Identification of Patient Profile Clusters and their Characteristics for ICU-admitted COVID-19 patients

B. Hoogerheide

Abstract

Introduction

A lot of research on the disease COVID-19 has already been performed, however most of these research papers held a overall high risk of bias. Additionally, the papers used multivariate logistic regression analysis on the entire research population. While useful for finding general risk- and prognostic factors, the effect of the parameters may be of little to no value to certain groups of people

Methods

Latent Profile Analysis (LPA) is proposed as a method to identify different profiles of patients within the research population. Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) were used to select the best fitting number of subpopulations. Entropy, Confidence Intervals, Odds of correct classification, posterior probabilities, and the correspondence between prior- and posterior probabilities were used for model assessment. The number of profiles was calculated for each day a patient had been in the ICU, for the first seven days of their stay.

Results

Most of the days had the best fit with three profiles, and two days (day 2 and day 5) had the best fit for two profiles. The characteristics of the profiles differ by day, making the profiles not one-on-one comparable. Age, gender and CRP values were the most recurring significant variables, although this differed by profile and day. The predictor's level of significance varied by day, from highly statistically significant ($p<0.001$) to not significant at all ($p>0.05$).

Due to missing data, severity of COVID-19 as classified by the WHO COVID-19 disease severity scoring model, could not be tested for significance as a predictor for survival chances.

Conclusion

The models for day two till day seven did not exceed the minimum of 100 events required to prevent overfitting of data, making the results not generalizable. However, the difference in the significance of the parameters, and difference in odds ratio's does suggest different relevant variables in different profiles, which is in line with the expectations. Repetition on a bigger database is required to internally validate the findings, and external validation is still needed for generalizability of the results.

Introduction

In December of 2019 there was an outbreak of unexplained pneumonia cases in the Hubei province in China. In hindsight it was found that these cases were caused by infection with the SARS-CoV-2 virus, the causative agent for the disease COVID-19 (1). The infection rate of this new virus is high, and on the 13th of January 2020 the first infection outside of

China was reported. The virus kept spreading further and on the 11th of March 2020, the WHO classified the COVID-19 disease as pandemic (1).

A lot of research has been done on COVID-19 in the past year, which generated insights into the risk factors and immunologic response. Moreover, this research has resulted in multiple different diagnostic tests, and even multiple vaccines with approval of the Food and Drug Administration (FDA) and the European Medicines Agency (EMA).

Besides research focussed on diagnosis, treatment and prevention, there has also been valuable research done into the prognosis of this disease. For example, prognostic research produced "risk calculators" for the chance of mortality and deterioration of COVID-19, like the Isaric4C calculator (2). While prognostic tools like this are useful for clinicians to aid in their decision-making process, they have some limitations.

The first complication of the current prognostic research on Covid-19 is the risk of bias. As shown in Figure 1, bias is a real problem in over two hundred reviewed research papers, especially in the analysis part of the papers (3). This figure shows the research type of the reviewed paper, with the cumulative bias score of all the reviewed papers of that research type. The papers were reviewed on their risk of bias in the participants, predictors, outcome, and analysis, which were bundled in an overall score.

So for the 107 reviewed prognostic papers, approximate 55% of the papers had a low risk of bias for their predictors, 35% had an unclear risk of bias, and 10% had an high risk of bias in their predictors. However, since basically all of the prognostic models had a high risk of bias in their analyses, their overall score (the left-most column) is very poor.

According to Wynants et al. (3), a high risk of bias on the participant domain indicates an unrepresentative model of the population. In the predictor domain, a high risk of bias indicates that the selected predictors were not available at the models' intended time of use, were not clearly defined, or were influenced by the outcome measurement. While most outcomes are relatively easy to asses, risk of bias still arose in this domain due to the use of subjective or proxy outcomes, like pneumonia or proxy outcomes taken from non-COVID-19 severe respiratory infection. Another risk in the outcome

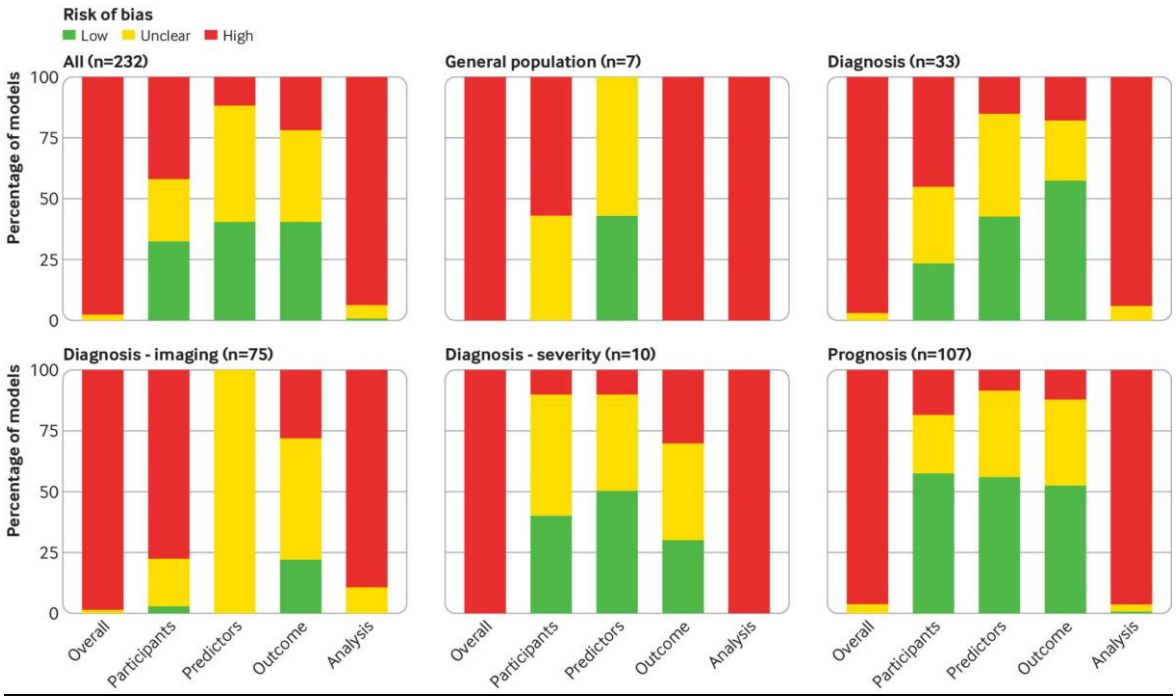

Figure 1: *Risk of bias in COVID-19 studies (n=232) calculated with PROBAST*

The figure shows the risk of bias in a total of 232 reviewed COVID-19 papers, and by research type. Each stacked bar represents a domain in which the risk of bias could be found, and shows the percentage of papers with their risk-levels. A mostly red coloured bar, indicate that the big majority of the reviewed papers showed a high risk of bias in the domain corresponding with that bar.

domain identified by Wynants et al. was the opaque or ambiguous reporting, which would have made any external validation of the findings pretty hard.

For the analysis domain, risk of bias was established due to small or modest sample sizes, resulting in overfitting of the data. Another problem for the analysis domain was the lack of model validation, both internal and external. A common concern for models that did use external validation, is that the datasets used for external validation were likely not representative of the target population, or the dataset did not meet the recommended minimum of 100 events, and preferably 200 or more. (4)

Abovementioned risk factors for bias were present in all research types, but the biases in the reviewed prognostic research papers were mainly due to dichotomisation of continuous predicting variables, and inappropriate inclusion or exclusion criteria of the patients. (3)

The second issue with the current prognostic research on COVID-19, is the identification and weighing of the predictors. More often than not, this identification and scoring is done with multivariate logistic regression. Although these types of calculations are useful for identifying the contributing variables, they do not necessarily portray the reality in an accurate manner. This has to do with magnitude of variables which may or may not contribute to a person's health, or in this case, the disease. Rothman's Causal Pie model (5), as seen in Figure 2, clearly illustrate how different combination of variables can lead to sufficient cause for the outcome of disease.

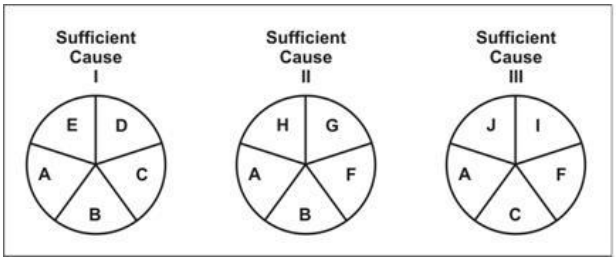

Figure 2: *Rothman's Causal Pie model. The model illustrates that multiple combinations of variables (depicted as letters), can lead to sufficient cause for effect. The model shows that not every variable is of equal importance. In this example, A is requires in all the models, however G occurs only once.*

The problem with linear and logistic regression is that these statistical methods require pre-determined combinations of variables applied to the to-be-researched (sub)population, meaning that effect of one variable is calculated based on all the cases. Referring back to Figure 2, this would mean that the effect of variable G is calculated over all the study participants, while it is only relevant in the context of "Sufficient Cause II". Therefore, the significance of variable G is likely to be underestimated, while it might be very significant in the subpopulation who correspond with "Sufficient Cause II".

Thus it seems that in order to predict the prognosis of COVID-19 patients in a more reliable and accurate manner, the identification of all the relevant subpopulations and their respective predictors is essential for aiding clinicians in their decision making.
Therefore, the aim of this study is to identify the different subpopulations within the data, their distinctive properties, and their corresponding outcomes.

Methods

A retrospective quantitative case study seems the most appropriate design in order to close the beforementioned knowledge gap. For this research, certain distinct steps have to be taken. First of all, the relevant data for this research has

to be collected, structured, and examined. For this research, coded data is collected from patients with COVID-19, who were admitted to the MST hospital in Enschede, The Netherlands. Patients, or their legal guardians, were given an opt-out possibility for the use of their data. This approach was deemed legal and ethical by COREON, the Dutch Committee for Research Regulations.

The gathered data from the participating patients is securely stored in Castor EDC and the entry for the database itself is done by multiple students and interns. The variables of the database are based on the ISARIC CRF. Data collection is done at admission of the patient, and at daily intervals on the ICU.

