen till modellen definierad på samma vis som i avsnitt 3.2. Vi antar dessutom att f är kvadratisk integrerbar över hyperkuben K^p . Under detta antagandet följer det att f kan uttryckas som en summa av kvadratisk integrerbara funktioner av färre variabler [15]:

$$f(\mathbf{X}) = f_0 + \sum_{i=1}^p f_i(X_i) + \sum_i \sum_{j>i} f_{ij}(X_i, X_j) + \dots + f_{12\dots p}(\mathbf{X}) \quad (3.9)$$

Denna expansion innehåller totalt $2^p - 1$ stycken summor och är unik under villkoret i ekvation (3.9) [15]:

$$\int_0^1 f_{i_1 \dots i_s}(X_{i_1}, \dots, X_{i_s}) dX_{i_k} = 0, \quad 1 \leq k \leq s, \quad \{i_1, \dots, i_s\} \subseteq \{1, \dots, p\} \quad (3.10)$$

Det följer från ekvation (3.10) att termerna i ekvation (3.9) är ortogonala och kan uttryckas som integraler av $f(\mathbf{X})$ genom [15]

$$\begin{aligned} \int f(\mathbf{X}) d\mathbf{X} &= f_0 \\ \int f(\mathbf{X}) \prod_{k \neq i} dX_k &= f_0 + f_i(X_i) \\ \int f(\mathbf{X}) \prod_{k \neq i, j} dX_k &= f_0 + f_i(X_i) + f_j(X_j) + f_{ij}(X_i, X_j) \end{aligned} \quad (3.11)$$

där varje integral ska tolkas som en bestämd multipelintegral över en hyperkub med lämplig dimension. En följd av ekvation (3.11) är att f_0 är konstant lika med väntevärdet av $f(\mathbf{X})$. Genom att kvadrera båda leden i ekvation (3.9) och integrera över K^p så följer det att

$$\int f^2(\mathbf{X}) d\mathbf{X} - f_0^2 = \sum_{s=1}^n \sum_{i_1 < \dots < i_s} \int_{K^p} f_{i_1 \dots i_s}^2 \prod_{k=1}^s dX_{i_k}. \quad (3.12)$$

Vänsterledet i ekvation (3.12) är lika med variansen enligt Steiners sats, som säger att $V(Y) = E(X^2) - (E(X))^2$. Därför erhåller vi att variansen av responsvariabel Y kan uttryckas på följande vis:

$$V(Y) = \sum_{i=1}^p V_i(Y) + \sum_i \sum_{j>i} V_{ij}(Y) + \dots + V_{12\dots p}(Y), \quad (3.13)$$

$$\begin{aligned} V_{ij}(Y) &= V(f_{ij}(X_i, X_j)) = V_{X_i X_j}(E_{X_{\sim ij}}(Y|X_i, X_j)) - V_{X_i}(E_{X_{\sim i}}(Y|X_i)) - V_{X_j}(E_{X_{\sim j}}(Y|X_j)) \\ &= \int f_{ij}^2 dX_i dX_j. \end{aligned} \quad (3.14)$$

$V_i(Y)$ kallas första ordningens varians, vilken är den förväntade minskningen av variansen om vi håller X_i fixerat. Om vi dividerar båda leden i ekvationen ovan med $V(Y)$ så får vi att

$$1 = \sum_{i=1}^p \frac{V_i(Y)}{V(Y)} + \sum_i \sum_{j>i} \frac{V_{ij}(Y)}{V(Y)} + \dots + \frac{V_{12\dots p}(Y)}{V(Y)}. \quad (3.15)$$

Sobols känslighetsindex definieras nu på följande vis

$$\begin{aligned} S_i &= \frac{V(E[Y|X_i])}{V(Y)} \\ S_{ij} &= \frac{V(E[Y|X_i, X_j]) - V_i - V_j}{V(Y)} \\ &\vdots \\ S_{T_i} &= 1 - \frac{V(E[Y|\mathbf{X}_{\sim i}])}{V(Y)}, \end{aligned} \quad (3.16)$$

där $\mathbf{X}_{\sim i} = (X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_p)$. S_i är Sobols index för första ordningens effekt och ger information om hur stor del av variansen i responsfunktionen som kan tillskrivas den

i :te variabeln isolerad från de andra variablerna. S_{ij} är index för första ordningens interaktionseffekt mellan den i :te och den j :te variabeln, och i en modell med p stycken variabler finns $\sum_{i=2}^p \binom{p}{i}$ möjliga interaktionseffekter. Vidare kommer det för ett stort p finnas ett stort antal interaktionstermer, och av beräkningsmässiga skäl uppskattas vanligen endast index av låg ordning. Till följd av ekvation (3.15) kommer summan av alla känslighetsindex vara 1.

S_{T_i} är index för den totala effekten, vilken är den förväntade variansen om alla variabler utom X_i hålls fixerade och kommer att innehålla information om första ordningens effekt samt alla möjliga interaktionseffekter. Detta index ger information om hur influerande den i :te variabeln är i modellen genom att samla alla interaktionseffeter i en storhet. Genom att slå samman alla interaktionseffekter på detta sätt till en term gör att man inte förlorar någon information då man empiriskt har sett att interaktioner mellan tre variabler eller fler oftast är obefintlig (Kerstin Wiklander, Uni.lektor Chalmers). Vidare uppskattas i regel endast första ordningens samt totala variansen för att hålla nere beräkningstiden, ty dessa två index ger god information om känsligheten i modellen.

Variansen som används vid beräkning av första ordningens index kan omskrivas [15]:

$$V_i = V(E[Y|X_i]) = \int f(\mathbf{X})f(X_i, \mathbf{X}'_{\sim i}) d\mathbf{X} d\mathbf{X}'_{\sim i} - f_0^2 \quad (3.17)$$

där X_i är det i :te elementet i \mathbf{X} , $\mathbf{X}'_{\sim i}$ är en vektor innehållandes alla variabler utom den i :te och $(X_i, \mathbf{X}'_{\sim i})$ utgör en komplett parameteruppsättning. Sobol visade också att den totala variansen kan omskrivas:

$$V_{\sim i} = V(E[Y|\mathbf{X}'_{\sim i}]) = V(Y) - \frac{1}{2} \int \left(f(\mathbf{X}) - f(X_i, \mathbf{X}'_{\sim i}) \right)^2 d\mathbf{X} d\mathbf{X}'_{\sim i}. \quad (3.18)$$

Integralerna i (3.17) och (3.18) uppskattas lämpligen med hjälp av Monte Carlo-simuleringar. Om $\mathbf{x}^{(k)} \in \mathbb{R}^p$, $\mathbf{y}^{(k)} \in \mathbb{R}^{p-1}$ och $z^{(k)} \in \mathbb{R}$ ($k = 1, 2, \dots, N$) är simulerade vektorer av oberoende likformigt fördelade slumpvariabler, skattas varianserna med följande summer:

$$\begin{aligned} \hat{f}_0 &= \frac{1}{N} \sum_{k=1}^N f(\mathbf{x}^{(k)}) \\ \hat{V} &= \frac{1}{N} \sum_{k=1}^N f^2(\mathbf{x}^{(k)}) - \hat{f}_0^2 \\ \hat{V}_i &= \frac{1}{N} \left(\sum_{k=1}^N f(\mathbf{x}^{(k)}) f(x_i^{(k)}, \mathbf{y}^{(k)}) \right) - \hat{f}_0^2 \\ \hat{V}_{\sim i} &= \hat{V} - \frac{1}{2N} \sum_{k=1}^N \left(f(\mathbf{x}^{(k)}) - f(z^{(k)}, \mathbf{x}'_{\sim i}^{(k)}) \right)^2. \end{aligned} \quad (3.19)$$

Uppskattade sobolindex \hat{S}_i och \hat{S}_{T_i} erhålls om varianserna i (3.17) byts mot de skattade i (3.19). Till följd av att felet i Monte Carlo-simuleringar avtar som $N^{-1/2}$ fordras ett stort N för att få bra resultat. Mer effektiva skattningar har emellertid utvecklats under de senare åren, vilka presenteras av Saltelli i Computer Physics Communication [14].

3.2.5 Fourier Amplitude Sensitivity Test – FAST

De integraler som behöver beräknas för Sobols index uppskattas vanligen med Monte Carlo-metoder, genom simuleringar av variabler som antas ligga i hyperkuben K^p . Ett alternativt tillvägagångssätt, som först presenterades av Cukier [16] är att generera slumpantal längs en kurva som fyller hyperkuben likformigt. Alltså, antas att alla variabler kan uttryckas $x_i = r_i(\sin \omega_i s)$ för någon funktion r_i och vinkelhastighet ω_i , och s tillåts variera i intervallet $[-\pi, \pi]$. Då kan f_0 approximeras som

$$f_0 = E(Y) = \int f(\mathbf{X}) d\mathbf{X} \approx \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\mathbf{r}(s)) ds \quad (3.20)$$

där $\mathbf{r}(s) = (r_1(\sin \omega_1 s), r_2(\sin \omega_2 s) \dots r_p(\sin \omega_p s))$. Denna approximation gäller endast om r_i och ω_i är väl specificerade, däribland att \mathbf{r} utrymmer hyperkuben likformigt, och att vinkelfrekvenserna är linjärt oberoende i heltalsmening, det vill säga att $\sum_{i=1}^p \lambda_i \omega_i = 0$ saknar heltalslösningar i λ_i [17]. Vidare kan variansen av responsen approximeras på följande sätt:

$$V(Y) = \int f^2(\mathbf{X}) d\mathbf{X} - (E(Y))^2 \approx \frac{1}{2\pi} \int_{-\pi}^{\pi} f^2(\mathbf{r}(s)) ds - \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} f(\mathbf{r}(s)) ds \right)^2. \quad (3.21)$$

