Ideation Phase

Literature Survey

Date	17 September 2022
Team ID	PNT2022TMID18388
Project Name	Analytics for Hospitals' Health-Care Data
Team Leader	Laxmi Priya S
Team Member 1	Kaviya M
Team Member 2	Keerthana R
Team Member 3	Madhumitha J

Abstract:

As healthcare organizations around the world are challenged to reduce costs, improve coordination with care teams, provide more with less, and focus on improving patient care, analytics will be especially important. Primary care physician and nursing shortages are requiring overworked professionals to be even more productive. Plus, new businesses entering the market and new approaches to healthcare delivery will increase competition in the industry. Building analytics competencies can help healthcare organizations harness big data to create actionable insights that can be used by healthcare providers, hospital and health system leaders, and those in government health and human services to improve outcomes deliver value for the people they serve.

Introduction:

The length of stay (LoS) is a key indicator of how efficiently hospitals are being managed and is used to assess the efficiency of hospital management and patient quality of care, and for functional evaluations. A shorter stay means that more beds are available for more patients and reduces hospital resource consumption; thus, it corresponds to a decrease in health-related expenditure [6]. Reducing LoS has been linked to lower risks of opportunistic infections and medication side effects [7], and improved treatment outcomes and lower mortality rates. In addition, a shorter hospital stay reduces the burden of medical fees and increases bed turnover, and thus increases hospitals' profit margins [7] while lowering the overall social costs [8,9,10]. Researchers must determine which characteristics are associated with longer or shorter hospital stays in patients [11].

Use Case:

Understanding and predicting hospital bed demand (as well as associated staff or equipment requirements) provide crucial evidence for decision-making and contingency planning. Predicting demand for hospital services requires an estimate of the number of patients requiring hospitalisation and an estimate of how long each person will require hospital care. It is possible to model the rate of hospitalisation in many settings based on estimated epidemic curves. However, estimating length of stay (LoS) in hospitals requires observation of individual patient pathways.

COVID-19 presents at varying levels of severity. Hospital care can vary from general ward-based care to high dependency units with oxygen support to intensive care where patients may be intubated for mechanical ventilation. The LoS is likely to depend on the level of care required, as well as the geographic setting due to varying COVID-19 care guidelines.

Moreover, patient characteristics—such as age and comorbidities—impact disease severity [8, 12–14] and are likely to influence LoS. If differences are significant, then capacity planning may need to account for these characteristics to provide accurate predictions of the number of beds required at each level of care. Modelling studies predicting bed occupancy published so far have broadly relied on very few sources of information for LoS estimates, which were often derived from very different settings [15–22]. Estimates for LoS can be obtained from a variety of studies, but are often an incidental result rather than a study's primary outcome, and typically, only summary statistics are reported. In general, LoS distributions are right-skewed due to a minority of patients with long hospital stays and are often modelled using gamma, log-normal, or Weibull distributions [23] (although log-normal is less preferred due to its heavier tails). A particular challenge is how to synthesise appropriate LoS distributions from a range of relevant sources in similar settings, capturing the variation both within and between them. Incorporating the uncertainty and stochasticity in parameters using a distribution, rather than fixed point estimates (such as the mean over all studies), allows for more realistic model predictions.

We performed a systematic review to identify the current evidence on LoS for COVID-19 patients worldwide. We also present a method for generating LoS summary distributions by combining information from different summary statistics (mean and medians) reported in multiple studies, and accounting for differences in sample sizes. This aims to include all the variation between studies, to obtain a distribution that covers all plausible LoSs. Although similar in the sense of synthesising multiple sources, this is unlike a classic meta-analysis which aims to get a more precise estimate of a quantity assumed as being a fixed point value. In doing this work, we aim to inform the efforts of modellers and policy makers to better anticipate healthcare needs during the evolving COVID-19 pandemic.

Keywords: count data model, length of stay, COVID-19, generalized linear model, Hurdle model, Vuong test, AIC, Rootograms, quantile regression.

