# Clinical utility and functionality of an artificial intelligence application to predict mortality in COVID-19: a mixed methods analysis.

Ahmed Abdulaal, Aatish Patel, Ahmed Al-Hindawi, Esmita Charani, Saleh A Alqahtani, Gary W Davies, Nabeela Mughal, Luke Stephen Prockter Moore

Submitted to: JMIR Formative Research
on: February 16, 2021

# *Table of Contents*

# Clinical utility and functionality of an artificial intelligence application to predict mortality in COVID-19: a mixed methods analysis.

Ahmed Abdulaal[1] MBBS, MRCS; Aatish Patel[1] MBChB, MRCP; Ahmed Al-Hindawi[1] MBBS, MRCA; Esmita Charani[2] PhD, MPharm; Saleh A Alqahtani[3, 4] MD; Gary W Davies[1] MD, FRCP; Nabeela Mughal[2, 5, 1] FRCPath, MRCP, MSc, BSc; Luke Stephen Prockter Moore[2, 5, 1] FRCPath, MRCP, PhD, MPH, MSc, DTM&H, MBChB

[1]Chelsea and Westminster NHS Foundation Trust London GB
[2]National Institute for Health Research Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance Imperial College London London GB
[3]King Faisal Specialist Hospital and Research Centre Riyadh SA
[4]Johns Hopkins University Baltimore US
[5]North West London Pathology Imperial College Healthcare NHS Trust London GB

**Corresponding Author:**
Luke Stephen Prockter Moore FRCPath, MRCP, PhD, MPH, MSc, DTM&H, MBChB
National Institute for Health Research Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance
Imperial College London
Commonwealth Building 8th Floor
Du Cane Road
London
GB

## Abstract

**Background:** Artificial neural networks (ANN) are an increasingly important tool in the context of solving complex medical classification problems. However, one of the principal challenges in leveraging AI technology in the healthcare setting has been the relative inability to translate complex models into clinician workflow at the point of care, in a time-efficient manner for end-users.

**Objective:** Here we delineate the development of a COVID-19 outcome prediction application which utilises an ANN and assess its usability in the clinical setting.

**Methods:** Usability assessment was conducted on the application using clinical vignettes followed by a semi-structured end-user interview. Usability was specified by effectiveness, efficiency, and satisfaction measures, reported with descriptive statistics. End-user interview data were analysed using a thematic framework, developing themes from the interview narratives.

**Results:** Thirty-one Nation Health Service (NHS) physicians at a London teaching hospital, ranging from first year post-graduate through to consultants (post-graduate year 20+). All participants were able to complete the assessment, with a mean time for each patient vignettes of 59.35 seconds (standard deviation (SD) = 10.35). Mean system usability scale (SUS) score was 91.94 (SD = 8.54), which corresponds with an adjective rating of “Excellent”. Thematic analysis described positive themes around (i) the intuitive user interface, and (ii) its utility as a clinical predictive tool. A negative theme was identified around (iii) The primary concern related to use of the application in isolation as opposed to in conjunction with other clinical parameters, yet most clinicians felt that the application could positively reinforce or validate their clinical judgement.

**Conclusions:** Translating AI technologies into the clinical setting remains an important but challenging task. We demonstrate the effectiveness, efficiency, and system usability of a web application designed to predict COVID-19 patient outcomes from an ANN.

(JMIR Preprints 16/02/2021:27992)
DOI: https://doi.org/10.2196/preprints.27992

### Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

Please make my preprint PDF available to anyone at any time (recommended).

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

✓ **No, I do not wish to publish my submitted manuscript as a preprint.**

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

# Original Manuscript

# Clinical utility and functionality of an artificial intelligence application to predict mortality in COVID-19: a mixed methods analysis.


## Authors:

Ahmed Abdulaal[1], Aatish Patel[1], Ahmed Al-Hindawi[1], Esmita Charani[2], Saleh A. Alqahtani[3,4], Gary W. Davies[1], Nabeela Mughal[1,2,5], *Luke SP Moore[1,2,5].

## Affiliations:

1. Chelsea and Westminster NHS Foundation Trust, 369 Fulham Road, London, SW10 9NH. UK.

2. National Institute for Health Research Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, Imperial College London, Hammersmith Campus, Du Cane Road, London. W12 0NN. UK.

3. King Faisal Specialist Hospital and Research Centre, Riyadh. Saudi Arabia.

4. Johns Hopkins University, Baltimore, Maryland. United States of America.

5. North West London Pathology, Imperial College Healthcare NHS Trust, Fulham Palace Road. London. W6 8RF. UK.

## *Corresponding author:

Dr Luke SP Moore, Consultant Infectious Diseases, Microbiology & Virology, Honorary Clinical Senior Lecturer, Imperial College London, Hammersmith Campus, Du Cane Road, London. W12 0NN. UK. Email: l.moore@imperial.ac.uk Telephone: 02033158273


## Running title:

Utility and functionality of COVID19 prognostic app.

## Abstract

### Background

The artificial neural network (ANN) is an increasingly important tool in the context of solving complex medical classification problems. However, one of the principal challenges in leveraging AI technology in the healthcare setting has been the relative inability to translate models into clinician workflow. Here we demonstrate the development of a COVID-19 outcome prediction application which utilises an ANN and assesses its usability in the clinical setting.

### Methods

Usability assessment was conducted on the application followed by a semi-structured end-user interview. Usability was specified by effectiveness, efficiency, and satisfaction measures. These data were reported with descriptive statistics. The end-user interview data were analysed using the thematic framework method, which allowed for the development of themes from the interview narratives.

