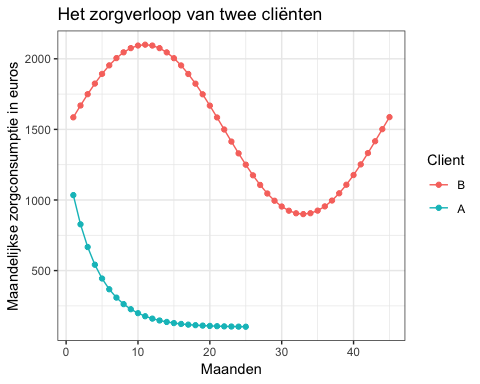
Deel 2 Statistiek & Doelmatigheid

# Het vraagstuk



In de grafiek hierboven zijn de evoluties van de zorgconsumptie van twee gesimuleerde cliënten (cliënt A en B) te zien. Cliënt A en B zijn dus geen echte cliënten, al zijn de simulaties wel gebaseerd op de evoluties die wij waarnemen.

Cliënt A heeft een relatief kort behandelingstraject. De kosten nemen gestaag af en stabiliseren na zo een anderhalf jaar. Het behandelingstraject van cliënt B duurt langer. Zijn zorgconsumptie intensiveert eerst, daalt daarna en lijkt na ongeveer drie jaar weer toe te nemen.

Het contrast tussen deze twee cliënten roept een vraag op: waarom verschillen de evoluties van deze twee gesimuleerde cliënten zo van elkaar? In bredere zin, waarom verschillen de zorgconsumptie-evoluties van elkaar?

Wij bij BuurtzorgT proberen verschillen tussen onze cliënten te verklaren. Daarvoor maken we gebruik van modellen die uit de biostatistiek komen. Deze modellen zijn onmisbaar bij het modelleren van processen die veranderen door de tijd heen omdat ze gebruik maken van **herhaalde metingen**. We kunnen aan de hand van deze modellen kijken welke factoren deze veranderingen teweeg brengen. In het beste geval hopen we deze evoluties te kunnen gaan voorspellen.

Zorgconsumptie zijn herhaalde metingen omdat we per cliënt per maand voor een langere periode kunnen meten wat de zijn zorgconsumptie is geweest.

# De onderdelen van het onderzoek

Statistici houden zich over het algemeen bezig met het verklaren van verschillen. Vanuit een puur statistisch perspectief verschillen de behandelingsstrajecten op twee manieren van elkaar:

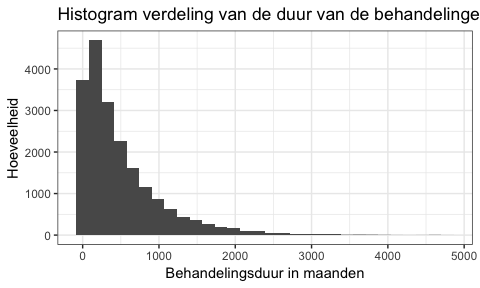
* De hoogte van de maandelijkse kosten ofwel de intensiteit van de behandeling: sommige cliënten kosten meer dan anderen en deze kosten ontwikkelen zich sterk verschillend.
* De duur van het behandelingstraject: de ene behandeling duurt korter/langer dan de andere.

De hoogte en de duur zijn de twee uitkomstvariabelen van het model. Deze moeten gezamenlijk gemodelleerd worden. Om de hoogte van de zorgconsumptie te modelleren belichten we verdeling van deze variabele en de hiërarchie in deze data. Daarna kijken we naar de duur van de behandelingsstrajecten en de variabelen die we gebruiken om de twee hoofduitkomstvariabelen te modelleren.

## Onderdeel 1: De verdeling van de maandelijkse zorgconsumptie

Ten eerste proberen wij te begrijpen waarom de ene cliënt meer kost dan de andere. Het is gebruikelijk om eerst naar de verdeling van de kosten, de uitkomstvariabele, te kijken middels een histogram. Een histogram laat zien hoe vaak een bepaalde waarde voorkomt.