Population

This study obtained data with permission from the local health authority (AUSL) of Bologna. Data were retrieved from 1 February 2020 to 10 May 2021. Our dataset referred to COVID-19 patients both in the ICU (intensive plus sub-intensive) and ordinary settings in 7 Bolognese hospitals. We also categorized the hospital stays as “regular” for 5 hospitals, and “long-term” for 2 hospitals. This distinction is important because hospitals providing extended hospitalization have mainly rehabilitation purposes.

Database Pre-processing

Individual patients were considered on the basis of their unique ID, and we merged the repeated IDs (patients who went to the hospital more than one time) via the following procedure:

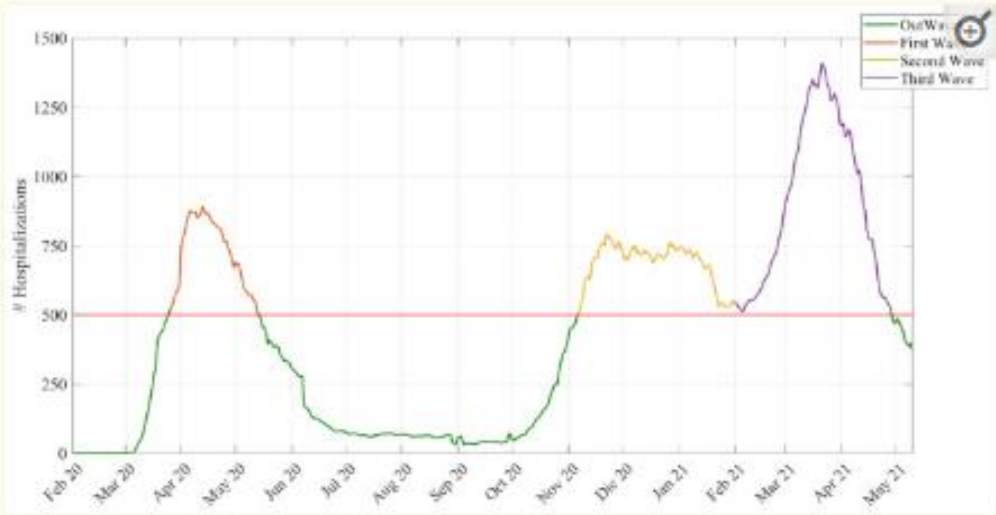


Figure 1

Number of hospitalizations over the study period and the empirical definitions of the waves.

Hospital: prevalent hospital

-

Setting: add a variable with two levels: “low-intensity” if the patient was only hospitalized in the low-intensity setting and “ICU” if the patient was in the intensive care unit at least one time.

We categorized the hospital stay information as regular hospitalization and long-term hospitalization based on the prevalent hospital.

Outcome and Covariates

The study’s main outcome was the LoS (a non-negative integer), defined as the number of days between inpatient admission and hospital exit or discharge, and it was the target outcome variable for which this study aimed to identify a proper count regression model. The explanatory variables were:

Clinical setting: ICU setting (intensive care + sub-intensive) of COVID-19 patients, and ordinary and low-intensity COVID-19 stays.

Age: An integer representing patient age in years, grouped in 10-year categories: [0, 10), [10, 20), [20, 30), [30, 40), [40, 50), [50, 60), [60, 70), [70, 80), and [80, 102).

First wave → 26 March 2020–13 May 2020 (48 days)

Second wave → 7 November 2020–1 February 2021 (86 days

Third wave → 2 February 2021–28 April 2021 (85 days)

Out-waves → other periods.

Hospital stay: A factor containing hospital name and categorized as long-term or regular hospitalization.