### Participants

Thirty-one Nation Health Service (NHS) physicians at a West London teaching hospital, including foundation doctors, senior house officers, registrars, and consultants.

### Results

All participants were able to complete the assessment, with a mean time to complete separate patient vignettes of 59.35 seconds (standard deviation (SD) = 10.35). Mean system usability scale (SUS) score was 91.94 (SD = 8.54), which corresponds with an adjective rating of “Excellent”. The clinicians found the application intuitive and easy to use, with the majority describing its predictions as a useful adjunct to their clinical practice. The main concern related to use of the application in isolation as opposed to in conjunction with other clinical parameters. However, most clinicians felt that the application could positively reinforce or validate their clinical judgement.

### Conclusion

Translating AI technologies into the clinical setting remains an important but challenging task. We demonstrate the effectiveness, efficiency, and system usability of a web application designed to predict COVID-19 patient outcomes from an ANN.



### Key words:

COVID-19, Graphical user interface, Coronavirus, Machine learning, Artificial intelligence

## Introduction

Clinical big-data being collated in many healthcare settings have enabled prognostic scores to be developed based upon classical regression analysis, but these models frequently rely on laboratory parameters (which are not available in many primary care settings and in some low and middle income settings)[1]. Furthermore, because of a priori assumptions made in these regression models, they may fail to leverage the data fully to create accurate prognostic models. Artificial intelligence (AI) techniques represent a potential solution[2], allowing fuller use of big-data, including potentially identifying proxy-indicators (such as symptomatology and co-morbidities) for laboratory parameters which may predict COVID-19 outcomes. Such systems have been shown to be accurate and reliable when compared to traditional regression models[3,5]. However, one of the principal challenges in leveraging AI clinically for COVID-19 has been in translating systems to the clinical setting[4].

Developing systems to accurately predict outcomes in severe acute respiratory syndrome-coronavirus disease (SARS-CoV-2) has several potential benefits at the patient, departmental and organisational level. On a patient level, predictive models would allow for early critical care reviews of high-risk patients and early discussions around treatment escalation plans. Medical departments could more accurately estimate bed requirements and account for intensive care unit (ICU) resource allocation issues. In turn, healthcare organisations could better manage staffing levels and healthcare resource procurement and distribution.

We describe here the clinical operationalisation of an artificial neural network (ANN) which produces patient-specific mortality predictions for COVID-19 patients[3,5], and explore the development of a graphical user interface (GUI) to facilitate use of the system at the bedside. Subsequently, we go on to assess the utility and functionality (measuring effectiveness, efficiency, and satisfaction) of the GUI which leverages this ANN, analysing the translational pathway for its integration and use in a clinical setting.

## Methods

### Development of the ANN

An ANN was developed as previously described to prognosticate for COVID-19 patients[3,5]. A multilayer perceptron was trained and validated on 398 patients from a single London hospital, with an input of 22 features selected in accordance to previously published work[6–8], in turn developed after review of existing evidence of contributory factors.[9,10] Demographics included gender and age. Smoking history was also included. Comorbidities included the presence or absence of asthma/chronic obstructive pulmonary disease/chronic respiratory disease, hypertension, diabetes, congestive cardiac failure, ischaemic heart disease, chronic kidney disease, hepatic cirrhosis, or a cerebrovascular event history. Symptom data included the presence of absence of fever, cough, dyspnoea, myalgia, abdominal pain, diarrhoea, vomiting, altered mentation, collapse, and olfactory change or ageusia, as well as the number of days of symptoms prior to hospital admission. Data was anonymised at point of extraction and encoded from patient electronic health records by three healthcare practitioners (by EC, AP and AA).

The model weights were initialised with Xavier normal initialisation, and a dropout of 20% and 40% were used on the two hidden layers, respectively. Euclidean (L2) regularisation was further added to the hidden layers to further prevent overfitting. The model was trained on 318 patients, and model hyperparameters were optimised based on 10-fold cross validation of the training set. The ANN was then trained on the full training set and validated on a held-out test set of 80 patients. For each patient input, the model produces a single output using a sigmoid activation function (which bounds the result to between 0 and 1). This output represents the probability of death during the current hospital admission for the patient. Discriminative ability was measured using the area under the receiver operating characteristic curve (AUROC), and calibration was assessed both visually and by using the Brier score.

Data was collected as part of routine care by the responsible clinical team. No patient-identifiable data was used in this analysis. The study protocol was approved by the antimicrobial stewardship group at Chelsea & Westminster NHS Foundation Trust. The need for written informed consent was waived by the Research Governance Office of Chelsea & Westminster NHS Foundation Trust. The study was conducted in accordance with the Helsinki declaration.

### Development of GUI

A web-application was developed using Node.js, an open-source, cross-platform, javascript runtime environment[11]. Express[12], a web framework for Node.js which provides a set of tools for application development, was used to build the backend of the application. A combination of Nielsen's and Shneiderman’s heuristics of user interface design were used in generating the initial GUI[13]. An iterative development process based on usability assessments (see results) throughout the design cycle was used to further develop the interface, ensuring its intuitiveness and ease of use. The application is currently developed as an English-language application.

