Research Notes – 1

Date: 19th September, 2024

Research Paper

Title: Prevalence of physical frailty, including risk factors, up to 1 year after hospitalization for COVID-19 in the UK: a multi-centre, longitudinal cohort study

Authors: PHOSP-COVID Study Collaborative Group

Link: [Prevalence of physical frailty, including risk factors, up to 1 year after hospitalisation for COVID-19 in the UK: a multicentre, longitudinal cohort study - eClinicalMedicine (thelancet.com)]()

Focus of the Logbook

Identification of hypothesis and parameters of data analysis

Field of Research

Data Analysis: This paper uses longitudinal cohort data to analyze frailty risk factors and recovery trends. It relies heavily on statistical data analysis to track health outcomes over time.

Hypothesis

- Physical frailty and pre-frailty are common following hospitalisation with COVID-19. Improvement in frailty was seen between 5 and 12 months although two-thirds of the population remained pre-frail or frail *[1, p2]*

- Both pre-existing frailty and acute worsening from the hospitalisation contribute to the high proportion of pre-frail and frail people at 5 months.

Parameters of Data Analysis

- Post-hospitalisation COVID-19 study (PHOSP-COVID) - Cohort study across 35 UK sites consisting of 2419 people who had survived hospitalisation with clinically diagnosed COVID-19 between 5th March 2020 and 31st March 2021 *[1, p2]*
- Burden of frailty measured using Fried's Frailty Phenotype (FFP) *[1, p2]*
- Five FFP criteria: Unintentional weight loss, Weakness, Exhaustion, Slowness and Low physical activity *[1, p3]*
- 3 FFP groups in the cohort: Robust (no FFP criteria), Pre- frail (one or two FFP criteria) and Frail (three or more FFP criteria
- Primary outcome from study based on the prevalence of the FFP groups at two time points: 5 months and 1 year *[1, p2]*
- Demographics included age, ethnicity, working status prior to hospitalisation, length of hospital stay, co-morbidities and severity of acute illness and quintiles for index of multiple deprivation *[1, p3]*
- Patient-reported outcome measures at the research visit included patient-perceived recovery and health-related quality of life (HRQoL)

Notable points

- Significant risk factors: Old age, the presence of two or greater comorbidities, the requirement for intubation and ventilation at the time of COVID illness, female sex and higher social deprivation *[1, p6]*
- Pre-frail to frail – Increased lack of physical activity and slowness
- Significant reduction between two time points: Unintentional weight loss
- Health Related Quality of Life, HRQoL – remained low, no increase between 5 months to a year *[1, p9]*
- Frailty was more prevalent than in a general population and was driven by both pre-existing conditions and COVID-19-related factors *[1, p10]*
- Targeted interventions like exercise and rehabilitation may help mitigate frailty.
- *Cohort study revealed that most COVID-19 survivors exhibited pre-frailty or frailty up to a year post-discharge, with some recovery seen between 5 months and 1 year.*
- *There is a need for targeted interventions to address frailty and support those affected, while also preventing frailty development in the pre-frail group.*

Reference

Hamish J.C. McAuley, Rachael A. Evans, Charlotte E. Bolton, Christopher E. Brightling, James D. Chalmers, Annemarie B. Docherty, Omer Elneima, Paul L. Greenhaff, Ayushman Gupta, Victoria C. Harris, Ewen M. Harrison, Ling-Pei Ho, Alex Horsley, Linzy Houchen-Wolloff, Caroline J. Jolley, Olivia C. Leavy, Nazir I. Lone, William D-C Man, Michael Mark, Dhruv Parekh, Krisnah Poinasamy, Jennifer K. Quint, Betty Raman, Matthew Richardson, Ruth M. Saunders, Marco Sereno, Aarti Shikotra, Amisha Singapuri, Sally J. Singh, Michael Steiner, Ai Lyn Tan, Louise V. Wain, Carly Welch, Julie Whitney, Miles D. Witham, Janet Lord, Neil J. Greening, Neil.Greening@leicester.ac.uk on behalf of the PHOSP-COVID Study Collaborative Group. s"Prevalence of physical frailty, including risk factors, up to 1 year after hospitalisation for COVID-19 in the UK: a multi-centre, longitudinal cohort study". eClinicalMedicine (Part of THE LANCET *Discovery Science*) Volume 57, 101896, March 2023. Prevalence of physical frailty, including risk factors, up to 1 year after hospitalisation for COVID-19 in the UK: a multicentre, longitudinal cohort study - eClinicalMedicine (thelancet.com). Accessed 19 September 2024