All patients with a hospital admission for a confirmed case of COVID-19 were enrolled in this study. The inclusion criterium is a positive diagnosis COVID-19, established by a PCR test and/or findings of CT abnormalities consistent with COVID-19. The second criteria for inclusion is an ICU admission. This second criteria was selected for the extensive data collection on daily intervals, which gives the dataset a time aspect on the level of individual patients.

Secondly, after all the data has been gathered, the patient clusters need to be identified in order to proceed with the analysis. Cluster identification and their corresponding properties can be identified through Latent Class Analysis (LCA) and Latent Profile Analysis (LPA)(6)(7). The difference between LCA and LPA is that LCA is limited to the use of categorical variables only, whereas LPA can use continuous variables to gain insight in the latent variables, in this case the observed patient profiles in the data. LPA has no need for the dichotomisation of continuous variables, thus reducing the risk of bias as identified by Wynants et al. (3). The difference between LCA, LPA, and relatable analyses is illustrated below in Figure 3 (7).

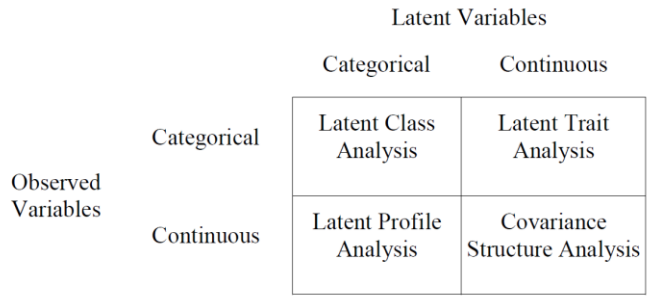

Figure 3: *Latent variable framework. The latent (hidden) variables and the observed variables can either be categorical or continuous. This framework shows which type of variables are accepted each of the analyses.*

This figure also shows the Latent Trait Analysis and the Covariance Structure Analysis for identification of continuous latent variables. In Covariance Structure Analysis, also called Factor Analysis, the variance in the observed variables can be explained by continuous latent variables, which are assumed to be normally distributed (7). Latent Trait Analysis refers to the framework where continuous variables can not be measured directly, but can be found with a set of discrete observed variables(7). For example someone's skill level which can be identified by the analysis of yes-or-no questions.

The LPA analysis will be performed with the use of the depmixS4 package (version 1.5-0) in R. For this research R version 4.0.5 was used, together with the IDE-program Rstudio (version 1.4.1103). For future LCA and LPA analysis in the MST hospital, the R script will be attached to this paper.

Statistical analysis plan

As mentioned above, he LCA and LPA analyses are suitable for identifying unobservable, yet distinct, subpopulations(6)(7). In the context of this research, subpopulations will be the different to-be-identified groups of patients within ICU of the MST

COVID-19 database. In order to use the model for predictions over time, it is possible to perform a Latent Transition Analysis (LTA). LTA is an extension of LCA, which would make it possible to predict the changes between the membership of subpopulations, meaning it can predict the shift of patients moving from group A to group B.

However, in this research LTA will not be performed to monitor the transition probabilities, due to limitations of the data, like the decreasing number of patients over time. Instead, LPA will be performed for each day of the ICU admission.
The numbers of patients in the ICU over time dwindle quickly at the first four days, and then decrease fairly slowly for the following days, e.g. the group of people staying six days in the ICU is considerably smaller than the group with one day ICU admission, but almost the same size as the group of patients with at least seven days of admission. With the slow decline and the fewer patients over time, it is more likely to find less distinct subpopulations due to the growing homogeneity. Because of the fast decline in patients at the beginning of ICU admission, the group of patients with multiple day on the ICU stays more or less the same. For example, the group of patients on the sixth day of ICU admission consists of 38 patients, of which 34 will be there on the next day, making the groups by day fairly homogeneous. Additionally, it is likely that the patients who suffer the longest form COVID-19 have certain characteristics in common, making the groups by day also internally more homogenous over time. This homogeneity combined with a decrease of the research population will make LTA too unstable.

Ultimately, LPA will also run into problems with this decreasing population and growing internal homogeneity, however this event will arise at a smaller dataset compared to LTA. This is because LPA can be applied on every admission day as if it were a new dataset, working around the issue of decreasing research participants. The problems with a small population will now arise when $n \leq p$, where p is the number of parameters in the model, and n the number of patients. However, the growing homogeneity will make it harder to identify different subpopulations or profiles of patients. Since the problems with groups sizes and homogeneity appear much later with LPA than LTA, LPA was selected for every ICU admission day rather than LTA on the entire dataset.
For the stability of the LPA models, a cut-off point has to be set. In the current dataset, ideally this point would be at day 10, because little change in the number of patients was observed after 10 days. This change population size can be seen at Figure 4.

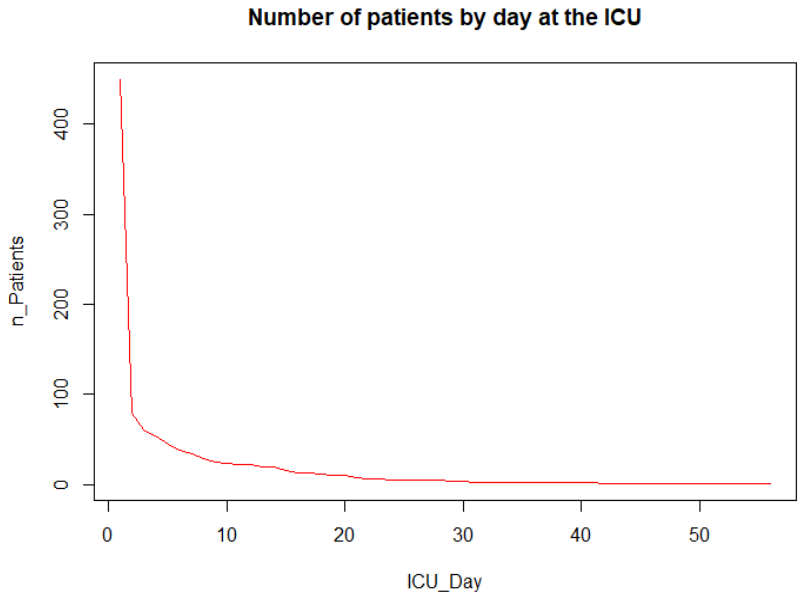

Figure 4: *Number of patients by admitted days on the ICU, within the dataset.*

However, the models became dysfunctional at day eight and later. This has to do with a parameter error for classification in the relative new package depmixS4, which should be fixed in a later version. A work-around was proposed in the form of removing random start in LPA, however this would less likely find the true number of profiles present within the dataset. For now, this means that the cut-off point has to be set at seven days due to limitations of the analytical software.

LCA and LPA are both unsupervised statistical methods, making it harder to validate the findings. However, certain methods for validation are available. Nagin & Odgers (6) identify the Bayesian Information Criteria (BIC), Akaike Information Criteria (AIC), Lo-Mendell-Rubin Likelihood Ratio Test (LMR-LRT), and Entropy as the most commonly used criteria to evaluate the models.

The AIC and BIC are selected to compare the number of parameters (model complexity) with the goodness of fit to the sample data (6). This can help with the selection of k numbers of classes, as opposed to $k + 1$ classes. LMR-LRT is a very useful method for finding the ideal number of classes, or profiles in this case. It provides a likelihood-ratio, where a low p-value indicates that a model with k classes is superior to a model with $k - 1$ classes (6).

Entropy is used to assess classification accuracy of the model. This is done by averaging the posterior probabilities of the different classes, after patients have been assigned to them. The result will be a range of 0 to 1, with values closer to 1 indexing higher levels of accuracy (6). With the averaging of the posterior chances, statements can be made about the accuracy of the model to appoint patients to the right group and corresponding trajectory. With a high accuracy, patients will be assigned to the right cluster, thus limiting the uncertainty overlap between groups and trajectories.

Simply put: AIC, BIC, and LMR-LRT can be used for validating both trajectory fit and the number of groups. They deliver the same information, only the other way around. (AIC and BIC for the preference of k+1 over k classes, and LMR-LRT for the preference of k-1 over k classes).
Entropy is used for validating distinguishable trajectories, with Posterior probabilities for validating the fit of individual patients in their cluster.

In this study, the best suitable model will be selected based on the AIC, BIC, and Entropy scores. The posterior probabilities will be used to assess the population share of each of the profiles.

With these statistical tests combined, the best suitable model can be selected. However, Nagin & Odgers (6) also mention that after selecting a model, its adequacy should still be tested. In order to assess the adequacy, certain criteria are provided:

- Each group should have a close correspondence between the estimated probabilities of group membership, and the proportion of the population assigned to the group, based on the posterior probability. This should be done for every group.
- The average of the posterior probabilities of group membership should have a minimum value of 0,7.
- The odds of correct classification of group membership, based on the posterior probabilities, should at least be 5.
- Observations of small Confidence Intervals (CI) around the estimated probabilities of group memberships.

Criteria	Score			
	0	1	2	3
Pulse, beats per min	51–100	41–50 or 101–110	≤ 40 or 111–130	≥ 131
Systolic blood pressure, mmHg	101–199	81–100	71–80 or ≥ 200	≤ 70
Respiratory rate, breaths per min	9–14	15–20	≤ 8 or 21–29	≥ 30
Body temperature, °C	36.1–37.4	35.1–36.0 or ≥ 37.5	≤ 35.0	NA
Level of consciousness	Normal	Voice reaction	Pain reaction	Non-reaction

COVID-19: coronavirus disease 2019; NA: not applicable.

Notes: Patients with a total score of 0–4 was classified as a mild case, a moderate case had a total score of 5–6 and severe and very severe case had a score ≤ 7. Very severe cases were patients needing renal replacement therapy or extracorporeal membrane oxygenation.

Figure 5: *WHO disease severity classification system for COVID-19, Republic of Korea, 2020. For each criteria, a score of 0 to 3 points can be given. The sum of all scores indicates the level of severity.*

After the different profiles are identified through LPA, multivariate logistic regression will be performed on each day, and on all the profiles within every day. Differences in effect of variables between profiles, indicate that not ever variable should be calculated on the entirety of the population. Instead, of levels of statistical significance for a certain variable vary between profiles, it illustrates that maybe a profile-based approach should be preferred over the population-based approach.