Den första termen i högerledet kan enligt *Parseval's formel* uttryckas genom

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} f^2(\mathbf{r}(s)) ds = \sum_{k=-\infty}^{\infty} |c_k|^2 \quad (3.22)$$

där c_k är den k :te Fourierseriekoeficienten till $f \circ \mathbf{r}$:

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\mathbf{r}(s)) e^{-jks} ds, \quad j^2 = -1. \quad (3.23)$$

I och med att $c_{-k} = \bar{c}_k \implies |c_{-k}| = |c_k|$, och att sista termen i högerledet av (3.21) är $|c_0|^2$, kan (3.21) omskrivas:

$$V(Y) \approx 2 \sum_{k=1}^{\infty} |c_k|^2 = D. \quad (3.24)$$

Dessutom, om ω_i , för $i = 1, 2 \dots p$, är heltal, kan $V(E(Y|X_i))$ uppskattas genom

$$V_i(Y) \approx 2 \sum_{k=1}^{\infty} |c_{k\omega_i}|^2 = D_i. \quad (3.25)$$

I praktiken måste dessa serier trunckas, men på grund av konvergens hos fourierserier, kan detta med fördel göras så länge tillräckligt många termer används. Heuristik visar att $M \geq 4$ termer brukar vanligtvis vara tillräckligt [17]. Dessutom kan c_k effektivt beräknas med hjälp av den diskreta fouriertransformen eller FFT, vilket innebär att det finns beräkningsmässiga fördelar med att tillämpa FAST. Dessutom har det visats att $S_i^{\text{FAST}} = D_i/D$ är ekvivalent med första ordningens Sobol index. Saltelli föreslog att åtminstone $N_s = 2M\omega_{\max} + 1$ simuleringar måste genomföras för att erhålla ett tillförlitligt resultat.

För att uppskatta den totala effekten för den i :te variabeln tilldelas den, i analogi med uppskattningen av första ordningens index, en frekvens ω_i . Till skillnad från tidigare tilldelas alla andra variabler en och samma vinkelfrekvens $\omega_{\sim i}$. Vidare kan man med hjälp av detta beräkna den varians "alla utom i " ger upphov till genom

$$V_{\sim i} \approx 2 \sum_{k=1}^{\infty} |c_{k\omega_{\sim i}}|^2 = D_{\sim i} \quad (3.26)$$

vilken används för att uppskatta totala effekten:

$$S_{T_i}^{\text{FAST}} = 1 - \frac{D_{\sim i}}{D}. \quad (3.27)$$

Saltelli [17] föreslog att kurvor på formen $r_i(\cdot) = \frac{1}{2} + \frac{1}{\pi} \arcsin(\cdot)$ kan användas som utfyllande banor i hyperkuben. Dessa har visat sig ge förhållandevis rättvisa resultat men nackdelen med denna kurva är att den alltid kommer börja i samma punkt. För att undvika potentiell systematisk avvikelse, kan istället kurvan $x_i = r_i(\sin(\omega_i s + \phi_i))$ användas, där $\phi_i \sim \text{unif}(0, 2\pi)$. Detta bör emellertid upprepas med N_r realiseringar av $(\phi_1, \phi_2, \dots, \phi_r)$, för att få ner den

systematiska avvikelsen ϕ_i ger upphov till. Då fås indexen med hjälp av aritmetiska medelvärden

$$S_i^{\text{FAST}} = \frac{\sum_{j=1}^{N_r} D_i^{(j)}}{\sum_{j=1}^{N_r} D^{(j)}} \quad (3.28)$$

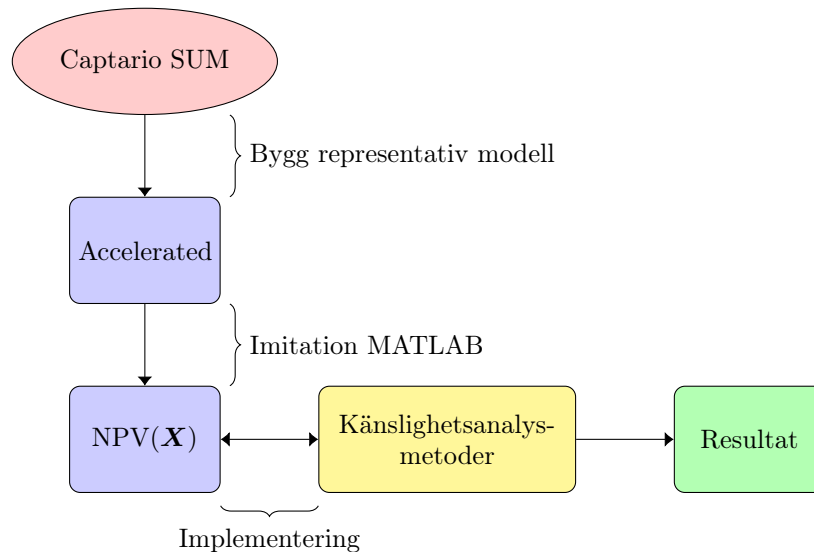
där $D_i^{(j)}$ och $D^{(j)}$ är de uppskattade varianserna associerade till den j :te realiseringen av $(\phi_1, \phi_2, \dots, \phi_r)$. Vidare kommer totala antalet simuleringar med denna kurva bli $(2M\omega_{\max} + 1)N_r$.

4 Metod

En användare i SUM bygger sin egen modell och gör antaganden kring sitt läkemedelsprojekt och därför finns det, i teorin, oändligt många unika modeller. För att utföra en känslighetsanalys på läkemedelsprojekt begränsades undersökningen till en modell som Captario byggt som kallas *Accelerated*. Den ansågs vara representativ för de flesta läkemedelsprojekt och tar vara på den sekvensiella strukturen av ett projekt som finns beskriven i avsnitt 1.1. Modellerna i SUM uppskattar flera projektvärdesstorheter men vi valde att bara fokusera på den mest centrala av dem; NPV. Andra projektvärdesstorheter kan analyseras helt analogt.

De flesta metoder som har valts för det här projektet, och som finns beskrivna i avsnitt 3.2, kan inte direkt använda sig av simuleringsdata från SUM. Metoderna behöver en *funktion* för Accelerated som givet en uppsättning variabler kan evaluera NPV. Evalueringfunktionen som SUM använder var inte tillgänglig under projektet så vi beslutade att skriva en funktion i MATLAB som imiterar beräkningen. Funktionen återfinns i [18] tillsammans med all källkod för projektet. Modeller i SUM kan vara komplexa och imitationen kräver grundlig förståelse för hur de är uppbyggda. Accelerated är en icke-linjär modell med 40 invariabler varav sju stycken är Bernoullifördelade. Fördelningarna av de resterande variablerna är specificerade i bilaga C, vilken innehåller en mer detaljerad beskrivning av accelerated. Vid beräkning av NPV nuvärdesjusteras dessutom alla kostnader över tid. Idéen bakom modeller i SUM finns beskriven i avsnitt 4.1.

Med en funktion på plats kunde ett antal olika metoder för känslighetsanalys appliceras. Valet av dessa metoder och hur de implementerades motiveras i avsnitt 4.2 och resultaten av implementeringarna finns i avsnitt 5. De förenklingar och vägval som gjorts under projektet är sammanställda i figur 2.



Figur 2: Schematisk figur över problemlösningstrategin i projektet. I programvaran SUM byggs den representativa modellen Accelerated, vars beräkning av NPV imiteras i MATLAB. Känslighetsanalysmetoderna implementeras sedan genom att anropa MATLAB-funktionen, och resultat erhålls.

4.1 Evaluering av NPV för modeller i SUM

För att kunna konstruera en funktion från modellen Accelerated behövde vi först förstå i allmänhet hur modeller i SUM fungerar och dessutom hur vi skulle hantera de Bernoullivariabler som finns. I vårt fall, när en projektvärdesstorhet i ett läkemedelsprojekt är av intresse att modellera, kommer variablerna i modellen vara händelser associerade till bland annat kliniska studier samt händelser i marknaden som påverkar försäljningsvolymen. Exempelvis kan elementen i \mathbf{X} vara kostnader, tider för olika kliniska studier och lanseringstider för kon-

kurrenter. För NPV, kan responsfunktionen för en allmän projektmodell explicit uttryckas enligt (4.1)

$$Y = h_1(\mathbf{X}) = -C_1 - I_1C_2 - I_1I_2C_3 - I_1I_2I_3C_{\text{reg.}} + I_1I_2I_3I_4(R - C_m) \quad (4.1)$$

där C_i och $R = R$ är den nuvärdesjusterade kostnaden för den i :te fasen respektive inkomsten givet variablerna i \mathbf{X} . Dessutom är $I_i \sim \text{Bernoulli}(p_i)$ och antas vara oberoende, vilken antar värdet 1 om läkemedelskandidaten tar sig igenom den i :te fasen och 0 annars. Vidare om $I_i = 0$ för något i kommer avkastningen bli negativ. Ett naturligt tillvägagångssätt att uppskatta $\text{eNPV} = E(Y)$ är genom att simulera ett stort antal utfall $\mathbf{X}_1, \mathbf{X}_2 \dots \mathbf{X}_N$ och beräkna medelvärdet

$$\widehat{\text{eNPV}}_1 = \frac{1}{N} \sum_{i=1}^N h_1(\mathbf{X}_i). \quad (4.2)$$