The application collects patient demographics, comorbidities, and symptomatology data[5]. The data is then converted into a normalised tensor (a multi-dimensional array of data which can be read by a machine learning algorithm[14]) in the browser. On the backend, this data is fed into the ANN[5] (the deep learning library, Tensorflow.js[15] was used to pass the data to the Node.js server), which makes a patient-specific mortality prediction, and the result is then returned to the user (**Figure 1**). The relative importance of patient-level factors with respect to the mortality prediction are displayed as a static figure on the results page. No patient data is stored by the application after a prediction is produced, and the application can be used for a new patient by navigating to the home screen.

## Study design

This was a between-subjects study with one condition: all participants used the application to predict mortality risk for several patients. Effectiveness was defined as successful completion of a task. This was measured by assessing whether participants were able to insert a complete patient dataset into the application GUI, and successfully navigate to the results screen. Efficiency was defined as the duration to complete a task. This was measured by timing participants for each patient-specific dataset they inserted into the application. This time-period was measured starting from when participants completed reading the introductory paragraphs until successful navigation to the results screen. Satisfaction was defined as a participant's perception of the effectiveness and efficiency of the application. Satisfaction was measured using the System Usability Scale (SUS)[16]. A semi-structured interview format was used after the SUS assessment to gather additional feedback on the application. This allowed for flexible data collection with open-ended responses whilst ensuring relevant topics were covered. [17,18]

## Participants

Several key informants[19] were chosen across different clinical settings and seniority levels to represent the varied roles in managing COVID-19 patients. For example, initial assessment of the patient might be carried out by a junior doctor in the emergency department, whilst a senior doctor could be involved with critical decision making, such as establishing treatment escalation plans.

Data saturation, defined as the point at which additional data would not add new information nor require changes to be made to the developed findings, was estimated to occur at 30 - 35 interviews[20]. Participants were recruited in person at a single hospital site. We used maximum variation and snowball sampling to increase the likelihood that findings represent a wide range of perspectives with regards to the semi-structured interviews[18,21].

## Materials and procedure

Informed written consent was obtained from all participants. Participants were made aware of their right to withdraw from the study at any point during data collection. Data was anonymised for all participants except for job title and age as this data was felt to be important for contextualising findings.

Demographic data and experience with electronics were collected verbally; baseline computer and phone application experience scores (on a scale of 1 representing novice experience, to 10 representing expert experience). Three fictitious patient datasets in the form of clerking sheets (medical histories) were given to each participant. Participants then entered the data into the application to generate a patient-specific mortality prediction on a computer device. This section of the assessment was timed.

Whilst participants were using the application, effectiveness and efficiency measures were collected. Once the tasks were completed, participants were given the SUS assessment on an online survey data-collection platform, and then the semi-structured interview was conducted. Audio recordings of the interviews were stored on a mobile device and transcribed (using Otter.ai, Los Altos, USA), and then analysed.

## Ethical approval and consent to participate

Data was collected as part of service development work by the responsible clinical team. Data was anonymised at the point of extraction by the care team. The analysis protocol was approved by the

Antimicrobial Stewardship Group at Chelsea & Westminster NHS Foundation Trust and this was confirmed as a service development.

## Data analysis

Usability, as measured by effectiveness, efficiency and satisfaction was reported with descriptive statistics. Interview data were analysed with a thematic framework method (by multiple researchers: AAb, AP, EC) which allowed for the development of themes from the interview narratives.[22]

## Results

Thirty-one healthcare workers were recruited from a single West London teaching hospital between June – August 2020. There were five (16.13%) foundation doctors (year 1-2 postgraduate), five (16.13%) senior house officers (years 3-4 postgraduate), 15 (48.39%) registrars or equivalent (year 5-10 postgraduate), five (16.13%) consultants (approximately year 10+ postgraduate) and one (3.2%) primary care general practitioner (GP). None were excluded from data analysis due to equipment failure or withdrawal from the study. Twelve (38.71%) were female. The mean age was 33.06 years, standard deviation (SD = 5.59). Mean baseline computer experience was 7.71 (SD = 2.07), and mean baseline smartphone experience was 8.58 (SD = 1.70).

## Effectiveness

All participants were able to complete the task. 78 of 93 vignettes (83.9%: 3 vignettes given to each participant) were completed correctly, producing the expected prediction results by the algorithm. The failure of participants to enter clinical parameters correctly into the GUI in 15 (16.1%) encounters was explored in the qualitative analysis below.

### Efficiency

The mean time to complete each vignette was 59.35 seconds (SD = 10.35). **Figure 2** shows the average duration of task completion for each patient vignette; participants completed the task more rapidly with each sequential attempt.

### Satisfaction

The mean SUS assessment score was 91.94 (SD = 8.54). This corresponds to an "A" on the University Grading Scale to help interpret SUS scores.[16] This score also corresponds with an adjective rating of "Excellent" on the adjective rating scale.[23]

### Semi-structured interview thematic analysis

## Uncertainty over COVID-19 prognostication underpin clinician concerns

When approaching the management of patients with COVID-19, there were a range of clinical concerns brought up by the physicians interviewed. Most revolved around patient care, with the majority "worried about the deterioration of patients and their treatment escalation plan" (P9: foundation doctor). Doctors on the front line found themselves asking "is this the correct setting for the patient?", and "found [themselves] predicting where to manage patients" (P4: consultant). This highlighted a difference in focus depending on specialty. Doctors working in the Emergency Department or community were more focussed on whether the patients "needed hospital admission" (P18: registrar) or if they could "be managed at home" (P31: general practitioner (GP)). In contrast, intensive care doctors' focus was on "the mode of oxygen delivery needed" (P25: foundation doctor) and "which patients were likely to need intubation" (P20: senior house officer).