The effects of the variables as calculated with logistic regression, will be translated to odds ratio's (OR) and corresponding confidence intervals for easier interpretation. With an OR of 1.00, there is no observed effect of the carriable on the outcome measure. Odds ratio's smaller than 1.00 indicate a negative effect of the observed value, compared to the baseline of the variable. The opposite is also true, where odds ratio's bigger than 1.00 hold a positive relation with the outcome variable, as compared to the baseline of the observed variable. For every OR, the 95% CI is also given. When a confidence interval contains 1.00, no conclusions can be made about the effect of the variable, for the relation could both be positive or negative. These variables are also not statistically significant in the corresponding model.

The model predictors will be based on WHO COVID-19 severity classification model (8), as seen in Figure 5. Severity of the disease is a likely indicator of the outcome on survival chances. The model will be extended with clinical relevant variables according to literature, which need to be easy to assess. The full selection of variables will be discussed in the next section.

Variable selection

Due to not having an dependent outcome variable for the classes directly, the relevant variables will be based on literature research. Variables which have been validated as being risk-, protective-, or prognostic factors for the severity of COVID-19 will be the prior selection for the model. This literature research resulted into the following list of relevant variables:

- Systolic blood pressure,(8)
- Patient consciousness, as coded in Alert, Verbal, Pain, or Unresponsive (AVPU). (8)
- Hypertension (9) (10) (11)(12)
- Diabetes (13)(9) (10) (11)(12)
- Chronic kidney disease (14)
- Cardiac disease (10)
- Chronic obstructive pulmonary disease (COPD) (14)(9) (10)
- Age (11) (10)(12)
- Gender (11)(9)(15) (10)

- Obesity (11)
- D-dimers (13)(16)
- Fever (13)(17)

Additional variables of interest are the other comorbidities and laboratory values measured by the WHO ISARIC CRF questionnaire. This questionnaire can be found as an attachment to this paper.

Since the research is of retrospective nature, it is also possible to enter the outcome at hospital discharge as one of the variables for LPA. Ideally this variable would also be measured as discharge of the ICU, however the information is still useful. This has to do with the fact that ICU admission mainly happens when patients are unstable or in critical conditions. Therefore, if a patient comes to pass away, this event is most likely to happen at the ICU. So in order gain some predicting value to the model, the outcome whether the patent lived or passed away is also added tot the model.

The inclusion of outcome variables decrease the usefulness for clinicians, and increases the risk of bias in the analysis domain. However, by using the outcome variable as a grouping parameter in the LPA models, it will more clearly illustrate the point that not every predictor holds the same relevance for every subpopulation, as explained in the introduction with Rothmans Casual Pie in Figure 2.

With the addition of the ISARIC questionnaire and the patient survival variable, the prior selection now holds 50 variables. However, due to model stability, it is advised that LCA models hold between the 5 and 12 variables (7). A filtering of the prior variable list is required and the final selection of variables implemented into the model is based on the ease of assessment of the found variables. and can be found as the columns of Table 2 of the results section.

Model selection

The AIC and BIC have been used to select the best fitting model for each ICU admission day. These scores were calculated with the depmixS4 package in R, and were compared for models with two to five classes. The average scores and entropy were also calculated for the models. Entropy was calculated as:

$$-\sum_{k}^{C} p(K=k)\log_2 p(K=k)$$

With C being the total number of classes, k equals the class number for which the calculation is done, and K is the class to which the patient has been assigned.

Since the model holds multiple classes, the entropy value will have a larger range than from 0 to 1, for every $k > 2$. The maximum entropy score for each model is the $\log_2(C)$, making this also the maximum of the range.

In order to make the entropy scores of models with differing ranges possible, by bringing them all back to a range from 0 to 1, relative entropy was calculated as:

$$\frac{-\sum_{k}^{C} p(K=k)\log_2 p(K=k)}{\log_2(C)}$$

The results of these calculations can be seen in Table 3 for every model by day. To ease the comparison of the AIC and BIC scores, a plot was made for visually aiding in the model selection, which can be found in Figure 6. AIC and BIC are relative scores, which make them only comparable to other scores within the same dataset, meaning that only scores within the same day can be compared.

Results

Baseline characteristics

The baseline characteristics of the population are shown in Table 1. The majority of the research population are males, with a population share of 60.9%. The research population are the patients with confirmed COVID-19, who are admitted to the ICU of the MST hospital. The mean age of the population at the first day is 66,78, with a standard deviation (sd) of 15.32.

The BMI scores vary between 15.43 and 70.00. Due the broad spread of these variables, they are described in Table 1 as a median of 27.38 and an Inter Quartile Range (IQR) of [24.8 – 31.18]. While the extreme outliers for BMI are just as with the age variable fairly uncommon, the high number of missing data on BMI in the population will strengthen the effect of outliers on the mean. Therefore, the median was preferred over the use of the mean to describe this variable.

Apart from the BMI scores, patients have also been assessed on whether they are obese or not, which is done by the clinical staff. This variable has less missing data, which would explain the higher number of obese people found in the population. BMI scores were calculated on the emergency rooms (ER) of the hospital, whereas the diagnosis for the variable obesity could also have been made at other hospital departments.

Table 1 *Description of the population at ICU admission (n=450). All characteristics, comorbidities, and missing values are listed as number of occurrences (n) and the corresponding percentage of the population, unless specifically stated otherwise.*

Variables	n	(%)	Missing data, n (%)
Patient characteristics			
Gender at birth			0 (0%)
- Male, n (%)	274	(60.9%)	
- Female, n (%)	176	(39.1%)	
Age in years, mean (sd)	66.78	(15.32)	0 (0%)
BMI, median [IQR]	27.38	[24.80 – 31.18]	198 (44%)
Weight indication, n (%)			
- Underweight (BMI < 18)	2	(0.4%)	
- Healthy weight (BMI 18 <> 25)	73	(16.2%)	
- Overweight (BMI 25 <> 30)	100	(22.2%)	
- Obese (BMI >30)	77	(17.1%)	
Smoking			148 (32.9%)
- Active smoker	23	(5.1%)	
- Stopped smoker	94	(20.9%)	

Main comorbidities/ risk factors				
Diabetes	116	(25.8%)	27	(6%)
- With complications	41	(9.1%)		
- Without complications	75	(16.6%)		
Chronic cardiac disease, including congenital heart disease (not hypertension)	106	(23.6%)	13	(2.9%)
Hypertension	170	(37.8%)	15	(3.3%)
Dyslipidaemia	37	(8.2%)	25	(5.6%)
Peripheral occlusive arterial disease (PAOD)	13	(2.9%)	19	(4.3%)
Chronic pulmonary disease (not asthma)	52	(11.6%)	16	(3.6%)
Asthma (physician diagnosed)	48	(10.7%)	18	(4.0%)
Chronic kidney disease	38	(8.4%)	20	(4.4%)
Mild liver disease	3	(0.6%)	20	(4.4%)
Moderate or severe liver disease	3	(0.6%)	20	(4.4%)
Chronic neurological disorder	31	(6.9%)	19	(4.3%)
Malignant neoplasm	63	(14.5%)	17	(3.8%)
Chronic hematologic disease	13	(2.9%)	17	(3.8%)
AIDS / HIV	1	(0.2%)	19	(4.3%)
Obesity	89	(19.8%)	71	(15.8%)
Rheumatologic disorder	48	(10.7%)	19	(4.3%)
Dementia	12	(2.7%)	20	(4.4%)
Malnutrition	9	(2%)	22	(4.9%)
Other risk factors	80	(17.8%)	25	(5.6%)

According to the WHO, malnutrition can be defined as deficiencies, excesses, or imbalances of energy and/or nutrition intake in a persons diet (18). In this study could also be described as deficiency, or "under-nutrition", and is defined as having a (to) low weight, combined with problematic unintended weight loss. The definition for a "to low weight" is having a BMI under 18,5 for patients under 69 years, and a BMI below 20 for patients of 70 years and older. More importantly, the problematic weight loss is defined as unintentionally losing 5% body weight or more in one month, and/or unintentionally losing 10% or more of a persons bodyweight in six months time. These definitions are in line with the guidelines on malnutrition from the Dutch Malnutrition Steering Group (DMG) (19).

Figure 6: *Visual model selection by day*

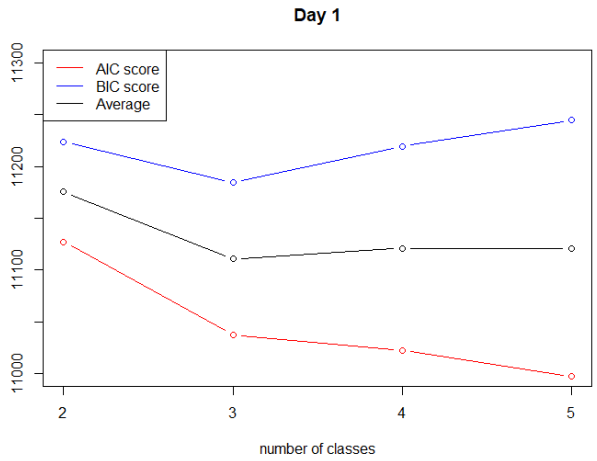

Figure 6a: *AIC, BIC, and Average (as defined by (AIC+BIC)/2) scores corresponding with the number of classes. Here the model with 3 classes suits the data the best.*

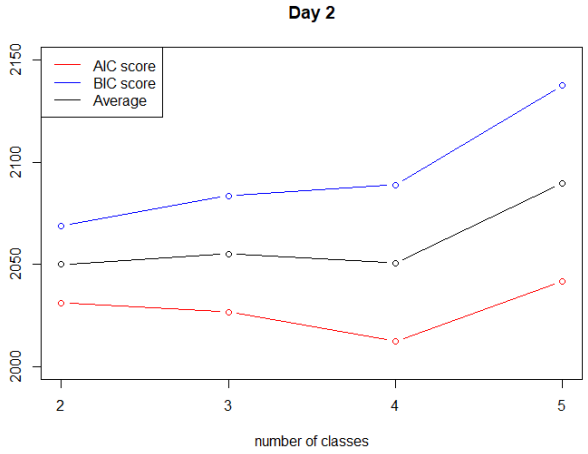

Figure 6b: *AIC, BIC, and their average plotted for the models of day 2. While the BIC-scores suggest the model with 2 classes, the average score seems indifferent for 2 or 4 classes. For model performance, the 2-class model was chosen*