Ett problem med skattningen (4.2) är att en stor del simuleringsdata inte används till följd av Bernoullivariablerna i (4.1). Exempelvis, om $I_1 = 0$ kommer alla variabler som inte finns i $\mathbf{X}^{(1)}$ att simuleras, men inte användas. En alternativ punktskattning som kringgår detta utnyttjar antagandet att Bernoullivariablerna är oberoende

$$\begin{aligned} \text{eNPV} &= E[h_1(\mathbf{X})] = E\left[-C_1 - I_1C_2 - I_1I_2C_3 - I_1I_2I_3C_{\text{reg.}} + I_1I_2I_3I_4(R - C_m)\right] = \\ &= -E(C_1) - p_1E(C_2) - p_1p_2E(C_3) - p_1p_2p_3E(C_4) + p_1p_2p_3p_4E(R - C_m) = E[h_2(\mathbf{X})] \end{aligned} \quad (4.3)$$

där h_2 defineras som

$$h_2(\mathbf{X}) = -C_1 - p_1C_2 - p_1p_2C_3 - p_1p_2p_3C_{\text{reg.}} + p_1p_2p_3p_4(R - C_m). \quad (4.4)$$

Vidare kan eNPV uppskattas med följande punktskattning, vilken använder all simuleringsdata:

$$\widehat{\text{eNPV}}_2 = \frac{1}{N} \sum_{i=1}^N h_2(\mathbf{X}_i). \quad (4.5)$$

4.1.1 Marknadsmodell

I många marknadsmodeller kan det vara intressant att modellera konkurrenternas anspråk på de tillgängliga marknadsandelarna. Vår analys begränsas till att en konkurrents beteende enbart bestäms av huruvida de lyckas lansera produkten överhuvudtaget, samt konkurrentens lanseringstid. Det förstnämnda modelleras med en oberoende Bernoullivariabel, och det senare med en oberoende kontinuerlig slumpvariabel med specificerad fördelning.

Mot samma bakgrund som tidigare, är det behändigt att få bort Bernoullivariablerna ur modellen, ty en simulering av en lanseringstid kommer inte till användning om konkurrenten i fråga inte lanserar. I allmänhet kommer ordningen aktörerna lanserar att ha betydelse för hur många marknadsandelar som vi erhåller och därför är det ofta svårt att formulera den årliga avkastningen $R_A(t)$ som en additiv modell, liksom (4.1). Vidare måste andra verktyg tillämpas. Vi låter lanseringstiden för projektet betecknas T_{RE} , vektorn med alla Bernoullivariabler för alla konkurrenter $I = (I_{c_1}, I_{c_2}, \dots, I_{c_n})$ och K alla möjliga utfall av I . Då gäller att

$$\begin{aligned} E[R_A(t)|T_{Re}] &= \sum_{\kappa \in K} E((R_A(t)|T_{Re})|I = \kappa)P(I = \kappa) = \\ &= \sum_{\kappa \in K} \left(E((R_A(t)|T_{Re})|I = \kappa) \prod_{i=1}^n p_i^{\kappa_i} (1 - p_i)^{1 - \kappa_i} \right) \end{aligned} \quad (4.6)$$

där κ_i är den i :te komponenten i κ och p_i är parametern för bernoullivariabeln som svarar mot den i :te konkurrenten. Om intäkterna diskonteras årsvis med räntan r gäller att

$$R|T_{RE} = \int_{T_{RE}}^T R_A(t)r^{-([t]+1)} dt \stackrel{(4.6)}{\implies} \quad (4.7)$$

$$E(R|T_{RE}) = \int_{T_{RE}}^T \left(\sum_{\kappa \in K} p_i^{\kappa_i} (1 - p_i)^{1 - \kappa_i} E((R_A(t)|T_{RE})|I = \kappa)r^{-([t]+1)} \right) dt \quad (4.8)$$

där T är den tid då projektet avslutas och $[t]$ är heltalsdelen av t . Används (4.8) vid uppskattningen av R i (4.4) kommer det inte att finnas några Bernoullivariabler kvar i modellen över huvudtaget, vilket utnyttjas nedan.

4.1.2 Val av responsfunktion

Det kanske mest naturliga valet av modell är $Y = h_1(\mathbf{X})$, ty simuleringarna av \mathbf{X} kommer efterlikna ett verkligt scenario, där någon aktivitet antingen misslyckas, eller alla studier påvisar positiva effekter. Om h_1 (4.1) studeras i en känslighetsanalys, så kommer i allmänhet Bernoullivariablerna bidra mest till osäkerhet i utdata. Detta på grund av att dessa variabler ger upphov till diskontinuitet i funktionen $h_1 \circ \mathbf{g}$, vilket är definierat i analogi med ekvation (3.2). Av den anledningen har vi valt att genomföra en känslighetsanalys på responsfunktionen h_2 definierad i ekvation (4.4), i vilken samtliga Bernoullivariabler är borttagna.

4.2 Val och implementering av känslighetsanalysmetoder

I det här projektet implementerade vi metoderna Morris metod, Monte Carlo-filtrering och Sobols index samt genomförde en grafisk känslighetsanalys. Den mest centrala av dessa metoder är Sobols index eftersom den ger oss både information från hela indata-rummet samt kvantifierar interaktionseffekter. Vår modell består av många variabler och indatan är alltid stokastisk och på grund av komplexiteten i modellen kan vi anta att det finns många interaktionseffekter. Nackdelen med att använda Sobol är att den är beräkningsmässigt krävande och för att behandla detta valde vi att avskärma insignifikanta variabler.

Att avskärma variabler innebär att fixera okänsliga variabler, oftast kring dess nominella värde. De variabler som avskärmades fixerades till dess väntevärde. Den vanligaste metoden för detta är Morris metod som dessutom kvalificerar som en typ av global känslighetsanalysmetod. Fördelarna med metoden i vårt fall är att vi både får ett mått på vilken effekt variablerna utgör på NPV, men även om de kan ingå i någon form av interaktionseffekter. Vid implementeringen utfördes transformen (3.2) till hyperkuben, innan en algoritm från en artikel i tidningen Acta Astronautica [19] tillämpades.

Monte Carlo-filtrering skulle potentiellt kunna användas som en avskärningsmetod men metoden förbiser helt interaktionseffekter. Det finns dock fördelar med MCF; den undersöker hela indata-rummet och den är enklare att genomföra eftersom realisationsdatan kan användas direkt från simuleringar i SUM. I det här projektet användes MCF som ett komplement till Morris metod för att jämföra resultaten. Vid implementering av MCF gjordes 10 000 realiseringar för att få tillförlitliga resultat och tröskelvärdet sattes till eNPV från vår modell. Även om det går att använda realiseringar direkt från SUM användes realiseringar från vår MATLAB-funktion för att få helt konsekventa resultat. För att avgöra om en variabel var signifikant eller inte användes KS-testet. Eftersom resultaten i MCF är så beroende av vilket värde tröskelvärdet har, gjordes ett test där KS-testet genomfördes för olika tröskelvärden som varierade likformigt från 2/3 eNPV till 4/3 eNPV med 100 värden.

Efter att flera variabler avskärmats med hjälp av Morris metod kunde *Sobols index* och *Sobols totala index* räknas ut under en väsentligt kortare beräkningstid. Vid implementeringen användes ett MATLAB-program som använder skattningarna presenterade i [14]. Det beräkningsmässiga felet uppskattades med hjälp av normal approximation med konfidensgrad 0,95.

Ett alternativt sätt att räkna ut Sobols index är att använda metoden FAST vilket innebär mindre beräkningskostnader men potentiellt mindre pålitliga resultat. I vår implementeringen av FAST skrev vi ett eget MATLAB-program som använde MATLAB:s funktion `FFT`. Vi valde att genomföra båda metoderna för att räkna ut Sobols index av första ordningen och sedan jämföra om FAST kan vara ett lika tillförlitligt alternativ.

Som komplement till Sobols index valde vi att genomföra en grafisk analys, som tillsammans med skalärerna från Sobols index kunde ge en ännu tydligare bild av hur en variabel påverkar responsen. Vi utförde detta genom att beräkna hur eNPV ändrades när vi fixerade en variabel X_i med ett antal punkter från dess teoretiska fördelning. Sedan illustrerade vi detta grafiskt med $E[Y|X_i] - E[Y]$ och X_i . Att utföra detta på alla variabler i modellen är beräkningsmässigt tidskrävande. Men eftersom vi redan hade avskärmat bort ett antal variabler med Morris metod kunde vi utföra detta på de variabler som var kvar.