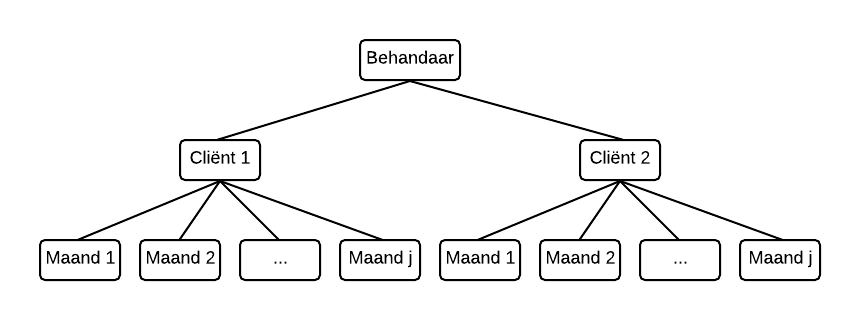
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



De histogram laat zien dat kosten scheef verdeeld zijn. Dat wil zeggen dat het leeuwendeel van de maanden gaan gepaard met zorgconsumptie ter hoogte van rond de 500 euro en dat er behoorlijke uitschieters zijn naar boven zijn.

Feitelijk komt geld voort uit tijd. Behandelaars besteden tijd aan een cliënt en dit zorgt voor kosten. We hebben het dus eigenlijk over een aantal minuten (of uren) per maand. Dat betekent dat we dus een aantal minuten per maand aan het modelleren zijn. Hiervoor moeten we zogeheten **count regressie** gebruiken.

Nu betrekken we het concept **herhaalde metingen** (longitudinaal is ongeveer hetzelfde) bij de analyse. Standaard (count) modellen nemen aan dat alle gewaarwordingen van de uitkomstvariabele (zorgconsumptie) onafhankelijk zijn van elkaar. Dit is niet zo omdat we meerdere metingen per cliënt hebben, maar ook omdat cliënten dezelfde behandelaars hebben. We spreken over een hiërarchie in de data. De modellen moeten dus worden uitgebreid om met de hiërarchie in de data om te kunnen gaan.

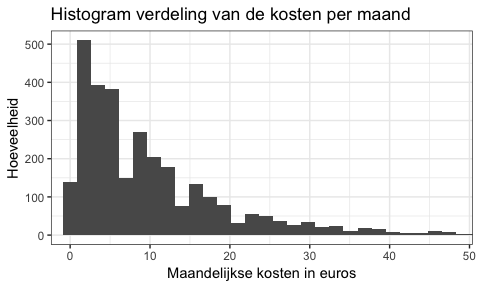


Hiërachie in de data

Zo een model wat rekening kan houden met hiërarchie is onderdeel van de **generalized linear mixed models**. Dit soort modellen worden breed ingezet in de medicijnontwikkeling en de beoordeling van de effectiviteit van (somatische) medische interventies.

## Onderdeel 2: De verdeling van de duur van de behandelingstrajecten

## Warning in which(len > 4) & which(len < 48): longer object length is not a  
## multiple of shorter object length



We zien aanzienlijke verschillen in de duur van de behandelingen, zelfs als we rekening houden met het feit dat sommige behandelingen nog gaande zijn terwijl we enkel data tot de afgelopen maand hebben. Enkele behandelingen duren zo’n vier jaar.

In de statistiek zijn speciale modellen ontwikkeld om de duur tot een bepaalde gebeurtenis te modelleren. Aan de hand van deze modellen, ook wel **time-to-event models** genoemd, kunnen we zien welke factoren de duur van een behandeling kunnen voorspellen.

De gebeurtenis in kwestie is dus het einde van de behandeling, om welke reden dan ook. Dit kan zijn wanneer een DBC wordt afgesloten, of de laatste maand met een zorgconsumptie. Of hebben behandelingsoverwegingen tot een staking van de behandeling geleid. Daarnaast kijken we ook of stabilisatie van de zorgconsumptie een statistisch relevant “einde” is.

## Onderdeel 3: Verklarende variabelen

Er zijn twee voorname categorieën van variabelen die verschillen in behandelingstrajecten moeten gaan verklaren: klinische gegevens uit het EPD en variabelen gerelateerd aan sociaal-economisch status. Een klankbordgroep bestaande uit BuurtzorgT-behandelaren denkt inhoudelijk met ons mee om zo veel mogelijk input uit de praktijk te krijgen.