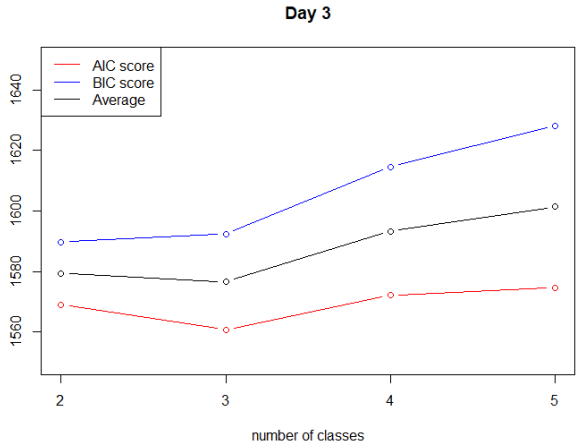

Figure 6c: *AIC, BIC, and their average plotted for the models of day 3. While the BIC suggests a slight preference for the 2 class-model over the 3 class-model, the 3-class model was selected due the combination with the AIC-score.*

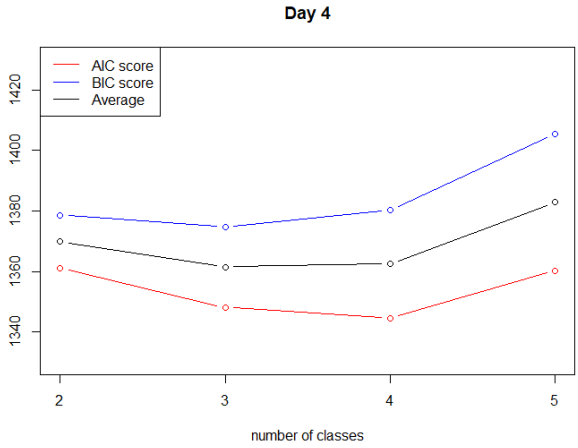

Figure 6d: *AIC, BIC, and their average plotted for the models of day 4. Here the BIC once again prefers the 3-class model*

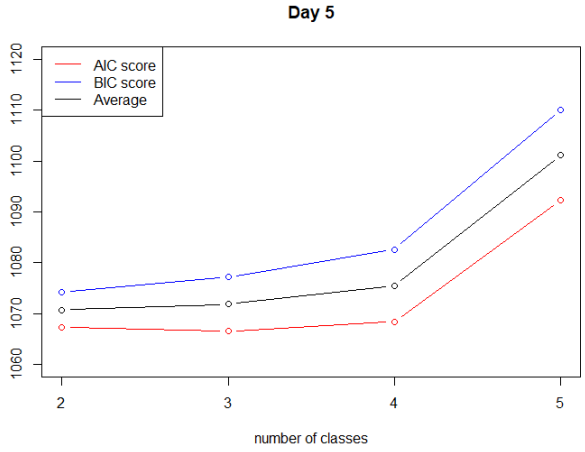

Figure 6e: *AIC, BIC, and their average for day 5. While the AIC score is basically indifferent for the first 3 models, the BIC and the average suggest the 2-class model would be the best fit for the data.*

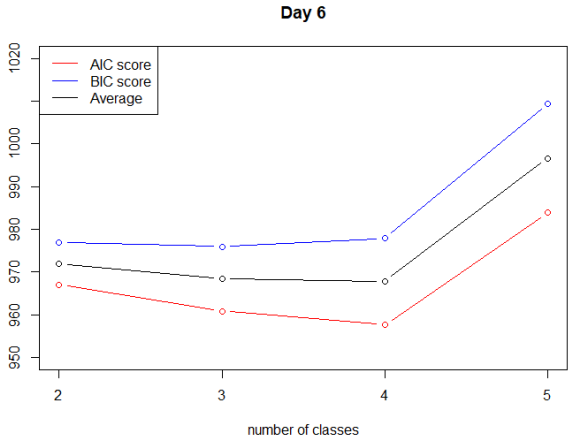

Figure 6f: *The AIC, BIC, and their average scores plotted for day 6. Both the AIC and BIC have a better fit with the 3-class model opposed to the 2-class model, although the scores are really getting close.*

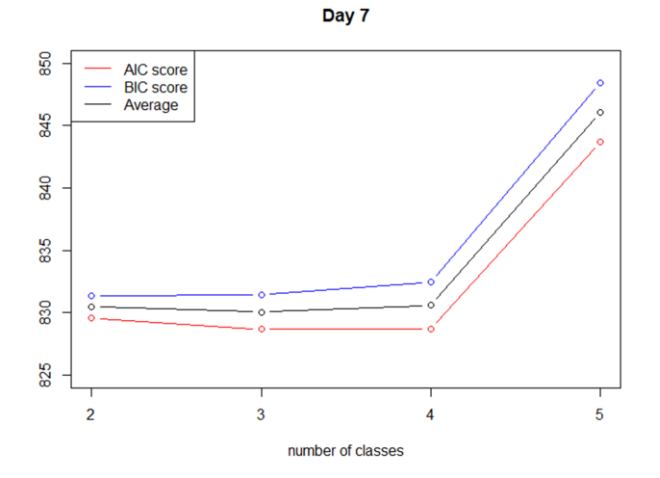

Figure 6g: *The AIC, BIC, and their average scores plotted for day 7. While hard to see, the average of the BIC and AIC fits best on the 3-class model.*

Then there is the relative big variable of "other risk factors", which affects quite a lot of people within this population. The most common comorbidity an potential risk factors in this list are Hypercholesterolemia (n=16), Transient Ischemic Attack (TIA) (n=6), Hyperthyroidism (n=5), and haemodialysis patients (n=4).

Model selection

Since the AIC and BIC scores are relative scores, comparisons can only be made for different models in the same dataset. The different admission days in this research are different subsets of the total database, and therefore unique datasets. Therefore, no conclusions can be drawn from the difference in scores between days. This relativity to the dataset also means there are no 'good' or 'bad' scores to judge the model with, only that lower scores within a certain dataset are better than higher AIC and BIC scores. Meaning for the plots of Figure 6, that the lowest point of the graph results in the best fitting number of classes for the model of that corresponding day. In case of (near) equal scores, the model with less classes is selected.

Model validation

In the statistical analysis plan, Nagins criteria for model adequacy were listed. The results of these calculations for these criteria can be found in Table 4.

The first criterion was the need for close correspondence between the estimated probabilities of group membership, and the proportion of the population assigned to the group, based on the posterior probability. For easier judgement, the expected group sizes based on the prior probabilities are calculated, and listed with the observed group sizes based on the posterior probabilities. As can be seen from these numbers, the expected and observed group sizes are close to each other. On day 1 in profile 2, there is a difference of 16 patients between the estimated and found group sizes. This is also the maximum difference found in all the models. On day 1 the research population holds data of 450 patients, which is significantly larger than the other days, making it also more likely to find bigger differences in absolute numbers. When looking at the other days, the biggest difference in group sizes for a profile, has a absolute value of 2.

The second criterion was that the average of the posterior probabilities of group membership should have a minimum value of 0,7. As can be seen from the column "Average classification probabilities", the minimum value over the models has a value of 0.78.

Next criterion in the list was the odds of correct classification of group membership, based on the posterior probabilities. These odds should at least be 5. Calculated odds take on a wide range of values, with the lowest odds being equal to 4.65.

These odds for correct classification are within profile 1 of the first day. This number does not exactly meets the adequacy criteria, however the rounded value does equal 5.

Furthermore, the model entropy is high with a sore of over 0.94. This means there is little uncertainty, or 'fuzziness' in the data, and the different profiles have little overlap. Therefore, the score of 4.65 is not deemed as problematic. The final criterion for adequacy was the observation of small Confidence Intervals (CI) around the estimated probabilities of group memberships. These intervals can be found in Table 3.

Comparison of the models

When looking at Figure 6, it is interesting to see that the models become more similar over time, with a growing indifference for 2 till 4 classes, and relative high scores for the model with 5 classes. Due to the decreasing population, a preference for less classes in the models was expected for the later days. It would make sense that in a bigger population, more groups will be present. Yet the data shows that over time, the 4-classes model becomes more in par with the models of 2 and 3 classes, hinting at similar performance.

Since not every model holds the same number of profiles, the profiles are not necessarily comparable between the all different days. However it is still interesting to look at the differences and comparability between the models of Table 2 with the same number of profiles. For example, the models with three states all hold a profile with a mean age around 55, and a (near) 100% survival rate, making them somewhat comparable.

However, this group of (relatively) young people hold different combinations of the probability of finding obesity and hypertension within the profile, underlining that these models, although comparable by mean age, are in fact different. This difference is also visible in the probabilities of finding a patient who were assigned the male gender at birth.

On a first glance obesity does not hold much value as a predictor for whether the patient is likely to survive. This can be seen in the majority of the different profiles, where an increase in the probability of finding someone with obesity, does not correspond with a lower chance of survival. For hypertension it seems to be the other way around, were an increase in the people with hypertension, seems to correlate with an decrease of survival probability.

The CRP values, as a signal variable for inflammation, does not seem to hold a relation to survival chance in this population. This can be observed from the fact that most days have a profile with a CRP value above 100 mg/L, and a profile of a CRP value around 40 mg/L., yet the profiles with a considerable higher CRP value, do not hold a lower survival rate. However, there does seem to be a relation between CRP and the Severity score.

In order to be able to verify the observed relations, and to support the theory that the groups are different from each other, multivariate logistic regression analysis was performed for each day, with all the profiles combined, and then for each profile per day. The "discharge alive" variable was selected as the outcome variable, meaning that negative coefficients decrease the chance of survival. The results from the logistic regression are translated into odds ratio's (OR), and the effect from these variables in can be found in Table 4.

In Table 4, the baseline for the variable gender is "Male", and for the AVPU score the baseline is "Alert". Effects of certain variables could not be calculated, and are shown as "NA". Failure of calculations was either due to certain variables not being observed, e.g. the AVPU-level "Pain" did not occur on the population of Day 1, and has therefore a prevalence of zero. In other words, these levels of consciousness were not observed within that subpopulation, or in this case, in that profile.

