



Hulpmiddel inschatten overlevingskans COVID-19 patiënten voor IC opname

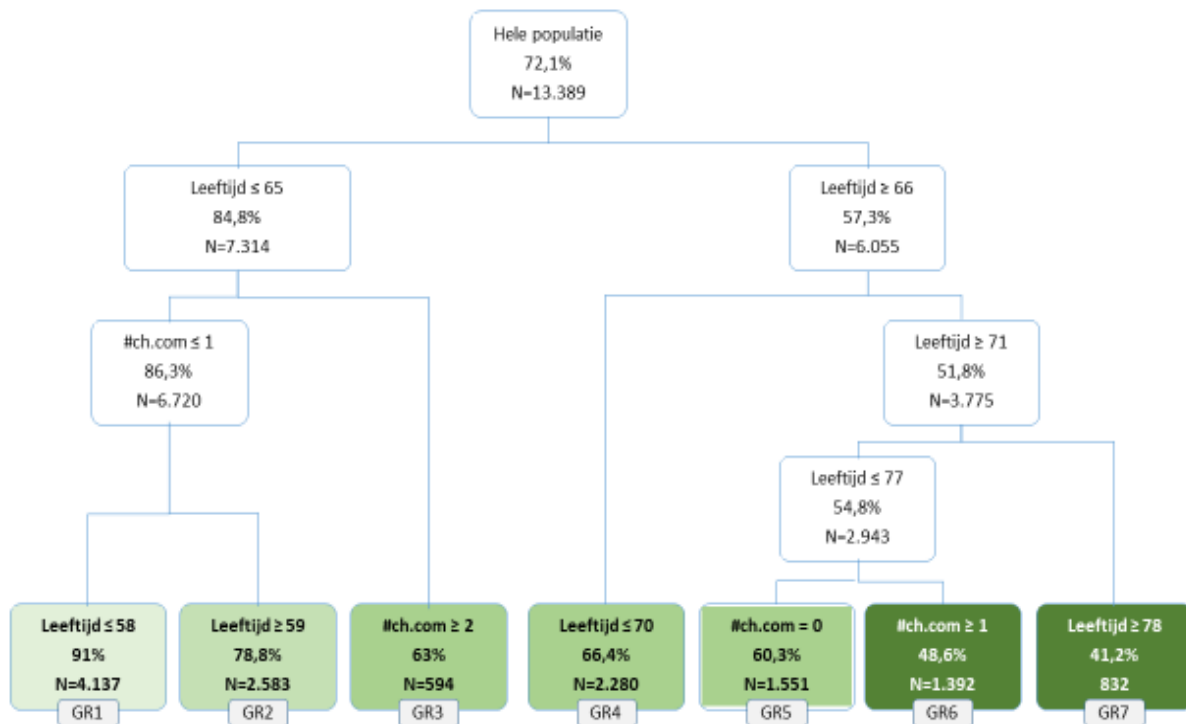
Doel: inzicht verkrijgen in de overlevingskans van COVID-19 patiënten vóór het moment van IC opname.

Populatie: Gegevens van 13.389 COVID-19 patiënten opgenomen op de IC tussen 27 februari 2020 en 23 november 2021. Gegevens bevatten NICE (MDS) data items die bekend zijn op het moment van IC opname: opname type, leeftijd, geslacht, chronische comorbiditeiten (en hun aantal), heropname, verwijzend specialisme, BMI, behandelduur voor IC-opname, en mortaliteit in het ziekenhuis. Onder het (aantal) chronische comorbiditeiten vallen: (1) immunologische insufficiëntie of aids; (2) chronisch nierfalen of chronische dialyse; (3) respiratoire insufficiëntie of COPD; (4) neoplasma of hematologische maligniteit; (5) cardiovasculaire insufficiëntie; (6) levercirrose; (7) diabetes.