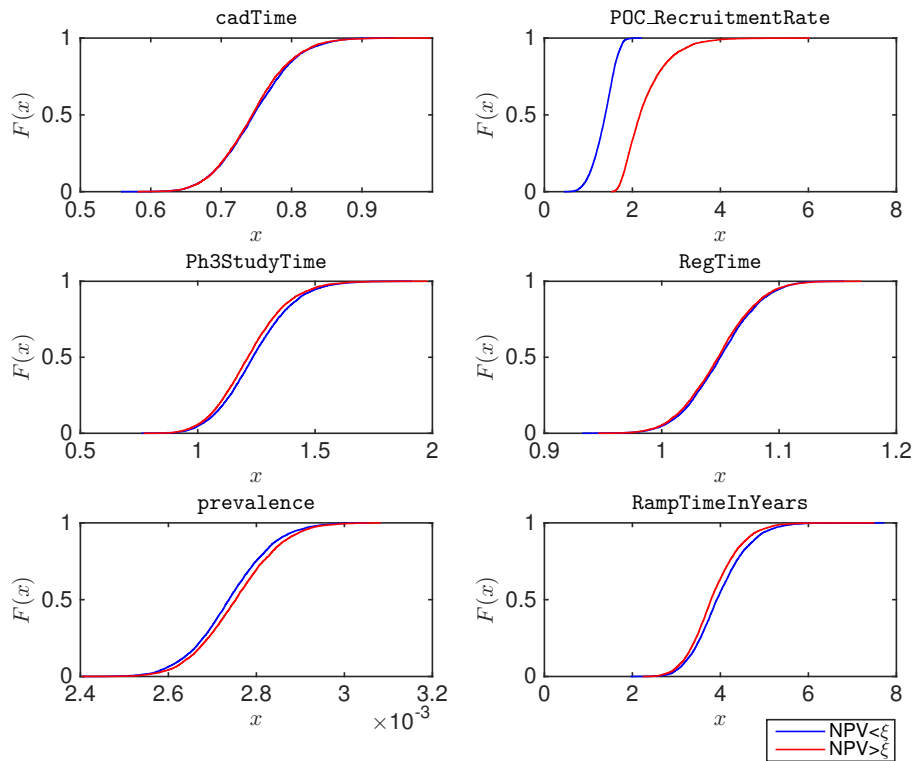
5 Resultat

I avsnitten nedan presenteras de figurer och tabeller som är relevanta för vårt projekt enligt den typ av modell som finns beskriven i avsnitt 4.1. I 4.1 motiveras även varför känslighetsanalysen har genomförts på modellen Accelerated utan Bernoullivariabler. Några av metoderna utfördes på modellen med Bernoullivariabler bara för att visa på att de dominerar känsligheten. Dessa resultat presenteras i bilaga A.

Samtliga metoder under känslighetsanalysen identifierade `POC_RecruitmentRate` som den dominerande variabeln. Dess värde beskriver rekryteringsfarten per klinik per månad för de tre kliniker som antas vara med och rekrytera patienter till POC-studien i fas 2. I bilaga C finns en mer detaljerad beskrivning av hur `POC_RecruitmentRate` samverkar i Accelerated.

5.1 Monte Carlo-filtrering

MCF utfördes på simuleringsdata innehållande 10 000 realiseringar av variablerna och beräkningar av NPV där tröskelvärdet ξ är satt till eNPV. Resultaten från MCF visas i figur 3.



Figur 3: De empiriska fördelningarna $F(x)$ för de variabler som fick ett signifikant utfall med KS-testet där signifikansnivån α sattes till 5 % och tröskelvärdet ξ sattes till eNPV. Den blå kurvan är den empiriska fördelningen för variabeln när responsen var lägre än eNPV och den orangea när utfallet på responsen var större än eNPV.

Av modellens 33 variabler så är det fem stycken som ger ett signifikant resultat med KS-testet. Utav dessa är `POC_RecruitmentRate` den variabel som visar på störst skillnad mellan de två olika empiriska fördelningarna. MCF utfördes även på simuleringsdata när ξ varierades i ett symmetriskt intervall kring eNPV. För varje unikt ξ inom intervallet genomfördes ett KS-test på alla variabler där resultatet av denna analys kan ses i tabell 1, i vilken andelen signifikanta test för en variabel anges.

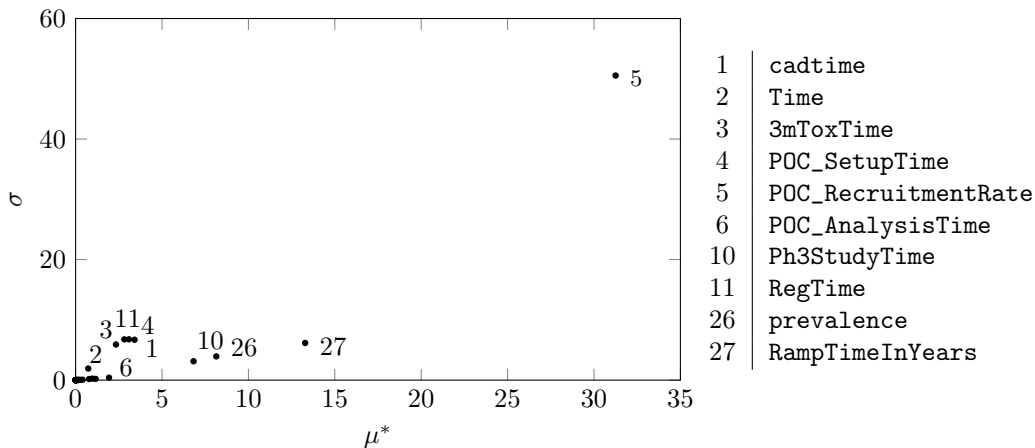
Tabell 1: Tabellen illustrerar andelen av KS-testet visade på signifikant avvikelse i de empiriska fördelningarna, med signifikansnivå α satt till 5%. Testet utfördes när ξ varierade från 2/3 av eNPV till 4/3 av eNPV med 100 värden. Enbart de variabler som någon gång visade på ett signifikant utfall med KS-testet återfinns i tabellen.

Variabelnamn	$\xi \leq eNPV$	$\xi > eNPV$
batchForPH3Time	0.12	0.60
POC_RecruitmentRate	1	1
Ph3StudyTime	1	1
RegTime	0.42	0.02
POC_CostPerCenter	0.14	0
Development_costs3	0.06	0.38
prevalence	1	1
RampTimeInYears	1	1
Comp1LaunchTime	0.12	0
Comp3LaunchTime	0.02	0

När tröskelvärdet varierar så börjar även en del andra variabler att påvisa skillnader i sina empiriska fördelningar. Men det återfinns fyra variabler som förblir signifikanta för varje test som görs, POC_RecruitmentRate, Ph3StudyTime, prevalence, RampTimeInYears. Något som även går att observera är att batchForPH3Time påvisar signifikant avvikelse i mer än hälften av testen när ξ var större än eNPV.

5.2 Morris metod

I figur 4 illustreras resultaten från implementeringen av Morris metod. De variabler som återfinns tillräckligt nära origo kommer inte betraktas i beräkningen av Sobols index för att minska den beräkningsmässiga kostnaden. Valet av kritiskt område i figur 4 är en ellips kring origo. Observera att Morris metod och MCF enas på vissa variablers inverkan på NPV, och att Morris uteslutande identifierar tider.



Figur 4: Diagram med absoluta medelvärde μ^* och standardavvikelse σ för de elementära effekterna associerade till alla variabler. Endast de variabler med relativt stora index är markerade med variabelnamn.

5.3 Sobols index och FAST

Beräknade värden av Sobols index för första ordningens effekter visas i tabell 2. Resultatet är framräknat från 1 000 simuleringar. Värden för första ordningens effekter framräknade med FAST visas även i tabellen. De variabler som visas är de som Morris metod identifierade som mest inflytserika.

Tabell 2: Beräknade värden på Sobols index för första ordningens effekter, S_i med tillhörande felmarginaler, ΔS_i , samt resultat från implementation av FAST, S_i^{FAST} .

Variabelnamn	$S_i \pm \Delta S_i$	S_i^{FAST}
POC_RecruitmentRate	0,9312 \pm 0.0591	0,9136
RampTimeInYears	0,0158 \pm 0.0093	0,0163
prevalence	0,0061 \pm 0.0057	0,0065
Ph3StudyTime	0,0046 \pm 0.0057	0,0054
Time	0,0016 \pm 0.0018	0,0005
3mToxTime	-0,0001 \pm 0.0016	0,0005
RegTime	-0,0012 \pm 0.0028	0,0003
POC_AnalysisTime	0,0009 \pm 0.002	0,0001
POC_SetupTime	0,0001 \pm 0.0027	0,0004
cadtime	-0,0009 \pm 0.0027	0,0010