Research Notes – 2

Date: 24th September, 2024

Research Paper

Title: The Role of Twitter Medium in Business with Regression Analysis and Statistical Modelling

Authors: Jia Cong, Zubair Ahmad, Basim S.O. Alsaedi, Osama Abdulaziz Alamri, Ibrahim Alkhairy, and Hassan Alsuhabi

Link: https://onlinelibrary.wiley.com/doi/full/10.1155/2021/1346994

Focus of the Logbook

Statistical Analysis of the Research Paper

Field of Research

Data Analysis: Regression analysis and statistical modeling are applied to Twitter data to explore its impact on business strategies, utilizing large data sets for correlation and prediction.

About this paper

- Studies the impact of Twitter advertising (the fourth most popular social media platform *[1, p1]*) on sales and marketing
- 53% of marketers worldwide used it as a channel for their marketing activities in 2020 *[1, p1]*
- 40% of Twitter users bought something after seeing it on this platform *[1, p1]*
- Recommendations from influencers resulted in a 20% increase in the sale of products *[2]*

- *The study indicates that Twitter is a powerful platform for businesses to communicate with customers and promote their products or services. Its real-time nature allows for rapid dissemination of information and feedback.*
- *Through regression analysis, the research demonstrates that there is a significant relationship between Twitter engagement metrics (like retweets and likes) and business outcomes, such as sales and customer acquisition.*
- *Statistical modelling analyzed data related to Twitter usage, revealing patterns that can inform business strategies.*
- *Recommendations to businesses to leverage Twitter more effectively to guide marketing decisions.*

Statistical Analysis

From [1, p2]

- Regression technique used to test the profitability of Twitter advertising.
- Hypothesis Tested: The impact of Twitter advertising on sales
- **Null Hypothesis (H_0):** Twitter advertising has no significant impact on sales
- **Alternate Hypothesis (H_1):** Twitter advertising has a significant impact on sales

- New Heavy-Tailed statistical model introduced. Termed 'Exponential T-X exponentiated exponential (ETXE-exponential) model'. *A combination of exponential distribution with exponential TX strategy.*

Regression Analysis

- Simple Linear Regression Approach:

  $Y = \theta 0 + \theta 1X + \varepsilon;$

  where, Y- sales, X- Twitter advertising, θ0 – intercept, θ1 – slope, ε – residual error

- For this research, the equation is:

  Sales = θ0 + θ1Twitter + ε *[1, p2]*

- Through regression,
  Sales = 5.621 + 0.193Twitter

  which confirmed there was a **positive linear relationship** between Twitter advertising and sales *[1, p3]* That is, the more money spent on Twitter advertising, the more the sales.

- Translating the above equation, the predicted sales in thousands of USD (United States dollars), when there is no investment on the Twitter medium is $5.621 * 1000 = 5621$ dollars *[1, p2]*
- With the Twitter medium as a marketing tool, the predicted/estimated sale is $5.621 + 0.193 * 1000 = 198.621$, representing a sale of 198621 dollars *[1, p3]*
- T-statistics: t-value is 12.652 and the p-value is 2e -16(less than 0.05) so NH is rejected *[1, p3]*
- Homoscedasticity (which confirms linearity) and normal distribution inferred from residual vs fitted plot, Q-Q plot and Scale-Location plot [1, p3]
- From correlation test, r = 0.5253 confirming positive relationship between Y= Sales and X = Twitter advertising *[1, p4]*
- Spearman rank correlation: Sp = 61 and p-value = 4.634e – 12 (>0.05), rejecting NH and confirming a significant relationship between X and Y.
- Normality Test:
  (1) Numerical method: Shapiro Wilk (SW) Test
    - For Twitter advertising data, SW = 0.96011 with p value = 0.000739
    - For the sale data, SW = 0.967900 with p value = 0.00356

    Both p-values less than 0.05, hence normality holds *[1, p5]*

  (2) Graphical method – Q – Q plot has a 45-degree angle confirming normality *[1, p5]*

Statistical Modeling

- Heavy-Tailed (HT) distributions show better performance in modeling data in the finance sector *[1, p5]*
  Because they better capture extreme variations or outliers, which are common in financial markets providing more accurate representations of the likelihood of extreme losses or gains.