Table 2: *Latent Profiles by day on the ICU*

	Gender = Male **n (%)**	**Age in years mean (sd)**	**Obese** **n (%)**	**CRP** **Mean (sd)**	**Hyper-tension** **n (%)**	**AVPU indication**		**COVID-19 Severity**		**Discharge alive** **n (%)**
Day 1 (n=450)	n= 274 (60,9%)	66,78 (15,32)	n= 89 (19,8%)	94,76 (84,29)	n= 170 (37.8%)	Alert: Verbal: Pain: Unresponsive *Missing*:	374 14 6 26 *30*	Mild: Moderate: Severe /critical *Missing*	277 52 2 *119*	n= 368 (81,8%)
Profile 1 (n=224)	54,0%	72,95 (9,74)	27,9%	73,84 (45,87)	57,6%	Alert: Verbal: Pain: Unresponsive	90,7% 5,6% 1,8% 1,9%	Mild: Moderate: Severe /critical	87,5% 12,5% 0,0%	78,6%
Profile 2 (n=97)	77,5%	65,92 (12,31)	24,7%	199,87 (90,16)	36,7%	Alert: Verbal: Pain: Unresponsive	74,9% 1,5% 2,3% 21,3%	Mild: Moderate: Severe /critical	64,5% 33,0% 2,5%	70,0%
Profile 3 (n=129)	57,9%	56,11 (19,68)	14,6%	34,52 (24,48)	7,0%	Alert: Verbal: Pain: Unresponsive	99,1% 0,9% 0,0% 0,0%	Mild: Moderate: Severe /critical	94,6% 5,5% 0,0%	99,7%
Day 2 (n=80)	n= 53 (66.3%)	64,38 (12,93)	n= 17 (21,3%)	144,20 (109,16)	n= 30 (37,5%)	Alert: Verbal: Pain: Unresponsive *Missing*:	54 1 0 18 *7*	Mild: Moderate: Severe / critical *Missing*	27 16 1 *36*	n= 62 (77,5%)
Profile 1 (n=23)	91,2%	52,91 (14,44)	13,1%	186,96 (118,62)	0,0%	Alert: Verbal: Pain: Unresponsive	58,0% 4,8% 0,0% 37,2%	Mild: Moderate: Severe /critical	41,8% 58,2% 0,0%	87,7%
Profile 2 (n=57)	56,0%	69,06 (8,46)	28,8%	127,63 (99,27)	53,7%	Alert: Verbal: Pain: Unresponsive	80,3% 0,0% 0,0% 19,7%	Mild: Moderate: Severe /critical	71,1% 25,5% 3,4%	73,4%
Day 3 (n=61)	n= 41 (67,2%)	63,48 (13,26)	n= 15 (24.6%)	142,20 (127,75)	n= 22 (36.1%)	Alert: Verbal: Pain: Unresponsive *Missing*:	32 1 0 18 *10*	Mild: Moderate: Severe /critical *Missing*	10 9 2 *40*	n= 47 (77,0%)

Profile 1 (n=20)	91,0%	70,40 (4,91)	28,0%	75,59 (40,46)	53,5%	Alert: Verbal: Pain: Unresponsive	87,0% 6,9% 0,0% 6,1%	Mild: Moderate: Severe /critical	100,0% 0,0% 0,0%	53,4%
Profile 2 (n=20)	42,2%	66,48 (7,34)	22,4%	244,09 (130,53)	45,8%	Alert: Verbal: Pain: Unresponsive	16,5% 0,0% 0,0% 83,5%	Mild: Moderate: Severe /critical	18,3% 65,4% 16,3%	75,2%
Profile 3 (n=21)	72,1%	54,17 (17,06)	30,5%	78,67 (86,79)	10,9%	Alert: Verbal: Pain: Unresponsive	93,3% 0,0% 0,0% 6,7%	Mild: Moderate: Severe /critical	86,6% 13,4% 0,0%	100,0%
Day 4 (n=54)	n= 39 (72.2%)	64,54 (12,85)	n= 15 (27,8%)	108,60 (122,51)	n= 19 (35,2%)	Alert: Verbal: Pain: Unresponsive *Missing*:	32 0 0 12 *10*	Mild: Moderate: Severe /critical *Missing*	10 9 0 *35*	n= 40 (74.1%)
Profile 1 (n=13)	85,2%	52,98 (18,59)	38,5%	63,86 (68,19)	0,0%	Alert: Verbal: Pain: Unresponsive	100,0% 0,0% 0,0% 0,0%	Mild: Moderate: Severe /critical	100,0% 0,0% 0,0%	100,0%
Profile 2 (n=14)	44,8%	65,59 (7,74)	41,1%	241,49 (141,90)	40,3%	Alert: Verbal: Pain: Unresponsive	15,8% 0,0% 0,0% 84,2%	Mild: Moderate: Severe /critical	10,1% 89,9% 0,0%	78,8%
Profile 3 (n=27)	81,4%	69,77 (5,65)	19,7%	50,03 (35,30)	50,1%	Alert: Verbal: Pain: Unresponsive	96,0% 0,0% 0,0% 4,0%	Mild: Moderate: Severe /critical	100,0% 0,0% 0,0%	58,3%
Day 5 (n=45)	n= 32 (71,1%)	65,89 (11,97)	n= 15 (33,3%)	70,45 (73,71)	n= 17 (37,8%)	Alert: Verbal: Pain: Unresponsive *Missing*:	28 3 0 7 *7*	Mild: Moderate: Severe /critical *Missing*	8 5 0 *32*	n= 33 (73,3%)
Profile 1 (n=36)	72,0%	66,86 (10,20)	37,8%	35,21 (25,26)	40,5%	Alert: Verbal: Pain: Unresponsive	87,6% 6,2% 0,0% 5,8%	Mild: Moderate: Severe /critical	100,0% 0,0% 0,0%	72,4%

Profile 2 (n=9)	68,0%	62,35 (16,00)	36,3%	178,79 (63,52)	32,1%	Alert: Verbal: Pain: Unresponsive	16,9% 13,4% 0,0% 69,7%	Mild: Moderate: Severe /critical	19,4% 80,6% 0,0%	76,9%
Day 6 (n=38)	n= 28 (73,7%)	64,34 (11,95)	n= 14 (36,8%)	69,42 (69,37)	n= 17 (44,7%)	Alert: Verbal: Pain: Unresponsive *Missing*:	28 1 1 4 *4*	Mild: Moderate: Severe /critical *Missing*	7 5 0 *26*	n= 27 (71,1%)
Profile 1 (n=10)	69,2%	69,70 (3,86)	21,7%	132,20 (83,49)	12,0%	Alert: Verbal: Pain: Unresponsive	52,2% 0,0% 9,6% 38,2%	Mild: Moderate: Severe /critical	33,5% 66,5% 0,0%	41,3%
Profile 2 (n=12)	76,3%	54,35 (15,17)	46,3%	36,26 (30,58)	0,0%	Alert: Verbal: Pain: Unresponsive	100,0% 0,0% 0,0% 0,0%	Mild: Moderate: Severe /critical	100,0% 0,0% 0,0%	100,0%
Profile 3 (n=16)	74,7%	68,22 (6,72)	50,4%	48,22 (41,51)	100,0%	Alert: Verbal: Pain: Unresponsive	93,2% 6,8% 0,0% 0,0%	Mild: Moderate: Severe /critical	66,5% 33,5% 0,0%	69,3%
Day 7 (n=34)	n= 25 (73,5%)	64,38 (12,39)	n= 11 (32,4%)	72,20 (67,75)	n= 14 (41,2%)	Alert: Verbal: Pain: Unresponsive *Missing*:	23 4 0 3 *4*	Mild: Moderate: Severe /critical *Missing*	7 2 0 *25*	n= 23 (67,6%)
Profile 1 (n=8)	87,8%	70,98 (5,24)	13,9%	119,75 (53,53)	57,6%	Alert: Verbal: Pain: Unresponsive	49,8% 12,6% 0,0% 37,7%	Mild: Moderate: Severe /critical	0,0% 100,0% 0,0%	12,5%
Profile 2 (n=11)	72,0%	58,58 (17,64)	46,5%	98,27 (72,55)	19,6%	Alert: Verbal: Pain: Unresponsive	69,8% 30,2% 0,0% 0,0%	Mild: Moderate: Severe /critical	100,0% 0,0% 0,0%	100,0%
Profile 3 (n=15)	66,0%	65,43 (4,66)	42,2%	20,87 (12,82)	51,6%	Alert: Verbal: Pain: Unresponsive	100,0% 0,0% 0,0% 0,0%	Mild: Moderate: Severe /critical	100,0% 0,0% 0,0%	73,3%

Table 3, *Model adequacy*

Profile	Prior probabilities	Expected group size	Posterior probabilities	95% CI of posterior probability	Observed group size	Average probability for correct classification	Odds of correct classification	Model Entropy
Day 1, n = 450								
1	0,488	220	0,49	[0.45-0.52]	224	0,82	4,65	
2	0,250	113	0,25	[0.22-0.28]	97	0,87	25,16	0,943
3	0,262	118	0,26	[0.23-0.30]	129	0,78	8,96	
Day 2, n = 80								
1	0,290	23	0,29	[0.21-0.37]	23	0,84	12,94	0,865
2	0,710	57	0,71	[0.63-0.79]	57	0,93	5,48	
Day 3, n = 61								
1	0,304	19	0,30	[0.20-0.41]	20	0,87	13,62	
2	0,355	22	0,35	[0.25-0.46]	20	0,98	121,04	0,999
3	0,341	21	0,34	[0.23-0.45]	21	0,90	17,69	
Day 4, n=54								
1	0,24	13	0,24	[0.14-0.35]	13	0,89	25,45	
2	0,28	15	0,25	[0.17-0.38]	14	0,98	165,25	0,946
3	0,48	26	0,48	[0.36-0.60]	27	0,91	9,93	
Day 5, n = 45								
1	0,79	35	0,79	[0.67-0.91]	36	0,98	12,82	0,722
2	0,21	10	0,21	[0.10-0.33]	9	1,00	1569,17	
Day 6, n = 38								
1	0,28	11	0,28	[0.13-0.42]	10	1,00	280016,65	
2	0,31	12	0,31	[0.17-0.45]	12	0,98	100,10	0,983
3	0,41	16	0,41	[0.25-0.58]	16	0,98	81,18	
Day 7, n = 34								
1	0,25	8	0,25	[0.10-0.40]	8	0,98	172,02	
2	0,35	12	0,35	[0.21-0.50]	11	0,99	148,60	0,971
3	0,40	13	0,40	[0.24-0.55]	15	0,88	9,43	

The calculation of odds ratio's and the corresponding confidence intervals could also fail due to lack of contrast, meaning a variable became singular. Or in other words, all individuals in the observed subpopulation have the same value for a predictor. This can be seen among other in profile 1 of the second day, where no patients have hypertension.