Among a group of doctors there was uncertainty with regards to communicating prognosis with patients and their relatives. “Communicating that risk to the family and to the patient themselves is my biggest concern” (P30: registrar).

Several doctors highlighted the fact that there was “a large amount of uncertainty in management and unpredictability in patient outcomes” (P7: registrar) in patients with COVID-19. This was thought to arise from the fact that “current knowledge [of COVID-19] was poorly understood” (P31: GP) and that this made “risk stratification in an unknown disease extremely difficult” (P23: senior house officer).

As well as concerns about the general care of the patient and being in the appropriate care setting, there were some more specific questions doctors had regarding “renal, thrombo-embolic, and cardiac events secondary to COVID-19” (P27: consultant).

## Experience of the Artificial Neural Network COVID-19 Prognostication Application

Most doctors gave positive feedback, commenting that the application was “very well designed” (P3: registrar), and “easy to pick up given I had never seen it before” (P16: registrar) and that “the [GUI] is very intuitive” (P1: registrar). Many found it simple to navigate the graphical interface and input patient data, with the application being “not too wordy, easy to use” (P9: senior house officer). One participant liked that the application did not “need biochemical parameters”, making it more “useful in [the] ED setting” (P22: foundation doctor), as it negated the need to wait for blood results and provided a more rapid quantification of the patient’s risk. One clinician commented that the application allows you to “cut through noise” (P24: senior house officer) when faced with a complicated patient and helped to “pull different aspects together”. The result was useful, as it was "nice to have numbers that are patient-specific" (P24: senior house officer).

## Interpretation of the Artificial Neural Network COVID-19 Prognostication Application predictions

The mortality risk predictions for the different scenarios elicited a range of reactions from participants. Twenty-nine percent (29%) of doctors felt surprised by the results of the application. "I was surprised by how high the first mortality prediction was" (P16: senior house officer). Some clinicians felt the predictions were lower than expected. "I was surprised by some of the results, one lower than I thought" (P2: registrar).

Others felt the scores reflected their experience of COVID-19 patients. "Those numbers were relatively reasonable to what I have seen" (P10: registrar). One commented on the fact that "despite two of the scenarios appearing fairly similar, they had significantly different mortality predictions" (P31: GP). Overall, six felt that the mortality predictions were higher than expected, with one doctor explaining they felt the predictions were lower than expected. Four doctors felt the predictions were closely aligned to their clinical judgement.

## Impact of the Artificial Neural Network COVID-19 Prognostication Application on clinical practice

Where cases were clear cut in the clinician's mind, the app worked to positively reinforce clinical judgement. Some physicians noted that "in clinical practice, it's quite obvious who's going to go off" (P3: registrar). Nonetheless, some underscored the potential benefit of concordance between their clinical judgement and the application's predictions, "If I was planning to admit someone to ICU, this app might be useful in helping me make that decision. I'd base my management on my clinical judgement, but this might be a useful adjunct” (P6: consultant). Others felt it gives them a sense of positive reinforcement, “I think it gives reassurance regarding your clinical judgement, especially if the app is roughly in agreement with your inclination” (P7: registrar). Several critical care physicians focussed on “integrating [the score]” into their “own clinical judgement”, and that if “the tool then validates [their] suspicion, then it gives [them] a good positive predictive value” (P17: registrar).

With strong disparities, most doctors commented that they would revisit the case. “It would help you take a step back and look at the patient again irrespective of the score - I think that’s the main use of predictive calculators to me” (P13: registrar). Many explained that where they strongly disagreed with the algorithm, they would base their management on their personal clinical judgement. “If I looked at the tool and it said to me ‘okay, she's got a 4% chance of mortality’, but I look at the patient at the end of the bed and they appear incredibly frail, in that instance my judgement would overrule the application’s prediction” (P17: registrar).

Where a case was felt to be borderline, the app helped as an “adjunct to the doctor” (P25: registrar), to aid form a general impression of the case. Going further, some felt the application could actively “help with clinical decision making in more complicated or borderline patients” (P23: senior house officer).

Several doctors commented that it would act as an additional tool in their decision-making process, and thereby adding to their clinical judgement. Fourteen doctors explained the results may help them stratify the risk to their patients more effectively, ensuring the right care setting. For example, “It would allow me to risk stratify patients who are coming in - I might contact ICU earlier on” (P16: registrar), and “it would be good as a screening tool to risk-stratify patients” (P19: foundation doctor), and “it would help me stratify future risk in an unknown disease” (P20: registrar).

Many doctors felt the predictions could be used to “better communicate patient outcomes” (P24: foundation doctor) to the patient and their family, as well as “between medical colleagues” (P26: registrar). Topics which doctors felt would benefit from the application’s results included “communicating disease severity” (P27: consultant) and “the need for intensive care” (P30: registrar) to the patient and their relatives.

Five doctors felt the use of this application would not impact their clinical management, and one felt unsure of the utility of the application. “It’s tricky - I’m not sure whether it would alter my decision making in any appreciable way, but the numbers are interesting to see” (P11: consultant). However, most agree that given SARS-CoV-2 is a “new disease, having any source of prediction would be useful for guiding management, and might help as an adjunct to decide on escalation” (P8: SHO).