Statistical Analysis

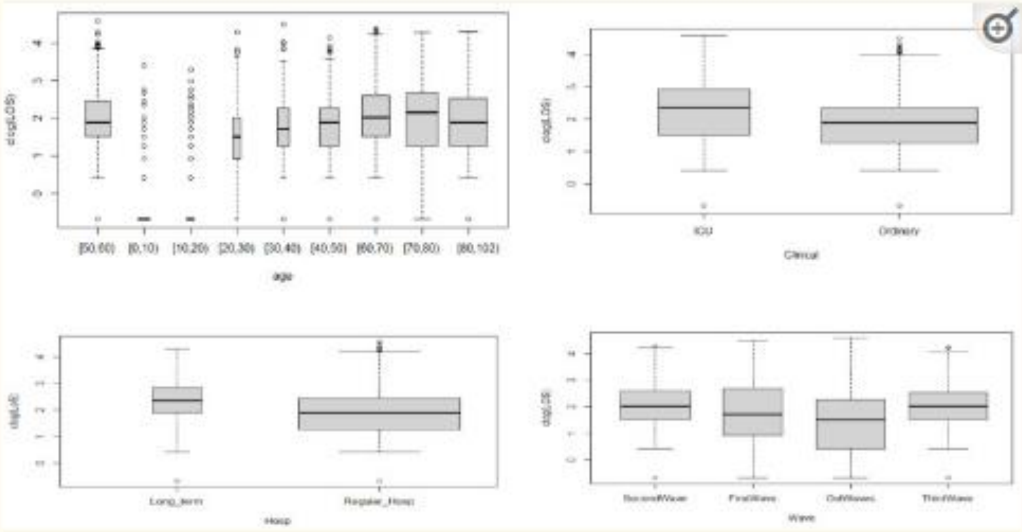


Figure 3

LoS plotted against the predictors used. All plots show the variation in LoS when the predictors were changed.

Vuong test:

The purpose of the Vuong test [38] is to compare two models (that are not nested) fitted to the same data by maximum likelihood, and it is based on a comparison of the predicted probabilities of two models that are not nested. Specifically, it tests the null hypothesis that the two models fit the data equally well. A large positive test statistic provides evidence of the superiority of Model 1 over Model 2, while a large negative test statistic is evidence of the superiority of Model 2 over Model 1, under the null hypothesis that the models are indistinguishable.

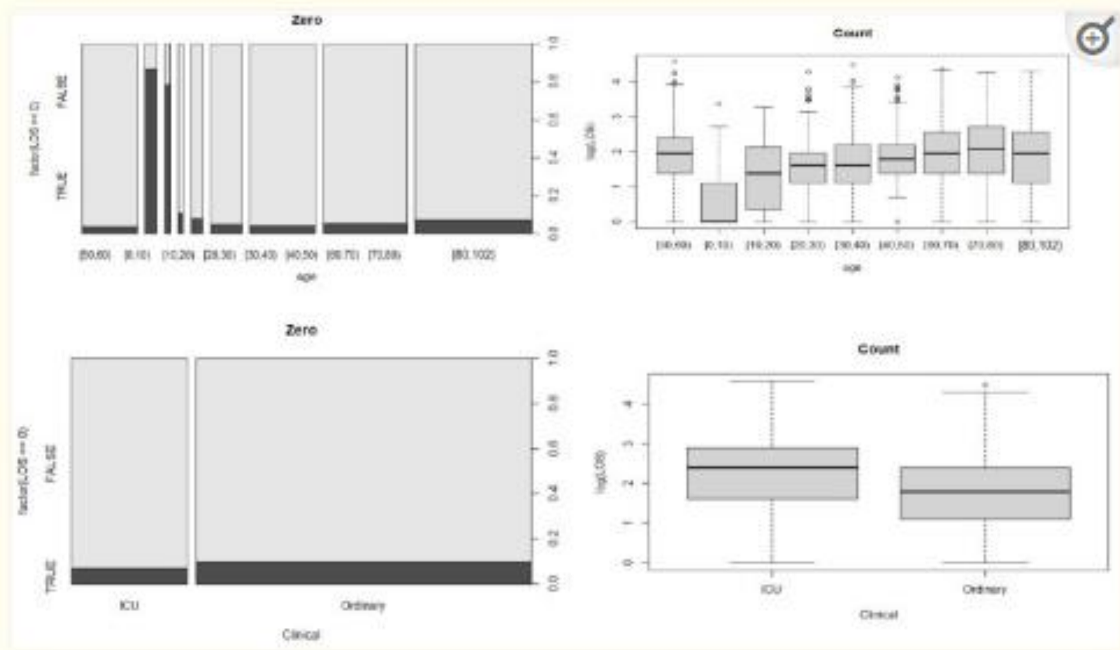
Let p1 be the predicted probabilities from Model 1, evaluated conditionally on the estimated MLEs. Let p2 be the corresponding probabilities from Model 2. The Vuong statistic is:

Estimates of coefficients: exponential (coef).