Due to the high rate of missing data on the Severity variable, it was decided to leave this variable out, to keep the number of deleted observations due to missingness at a minimum. In certain regression analysis the AVPU score also had to be deleted, due to missingness in a small population. Without exclusion of this variable, certain models would hold more predictors than observations, making the regression analysis impossible.

As can be seen in Table 4, on the total population (Day 1, all profiles), there are three statistically significant variables, which are marked as bold text. These variables are gender ($p < 0.05$), age ($p < 0.001$), and CRP value ($p < 0.01$). Of these variables, a female gender holds a protective value (OR > 1.00), whereas an increase in age and CRP values worsen the

survival rate (OR < 1.00). However, when looking at the different profiles within the first day, gender and CRP lose their statistical significance. Odds ratio's of profile 3 could not be computed due to the 100% survival rate, after logistic regression deleted certain observations due to missing data. With no deceased patients, effects of variables on survival chances can not be calculates and there were no different odds which could be compared.

On the second day, age still holds its statistically significant ($p < 0.05$) negative relation on the survival chances. It is also the only significant predictor for the population of this day. Yet, this significance disappears within the different profiles of day two, resulting in no significant predictors within the profiles of the second day.

For the third day, the CRP value is statistically significant in the entire population, but this significance is lost within the different profiles. Profiles one hold no significant predictors at all, and profile three could not be calculated since there was a 100% chance of survival. On the entire day three and in the second profile, patient unresponsiveness becomes statistically significant. Counterintuitively, this effect in both models is positive (OR > 1.00), meaning an increase in survival chances for unresponsive patients compared to alert patients.

Day number four only holds age as a predictor with significance ($p<0.05$, OR = 0.93 [0.86 – 0.99]). The odds ratio's of the first profile of could not be calculated due to a lack of deceased patients, indicating another group for which none of the selected model predictors hold any predictive value. The second profile of this holds no statistically significant predictors, but interestingly enough, the third profile does. This is interesting, because the third profile of the fourth day holds CRP as a significant variable ($p<0.05$, OR = 0.91 [0.86 – 0.98]), whereas this variable was not relevant in the entirety of the patients on their fourth day. This is effect can also be seen in the first profile of the fifth day, which also holds a significance for CRP ($p<0.01$, OR = 0.87 [0.82 – 0.93]), while the entirety of the fifth day does not.

The sixth day does not hold any significant variables, nor does the entirety of the seventh day. Profile one of day seven holds a significance for gender ($p < 0.001$, OR = 2.72 [2.72 – 2.72]) . When taking a closer look, this sudden significance for gender and effect size can easily be explained by chance, since this profile only holds one survivor, who happens to be female. All the deceased patients of profile one were male, hence the survival rate seems to solely depend on gender. This can also be seen in the odds ratio's of the other variables, which are all equal to 1.00, meaning they have no effect on the outcome variable.

The third profile hold a negative effect (OR = 0.29 [0.10 – 0.84]) for the female gender. Interestingly enough, the logistic regression showed no statistical significance ($p > 0.05$), even though the OR and corresponding 95% CI suggest an effect with a certain level of significance.

Table 4, Oddsratio's of model parameters and corresponding 95% Confidence Intervals, based on logistic regression. *Bold variables indicate statistical significant relation (p < 0.05) with the outcome variable of discharged alive. OR > 1.00 indicates a positive effect*

		Variable	Odds ratio	[95% CI]
Day 1	Gender	(Female vs Male)	1.10	[1.01 – 1.19]
Entire population	Age	(Increase of 5 years)	0.97	[0.96 – 0.98]
	Obesity	(Obese vs Non-obese)	1.01	[0.92 – 1.11]
	CRP	(Increase of 10)	0.99	[0.99 – 1.00]
	Hypertension	(Hypertension vs No hypertension)	0.99	[0.91 – 1.08]
	AVPU	(Verbal vs Alert)	0.83	[0.66 – 1.05]
	AVPU	(Pain vs Alert)	*NA due to missing data*	
	AVPU	(Unresponsive vs Alert)	1.00	[0.84 – 1.20]
Day 1	Gender	(Female vs Male)	1.12	[0.99 – 1.26]
Profile 1	Age	(Increase of 5 years)	0.95	[0.92 – 0.98]
	Obesity	(Obese vs Non-obese)	1.06	[0.93 – 1.20]
	CRP	(Increase of 10)	1.00	[1.00 – 1.02]
	Hypertension	(Hypertension vs No hypertension)	1.10	[0.97 – 1.25]
	AVPU	(Verbal vs Alert)	0.80	[0.61 – 1.06]
	AVPU	(Pain vs Alert)	*NA due to missing data*	
	AVPU	(Unresponsive vs Alert)	0.79	[0.50 – 1.24]
Day 1	Gender	(Female vs Male)	1.23	[0.92 – 1.64]
Profile 2	Age	(Increase of 5 years)	0.93	[0.89 – 0.98]
	Obesity	(Obese vs Non-obese)	0.96	[0.72 – 1.27]
	CRP	(Increase of 10)	1.01	[0.99 – 1.02]
	Hypertension	(Hypertension vs No hypertension)	0.96	[0.74 – 1.25]
	AVPU	(Verbal vs Alert)	1.54	[0.61 – 3.87]
	AVPU	(Pain vs Alert)	*NA due to missing data*	
	AVPU	(Unresponsive vs Alert)	1.17	[0.89 – 1.53]
Day 1	Gender	(Female vs Male)		
Profile 3	Age	(Increase of 5 years)	*Oddsratio's and Confidence*	
	Obesity	(Obese vs Non-obese)	*Intervals could not be calculated*	
	CRP	(Increase of 10)	*due to lack of contrast. This profile*	
	Hypertension	(Hypertension vs No hypertension)	*holds 100% survival rate after*	
	AVPU	(Verbal vs Alert)	*deletion of missing data for logistic*	
	AVPU	(Pain vs Alert)	*regression.*	
	AVPU	(Unresponsive vs Alert)		
Day 2	Gender	(Female vs Male)	1.19	[0.95 – 1.49]
Entire population	Age	(Increase of 5 years)	0.95	[0.91 – 0.99]
	Obesity	(Obese vs Non-obese)	1.05	[0.82 – 1.36]
	CRP	(Increase of 10)	0.99	[0.98 – 1.00]

	Hypertension	(Hypertension vs No hypertension)	1.02	[0.81 – 1.29]
	AVPU	(Verbal vs Alert)	1.41	[0.65 – 3.07]
	AVPU	(Pain vs Alert)	*NA due to missing data*	
	AVPU	(Unresponsive vs Alert)	1.20	[0.96 – 1.51]
Day 2	Gender	(Female vs Male)	0.80	[0.35 – 1.79]
Profile 1	Age	(Increase of 5 years)	0.94	[0.86 – 1.03]
	Obesity	(Obese vs Non-obese)	0.88	[0.48 – 1.63]
	CRP	(Increase of 10)	0.99	[0.97 – 1.00]
	Hypertension	(Hypertension vs No hypertension)	*NA due to singularity of data*	
	AVPU	(Verbal vs Alert)	1.25	[0.56 – 2.80]
	AVPU	(Pain vs Alert)	*NA due to missing data*	
	AVPU	(Unresponsive vs Alert)	1.11	[0.78 – 1.58]
Day 2	Gender	(Female vs Male)	1.26	[0.96 – 1.66]
Profile 2	Age	(Increase of 5 years)	0.98	[0.89 – 1.08]
	Obesity	(Obese vs Non-obese)	1.08	[0.80 – 1.45]
	CRP	(Increase of 10)	0.99	[0.98 – 1.00]
	Hypertension	(Hypertension vs No hypertension)	1.08	[0.79 – 1.46]
	AVPU	(Verbal vs Alert)	*NA due to missing data*	
	AVPU	(Pain vs Alert)	*NA due to missing data*	
	AVPU	(Unresponsive vs Alert)	1.21	[0.88 – 1.65]
Day 3	Gender	(Female vs Male)	1.07	[0.84 – 1.37]
Entire population	Age	(Increase of 5 years)	0.97	[0.92 – 1.02]
	Obesity	(Obese vs Non-obese)	1.21	[0.92 – 1.59]
	CRP	(Increase of 10)	0.99	[0.98 – 1.00]
	Hypertension	(Hypertension vs No hypertension)	0.9	[0.72 – 1.20]
	AVPU	(Verbal vs Alert)	0.46	[020 – 1.03]
	AVPU	(Pain vs Alert)	*NA due to missing data*	
	AVPU	(Unresponsive vs Alert)	1.34	[1.02 – 1.76]
Day 3	Gender	(Female vs Male)	0.57	[0.16 – 1.98]
Profile 1	Age	(Increase of 5 years)	1.05	[0.75 – 1.48]
	Obesity	(Obese vs Non-obese)	1.35	[0.61 – 2.98]
	CRP	(Increase of 10)	0.94	[0.85 – 1.03]
	Hypertension	(Hypertension vs No hypertension)	0.99	[0.46 – 2.13]
	AVPU	(Verbal vs Alert)	0.51	[0.13 – 1.97]
	AVPU	(Pain vs Alert)	*NA due to missing data*	
	AVPU	(Unresponsive vs Alert)	0.77	[0.20 – 1.94]
Day 3	Gender	(Female vs Male)	1.31	[0.81 – 2.11]
Profile 2	Age	(Increase of 5 years)	0.97	[0.84 – 1.13]
	Obesity	(Obese vs Non-obese)	1.05	[0.62 – 1.76]