## User-driven evolution of the Artificial Neural Network COVID-19 Prognostication Application

Many noted that it would be more intuitive to elicit symptoms before comorbidities, as this workflow more closely aligns with many doctors’ clinical practice. “I found myself scrolling down to fill in some details and then scrolling up to fill in the rest” (P1: registrar). However, others tended to prefer inserting comorbidity data prior to symptomatology. “The flow makes more sense for my clinical practice” (P2: registrar). Two doctors felt there were many required variables for use of the app. “It might be easier to reduce the number of variables from 20 without reducing the model’s predictive power too dramatically. This might make it easier to use” (P3: registrar). However, one explains that this was not an important issue as the data was easy to accrue from the initial clerking. “There are a lot of yes/no boxes relative to other medical calculators, but that was alright because they were very easy to answer - data entry is elicitable from the clinical history” (P4: consultant).

One doctor noted they could not find a disclaimer to explain that the app should only be used in patients with known SARS-CoV-2 infection. Similarly, one doctor suggested the inclusion of a “disclaimer regarding the use of the app on first use” (P17: registrar), with the same physician noting the app should not be used in “isolation”. Another suggested the “addition of ethnicity in future” (P6: consultant) iterations of the model as an important prognostic factor. Yet another suggested “linking trust-based guidelines for COVID-19 management” on the results page of the application, or “integrating the results into the patient’s electronic health records” (P13: registrar).

Two doctors noted that it should be made clear that duration of symptoms is always taken from onset of first symptom by the app. “I think you should specify that the duration of symptoms is from the first symptom, as sometimes symptoms develop at different time points” (P16: registrar). Finally, being able to predict “intensive care requirements” (P6: consultant) and “prolonged hospital stay” (P4: consultant) were discussed as useful improvements to the algorithm.

## User-derived concerns over the Artificial Neural Network COVID-19 Prognostication Application

The principal concern expressed by users was the use of the predictions as an exclusive decision-making tool, for example by making “management setting and treatment escalation decisions based solely on the results” (P5: senior house officer) obtained from the application. “I think a discussion may be required with ICU before deciding on ward-based care, and I’d worry if a high mortality prediction led to an automatic decision to not admit to ICU” (P2: registrar). There were concerns that “generalizability would be difficult” (P1: registrar) given the data is accrued from admissions to a single centre. “Different patients in the UK will have different cohorts and so it should be generalized with caution” (P8: senior house officer).

The model underlying this application was trained on patients in the first high-prevalence period of COVID-19 in the UK. There were no established management guidelines or prognostic scoring system relating to this disease. Several doctors noted the importance of retraining the model with more recent data from COVID-19 patients to reflect recent developments in the management of this condition. “The guidelines are changing, and so the data itself may change” (P14: consultant), therefore “the application may not be calibrated to new waves, given newer treatments” (P29: registrar). The same physician notes that “there is little concern if this is used as part of the big picture but shouldn’t be used in a binary sense” (P29: registrar). This sentiment was echoed by several other physicians who felt that “you have to be responsible and realize no predictive calculator is a substitute for clinical judgement- I don’t think anyone should be under the impression that a calculator can replace their judgement entirely” (P7: registrar).

## Discussion

We tested the clinical utility of a responsive web application/GUI which interfaces to an ANN to predict COVID-19 patient outcomes at the bedside. All clinician-users were able to use the GUI with a mean time to derive a mortality prediction of 59.35 seconds (standard deviation (SD) = 10.35). We found clinician-users gave a mean system usability scale (SUS) score of 91.94 (SD = 8.54), which corresponds with an adjective rating of “Excellent”. Clinician-users found the app intuitive and easy to use, with the majority describing its predictions as a useful adjunct to their clinical practice. The main concerns related to use of the application in isolation as opposed to in conjunction with other clinical parameters. However, most clinicians felt that the application could positively reinforce or validate their clinical judgement. Effectiveness and efficiency measures showed that the application could be used easily with little technical support, or prior explanation with respect to the function of the system. The application is therefore highly productive, whilst maintaining low costs and learnability times. No participant took longer than 2.2 minutes to successfully input all required patient data and retrieve a prediction.

Thematic framework analysis provided further insight into the implications of the application use. Identifying deteriorating COVID-19 patients was a key concern for most physicians. From a clinical perspective, accurate risk stratification underpins hospital admission decisions, as well as appropriate ceilings of patient management. Furthermore, understanding risk allows physicians to better communicate prognoses to the patient and their relatives. For these reasons, a significant majority of participants in this study felt that a scoring system can be a useful adjunct to their clinical workflow and could aid in communicating risk to patients and their families. However, most agreed that the use of a predictive scoring system alone cannot surmount the clinical judgement of a clinician.

The spectrum of opinions about the mortality risk predictions when faced with the same clinical scenario highlighted the variation between clinicians. This emphasises the potential role of an easy to use, widely accessible predictive system in minimising biases such as experiential bias and the availability heuristic in prognostication.

A strength of this work is that both usability assessments and a qualitative framework were used to evaluate the application, thereby developing a deeper insight into all aspects of its use and implications. In addition, multiple researchers analysed the thematic framework data, ensuring consensus with regards to the results and their interpretation.

However, there are limitations to consider in this analysis. The participants in this study had high self-reported levels of expertise in using computers and smartphones. If this application were to be used in settings where user experience of clinical decision tools was limited, this may impact usability, and subsequently affect results and result interpretation. Furthermore, the underlying algorithm is trained on patients from a single hospital site during the first high-prevalence period of SARS-CoV-2 infection in London. The generalizability of its predictions is therefore reduced for other populations, and further work needs to be undertaken to evaluate the application in other health care settings.