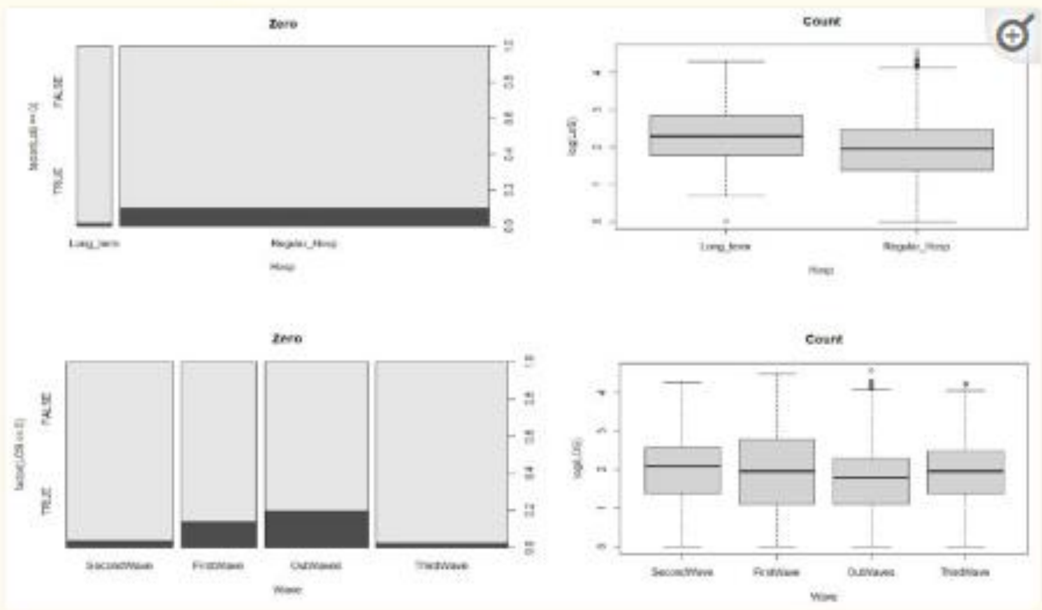
Hurdle-NB			NB
	Count_Model	**Zero_Hurdle_Model**	-
Intercept	6.72	47.57	7.20
Clinical–ICU	1.84	***0.90***	1.77
Long-term	1.58	3.61	1.61
Age [0, 10)	0.29	0.01	0.06
Age [10, 20)	0.64	0.02	0.17
Age [20, 30)	***0.87***	0.40	0.84
Age [30, 40)	0.88	0.49	0.86
Age [40, 50)	0.87	***0.71***	0.87
Age [60, 70)	1.12	***0.85***	1.10
Age [70, 80)	1.19	0.64	1.15
Age [80, 102)	1.19	0.44	1.12
First wave	1.21	0.34	1.11
Third wave	0.95	***1.30***	***0.96***
Out-wave	0.84	0.30	0.77