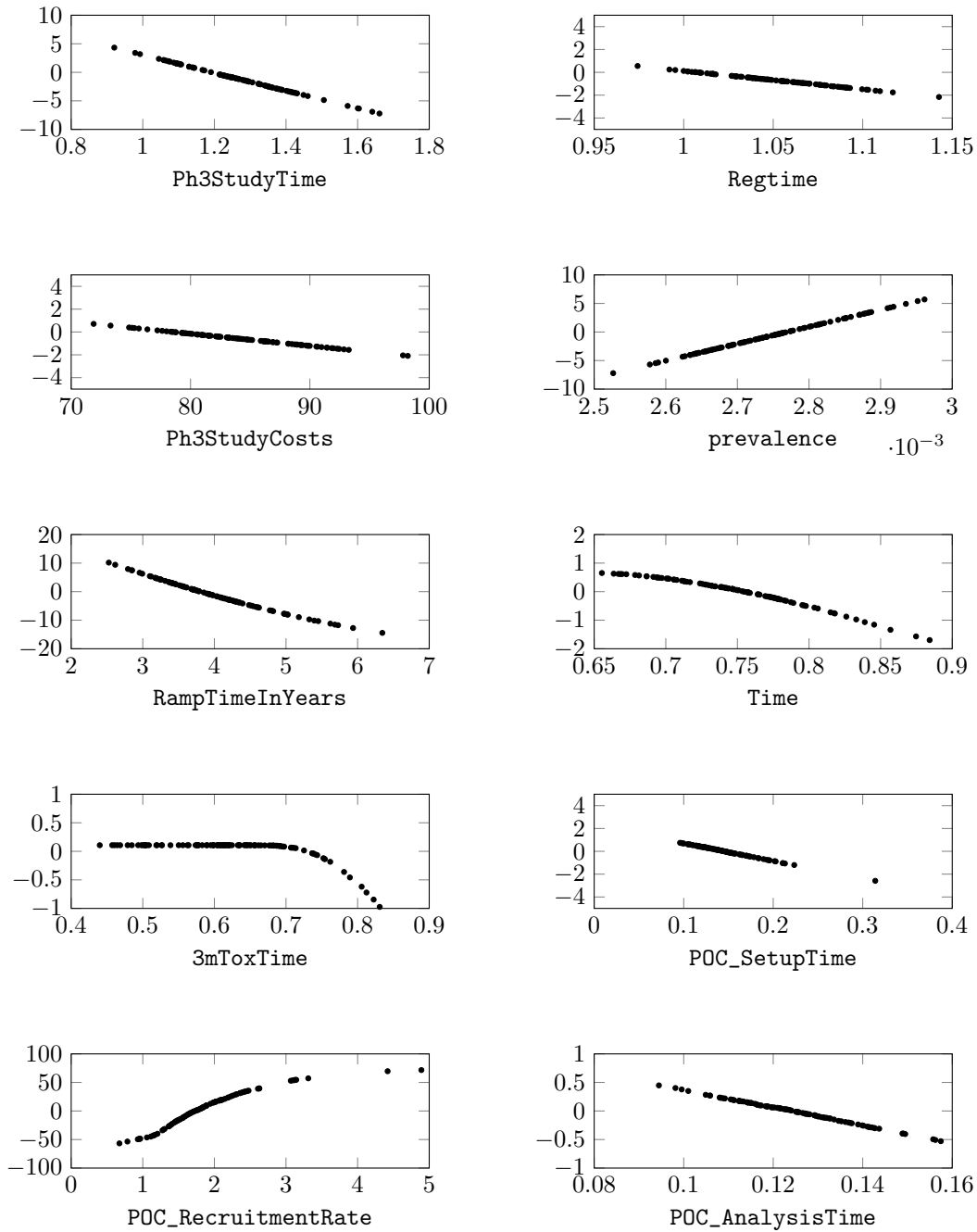
POC_RecruitmentRate är förklarar cirka 93 procent av variansen i NPV för modellen utan Bernoullivariabler, och prevalence bidrar näst mest med känslighet. Övriga variabler har relativt liten inverkan. POC_RecruitmentRate var dominant även vid implementation av FAST, men den inbördes rangordningen skiljer sig något för övriga variabler. I tabell 3 visas motsvarande beräkningar för Sobols totala index, S_{T_i} . Notera de stora beräkningsmässiga felmarginallerna.

Tabell 3: Beräknade värden på Sobols totala index, S_{T_i} , med tillhörande felmarginal, ΔS_{T_i} , för de variabler som Morris metod identifierade som mest inflytelserika.

Variabelnamn	$S_{T_i} \pm \Delta S_{T_i}$
POC_RecruitmentRate	0,9546 \pm 0.0246
RampTimeInYears	0,0246 \pm 0.008
Ph3StudyTime	0,0095 \pm 0.0004
prevalence	0,0091 \pm 0.0033
3mToxTime	0,0007 \pm 0.001
Time	0,0009 \pm 0.001
RegTime	0,0022 \pm 0.0002
POC_AnalysisTime	0,0011 \pm 0.0001
POC_SetupTime	0,0021 \pm 0.002
cadtime	0,0021 \pm 0.002

5.4 Grafisk känslighetsanalys

Som komplement till variablerna i Sobols index utfördes en grafisk analys. I figur 5 återfinns resultatet från denna. Notera att värden på POC_RecruitmentRate bidrar till en markant skillnad jämfört med ändringar för de andra variablerna. Det kan även ses i graferna att skillnaden i väntevärde av NPV är nästan konstant för variablerna ph3studyCosts och RegTime medan andra variabler i graferna har större inverkan på väntevärdet.



Figur 5: I graferna ovan visas skillnaden $E[\text{NPV}|X_i] - E[\text{NPV}]$ för ett stickprov av storlek 100 från variabelns fördelning. För varje värde av X_i har 10.000 beräkningar av NPV gjorts för att uppskatta $E[\text{NPV}|X_i]$

5.5 Relevans för Captario

En del av frågeställningen i det här projektet är huruvida Captario kan använda de känslighetsanalysmetoder vi har använt. Rent praktiskt innebär detta att för att Captario skall kunna använda sig av metoderna bör de vara effektiva. Den mest centrala metoden, Sobols index, har visat sig vara beräkningsmässigt tidskrävande och tog flera timmar att beräkna. Indexen beräknades betydligt snabbare, under en minut, när FAST implementerades. Den grafiska känslighetsanalysen tog även den flera timmar att beräkna. Morris metod och MCF beräknades snabbt.

6 Diskussion

I samtliga presenterade resultat är `POC_RecruitmentRate` den mest influerande variabeln. Denna variabel är approximativt inverst proportionell till tiden för en aktivitet i den andra fasen och vidare kommer NPV vara högst olinjär i detta argument. Dessutom är ett antagande i modellen att samtliga tre deltagande kliniker i denna modell har samma rekryteringstakt. Vidare vore ett lämpligt sätt att få ned känsligheten i modellen att modellera rekryteringstakten på ett annat sätt, exempelvis genom att låta varje klinik ha en egen rekryteringstakt.

En annan tydlig genomgående tendens är att variabler associerade till tider är mer påverkande än de andra. Detta har att göra med att en sänkt projekttid har flera effekter på lönsamheten. Dels kommer en tidig lansering innebära att produkten säljs på marknaden under en längre tid, och vidare blir den totala försäljningsvolymen högre. Dessutom kommer en kortare projekttid göra det mer sannolikt att företaget lanserar läkemedlet före konkurrenterna, vilket resulterar i att en större marknadsandel erhålls. En tredje effekt är att en tidig lansering gör att de första intäkterna inkasseras närmare projektstarten, och om alla intäkter och kostnader nuvärdesjusteras kommer detta göra att eNPV ökar. Därmed kommer eNPV vara icke-linjär, eller svagt icke-linjär i dessa variabler, till skillnad från exempelvis kostnadsvariabler, som förväntas öka eNPV linjärt. Detta framgår tydligt i den grafiska känslighetsanalysen, men även i graferna som erhöles från Morris metod, där $\sigma \ll 1$ för samtliga kostnadsvariabler.

Den variabel som avgör hur stor andel av populationen som insjuknar och behöver vår läkemedelskandidat, `Prevalence`, kommer trots dess låga varians ha ett väsentligt bidrag till känsligheten i modellen. Detta beror på att `prevalence` skalas med en faktor av storleksordning 10^8 på samtliga ställen där den används.

Det bör poängteras att känsligheter bör hanteras på olika sätt av en modellskapare beroende på vad den känsliga variabeln i fråga representerar. Variabler såsom `prevalence` och `POC_RecruitmentRate` kan vara omöjliga att kontrollera. Vidare bör den som konstruerar modellen se till att de antagna fördelningarna av dessa variabler är så exakta som möjligt. Annars måste man överväga ett annat sätt att modellera vad variabeln i fråga representerar. Exempelvis, i fallet med `POC_RecruitmentRate` vore det fördelaktigt att dela upp denna hastighet, och låta varje klinik ha en egen rekryteringstakt. Däremot finns variabler som går att kontrollera rent praktiskt. De resultat som presenterats tyder på att det är viktigare att lansera tidigt än att få ned kostnaden för vardera studie. Alltså är det avgörande för att öka sannolikhet för vinst på ett läkemedelsprojekt, att planera projektet på ett sådant sätt att den förväntade lanseringstiden sker så tidigt som möjligt.

6.1 Hantering av modeller

Genom att skapa vår egen modell, med Accelerated i SUM som utgångspunkt, har vi kunnat välja hur vi bäst kan behandla de Bernoullivariabler som finns. Istället för att inkludera Bernoullivariablerna i modellen valde vi att integrera dess sannolikheter i funktionen. Resultatet blev att vi tydligare kunde se vilka andra variabler som bidrog till känsligheten i modellen. Detta är möjligt tack vare antagandet att variablerna i Accelerated är oberoende. Detta kommer i regel att gälla, eftersom det enda som utnyttjas i avsnitt 4.1, ekvation (4.2) är att Bernoullivariablerna är oberoende av varandra. Dessutom att Bernoullivariablerna som avgör ifall projektet går vidare till nästa fas är oberoende av kostnaden i nästa fas. Alltså gäller att, även om modellen hade utvecklats så att Bernoullivariablerna är beroende av aktiviteter i fasen innan, hade vår uppdelning fortfarande varit giltig.

Vi begränsade känslighetsanalysen till endast NPV som utvariabel. SUM returnerar fler projektvärdesstorheter som inte behandlas alls i detta kandidatarbete. Identifieringen av de mest känsliga variablerna hade möjligtvis blivit annorlunda om vi hade valt att undersöka en annan projektvärdesstorhet. Trots detta kan samtliga metoder vi studerat användas, oberoende av projektvärdesstorhet, eftersom vår hantering av Bernoullivariabler är applicerbar på al-

la projektvärdesstorheter. Exempelvis om sannolikheten att göra förlust undersöktes skulle följande kunna utnyttjas:

$$\begin{aligned} P(NPV < 0) &= P(NPV < 0 | I_1 I_2 I_3 I_4 = 1)p + \underbrace{P(NPV < 0 | I_1 I_2 I_3 I_4 = 0)}_{=1}(1 - p) = \\ &= P(NPV < 0 | I_1 I_2 I_3 I_4 = 1)p + (1 - p) \quad (6.1) \end{aligned}$$

där $p = P(I_1 I_2 I_3 I_4 = 1)$. Vidare, om (6.1) utnyttjas exkluderas alla Bernoullivariablerna ur modellen.

Det bör även poängteras att en känslighetsanalys kan utföras på flera utvariabler samtidigt. Detta hade möjligen gett en bättre beskrivning av känsligheten i de ingående variablerna. Nackdelen med detta är naturligtvis att många fler beräkningar krävs.