Given that treatment for COVID-19 has progressed, for example the recent finding that dexamethasone reduces mortality in hospitalised COVID-19 patients[24], it's important to retrain or update the algorithm with new data to maximise the prognostic accuracy of the application. The adaptive nature of ANNs with their ease of re-trainability, and the continued deposition of clinical big-data for COVID-19 patients, means these latter limitations can be mitigated with future iterations.

The principal challenges in deploying AI technologies in a pandemic include the rapidly shifting clinical needs which the models need to address, and in translating these models to local environments.[4] Whilst numerous studies have now emerged using machine learning processes for aspects of COVID-19 clinical care in multiple different settings,[25-29] few use co-design as we have here to optimise utility for clinicians. Further still, beyond user interface and utility challenges lie ethical and legal issues inherent when smartphone applications are used as healthcare decision support systems.[30] The ethical aspects of integrating computerised decision support systems into management of infectious diseases remain unclear, but there is precidence for the improtance in co-design with clinican-users early on in the pre-implementation phase (as we have done here) to ensure clinicans use them as part, rather than the whole, of their overarching clinical assessement.[31,32]

Based upon our development of the ANN[5], and the clinical utility and feasibility assessment undertaken in this analysis, we propose an adaptive translational pathway for predictive systems in

COVID-19 (**Figure 3**). This workflow recognises the need for feedback mechanisms in the development and deployment of both the GUI and its underlying AI algorithm. As management strategies shift, new data must be incorporated by online learning or retraining of the algorithm to maintain accurate predictions. The new models then require further validation on test datasets to ensure reliability. In tandem, the application must be actively monitored for usability and security issues and updated as appropriate. Utilising interconnected feedback mechanisms in this way can ensure that both the algorithm and the interface to it remain robust to changing trends in patient cohorts and the management of COVID-19.

Following this framework, because of the usability assessment and thematic framework analysis, our current application was modified to include several of the suggested improvements. These included, but were not limited to, the addition of a disclaimer on the index page, and retraining the algorithm to estimate (i) mortality, (ii) probability of admission to an intensive care unit, and (iii) probability of a prolonged hospital stay (defined as a stay of at least one week). These changes are shown in **figure 4**. Future improvements include model retraining from patient samples across multiple hospital sites, and the potential integration of the application to patient electronic health records to facilitate application use in the context of clinicians' workflow, although the barriers to integration into electronic medical records are numerous.[33]

## Conclusion

Developing, validating, and deploying AI technologies in healthcare is associated with a variety of challenges. In this single hospital study, we tested a responsive web application, which leverages an ANN to produce multiple outcome predictions for COVID-19 patients without the need for laboratory parameters. It demonstrates potential utility for patients with an initial presentation of COVID-19 and for those in the community without diagnostic capability. The application is intuitive and requires minimal training for use. Clinicians find the system represents a useful adjunct to their daily clinical practice, and we propose a translational workflow for future predictive systems leveraging similar technologies. We demonstrate that both model and interface adaptation can be used to meet the developing needs of clinicians in the context of a pandemic.

## Declarations

*Consent for publication*

All authors consent to publication of this manuscript in this journal. This work has not been previously published in any other journal.

*Availability of data and materials*

The datasets analysed during the current study and further details on gaining access to the intervention reported within this study are available from the first author (AA ahmed.abdulaal@nhs.net) on reasonable request, as long as this meets local ethics and research governance criteria. The application is available in alpha at [https://desolate-springs-37403.herokuapp.com/](https://desolate-springs-37403.herokuapp.com/)

*Competing interests*

All authors have completed the ICMJE form for uniform disclosure Form for Disclosure of Potential Conflicts of Interest and declare the following: EC has received speaker fees from bioMerieux (2019). NM has received speaker fees from Beyer (2016) and Pfizer (2019) and received educational support from Eumedica (2016) and Baxter (2017). LSPM has consulted for/received honoraria from DNAelectronics (2015-18), Dairy Crest (2017–2018), Profile Pharma (2018-2019), Umovis Lab (2020), bioMerieux (2013-2021), Pfizer (2018-2021), Eumedica (2016–2021), and Shionogi (2021) and received research grants from Leo Pharma (2016), National Institute for Health Research (2013-2020), and CW+ Charity (2018-2021). AA and AP none to declare.

*Funding*

This research did not receive any grant from funding agencies in the public or commercial sectors.

*Author Contributions*

Conceptualization: AA, AP, LSPM. Data curation: AA, AH, AP, EC. Formal Analysis: AA, AP, LSPM. Methodology: AA, AP, LSPM. Validation: AA, AP, LSPM. Visualization: AA, AP. Writing – original draft: AA, AP. Writing – review and editing: AA, AP, AH, EC, NM, LSPM. All authors have read and approved this manuscript and agree as to its contents.

*Acknowledgements*

EC is supported by Economic and Social Science Research Council (ESRC) and the National Institute for Health Research Health Protection Research Unit (HPRU) in Healthcare Associated Infections and Antimicrobial Resistance, UK Department of Health [HPRU-2012-10047] in partnership with Public Health England. LSPM acknowledges support from the National Institute of Health Research (NIHR) Imperial Biomedical Research Centre (BRC) and the National Institute for Health Research HPRU in Healthcare Associated Infection and Antimicrobial Resistance at Imperial College London in partnership with Public Health England. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, or the UK Department of Health.

References

1. Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM, et al. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: Development and validation of the 4C Mortality Score. BMJ. 2020;370:22.