Klinische gegevens uit het EPD zijn in eerste instantie gerelateerd aan de assen uit het DSM. Denk bijvoorbeeld aan comorbiditeit. In een later stadium gaan we kijken of we de notities van de behandelaars kunnen analyseren met behulp van **natural language processing**.

We verwachten dat variabelen gerelateerd aan sociaal-economische status meer verschillen kunnen gaan verklaren dan klinische gegevens. Behandelaars hebben aangegeven dat ze op het moment dat ze een wijk binnenlopen al een idee hebben van het behandelingstraject dat komen gaat.

Verschillen tussen (teams van) behandelaars verklaren waarschijnlijk ook verschillen tussen behandelingstrajecten. Een cliënt zal bij de ene behandelaar een ander traject dan bij een andere behandelaar. Statistici noemen dit **behandelaarseffecten** en generalized linear mixed models kunnen hier goed mee omgaan.

# De theorie der doelmatigheid

Hoe kunnen we aan de hand van de twee de uitkomstkvariabelen en de verklarende variabelen doelmatigheid definiëren?

De totale zorgconsumptie van een cliënt is een functie van de duur van het behandelingstraject een de maandelijkse hoogte van de zorgconsumptie. De behandeling van cliënt B in de eerste grafiek brengt duidelijk meer kosten voort dan de behandeling van cliënt A.

Zeer recente ontwikkelingen in de beschikbaarheid van software stellen ons in staan om de duur van een behandelingstraject en de evolutie van de zorgconsumptie gezamenlijk te modelleren, met de toepasselijke naam **joint models**. Het gebruik van dit model is belangrijk wegens twee redenen:

* We vermoeden dat een intensivering van de zorgconsumptie verband houdt tot de duur van een behandelingstraject.
* We denken ook dat cliënten met “zwaardere diagnoses” langer bij ons in behandeling zijn dan cliënten met kortere behandelingen. Als we niet corrigeren voor de duur van een behandelingstraject leidt dit tot bias. De bias komt voort uit de mogelijkheid dat bijvoorbeeld na twee jaar enkel cliënten met “zwaardere diagnoses” in onze longitudinale data overblijven.

Wij zijn druk bezig met het **definiëren van doelgroepen** aan de hand van beschikbare variabelen. Deze definitie moet verschillen tussen de evoluties van cliënten verklaren. We zijn dus op zoek naar definities die statistisch en klinisch relevant zijn.

Aan de hand van joint models kunnen wij tot twee inzichten komen:

1. Hoe ontwikkelt de gemiddelde zorgconsumptie-evolutie zich per doelgroep?
2. Wat is het effect van een intensivering van de zorgconsumptie op de resterende duur van de behandeling per doelgroep?

Als we dan doelgroepen met elkaar vergelijken, kunnen we zien welke doelgroepen doelmatiger behandeld worden. Doelmatigheid komt dan voort uit de mate waarin de gemiddelde zorgconsumptie afneemt. Een doelmatige behandeling ontwikkelt zich ook op een manier die de behandelingsduur verkort.

Dit betekent niet dat een doelmatige behandeling per se een laag niveau van zorgconsumptie heeft en korter van duur is. Per slot van rekening is er veel heterogeniteit tussen de cliënten en onze modellen proberen daar recht aan te doen.

Doelmatigheid betreft de ontwikkeling ten opzichte van een verwachtingspatroon dat invulling krijgt aan de hand van de joint models.

Software-matig maken wij gebruik van R Team (2019). In het bijzonder: JMbayesRizopoulos (2016), rstanarmGoodrich et al. (2018) en stan\_jmBrilleman et al. (n.d.).

Brilleman, SL, MJ Crowther, M Moreno-Betancur, J Buros Novik, and R Wolfe. n.d. “Joint Longitudinal and Time-to-Event Models via Stan.” <https://github.com/stan-dev/stancon_talks/>.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2018. “Rstanarm: Bayesian Applied Regression Modeling via Stan.” <http://mc-stan.org/>.

Rizopoulos, Dimitris. 2016. “The R Package Jmbayes for Fitting Joint Models for Longitudinal and Time-to-Event Data Using Mcmc.” *Journal of Statistical Software* 72 (7): 1–45. <https://doi.org/10.18637/jss.v072.i07>.

Team, R Core. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.