Quantile Regression Model

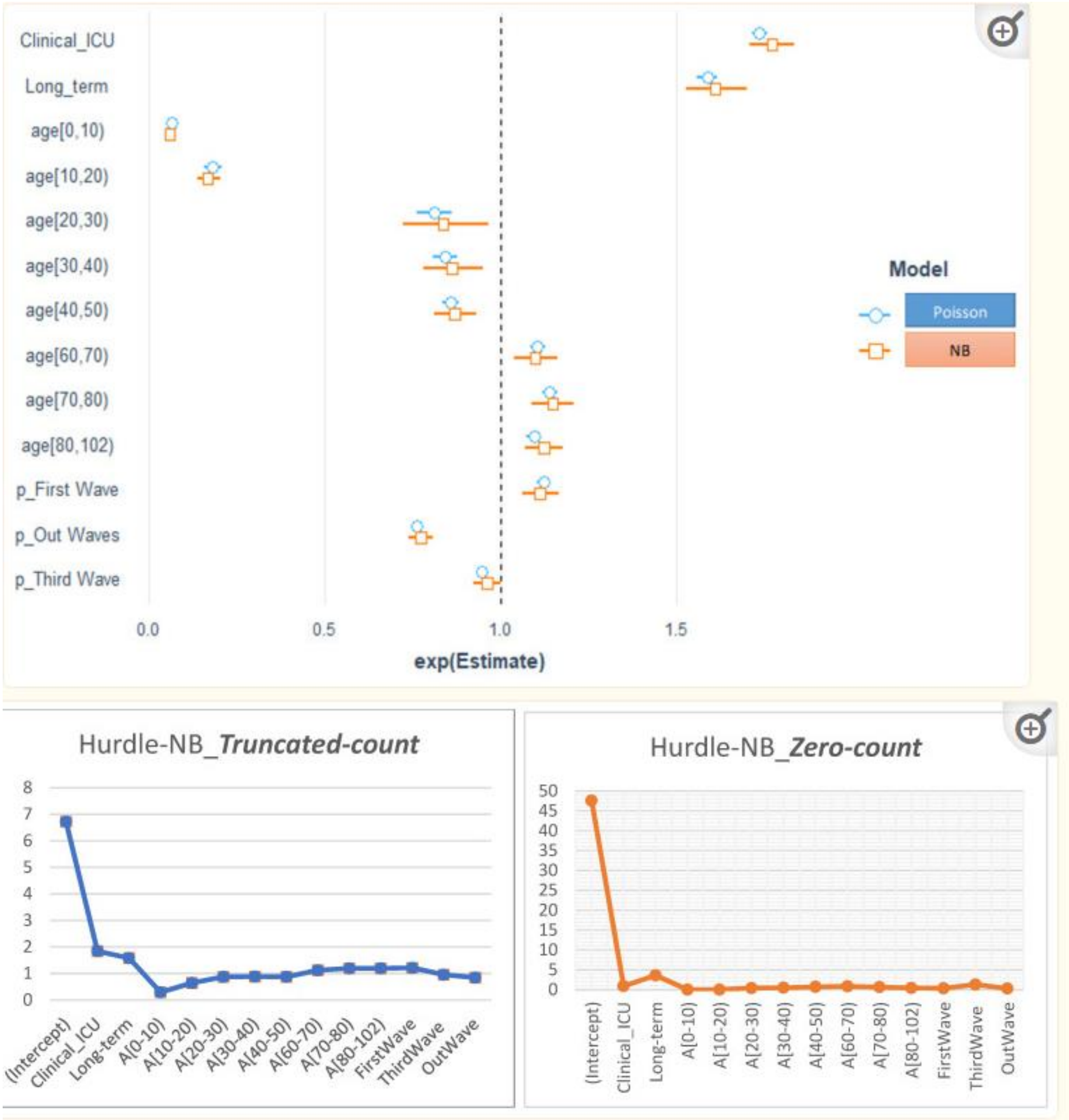
A quantile regression model can be used to explore the relationships between the quantiles of the response distribution and the available covariates. By comparing such quantiles, we can obtain a more complete picture of the conditional distribution than we can with regression models that consider the mean. In addition, quantile regression allows researchers to explore a range of conditional quantile functions, thereby exposing various forms of conditional heterogeneity and controlling for unobserved individual characteristics.