6.2 Applicering i SUM

Captario kan använda våra resultat för att förstå sina modeller bättre och till och med välja att bygga dem annorlunda. En känslighetsanalys i deras fall behöver inte bara vara för att göra prognoser utan även för att förbättra modeller. Ett exempel som nämndes tidigare är att `POC_RecruitmentRate` är den mest influerande variabeln. Detta beror troligtvis på att den används i alla tre kliniker i modellen, och om variabelns fördelning skattas individuellt för varje klinik så kommer känsligheten i den att minska.

Från våra resultat har vi kunnat hjälpa Captario att komma till insikter kring en modell i SUM, men en viktig fråga är om Captario, tillsammans med deras kunder, kan använda de känslighetsanalysmetoder vi har använt oss av i framtida modeller. Både Sobols index och Morris metod kräver evaluering av modellen i specificerade simulerade variabler. I exempelvis Morris metod krävs detta för att beräkna varje elementär effekt. Idag har inte SUM den funktionen, och vidare måste Captario utveckla nya applikationer till SUM för att genomföra nya simuleringar. Den metoden som vi ansåg var mest informativ – Sobol index – är för beräkningmässigt tidskrävande för Captario om programmen som vi tillämpade används. Här finns utrymme för effektivisering av koden, och vid integrering av denna metod i SUM, skulle potentiellt de simuleringar kunna anpassas på ett sådant sätt att simuleringar kan återanvändas i flera metoder. Ett annat alternativ vore att använda FAST, vilket drar ned beräkningstiden väsentligt, framförallt för första ordningens Sobol index.

Förutom att MCF kan användas med simuleringsdata direkt från SUM är metoden snabb och dessutom anpassningsbar. Genom att göra ändringar i SUM och justera tröskelvärden som delar upp simuleringsdatan kan en användare i SUM använda MCF för att kolla på olika fall. Exempelvis kan en användare välja att bara kolla på projekt som når lansering och göra en känslighetsanalys för att ta reda på vilka ingående variabler som driver upp ett väldigt lönsamt projekt.

6.3 Framtida frågeställningar

Många metoder vi har använt har varit beräkningmässigt tidskrävande, framförallt beräkningen av Sobols index och den grafiska känslighetsanalysen. Detta beror huvudsakligen på att funktionen vi har använt i implementeringen inte är beräkningmässigt effektiv. För att göra metoderna lättare att använda för Captario hade vår kod kunnat effektiviseras så att beräkningstiderna hade minskats.

Vårt huvudsakliga fokus har varit på variationer i ingående variabler. Eftersom många av de ingående variablerna tar sina värden från skattade fördelningar kan det vara intressant att utföra en känslighetsanalys på parametrarna från dessa fördelningar. På så sätt kan man bygga

mer robusta modeller genom att förstå hur responsen påverkas av de antaganden man gjort. En sådan analys hade varit mer beräkningsmässigt krävande då eNPV behöver uppskattas för varje parameteruppsättning. Av de metoder som behandlats i detta kandidatarbete vore Morris metod den mest lämpade för en sådan typ av analys, eftersom den kräver förhållandevis få funktionsanrop. Dessutom kan man anta att variabler som inte ger något utslag på en känslighetsanalys inte har någon känslighet i sin parameteruppsättning. Man kan alltså avskärma icke-signifikanta variabler inför en sådan analys.

I Accelerated antas att alla ingående variabler är oberoende, vilket dessutom varit ett genomgående antagande i samtliga metoder vi betraktat. I mer verklighetstroga modeller bör detta inte vara fallet. Exempelvis om en aktivitet tar lång tid bör detta innebära en högre kostnad. För att få mer verklighetstroga modeller och kunna utföra en känslighetsanalys kan modellerna byggas så att exempelvis tid och kostnad för en viss aktivitet slås samman till variabler som är oberoende i sig. På detta vis skulle känslighetsanalyser kunna utföras på mer verklighetstroga modeller i SUM.

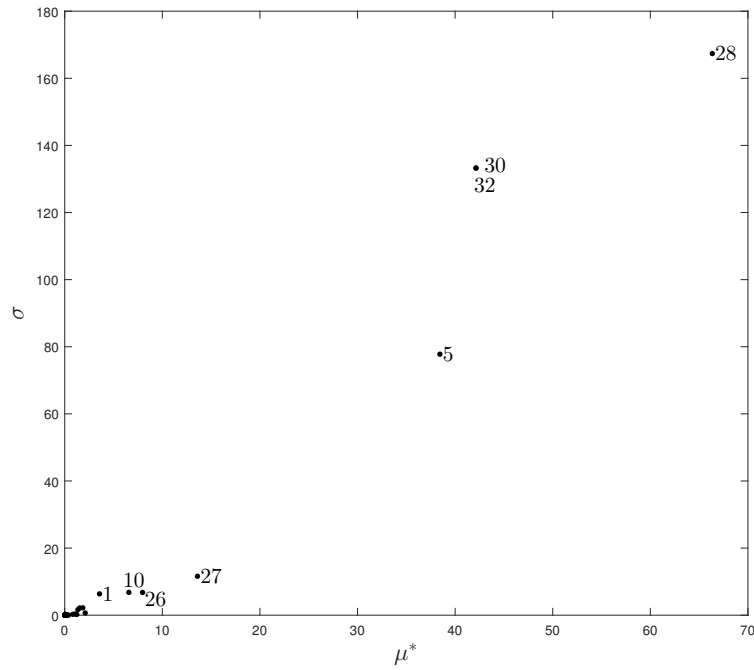
7 Referenser

- [1] Ciociola AA, Cohen LB, Kulkarni P. *How Drugs are Developed and Approved by the FDA: Current Process and Future Directions*, The American Journal of Gastroenterology. 2014 Maj;109(5):620-3.
doi:10.1038/ajg.2013.407
- [2] DiMasi JA, Grabowski HG, Hansen RW. *Innovation in the pharmaceutical industry: New estimates of R&D costs*. Journal of Health Economics. 2016 Maj;47:20-33.
doi:10.1016/j.jhealeco.2016.01.012
- [3] Paul SM, Mytelka DS, Dunwiddle CT, Persinger CC, Munos BH, Lindborg SR, et al. *How to improve R&D productivity: the pharmaceutical industry's grand challenge*. Nat Rev Drug Discov. 2010 mars;9(3):203-14.
doi:10.1038/nrd3078
- [4] Investopedia. *Net Present Value - NPV* [Internet]. New York: Investopedia, LLC; 2017 [citerad 30 mars 2017]. Hämtad från: <http://www.investopedia.com/terms/n/npv.asp>
- [5] Iooss B, Saltelli A. *Introduction to Sensitivity Analysis*. I: Ghanem R, Higdon D, Owhadi H, redaktörer. Handbook of Uncertainty Quantification. 1:a uppl. Zürich. Springer International Publishing; 2015. 1-17.
- [6] Saltelli A, Tarantola S, Campolongo F, Ratto M. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. 1:a uppl. Chichester, England: John Wiley Sons Ltd; 2002. 45.
- [7] Iooss B, Lema P. *A Review on Global Sensitivity Analysis Methods*. I: Dellino G, Meloni C, redaktörer. Uncertainty Management in Simulation Optimization of Complex Systems. 59 uppl. New York. Springer; 2015. 101-22.
- [8] Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, et al. *Global Sensitivity Analysis*. 1:a uppl. Chichester, England: John Wiley & Sons Ltd; 2008.
- [9] Saltelli A, Chan K, Scott EM. *Sensitivity analysis*. 4:e uppl. Chichester, England: John Wiley Sons Ltd; 2006.
- [10] Daniel C. *One-at-a-Time Plans*. Journal of the American Statistical Association. 1973 jun;68(342):353-60.
- [11] Morris MD. *Factorial Sampling Plans for Preliminary Computational Experiments*. Technometrics. 1991 maj;33(2):161-74.
- [12] Campolongo F, Cariboni J, Saltelli A. *An effective screening design for sensitivity analysis of large models*. Environmental Modelling and Software. 2007;22(10):1509–18.
Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, et al. *Global Sensitivity Analysis*. 1:a uppl. Chichester, England: John Wiley & Sons Ltd; 2008.
- [13] Saltelli A, Tarantola S, Campolongo F, Ratto M. *Sensitivity analysis in practice*. 1:a uppl. Chichester, England: John Wiley & Sons Ltd; 2004.
- [14] Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S. *Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index*. Computer Physics Communications. 2010 feb;181(2):259-70.
- [15] Sobol IM. *Sensitivity Estimates for Nonlinear Mathematical Models*. Mathematics and Computers in Simulation 2001;55(1):271-280.
- [16] Cukiert RI, Fortuin CM, Shuler KE, Petschek AG, Schaibly JH. *Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory*. The Journal of Chemical Physics. 1973;59(8):3873-8.

- [17] Saltelli A, Tarantola S, Chan K. P.-S. *A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output*. *Techometrics*. 1999;41(1):39-56.
- [18] Deniz A, Ferreira E, Johansson E, Johansson H, Lindbäck J, Wiskman L. *sa-accelerated* GitHub repository, 2017.
<https://github.com/wiskman/sa-accelerated>
- [19] Sohier H, Piet-Lahanier H, JL Farges. *Analysis and optimization of an air-launch-to-orbit separation* *Acta Astronautica*. 2015 mar/apr ;108:18-29.
<http://doi.org/10.1016/j.actaastro.2014.11.043>
- [20] Saltelli A. *Sensitivity Analysis for importance Assessment*. *Risk Analysis*. 2002;22(3):579-90.