2. Wynants L, Van Calster B, Bonten MMJ, Collins GS, Debray TPA, De Vos M, et al. Prediction models for diagnosis and prognosis of covid-19 infection: Systematic review and critical appraisal. BMJ. 2020;369.

3. Abdulaal A, Patel A, Charani E, Denny S, Alqahtani SA, Davies GW, et al. Comparison of deep learning with regression analysis in creating predictive models for SARS-CoV-2 outcomes. BMC Med Inform Decis Mak [Internet]. 2020;20(1):299.

4. Hu Y, Jacob J, Parker GJM, Hawkes DJ, Hurst JR, Stoyanov D. The challenges of deploying artificial intelligence models in a rapidly evolving pandemic. Nat Mach Intell [Internet]. 2020;2(6):298–300.

5. Abdulaal A, Patel A, Charani E, Denny S, Mughal N, Moore L. Prognostic modeling of COVID-19 using artificial intelligence in the United Kingdom: Model development and validation. J Med Internet Res [Internet]. 2020;22(8):e20259.

6. Shi Y, Yu X, Zhao H, Wang H, Zhao R, Sheng J. Host susceptibility to severe COVID-19 and establishment of a host risk score: Findings of 487 cases outside Wuhan. Crit Care [Internet]. 2020;24(1):108.

7. Caramelo F, Ferreira N, Oliveiros B. Estimation of risk factors for COVID-19 mortality - preliminary results. medRxiv [Internet]. 2020 Feb 25 [cited 2020 May 14];19:2020.02.24.20027268. Available from: https://www.medrxiv.org/content/10.1101/2020.02.24.20027268v1

8. Patel A, Charani E, Ariyanayagam D, Abdulaal A, Denny SJ, Mughal N, et al. New-onset anosmia and ageusia in adult patients diagnosed with SARS-CoV-2 infection. Clin Microbiol Infect. 2020;26(9):1236-1241.

9. Syeda HB, Syed M, Sexton KW, Syed S, Begum S, Syed F, Prior F, Yu Jr F. Role of Machine Learning Techniques to Tackle the COVID-19 Crisis: Systematic Review.JMIR Med Inform 2021;9(1):e23811

10. Pairo-Castineira E, Clohisey S, Klaric L, et al. Genetic mechanisms of critical illness in COVID-19. Nature. 2021;591:92–98.

11. Node.js [Internet]. [cited 2020 Nov 8]. Available from: https://nodejs.org/en/

12. Express - Node.js web application framework [Internet]. [cited 2020 Nov 8]. Available from: https://expressjs.com/

13. Miller K, Capan M, Weldon D, Noaiseh Y, Kowalski R, Kraft R, et al. The design of decisions: Matching clinical decision support recommendations to Nielsen's design heuristics. International Journal of Medical Informatics. 2018;117:19–25.

14. Introduction to Tensors | TensorFlow Core [Internet]. [cited 2020 Nov 8]. Available from: https://www.tensorflow.org/guide/tensor

15. TensorFlow.js | Machine Learning for Javascript Developers [Internet]. [cited 2021 Jan 17]. Available from: https://www.tensorflow.org/js

16. Xiong J, Ziegler C, Adjunct A, Kortum P. SUSapp: A Free Mobile Application That Makes the System Usability Scale (SUS) Easier to Administer. Vol. 15, Journal of Usability Studies. 2020.

17. Britten N. Qualitative Research: Qualitative interviews in medical research. BMJ [Internet]. 1995;311(6999):251.

18. Heath G, Cameron E, Cummins C, Greenfield S, Pattison H, Kelly D, et al. Paediatric "care closer to home": Stake-holder views and barriers to implementation. Heal Place. 2012;18(5):1068–73.

19. Marshall. The key informant technique MN Marshall. Vol. 13, Family Practice © Oxford University Press. 1996.

20. Morse JM. Determining Sample Size. Qualitative Health Research. 2000;10:3–5.

21. Creswell J, Poth C. Qualitative inquiry and research design: Choosing among five approaches [Internet]. 2016 [cited 2020 Nov 8]. Available from: https://books.google.com/books?hl=en&lr=&id=DLbBDQAAQBAJ&oi=fnd&pg=PP1&ots=-hn4adFSOs&sig=BNg_b60DcH7NPLiA8TupPMHjMwk

22. Rapport F, A Doel M, Elwyn G. Snapshots and snippets: General practitioners' reflections on professional space. Heal Place. 2007;13(2):532–44.

23. Bangor A, Kortum P, Miller J. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. Vol. 4, Journal of Usability Studies. 2009.

24. Matthay MA, Thompson BT. Dexamethasone in hospitalised patients with COVID-19: addressing uncertainties. Lancet Respir Med. 2020;8(12):1170-1172.

25. Pan P, Li Y, Xiao Y, Han B, Su L, Su M, Li Y, Zhang S, Jiang D, Chen X, Zhou F, Ma L, Bao P, Xie L. Prognostic Assessment of COVID-19 in the Intensive Care Unit by Machine Learning Methods: Model Development and Validation. J Med Internet Res 2020;22(11):e23128.