- For this analysis, a new HT model called ETXE-exponential distribution which offers to model the data that are skewed to the right with a heavy tail is used *[1, p5]*
- Twitter advertising and sales data from https://data.world/datasets/twitter *[1, p7]*
- ETXE- exponential method found to be the best for financial data modeling based on comparison with other modeling methods (MOW, FW, APTW *[1, p8]*)
- Measures used were Information Criterion (IC) and Goodness of fit (g-o-f), which confirmed the suitability of this model for this specific research.

References

[1] Jia Cong, Zubair Ahmad, Basim S.O. Alsaedi, Osama Abdulaziz Alamri, Ibrahim Alkhairy, and Hassan Alsuhabi. "The Role of Twitter Medium in Business with Regression Analysis and Statistical Modelling". Google Scholar. https://onlinelibrary.wiley.com/doi/full/10.1155/2021/1346994. Accessed 24 September, 2024

[2] V. Israel-Turim, J. L. Mico-Sanz, and E. Ordeix-Rigo. "Who did the top media from Spain started following on twitter? An exploratory data analysis case study," American Behavioral Scientist, vol. 65, no. 3, pp. 512–539, 2021.

Research Notes – 3

Date: 2nd October, 2024

Research Paper

Title: Research on Optimization of Export Cross-border E-commerce Model Based on Big Data Analysis

Author: Yang Lihua, Guangdong Polytechnic of Science and Technology, Guangzhou, 510640, China

Link: https://ieeexplore.ieee.org/document/10129519

Focus of the Logbook

Optimization Techniques

Field of Research

Data Analysis: This research utilizes big data analysis to optimize e-commerce models, leveraging large-scale data to improve decision-making in export markets.

About this paper

- Focus on the intersection of big data and business model innovation. Essential for developing competitive business models in e-commerce *[1, p 487]*
- Advancement Over Traditional Methods: Big data analysis incorporates unstructured data types (e.g., images, videos) for more comprehensive insights *[1, p 489]*

- Analysis Scope: Broader analysis capabilities compared to conventional methods, allowing full data set examination
- Comprehensive Services: Provides interactive business services for global manufacturers and traders.
- Consumer Insights: Analyzes consumer preferences for product recommendations and after-sales services.
- Market Forecasting: Helps in precise market positioning and dynamic decision-making.

- Identifies key elements driving business model innovation through data analysis.
- Establishing indicator weights for business value impact.

- **Null Hypothesis (H_0):** The optimization of the cross-border e-commerce model using big data analysis does not significantly improve enterprise resource allocation or commercial value.
- **Alternate Hypothesis (H_1):** The optimization of the cross-border e-commerce model using big data analysis significantly improves enterprise resource allocation and commercial value

- Uses a pairwise comparison method to build a judgment matrix (Table I) *[1, p 489]*
- Three major business values of data analysis: High transparency and wide availability of data, the impact of decision validation on the way of competition, and the establishment of a data based business model.
- Evaluation of these three elements – Table II *[1, p 489]*

Mathematical Framework:

- Formulation of the random disturbance term: Facilitates accurate determination of the influence of various business values *[1, p 489]*
- Consistency Checks: Ensures the reliability and validity of the evaluation model *[1, p 489]*

Optimization through Big Data Analysis

From [1, p 489]

- **Fuzzy Comprehensive Evaluation**: Utilizes fuzzy matrices to evaluate components of business models, highlighting significant impacts of big data on transaction structures, market segmentation, and marketing channels.
- **High Influence**: Big data's commercial value impacts business model elements significantly, driving innovation.
- **Increased Transparency**: Information sharing is enhanced among enterprises.
- **Real-Time Insights**: Allows quick adaptation to market trends.
- **Enhanced Decision-Making**: Promotes accurate and data-driven business decisions.