Bilagor

A Morris metod på modell med Bernoullivariabler



Figur A.1: Diagram med absoluta medelvärde μ^* och standardavvikelse σ för de elementära effekterna associerade till alla variabler. Endast de variabler med höga index är markerade med variabelnamn.

B Morris Implementeringsalgoritmen

Morris implementeringsmetod [11] bygger på att hyperkuben diskretiseras likformigt med n steg i varje dimension. Alltså betraktas den diskreta delmängden $\Omega = \{\frac{1}{n+1}, \frac{2}{n+1}, \dots, \frac{n}{n+1}\}^p$ istället för hela parameterrummet. Δ förbestäms till en heltalsmultipel av $\frac{1}{n+1}$. I nätet Ω evalueras f i varje nod. För att skatta en elementär effekt för varje variabler föreslog Morris följande tillvägagångssätt:

- i) Specificera en designmatris \mathbf{B} av storlek $l \times p$ där $l = p + 1$ och \mathbf{B} innehåller enbart ettor och nollor. Denna ska även uppfylla att i varje kolonn $i = 1, 2, \dots, p$ så måste de i första elementen vara samma - vilka ofta sätts till 0. Exempelvis kan designmatrisen väljas till en nedåt triangulär matris med $b_{ij} = 0$ om $i \leq j$ och $b_{ij} = 1$ annars. $\Delta \mathbf{B}$ kan i princip användas för att skatta en elementär effekt för $(\Delta \mathbf{B})_{i+1}$ och $(\Delta \mathbf{B})_i$ skiljer sig med Δ i en komponent. Men i och med att $\Delta \mathbf{B}$ inte är slumpmässig, måste denna uppdateras.
- ii) Generera en slumpmässig $p \times p$ matris \mathbf{D}^* med 1 och -1 på diagonalen och de resterande elementen satta till 0. Diagonalelementen generas oberoende under villkoret att $P(d_{ii}^* = 1) = P(d_{ii}^* = -1) = \frac{1}{2}$. Låt sedan $\mathbf{N} = \frac{1}{2}((2\mathbf{B} - \mathbf{1}^{l \times p})\mathbf{D}^* + \mathbf{1}^{l \times p})$. Den första faktorn $(2\mathbf{B} - \mathbf{1}^{l \times p})$ kommer ha samma värden som \mathbf{B} i alla dess nollskilda element. De andra elementen är blir -1 . Om sedan \mathbf{D}^* verkar på denna uppdaterade matris byts tecken på dess kolonner enligt diagonalelementen i \mathbf{D}^* . Utförs sedan de sista operationerna för att erhålla \mathbf{N} kommer de element i $(2\mathbf{B} - \mathbf{1}^{l \times p})\mathbf{D}^*$ som är 1 bevaras, medan de element som är negativa bli 0. Vidare kommer varje kolonn i \mathbf{N} vara lika med motsvarande kolonn i \mathbf{B} eller i $\mathbf{1}^{l \times p} - \mathbf{B}$ enligt simuleringen av \mathbf{D}^* . Två rader på varandra i \mathbf{N} kommer endast skiljas i ett element med ± 1 . Detta är på grund av att två rader på varandra i \mathbf{B} skiljer sig åt i exakt en kolonn. Efter uppdateringen kommer dessa element fortfarande skiljas åt, potentiellt i omvänd ordning. I de andra kolonnerna, där båda raderna är lika med 1 eller 0 samtidigt, kommer båda dessa element bevaras, eller ändras samtidigt efter uppdateringen till \mathbf{N} .
- iii) Skapa en startvektor $\mathbf{x}^* \in [0, 1]^p$ där alla element är slumpade från $\{\frac{1}{n+1}, \frac{2}{n+1}, \dots, \frac{n}{n+1} - \Delta\}$.
- iv) En slumpmässig inverterbar permutationsmatris \mathbf{P}^* av storlek $p \times p$ genereras i vilken varje kolonn innehåller exakt en etta och resten nollor. Med hjälp av denna beräknas en så kallad slumpmässig orientering, vilken är definerad: $\mathbf{B}^* = (\mathbf{X}^* + \Delta \mathbf{N})\mathbf{P}^*$, där \mathbf{X}^* är en $l \times p$ -matris vars rader är \mathbf{x}^* . Till följd av \mathbf{N} 's egenskap, kommer skillnaden mellan två rader i $(\mathbf{X}^* + \Delta \mathbf{N})$ endast skilja med $\pm \Delta$. Om \mathbf{P}^* verkar på denna matris kommer dess kolonner att byta plats slumpmässigt. Vidare kommer även \mathbf{B}^* att ha denna egenskap.

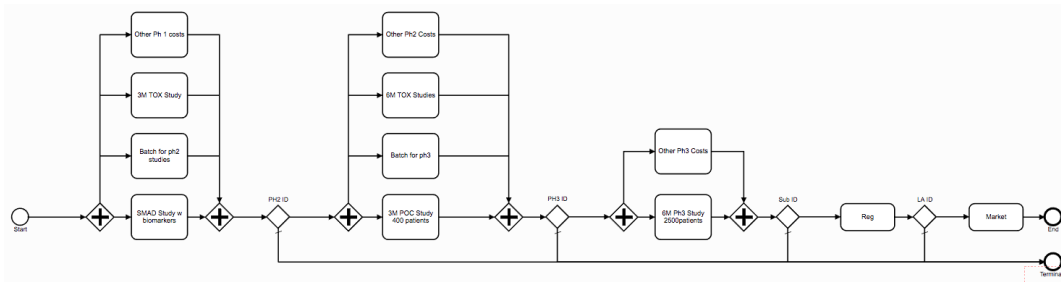
En slumpmässig orientering kan användas för att beräkna en elementäreffekt för varje variabel, i vilken raderna parvis utgör den data som fordras i (3.3). Detta genom

$$E_i(\mathbf{x}^{(i)}) = \pm \frac{f(\mathbf{x}^{(i+1)}) - f(\mathbf{x}^{(i)})}{\Delta} \quad (\text{B.1})$$

där $\mathbf{x}^{(i)}$ är den i :te raden i \mathbf{B}^* och tecknet motsvarar det tecken som det enda nollskilda element i $\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}$ har. Om man är intresserad av m stycken elementära effekter, för att uppskatta storheterna (3.4) och (3.5), genereras m stycken olika slumpmässiga orienteringar oberoende från varandra, vilka används för att beräkna m stycken elementära effekter för p variabler.

C Modellbeskrivning av Accelerated

Nedan följer en mer detaljerad beskrivning av Captarios exempelmodell Accelerated som vår känslighetsanalys är utförd på. Accelerated-modellen består av 4 faser, varav varje fas består av olika aktiviteter som pågår parallellt. Projektet antas starta i mars 2016 (år 2016.3) och avslutas i början av 2047. Nedan är en bild som illustrerar starten till slutet av ett läkemedelsprojekt med accelerated-modellen:



där varje rektangel representerar en aktivitet, och varje romb är ett *vägskäl*. En ifylld romb innebär att flera aktivitet startas alternativt avslutas och de tomma romberna avgör om läkemedlet går igenom fasen till vänster. Här nedan kommer vi beskriva de olika faserna med tillhörande variabler, såsom kostnader och tider. Vi låter $\text{LogN}(a, b)$ beteckna en LogNormalfördelning med a som 5%-kvantil och b som 95%-kvantil.

C.1 Variabelförteckning

Första fasen består av 4 olika aktiviteter: *SMAD Study w biomarkers*, *Batch for ph2 studies*, *3M TOX Study* och *Other ph1 activities*. De variabler som tillhör vardera aktivitet är. Vi låter dessutom T_{ij} vara tiden det tar att genomföra aktivitet j i den t :te fasen. C_{ij} och \hat{C}_{ij} betecknar den diskonterade, respektive den icke-diskonterade kostnaden för just denna aktivitet.

Aktivitet	Variabelnamn	Fördelning	Förklaring
SMAD Study w biomarkers	cadtime Development_costs1	$\text{LogN}(2/3, 5/6)$ $\text{LogN}(2, 11/5)$	tiden kostnaden
Batch for ph2 studies	Time Development_costs2	$\text{LogN}(3/5, 4/5)$ $\text{LogN}(3/10, 2/5)$	tiden kostnaden
3M TOX Study	3mToxTime 3mToxCost	$\text{LogN}(1/2, 4/5)$ $\text{LogN}(1/4, 3/10)$	tiden kostnaden
Other ph1 activities	cadtime Other_Ph1_Costs	– $\text{LogN}(3/2, 9/5)$	tiden (se SMAD) kostnader

Dessutom gäller att sannolikheten att projektet går vidare till nästa fas 50%. Det vill säga $I_1 \sim \text{Bernoulli}(0.5)$.