Aanpak: We hebben een **regressieboom** geleerd uit de data. Dit is een stroomdiagram dat uit knopen en vertakkingen bestaat, met de hele patiëntengroep als startpunt. Deze groep wordt gesplitst in twee knopen (subgroepen). Elke knoop mag weer gesplitst worden in twee knopen (subgroepen). De splitsing is gebaseerd op een waarde van een variabele. Bijvoorbeeld “leeftijd \leq 78” splitst een patiëntengroep in twee subgroepen: jonger dan of gelijk aan 78 jaar en ouder dan 78 jaar. De variabele en de waarde worden automatisch door het leeralgoritme bepaald, en niet alle mogelijke variabelen hoeven gekozen te worden in de uiteindelijke boom. Bij elke knoop staat het overlevingspercentage in de betreffende groep alsook het aantal patiënten in die groep. Een knoop die niet meer gesplitst wordt heet een blad. Elk pad van het startpunt (hele studiepopulatie) tot een blad geeft een regel weer. De complexiteit van de boom (hoe diep de vertakkingen in de boom mag zijn) werd bepaald door 10-voudige kruisvalidatie. We presenteren de boom zowel als stroomdiagram, tabel en als losse regels.

Resultaat:

A. Regressieboom voor overleving (#ch.com staat voor aantal chronische comorbiditeiten):



Weergave in tabelvorm

| Groep | Leeftijd | Aantal comorbiditeiten | Overlevingskans |
|-------|--------------|------------------------|-----------------|
| 1 | ≤ 58 | ≤1 | 90% |
| 2 | ≥59 ≤ en ≤65 | ≤1 | 78,8% |
| 4 | ≥66 en ≤70 | | 66,4% |
| 3 | ≤ 65 | ≥ 2 | 62,5% |
| 5 | ≥71 en ≤77 | 0 | 60,3% |
| 6 | ≥71 en ≤77 | ≥1 | 48,6% |
| 7 | ≥ 78 | | 41,2% |

B. De expliciete regels zijn:

Groep 1: N=4137 (31% van dataset)

ALS leeftijd \leq 58 EN #ch.com \leq 1 DAN overlevingskans = 91%

Groep 2: N=2583 (19% van dataset)

ALS 59 \leq leeftijd \leq 65 EN #ch.com \leq 1 DAN overlevingskans = 78,7%

Groep 3: N=594 (4% van dataset)

ALS Leeftijd \leq 65 EN #ch.com \geq 2 DAN overlevingskans = 63%

Groep 4: N=2280 (17% van dataset)

ALS 66 \leq Leeftijd \leq 71 DAN overlevingskans = 66,4%

Groep 5: N=1.551 (12% van dataset)

ALS 71 \leq Leeftijd \leq 77 EN #ch.com =0 DAN overlevingskans = 60,3%

Groep 6: N = 1392 (10% van dataset)

ALS 71 \leq Leeftijd \leq 77 EN #ch.com \geq 1 DAN overlevingskans = 48,6%

Groep 7: N=832 (6% van dataset)

ALS Leeftijd \geq 78 DAN overlevingskans = 41,2%

Sensitiviteitsanalyse en bevindingen

We hebben gekeken naar:

- 1) de stabiliteit van de **gekozen variabelen** in de boom.
Aanpak: 200 bootstrap steekproeven.
Bevinding: de variabelen in de getoonde boom werden in de grote meerderheid van de bootstrap steekproeven gekozen en daarom zijn ze vrij stabiel.

- 2) de **variantie van overlevingskans** in de bladeren.
Aanpak: 200 bootstrap steekproeven.
Bevinding: de variantie was relatief klein behalve in groep 3
Groep 1 0.0824439 - 0.1001229
Groep 2 0.1969615 - 0.2280846
Groep 3 0.3311213 - 0.4124388
Groep 4 0.3131909 - 0.3546500
Groep 5 0.3778463 - 0.4239171
Groep 6 0.4888061 - 0.5344138
Groep 7 0.5601203 - 0.6182145

3) de **variatie in kansen** in de bladeren **volgens het beste logistische regressie model**.

Aanpak: We ontwikkelen eerst een logistisch regressie model met alle mogelijke huidige voorspellers. We selecteren vervolgens de beste voorspellers middels een achterwaartse variabele selectie procedure via de AIC (Akaike Information Criterion). We voorspellen de overlevingskansen van alle patiënten binnen elk blad en kijken naar het gemiddelde, de mediaan, en de 1^{ste} en 3^{de} kwartielen van deze voorspellingen. Opmerking: omdat logistische regressie, anders dan de boom, niet met missende waarden kan omgaan, is de analyse uitgevoerd op alleen de patiënten die geen missende waarden hadden voor de voorspellers van het verkregen logistische regressie model.