	CRP	(Increase of 10)	1.00	[0.98 – 1.02]
	Hypertension	(Hypertension vs No hypertension)	1.18	[0.73 – 1.93]
	AVPU	(Verbal vs Alert)	*NA due to missing data*	
	AVPU	(Pain vs Alert)	*NA due to missing data*	
	AVPU	(Unresponsive vs Alert)	2.35	[1.15 – 4.77]
Day 3	Gender	(Female vs Male)		
Profile 3	Age	(Increase of 5 years)	*Oddsratio's and Confidence*	
	Obesity	(Obese vs Non-obese)	*Intervals could not be calculated*	
	CRP	(Increase of 10)	*due to lack of contrast. This profile*	
	Hypertension	(Hypertension vs No hypertension)	*holds 100% survival rate after*	
	AVPU	(Verbal vs Alert)	*deletion of missing data for logistic*	
	AVPU	(Pain vs Alert)	*regression.*	
	AVPU	(Unresponsive vs Alert)		
Day 4	Gender	(Female vs Male)	1.18	[0.86 – 1.63]
Entire population	Age	(Increase of 5 years)	0.93	[0.86 – 0.99]
	Obesity	(Obese vs Non-obese)	1.05	[0.76 – 1.47]
	CRP	(Increase of 10)	1.00	[0.98 – 1.01]
	Hypertension	(Hypertension vs No hypertension)	1.08	[0.78 – 1.49]
	AVPU	(Verbal vs Alert)	*NA due to missing data*	
	AVPU	(Pain vs Alert)	*NA due to missing data*	
	AVPU	(Unresponsive vs Alert)	1.13	[0.76 – 1.68]
Day 4	Gender	(Female vs Male)		
Profile 1	Age	(Increase of 5 years)	*Oddsratio's and Confidence*	
	Obesity	(Obese vs Non-obese)	*Intervals could not be calculated*	
	CRP	(Increase of 10)	*due to lack of contrast. This profile*	
	Hypertension	(Hypertension vs No hypertension)	*holds 100% survival rate after*	
	AVPU	(Verbal vs Alert)	*deletion of missing data for logistic*	
	AVPU	(Pain vs Alert)	*regression.*	
	AVPU	(Unresponsive vs Alert)		
Day 4	Gender	(Female vs Male)	1.31	[0.70 – 2.44]
Profile 2	Age	(Increase of 5 years)	0.88	[0.73 – 1.06]
	Obesity	(Obese vs Non-obese)	0.79	[0.43 – 1.47]
	CRP	(Increase of 10)	1.00	[0.98 – 1.03]
	Hypertension	(Hypertension vs No hypertension)	1.38	[0.67 – 2.84]
	AVPU	(Verbal vs Alert)	*NA due to missing data*	
	AVPU	(Pain vs Alert)	*NA due to missing data*	
	AVPU	(Unresponsive vs Alert)	1.93	[0.78 – 4.79]
Day 4	Gender	(Female vs Male)	1.25	[0.77 – 2.01]
Profile 3	Age	(Increase of 5 years)	1.14	[0.94 – 1.37]

	Obesity	(Obese vs Non-obese)	1.40	[0.86 – 2.26]
	CRP	(Increase of 10)	0.91	[0.86 – 0.98]
	Hypertension	(Hypertension vs No hypertension)	1.27	[0.83 – 1.92]
	AVPU	(Verbal vs Alert)	*NA due to missing data*	
	AVPU	(Pain vs Alert)	*NA due to missing data*	
	AVPU	(Unresponsive vs Alert)	0.54	[0.22 – 1.33]
Day 5	Gender	(Female vs Male)	1.20	[0.79 – 1.84]
Entire population	Age	(Increase of 5 years)	0.90	[0.80 – 1.02]
	Obesity	(Obese vs Non-obese)	0.96	[0.62 – 1.49]
	CRP	(Increase of 10)	0.99	[0.96 – 1.02]
	Hypertension	(Hypertension vs No hypertension)	1.08	[0.73 – 1.61]
	AVPU	(Verbal vs Alert)	1.47	[0.75 – 2.88]
	AVPU	(Pain vs Alert)	*NA due to missing data*	
	AVPU	(Unresponsive vs Alert)	1.05	[0.63 – 1.76 [
Day 5	Gender	(Female vs Male)	1.05	[0.72 – 1.53]
Profile 1	Age	(Increase of 5 years)	0.96	[0.86 – 1.08]
	Obesity	(Obese vs Non-obese)	1.00	[0.68 – 1.46]
	CRP	(Increase of 10)	0.87	[0.82 – 0.93]
	Hypertension	(Hypertension vs No hypertension)	0.98	[0.68 – 1.41]
	AVPU	(Verbal vs Alert)	1.35	[0.78 – 2.32]
	AVPU	(Pain vs Alert)	*NA due to missing data*	
	AVPU	(Unresponsive vs Alert)	0.90	[0.54 – 1.50]
Day 5	Gender	(Female vs Male)	1.60	[0.41 – 6.24]
Profile 2	Age	(Increase of 5 years)	1.04	[0.67 – 1.63]
	Obesity	(Obese vs Non-obese)	0.62	[0.22 – 1.74]
	CRP	(Increase of 10)	0.96	[0.85 – 1.07]
	Hypertension	(Hypertension vs No hypertension)	1.45	[0.85 – 1.07]
	AVPU	(Verbal vs Alert)	*NA due to singularity of data*	
	AVPU	(Pain vs Alert)	*NA due to singularity of data*	
	AVPU	(Unresponsive vs Alert)	*NA due to singularity of data*	
Day 6	Gender	(Female vs Male)	1.14	[0.77 – 1.71]
Entire population	Age	(Increase of 5 years)	0.94	[0.83 – 1.06]
	Obesity	(Obese vs Non-obese)	1.14	[0.76 – 1.71]
	CRP	(Increase of 10)	1.00	[0.97 – 1.03]
	Hypertension	(Hypertension vs No hypertension)	1.03	[0.67 – 1.58]
	AVPU	(Verbal vs Alert)	1.34	[0.45 – 4.02]
	AVPU	(Pain vs Alert)	0.61	[0.21 – 1.79]
	AVPU	(Unresponsive vs Alert)	0.90	[0.50 – 1.61]
Day 6	Gender	(Female vs Male)	1.09	[0.09 – 13.92]

Profile 1	Age	(Increase of 5 years)	1.18	[0.26 – 5.32]
	Obesity	(Obese vs Non-obese)	0.98	[0.08 – 12.01]
	CRP	(Increase of 10)	1.01	[0.86 – 1.20]
	Hypertension	(Hypertension vs No hypertension)	1.77	[0.03 – 108.62]
	AVPU	(Verbal vs Alert)	*NA due to missing data*	
	AVPU	(Pain vs Alert)	0.61	[0.03 – 12.28]
	AVPU	(Unresponsive vs Alert)	0.97	[0.10 – 9.12]
Day 6	Gender	(Female vs Male)		
Profile 2	Age	(Increase of 5 years)	*Oddsratio's and Confidence*	
	Obesity	(Obese vs Non-obese)	*Intervals could not be calculated*	
	CRP	(Increase of 10)	*due to lack of contrast. This profile*	
	Hypertension	(Hypertension vs No hypertension)	*holds 100% survival rate after*	
	AVPU	(Verbal vs Alert)	*deletion of missing data for logistic*	
	AVPU	(Pain vs Alert)	*regression.*	
	AVPU	(Unresponsive vs Alert)		
Day 6	Gender	(Female vs Male)	1.13	[0.57 – 2.22]
Profile 3	Age	(Increase of 5 years)	0.90	[0.72 – 1.13]
	Obesity	(Obese vs Non-obese)	1.06	[0.53 – 2.12]
	CRP	(Increase of 10)	0.92	[0.84 – 1.02]
	Hypertension	(Hypertension vs No hypertension)	*NA due to singularity of data*	
	AVPU	(Verbal vs Alert)	3.39	[0.64 – 17.94]
	AVPU	(Pain vs Alert)	*NA due to missing data*	
	AVPU	(Unresponsive vs Alert)	*NA due to missing data*	
Day 7	Gender	(Female vs Male)	1.08	[0.70 – 1.66]
Entire population	Age	(Increase of 5 years)	0.89	[0.77 – 1.02]
	Obesity	(Obese vs Non-obese)	1.11	[0.70 – 1.75]
	CRP	(Increase of 10)	0.99	[0.96 – 1.02]
	Hypertension	(Hypertension vs No hypertension)	1.04	[0.67 – 1.60]
	AVPU	(Verbal vs Alert)	1.46	[0.76 – 2.81]
	AVPU	(Pain vs Alert)	*NA due to missing data*	
	AVPU	(Unresponsive vs Alert)	0.88	[0.45 – 1.70]
Day 7	Gender	(Female vs Male)	2.72	[2.72 – 2.72]
Profile 1	Age	(Increase of 5 years)	1.00	[1.00 – 1.00]
	Obesity	(Obese vs Non-obese)	1.00	[1.00 – 1.00]
	CRP	(Increase of 10)	1.00	[1.00 – 1.00]
	Hypertension	(Hypertension vs No hypertension)	1.00	[1.00 – 1.00]
	AVPU	(Verbal vs Alert)	1.00	[1.00 – 1.00]
	AVPU	(Pain vs Alert)	*NA due to missing data*	
	AVPU	(Unresponsive vs Alert)	1.00	[1.00 – 1.00]

Day 7	Gender	(Female vs Male)		
Profile 2	Age	(Increase of 5 years)	*Oddsratio's and Confidence*	
	Obesity	(Obese vs Non-obese)	*Intervals could not be calculated*	
	CRP	(Increase of 10)	*due to lack of contrast. This profile*	
	Hypertension	(Hypertension vs No hypertension)	*holds 100% survival rate after*	
	AVPU	(Verbal vs Alert)	*deletion of missing data for logistic*	
	AVPU	(Pain vs Alert)	*regression.*	
	AVPU	(Unresponsive vs Alert)		
Day 7	Gender	(Female vs Male)	0.29	[0.10 – 0.84]
Profile 3	Age	(Increase of 5 years)	1.14	[0.81 – 1.61]
	Obesity	(Obese vs Non-obese)	1.61	[0.82 – 3.19]
	CRP	(Increase of 10)	1.17	[0.90 – 1.53]
	Hypertension	(Hypertension vs No hypertension)	0.43	[0.16 – 1.15]
	AVPU	(Verbal vs Alert)	*NA due to singularity of data*	
	AVPU	(Pain vs Alert)	*NA due to singularity of data*	
	AVPU	(Unresponsive vs Alert)	*NA due to singularity of data*	

Discussion

The performed research used latent profile analysis to identify the distinct subpopulations, or patient profiles, within each day of ICU admission. The number of profiles by day, were found with the help of the Akaike information criteria and the Bayesian information criteria. Lower scores for AIC and BIC indicate a bitter fit with the data than higher scores. Entropy scores were calculated to indicate the overlap between different profiles, where a higher score means better distinction between profiles.