Andra fasen består av 4 olika aktiviteter: *3M POC Study 400 patients*, *Batch for ph3*, *6M TOX Studies* och *Other ph2 Costs*. Associerade variabler till dessa är:

Aktivitet	Variabelnamn	Fördelning	Förklaring
3M POC Study 400 patients	POC_SetupTime POC_AnalysisTime POC_RecruitmentRate POC_SetupCost POC_AnalysisCost POC_CostPerPatient POC_CostPerCenter	LogN(1/10, 1/5) LogN(1/10, 3/20) LogN(1, 3) LogN(1/10, 3/10) LogN(1/10, 1/5) LogN(3/100, 1/25) LogN(1, 6/5)	
Batch for ph3	batchForPH3Time batchForPH3Cost	LogN(0.6, 1) LogN(1, 3)	tiden kostnaden
6M TOX Studies	6MToxTime 6MToxCost	LogN(1, 1.1) LogN(0.3, 0.4)	tiden kostnaden
Other ph2 Costs	3MPOCTime Development_costs_3	LogN(0.75, 1) LogN(2, 3)	tiden kostnaden

Tiden för 3M POC Study 400 patients fås genom

$$T_{2,1} = \text{POC_SetupTime} + \frac{400}{3 \times 12 \times \text{POC_RecruitmentRate}} + \text{POC_AnalysisTime} + 0.25$$

där 400 är antalet patienter i studien, 3 är antalet kliniker och 12 konverterar enheten till år, eftersom $\text{POC_RecruitmentRate}$ är rekryteringstakt i patient/(klinik \times månad). Den sista termen är en fix behandlingstid. Kostnaden för denna aktivitet fås av:

$$\hat{C}_{2,1} = \text{POC_SetupCost} + \text{POC_AnalysisCost} + \text{POC_FSI-LSO_Cost},$$

$$\text{POC_FSI-LSO_Cost} = 400 \times \text{POC_CostPerPatient} + 3 \times \text{POC_CostPerCenter}.$$

Projektet går vidare till nästa fas med sannolikheten 0.5. Alltså gäller att $I_2 \sim \text{Bernoulli}(0.5)$. Tredje fasen består av 2 aktiviteter: *6M Ph3 Study 2500 patients* och *Other Ph3 Costs*. Här är en lista på variablerna som tillhör samtliga aktiviteter

Aktivitet	Variabelnamn	Fördelning	Förklaring
6M Ph3 Study 2500 Patients	Ph3StudyTime ph3studyCosts	LogN(1, 1.5) LogN(75, 95)	tiden kostnaden
Other Ph3 Costs	Ph3StudyTime ph3OtherCosts	– LogN(10, 20)	tiden (se 6M Ph3) kostnaden

Sannolikheten att gå igenom fas 3 är 70%, vidare gäller att $I_3 \sim \text{Bernoulli}(0.7)$.

Registreringsfasen är bara en aktivitet. Vi går alltid igenom registreringsfasen. Här är en lista på variablerna som tillhör registreringsfasen:

Variabelnamn	Fördelning	Förklaring
RegTime	LogN(1, 1.1)	tiden
RegCosts	LogN(4, 6)	kostnaden

Sannolikheten att gå igenom registreringsfasen är 90%, och därför blir den sista bernoullivariabeln $I_4 \sim \text{Bernoulli}(0.9)$.

C.2 Fasernas kostnader och tider

Tiden att genomföra den i :te fasen är $T_i := \max\{T_{i,1}, T_{i,2}, \dots, T_{i,n}\}$. Den diskonterade kostnaden för en fas fås genom att summera alla diskonterade kostnader för alla aktiviteter i fasen, det vill säga $C_i = \sum_j C_{i,j}$. För att diskontera kostnaden för en aktivitet antas att kostnaden är likformigt fördelad över hela tidsintervallet aktiviteten äger rum. För årsvis diskontering blir den nuvärdesjusterade kostnaden

$$C_{i,j} = \int_{T_i^{(s)}}^{T_{i,j} + T_i^{(s)}} \frac{\hat{C}_{i,j}}{T_{i,j}} r^{[t]+1} dt \quad (\text{C.1})$$

där fasens starttid $T_i^{(s)} = \sum_{k < i} T_k$ och $[t]$ är heltalsdelen av t .

C.3 Marknadsmodellen

Vi antar att 3 konkurrenter är verksamma på samma marknad som läkemedelskandidaten kommer säljas på. De variabler som används i marknadsfasen är ordnade nedan.

Variabelnamn	Fördelning
Prevalence	LogN(0.0026, 0.0029)
RampTimeInYears	LogN(3, 5)
Comp1DoesLaunch	Bernoulli(0.2)
Comp1LaunchTime	Tria(2021.2, 2022.3, 2023.1)
Comp2DoesLaunch	Bernoulli(0.2)
Comp2LaunchTime	Unif(2021.2, 2022.5)
Comp3DoesLaunch	Bernoulli(0.2)
Comp3LaunchTime	LogN(2020.8, 2023.5)

där $\text{Unif}(a, b)$ står för likformiga fördelningen med parametrar a och b , och $\text{Tria}(a, b, c)$ står för triangelulärfördelningen med parametrar a och b och c . Kostnaden under marknadsfasen antas vara 10% av den årliga intäkten. Vidare kommer den totala diskonterade nettointäkten vara

$$R - C_m = \int_{T_{RE}}^{T_e} (\hat{R}_A(t) - \hat{C}_{m,A}(t)) \cdot r^{[t]+1} dt = 0.9 \int_{T_{RE}}^{T_e} \hat{R}_A(t) \cdot r^{[t]+1} dt \quad (\text{C.2})$$

där lanseringstiden definieras $T_{RE} = \sum T_i$, $\hat{R}_A(t)$ och $\hat{C}_{m,A}$ är den årliga intäkten, respektive årliga kostnaden under marknadsfasen. Dessutom är T_e den tid från start då projektet avslutas, det vill säga $T_e = 2047 - 2016.3 = 30.7$. Den årliga avkastningen fås genom

$$\hat{R}_A(t) = 0.00365 \times \text{EligibleMarket}(t) \times \text{MarketShare}(t)$$

där 0.00365 är försäljningspriset, och de resterande faktorerna utgör försäljningsvolymen. EligibleMarket är den totala tillgängliga marknaden, vilken modelleras enligt

$$\text{EligibleMarket}(t) = 0.25 \times \text{Prevalence} \times 321000000 \times 1.0078^{t+2.3}$$

där 0.25 är de andelar av alla insjukna som kommer ha behov av medicinen som utvecklas, Prevalence beskriver hur stor utbredning sjukdomen får och sista faktorerna är en modell

för hur USA:s befolkning ökar över tid. `MarketShare`, beskriver hur många marknadsandelar som erhålls, vilken bestäms genom

$$\text{MarketShare}(t) = \begin{cases} \frac{\text{PeakMarketShare} \times (t - T_4)}{\text{RampTimeInYears}}, & t \leq (\text{RampTimeInYears} + T_4) \\ \text{PeakMarketShare}, & (\text{RampTimeInYears} + T_4) < t \leq 16.9 \\ 0.1 \times \text{PeakMarketShare}, & t > 16.9 \end{cases}$$

där 16.9 tiden till patentet av produkten går ut. Under antagandet att `PeakMarketShare` är konstant, modelleras alltså att företagets marknadsandelar ökar linjärt upp till en den maximala marknadsandelen `PeakMarketShare`. Efter att patentet gått ut, förloras 90% av alla marknadsandelar. I allmänhet kommer `PeakMarketShare` inte vara konstant, utan kommer vara beroende på vilka konkurrenter som lanserat. Denna variabel fås av

PeakMarketShare	NumberOfEntries och OrderOfEntry
0.5	NumberOfEntries = 1, OrderOfEntry = 1
0.4	NumberOfEntries = 2, OrderOfEntry = 1
0.2	NumberOfEntries = 2, OrderOfEntry = 2
0.3	NumberOfEntries = 3, OrderOfEntry = 1
0.2	NumberOfEntries = 3, OrderOfEntry = 2
0.15	NumberOfEntries = 3, OrderOfEntry = 3

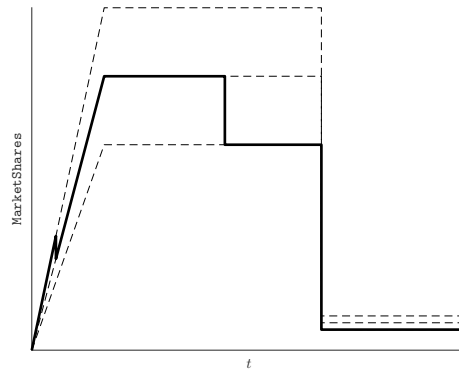
där `OrderOfEntry` och `NumberOfEntries` fås av:

$$\text{OrderOfEntry} = 1 + \sum_i \text{Comp}(i)\text{DoesLaunch} \times I_{A_i}(T_4)$$

$$\text{NumberOfEntries} = 1 + \sum_i \text{Comp}(i)\text{DoesLaunch} \times I_{A_i}(t)$$

$$I_{A_i}(t) = \begin{cases} 1, & \text{Comp}(i)\text{LaunchTime} - 2016.3 \leq t \\ 0, & \text{Comp}(i)\text{LaunchTime} - 2016.3 > t \end{cases}$$

Ett sätt utvecklingen i erhållna marknadsandelar kan bli illustreras i C.1



Figur C.1: Den tjockare linjen är ett scenario av vilka marknadsandelar som erhålls över tid, i vilken en konkurrent lanserar innan $(\text{RampTimeInYears} + T_4)$ och en efter. De streckade graferna är den utveckling som hade skett om inga konkurrenter lanserat efter företaget lanserat läkemedlet.

C.4 Andra projektvärdesstorheter

Tabell C.1: Lista över andra projektvärdesstorheter Captario SUM beräknar

	Förklaring
NPV	Summan av all framtida nuvärdesjusterad kassaflöde.
eNPV	Väntevärdet av NPV.
Nominell kostnad	Summan av alla kostnader under projektets gång.
Nominell avkastning	Summan av alla intäkter under projektets gång.
ROINom	Nominell avkastning på investering: $\frac{\text{Nominell avkastning} - \text{Nominell kostnad}}{\text{Nominell kostnad}}$.
ROIPV	Nuvärdesjusterad avkastning på investering.
p_f	Risken att göra förlust: $P(\text{NPV} < 0)$.
T_{RE}	Lanseringstid.
T_A	Tid tills kliniska studier avslutar. Speciellt är $T_A = T_{RE}$ om läkemedlet lanserar.