Zero
Count
age
ICU
Ordinary
Clinical



Zero
Count
SecondWave
FirstWave
OutWaves
ThirdWave
Wave



Discussion

Researchers in the medical field are currently working to improve the quality and efficiency of health care systems and services in various ways, with LoS [26] being one of the efficiency and quality indicators. To the best of our knowledge, no previous study has examined the LoS in Bologna, Italy, among COVID-19 patients, and this is the first study to consider analyzing LoS using the best count fit model and comparing several models. This study also aimed to explore the hospital admission risk factors associated with LoS and presents a relatively novel method for modeling LoS using predictors.

The necessity of carefully picking a model that effectively describes the observed count data is demonstrated by the fact that the different statistical techniques produce mixed findings. First, we used the Poisson, negative binomial, Hurdle–Poisson, and Hurdle–NB regression models to model the effects of covariates on LoS. Second, we used quantile regression to model the impact of the variables on the quantile values of the response variable LoS. Compared with Poisson regression, the fit of the NB regression better tolerated the overdispersion in the data [20,24,25]. In addition to the overdispersion, the Hurdle models accounted for zero LoS more thoroughly [26]. The Hurdle–NB model was finally chosen on the basis of three criteria, including the AIC, the Vuong test, and Rootograms to understand the impacts of the predictors on the average LoS.

In our analysis, the median LoS was 6 days, which is comparable with the value reported in a similar study in Europe, USA, and the UK, but shorter than one in a study on China [12]. In this systematic review, the median hospital LoS ranged from 4 to 53 days within China (45 studies surveyed), and from 4 to 21 days outside China (eight studies surveyed). Similarly, a shorter LoS was also documented by the ISARIC (International Severe Acute Respiratory and Emerging Infections Consortium) report [40]. This report (which included data from 25 countries) described a median and IQR LoS of 4 and 1–9 days, respectively, which are substantially lower than in our study. Differences in LoS duration between nations can be explained by the different policies or strategies applied to control COVID-19 infection, or by the different population samples involved in the studies. Knowing the LoS or other adverse events in advance can help te health care systems organize the allocation of limited resources more efficiently.

Our results in Bologna are entirely consistent with those from studies which observed that older age (≥60 years) was associated with prolonged LoS [41,42,43]. Patients with ICU admissions had a longer LoS than those with only ordinary hospitalization, as these patients might need additional treatment or time when their disease reaches severe stages and they need more complex treatments. Staying in a long-term hospital was another contributing factor for a LoS longer than that in regular hospitals. This might be explained by the higher percentage of surgical operations and transfer rates, as well as restricted antibiotic use compared with other patients.

Usually, public hospitals manage acute patients' LoS efficiently [44] in Italy. However, due to the new variants of the virus that causes COVID-19, the ability to assess or predict LoS will be more and more crucial in the future as a higher LoS is also associated with higher costs [45] and reduced capacity for other sanitary needs [46]. From this perspective, it is worth noting that we found that the LoS in Bologna was higher during the first wave—when the government proclaimed a state of emergency in response to an increase in the number of infection cases—and peaked around April 2021. However, during the third wave and out-wave, there was a drop, indicating that the chosen policies and strategies of the government and the health department, together with better clinical knowledge of the disease, have had a positive impact on the management of the patients.

We also used quantile regression to model the effects of the covariates on the quantile values of the response variable LoS, using a Poisson distribution, and to explore a range of conditional quantile functions, thereby exposing various forms of conditional heterogeneity and controlling for unobserved individual characteristics. The results from the quantile regression showed that the covariates related to the ICU setting, long-term hospitals, age groups, the first wave, and the out-waves were statistically significant for all the modeled quantiles. Specifically, for the ICU setting, the coefficients were higher for higher quantile values than ordinary hospitalization, meaning a longer LoS, and the effect was more pronounced for the higher quantiles. All the age groups, except for the [40, 50) group, showed an increasing trend with the quantiles. The wave also played a role: during the first wave, short LoSs were shorter (negative coefficients) and long LoSs were longer (positive coefficients) compared with those of the second wave. LoSs recorded in the out-wave period were shorter for all the quantiles (negative coefficients), and for the third wave, longer LoSs shortened (negative coefficients). The coefficients for long-term hospitalization became closer to zero as the quantile values increased, representing a less marked difference with respect to regular hospitalization.