For five of the seven days, models with three profiles resulted into the best fit. The other two days had the best fit for models with two profiles. The different profiles were not one-on-one comparable between the different days, although they did hold some similarities. However, they did show a differing effect and statistical significance of the models parameters, meaning not every variable has the same effect and is not of equal importance in the observed profiles. This result was also suggested by Rothmans causal pie model and is in line with the literature.

On a first glance of the results, CRP did seem to hold a relations with survival chances. The odds ratio's showed that is only the case in four of the models. This is an odd observation, because CRP is a signal variable for inflammation. However, the absence of a relation could be explained due the overall high CRP value in the populations, and thus losing its predictive value.

Strengths and weaknesses:

The main strength of this research is the use of Latent Profile Analysis as an objective way to stratify to the population into observed subpopulations, before the logistic regression analysis is used. By using LPA prior to regression analysis, it became clear that distinct subpopulations could be observed, in which not every variable holds the same effect and statistical significance. This is in line with the turn opens the road for more accurate prediction models based on specific patient profiles, instead of predictions based on the entire population.

Another strength of this research is the relatively large dataset, which held a total of 1126 observations from 450 patients, where one observation is one day a patient has been reviewed. The entirety of the dataset holds 370 variables, which have been measured accurately.

The patients in this research were enrolled with an opt-out approach, as opposed the more common informed consent procedure. This opt-out approach was deemed ethical, and in itself lowers the risk of selection bias, since there will not be an overrepresentation of certain groups who are more likely to consent for medical research. However, the generalizability of this research might be reduced, due to the analysed data only being drawn from the local population in the Enschede area. Therefore, after developing the model, validation on a bigger, more diverse database will be required to fully assess the usability.

This need for external validation becomes all the stronger when looking at the population size by ICU admission day. While a diminishing population size is also to be expected from external databases, the total population by ICU day needs to exceed the minimum of 100 events, or patients in this case, which was identified as a risk factor for bias in the analysis. A bigger populations size by day, would most likely return smaller confidence intervals around the probability of group membership, boosting the certainty of the group size of each profile. Additionally, this increase in the research populations, would make the regression analyses more reliable, opening the path to draw generalizable conclusions for the subpopulations.

Furthermore, model stability decreased over time with lower number of research participants, which resulted in a relatively low number of variables. Also, certain comorbidities and risk factors were pretty rare in the database. For example, when liver disease were to be taken into the model, the low prevalence of this disease in the research population would make it unlikely that it holds much weight in the grouping of the profiles. Meaning, that the probability of finding this variable in any of the profiles would be low, adding little to no useful information to the model. Currently, only conclusions can be drawn for the first day and its profile, since this population exceeds the limit of 100 events. However, external research on a new dataset would still be required to validate the different profiles and their characteristics.

Then there is the issue of missing data. Most patient characteristics as described by Table 1, vary between 13 and 27 missing data inputs, which is actually quite good for such a large database. However, three variables have considerably more missing data. These variables are BMI scores, and indication of obesity, with missing values of 198 and 71 patients, respectively. Data on smoking habit had a total of 148 missing values, which is also considerably bad. However, this is less of an issue, since smoking behaviour was not selected as a model parameter.

The input of data within the database was done by different people than the clinicians who see the patients. Additionally, the measures are either not standardized, or not clearly reported. For example, if a patients is clearly not obese based on visual assessment, but there is no note of BMI-scores or weight in the patients records, whatsoever. Without any documented indication for obesity, it would greatly increase the risk of bias to assume their status, and therefore information about a patient's obesity will not be included in the database. This will result into missing data of a known measure.

Model validation

As to be expected with the smaller population sizes, the error margin increases by day, leading to bigger confidence intervals around the estimated probabilities for group membership. This resulted in relatively large uncertainties for the group probabilities on later days. Nagin's criteria listed a small confidence interval around these probabilities for group

membership, however there was no clear definition given for small confidence intervals, leaving room for interpretation on which values of the maximum margin error are still acceptable.

Future research

Since this research uses an ISARIC CRF-inspired database, a lot of factors are at the disposal of this research. Besides the goal of identifying the different subpopulations, another goal of this research was to find the relevant prognostic risk factors. While the first goal was achieved, for the secondary goal was not fully achieved. The available variables in the database might still not give the desired holistic view of COVID-19 prognostic factors. For example, genetic factors might be factors which affect the prognosis of the COVID-19 disease, but are not included into this database.

While an increase in potential parameters could provide more information, an increase in model parameters will decrease the model stability. A solution to gain more information for future research, could be adding conditional variables to the models, for example by performing LPA only for patients with a positive diagnosis for diabetes.

Conclusion

Due to a relatively small database after the first day of ICU admission, and the large amount of missing data in the database, no definitive conclusions can be drawn for the different profiles by day. Additionally, the COVID-19 severity score had to be dropped from the regression analyses, making it impossible to suggest improvements to the classification model of the WHO.

Multiple profiles delivered regression models that hold no statistically significant parameters, meaning that none of the selected variables has a predictive value on the outcome. Comparing this sudden absence of significance to the statistical significance of the same variables on the total population, a conclusion about differences in subpopulation can be drawn. It means that while certain variables are of importance on the entire population (e.g. age), the same variable becomes negligible within certain subpopulations.

While not all LPA and regression models are reliable due to the small number of patients, it does illustrate the fact that different subpopulations exist within the research population, and that these different subpopulations hold different effects and significance levels on their predictors. The additional value of Latent Profile Analysis prior to regression analysis, in order to identify the effect and relevance of predictors within the subpopulations, as opposed to the entire population, can still be seen.

In order to draw definitive conclusions on the days later than the ICU admission day, and to further validate the value of this approach, additional research with a database consisting of more participants is needed.

Finally, the taken approach should be recreated with a different set of variables in the model, in order to fully understand the COVID-19 prognostic factors.

References

1. Coronavirus (COVID-19) events as they happen [Internet]. [geciteerd 19 maart 2021]. Beschikbaar op: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen

2. ISARIC 4C consortium [Internet]. [geciteerd 19 maart 2021]. Beschikbaar op: https://isaric4c.net/risk/

3. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, e.a. Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. BMJ [Internet]. 7 april 2020 [geciteerd 19 maart 2021];369:26. Beschikbaar op: https://www.bmj.com/content/369/bmj.m1328

4. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: A resampling study. Vol. 35, Statistics in Medicine. 2016. p. 214–26.

5. ROTHMAN KJ. CAUSES. Am J Epidemiol [Internet]. 1 december 1976 [geciteerd 19 maart 2021];104(6):587–92. Beschikbaar op: https://academic.oup.com/aje/article/139202/CAUSES

6. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. Annu Rev Clin Psychol. 2010;6(March 2010):109–38.

7. Lanza ST, Bray BC, Collins LM. LATENT CLASS ANALYSIS An introduction to latent class and latent transition analysis. In: Handbook of Psychology. p. 814–65.

8. Son KB, Lee TJ, Hwang SS. Disease severity classification and covid-19 outcomes, Republic of Korea. Bull World Health Organ. 2021;99(1):62–6.

9. Abate BB, Kassie AM, Kassaw MW, Aragie TG, Masresha SA. Sex difference in coronavirus disease (COVID-19): a systematic review and meta-analysis. BMJ Open [Internet]. 6 oktober 2020 [geciteerd 17 februari 2021];10(10):e040129. Beschikbaar op: https://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2020-040129

10. Caramelo F, Ferreira N, Oliveiros B. Estimation of risk factors for COVID-19 mortality - preliminary results. medRxiv. 25 februari 2020;

11. Chen Y, Klein SL, Garibaldi BT, Li H, Wu C, Osevala NM, e.a. Aging in COVID-19: Vulnerability, immunity and intervention. Vol. 65, Ageing Research Reviews. Elsevier Ireland Ltd; 2021. p. 101205.

12. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, e.a. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. JAMA [Internet]. 17 maart 2020 [geciteerd 17 februari 2021];323(11):1061. Beschikbaar op: https://jamanetwork.com/journals/jama/fullarticle/2761044

13. Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, e.a. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. JAMA Intern Med [Internet]. 1 juli 2020 [geciteerd 17 februari 2021];180(7):934. Beschikbaar op: https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2763184

14. Yang C, Chapman KR, Wong A, Liu M. α1-Antitrypsin deficiency and the risk of COVID-19: an urgent call to action. Lancet Respir Med. 2021;9(4):337–9.

15. Kopel J, Perisetti A, Roghani A, Aziz M, Gajendran M, Goyal H. Racial and Gender-Based Differences in COVID-19. Vol. 8, Frontiers in Public Health. Frontiers Media S.A.; 2020.

16. Contou D, Fraissé M, Pajot O, Tirolien JA, Mentec H, Plantefève G. Comparison between first and second wave

among critically ill COVID-19 patients admitted to a French ICU: no prognostic improvement during the second wave? Crit Care [Internet]. 1 december 2021 [geciteerd 17 februari 2021];25(1):3. Beschikbaar op: https://ccforum.biomedcentral.com/articles/10.1186/s13054-020-03449-6

17. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, e.a. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet [Internet]. 15 februari 2020 [geciteerd 17 februari 2021];395(10223):507–13. Beschikbaar op: https://linkinghub.elsevier.com/retrieve/pii/S0140673620302117

18. World Health Organization. Fact sheets - Malnutrition [Internet]. [geciteerd 7 juli 2021]. Beschikbaar op: https://www.who.int/news-room/fact-sheets/detail/malnutrition

19. Kruizenga H, Beijer S, Huisman-de Waal G, Jonkers-Schuitema C, Klos M, Remijnse-Meester W, e.a. Guideline on Malnutrition: recognising, diagnosing and treating malnutrition in adults. Dutch Malnutrition Steer Gr. 2017;(August):49–64.