- **Operational Improvements**:
  - Feedback mechanisms help refine services and strategies.
  - Reduces operational loopholes and inefficiencies.

Quantitative Analysis:

- **Resource Allocation Efficiency**: Improved by **18.47%** *[1, p 491]*
- **Business Value Increase**: Enhanced by **27.94%** *[1, p 490]*

E-commerce Optimization:

- **Consumer-Centric Approach**: Data-driven product recommendations lead to higher consumer satisfaction.
- **Cost Reduction**: Overall operational costs decrease, improving competitiveness.
- **Website Optimization**: Detailed, accurate information improves consumer understanding.
- **Feedback Loop**: Customer evaluations provide insights into platform performance.
- **Shift in Business Model Analysis**: Traditional methods are no longer sufficient; big data becomes a core production factor *[1, p490]*

- **Competitive Advantage through Big Data**:
  From [1, p491]
  - The research concludes that big data serves as a driving mechanism for developing competitive business models.
  - Emphasizes the increasing value of data assets in decision-making processes.

- **Recommendations for Enterprises**:
    - Enterprises should harness big data analysis tailored to industry-specific characteristics to drive product and service innovation.
    - By redefining their business models through a big data lens, companies can differentiate themselves and enhance operational efficiency *[1, p491]*

- **Sustainable Growth**: The paper advocates for leveraging big data for sustainable competitive advantage

Reference

[1] Yang Lihua. “Research on Optimization of Export Cross-border E-commerce Model Based on Big Data Analysis”. IEEEXplore. https://ieeexplore.ieee.org/document/10129519 . Accessed 2 October, 2024

Research Notes – 4

Date: 9th October, 2024

Research Paper

Title: Assessing the Risk of Software Development in Agile Methodologies Using Simulation.

Authors: Maria Ilaria Lunesu, Roberto Tonelli, Lodovica Marchesi, Michele Marchesi.

Link: https://ieeexplore.ieee.org/abstract/document/9548910

Focus of the Logbook

Simulation and Data Analysis Techniques

Field of Research

Data Analysis: Simulation methods are employed to analyze risk in Agile methodologies, using data to model project uncertainties and outcomes.

About this paper

- Software Process Simulation Modeling (SPSM) and Monte Carlo stochastic approach to model key risk factors in agile development namely project duration, number of implemented issues and issue completion time *[1, p1]*

  Monte Carlo simulations - to model the probability of various outcomes in a process that cannot easily be predicted due to the presence of random variables. Helps in decision-making by analyzing the range of possible outcomes and their probabilities.

- **Null Hypothesis (H_0):** The simulator does not accurately predict project completion times or provide reliable risk assessments for Agile software development projects.

- **Alternate Hypothesis (H_1):** The simulator accurately predicts project completion times and provides reliable risk assessments for Agile software development projects, improving the understanding and management of risks related to effort estimation and developer assignment.

- Addresses Requirements Risk, Project Complexity Risk and Planning & Control Risk *[1, p5]*

- Basic components are *[1, p6]*:

  1. Issues/Features
  2. Activities (Planning, Analysis, Coding, and Testing)
  3. Team members
  4. Events

- Proposed approach is relevant for project managers, who get a tool to quantitatively evaluate the risks, provided that the process and the project's data are properly modelled *[1, p17]*

Simulation and Data Analysis

Simulation Process *[1, p6]*

- Simulation starts at time t = 0 and proceeds by time steps of one day. Each day has 8 working hours.
- Issues are entered at given days drawn from a random distribution or real data and then each issue passes through the activity phase.
- Sum of all activities must be 100% which is equal to the efforts during one sprint.
- If a developer works on the same task they worked on the previous day, there is no penalty (p = 1)
- For team/developer/task switches, a multiplicative penalty factor p, with $p \geq 1$, is applied to compute the time effort, to model time waste due to task switching (extra time needed to study the issue, which is proportional to the size of the task)
- Penalty factor, p = 1.15, meaning a 15% increment in the work to be done when a developer changes the issue s/he is working on. Figure is consistent with the data by Tregubov et al. [2] who estimate that 'developers who work on 2 or 3 projects spend on average 17% of their effort on context switching between tasks from different projects' *[1, p9]*