The quantile regression models provided a more comprehensive overview of the effect of the covariates on a given phenomenon. As appears evident from this work, the same covariate can have a non-constant effect on different quantiles. This is valuable information for better modeling the LoS, which would not be controlled by simply evaluating the effect of the covariates on the mean values.

One limitation of our study is the absence of information on additional demographic and socio-economic variables, clinical characteristics, medical conditions, and laboratory tests for the patients, which might be among the most important factors that tend to increase the LoS. However, we have now started to analyze an extended, non-aggregated version of this dataset including all these variables.

COVID-19 cases are complex and still increasing worldwide, and it is difficult to predict when it will stop completely. Countries should plan and prepare for the worst-case scenarios, such as when different variants come into play (as in the case of Omicron at this time). In Italy, we have now entered into a fourth wave, so the wise use of limited health care resources is one of the most important priorities. Studies such as the present one might help health policymakers and managers to better plan the logistics of hospital settings, define priorities, and carry out a more accurate cost analysis. In addition, the information we have obtained and the method we used might serve as a tool or as a reference in the case of similar epidemics in the future.

Social Impact:

Access to primary healthcare, Less Casualty.

Business Model/Impact:

Pharmacy companies will sell their medical products to generate more revenue.

Insurance companies will sell their health policies to needed people.

Existing Solutions:

https://www.boldbi.com/dashboard-examples/healthcare
https://www.orangemantra.com/industries/healthcare/

Recommended Technology Stack:

Cognos Analytics, Tableau, Data Analysis with Python, Power-BI, etc.

References:

https://www.researchgate.net/publication/348834045_Development_of_the_Health_Information_Analytics_Dashboard_Using_Big_Data_Analytics

World Health Organization Coronavirus Disease (COVID-19) 2020. [(accessed on 25 October 2021)]. Available online: https://www.who.int/health-topics/coronavirus#tab=tab_1

World Health Organization WHO Director-General's Opening Remarks at the Media Briefing on COVID-19. 2020. [(accessed on 25 October 2021)]. Available online: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

Worldometers COVID-19 Coronavirus Pandemic. 2021. [(accessed on 21 December 2021)]. Available online: https://www.worldometers.info/coronavirus/

Specchia M., Di Pilla A., Sapienza M., Riccardi M., Cicchetti A., Damiani G. Dealing with COVID-19 Epidemic in Italy: Responses from Regional Organizational Models during the First Phase of the Epidemic. Int. J. Environ. Res. Public Health. 2021;18:5008. doi: 10.3390/ijerph18095008 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

Armocida B., Formenti B., Ussai S., Palestra F., Missoni E. The Italian health system and the COVID-19 challenge. Lancet Public Health. 2020;5:e253. doi: 10.1016/S2468-2667(20)30074-8. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

Feng C.X., Li L. Modeling Zero Inflation and Overdispersion in the Length of Hospital Stay for Patients with Ischaemic Heart Disease. In: Chen D.G., Chen J., Lu X., Yi G., Yu H., editors. Advanced Statistical Methods in Data Science. Springer; Singapore: 2016. (ICSA Book Series in Statistics). [CrossRef] [Google Scholar]

Baek H., Cho M., Kim S., Hwang H., Song M., Yoo S. Analysis of length of hospital stay using electronic health records: A statistical and data mining approach. PLoS ONE. 2018;13:e0195901. doi: 10.1371/journal.pone.0195901. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

Bueno H., Ross J.S., Wang Y., Chen J., Vidan M.T., Normand S.L., Curtis J.P., Drye E.E., Lichtman J.H., Keenan P.S., et al. Trends in length of stay and short-term outcomes among Medicare patients hospitalized for heart failure, 1993–2006. JAMA. 2010;303:2141–2147. doi: 10.1001/jama.2010.748. [PMC free article] [PubMed] [CrossRef] [Google Scholar]