Bevinding: Zie tabel hieronder. De kansen van de regressieboom komen aardig overeen met de gemiddelden en medianen van de voorspellingen van het logistische regressie model.

| Groep | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---------------------|---------------------|-------------------|-----------------|---------------------|---------------------|---------------------|
| N | 4017 | 2496 | 578 | 2206 | 1504 | 1353 | 794 |
| Overlevingskans in groep (%) | 91,3 | 78,9 | 63,5 | 66,8 | 60,3 | 48,5 | 40,7 |
| Gem. LR model voorspelling in blad (%) | 90,4 | 78,2 | 67,5 | 66,3 | 61,9 | 48,7 | 42,5 |
| Mediaan LR model voorspelling en 1 ^{ste} en 3 ^{de} kwartielen (%) | 90,9 [87,1-94,8] | 79,5 [75,2-83,2] | 69 [58,3-78,1] | 69,3 [62-74] | 63,0 [57,4-67,1] | 50,8 [41,7-57,9] | 44,8 [35,2-52,5] |

4) het **onderscheidend vermogen** van de boom

Aanpak: Berekenen van de bias-gecorrigeerde AUC o.b.v. 200 bootstrap steekproeven.

Bevinding: de bias-gecorrigeerde AUC was 0.72 [0.71 - 0.74]

5) **stabiliteit** van de **overlevingspercentages** in zes COVID-golven.

Aanpak: We kijken naar de overlevingskansen binnen de COVID-golven in elk blad.

- a) golf 1: 01 feb 2020 – 15 mei 2020
- b) in between (golf 0): 16 mei 2020 - 30 sept 2020
- c) golf 2: 1 okt 2020 - 1 dec 2020
- d) golf 3: 1 dec 2020 - 31 jan 2021
- e) golf 4: 1 febr 2021 - 1 juli 2021
- f) golf 5: 1 juli 2021 – 30 sept 2021
- g) golf 6: 1 okt 2021 - heden

Bevinding: Binnen elk blad (groep) had elke COVID-golf een overlevingskans die over het algemeen vergelijkbaar was met de overlevingskans in het blad als geheel.

- 6) de **additionele voorspellende waarde van de bladeren** bovenop het beste (interactie-vrije) logistische regressie model.

Aanpak: We ontwikkelen een logistisch regressie model als in stap 3 met alle mogelijke huidige voorspellers en selectie van variabelen middels de AIC. Vervolgens voegen we aan onze dataset een “groep variabele” toe waarin per patiënt staat bij welke van de 7 subgroepen (bladeren) hij hoort. We vergelijken via een ANOVA analyse ons logistische regressie model met een tweede model dat ook de beschikking heeft tot de ‘groep’-variabele. Ook doen we variabele selectie middels de AIC met het tweede logistische regressie model om te kijken of de ‘groep’-variabele wordt geselecteerd.

Bevinding: Het model met de ‘groep’-variabele was significant beter dan het model zonder deze variabele. Ook werd deze ‘groep’-variabele geselecteerd na achterwaartse selectie gebaseerd op de AIC. Dit wil zeggen dat de groepen additionele voorspellende waarde hebben op het logistische regressie model zonder deze variabele.

- 7) **andere methoden voor het verkrijgen van regels**. Een nadeel van een boom is dat alle regels per definitie dezelfde variabele gebruiken die in de “wortel” van de boom (bovenaan) staat (leeftijd in ons geval).

Aanpak: Na het leren van de eerste boom verwijderen we daarna de gegevens van de patiënten in de hoogste risicogroep. We leren daarna weer een boom op de rest van de gegevens. We herhalen deze procedure. Op deze manier geven we kans aan andere variabelen om in de volgende bomen geselecteerd te worden.

Bevinding: In essentie kregen we een reeks van bomen die gezamenlijk toch dezelfde groepen hebben als onze oorspronkelijke boom.