Simulation Design *[1, p7]*

- Simulator is object oriented and is implemented using SmallTalk
- Input given: Process info, Team members' Data, Issue Data
- Output from the Simulator: Project Duration, Project Status at a given time, Issue-cycle time statistics
- Each developer i is characterized by a skill array (one skill for each Activity), and a productivity factor at time t, qi(t), obtained as the ratio between the number of closed issues of i-th developer at time t, Ci(t) and the number of project days elapsed:

$$\mathbf{qi(t) = Ci(t) / t}$$

Simulator Assessment *[1, p11]*

Three open-source projects analyzed to assess the performance of the Simulator:

1. **Test Engineering (TE)**
    - **Duration**: 570 working days
    - **Team Size**: 13 developers
    - **Issues Analyzed**: 675, categorized as bug, epic, story, and task.
    - **Effort Distribution**: Follows a Pareto distribution with a shape parameter $b \approx 1.35$

2. **Platform**
    - **Duration**: 622 working days
    - **Team Size**: 65 developers (average active size ~14-15).
    - **Issues Analyzed**: 853, including subtask, bug, story, and epic.
    - **Effort Distribution**: Approximates a Pareto distribution with b=1.38

3. **CORD (Central Office Re-architected as a Datacenter)**
    - **Duration**: 192 working days
    - **Team Size**: Not specified, but similar to others in terms of active developers.

    - **Issues Analyzed**: 523, classified as subtask, feature, bug, story, and epic.
    - **Effort Distribution**: Follows a Pareto distribution with b=1.51

- The simulator developed for assessing project durations was validated by simulating the time taken to complete issues based on JIRA data. Results showed good alignment with actual project durations, indicating the simulator's effectiveness in modeling real-world scenarios *[1, p11]*

- **Effort Estimation**: The percentage difference between estimated and actual times to close issues was minimal, with a standard deviation of about 0.22 *[1, p12]*
- **Project Duration Analysis**: The simulator demonstrated an ability to predict project durations with an error margin of 6% to 11% *[1, p15]*
- In 60-day intervals, the average percentage **error was typically under 10%,** with total completed issues differing less than 1.5% from real counts *[1, p15]*

- **RQ1:** To what extent it is possible to automatically import into the simulator data of real projects, from issue management systems?
  JIRA is used to import real-time project but manual intervention is needed for modelling activities and effort estimation *[1, p15]*

- **RQ2:** How accurate can the simulator be in predicting project completion times?
  Simulations were almost accurate predicting completion times with a 10% margin. Shortened intervals (60 days) improved accuracy with only 1 - 2% error over 100 simulations *[1, p15]*

- **RQ3:** Can the simulator be useful to estimate project risk (induced by errors in efforts estimation, and random developer issues assignment) with a Monte Carlo approach?
  Yes, by varying issue estimation errors and developer availability, the results can be fine-tuned to forecast completion times and risks accurately *[1, p16]*

- *The study concludes that the simulator is suitable for risk analysis in agile methodologies, providing insights into project performance and potential areas for intervention. The approach allows for better management of expectations and resources in software development.*

Feedback from Experts

Experts encouraged continued development of the simulator, highlighting the need for an interface to tune simulation parameters and emphasize risks around time and costs *[1, p16]*

References

[1] Maria Ilaria Lunesu, Roberto Tonelli, Lodovica Marchesi, and Michele Marchesi. “Assessing the Risk of Software Development in Agile Methodologies Using Simulation”. IEEEXplore. https://ieeexplore.ieee.org/abstract/document/9548910 . Accessed 9 October, 2024

[2] A. Tregubov, B. Boehm, N. Rodchenko, and J. A. Lane, “Impact of task switching and work interruptions on software development processes,” in Proc. Int. Conf. Softw. Syst. Process, Jul. 2017, pp. 134–138.