26. Domínguez-Olmedo JL, Gragera-Martínez Á, Mata J, Pachón Álvarez V. Machine Learning Applied to Clinical Laboratory Data in Spain for COVID-19 Outcome Prediction: Model Development and Validation. J Med Internet Res 2021;23(4):e26211

27. Quiroz JC, Feng Y, Cheng Z, Rezazadegan D, Chen P, Lin Q, Qian L, Liu X, Berkovsky S, Coiera E, Song L, Qiu X, Liu S, Cai X. Development and Validation of a Machine Learning Approach for Automated Severity Assessment of COVID-19 Based on Clinical and Imaging Data: Retrospective Study. JMIR Med Inform 2021;9(2):e24572

28. Patrício A, Costa RS, Henriques R. Predictability of COVID-19 Hospitalizations, Intensive Care Unit Admissions, and Respiratory Assistance in Portugal: Longitudinal Cohort Study. J Med Internet Res 2021;23(4):e26075

29. Schwab P, DuMont Schütte A, Dietz B, Bauer S. Clinical Predictive Models for COVID-19: Systematic Study. J Med Internet Res 2020;22(10):e21439

30. Charani E, Castro-Sanchez E, Moore LSP, Holmes AH. Do smartphone applications in healthcare require a governance and legal framework? It depends on the application! BMC Medicine. 2014;12:29

31. Peiffer-Smadja N, Poda A, Ouedraogo A, Guiard-Schmid J, Delory T, Le Bel J, Bouvet E, Lariven S, Jeanmougin P, Ahmad R, Lescure F. Paving the Way for the Implementation of a Decision Support System for Antibiotic Prescribing in Primary Care in West Africa: Preimplementation and Co-Design Workshop With Physicians. J Med Internet Res. 2020;22(7):e17940

32. Bezemer T, de Groot MCH, Blasse E, ten Berg MJ, Kappen TH, Bredenoord AL, van Solinge WW, Hoefer IE, Haitjema S. A Human(e) Factor in Clinical Decision Support Systems. J Med Internet Res. 2019;21(3):e11732

33. Jung SY, Hwang H, Lee K, Lee H, Kim E, Kim M, Cho IY. Barriers and Facilitators to Implementation of Medication Decision Support Systems in Electronic Medical Records: Mixed Methods Approach Based on Structural Equation Modelling and Qualitative Analysis. JMIR Med Inform. 2020;8(7):e18758

## Figures

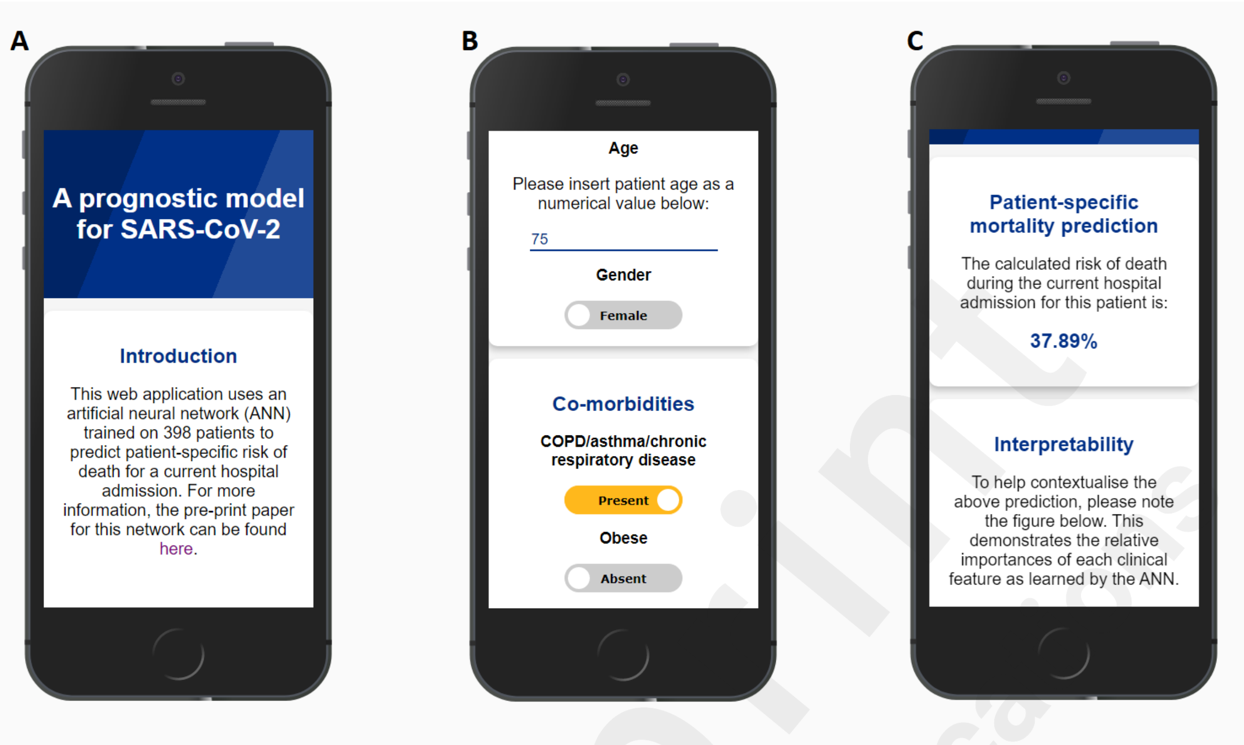



**Figure 1** - Screenshots of the Initial Artificial Neural Network COVID-19 Prognostication Application. **A** shows the introductory screen with a hyperlink to access more data regarding the ANN and its development. **B** demonstrates the data input process with examples of numerical and categorical features. Selected categorical features are coloured and labelled. Numerical features have input instructions above the data collection field. **C** shows a portion of the results screen. Patient mortality data is presented as a human-readable percentage.

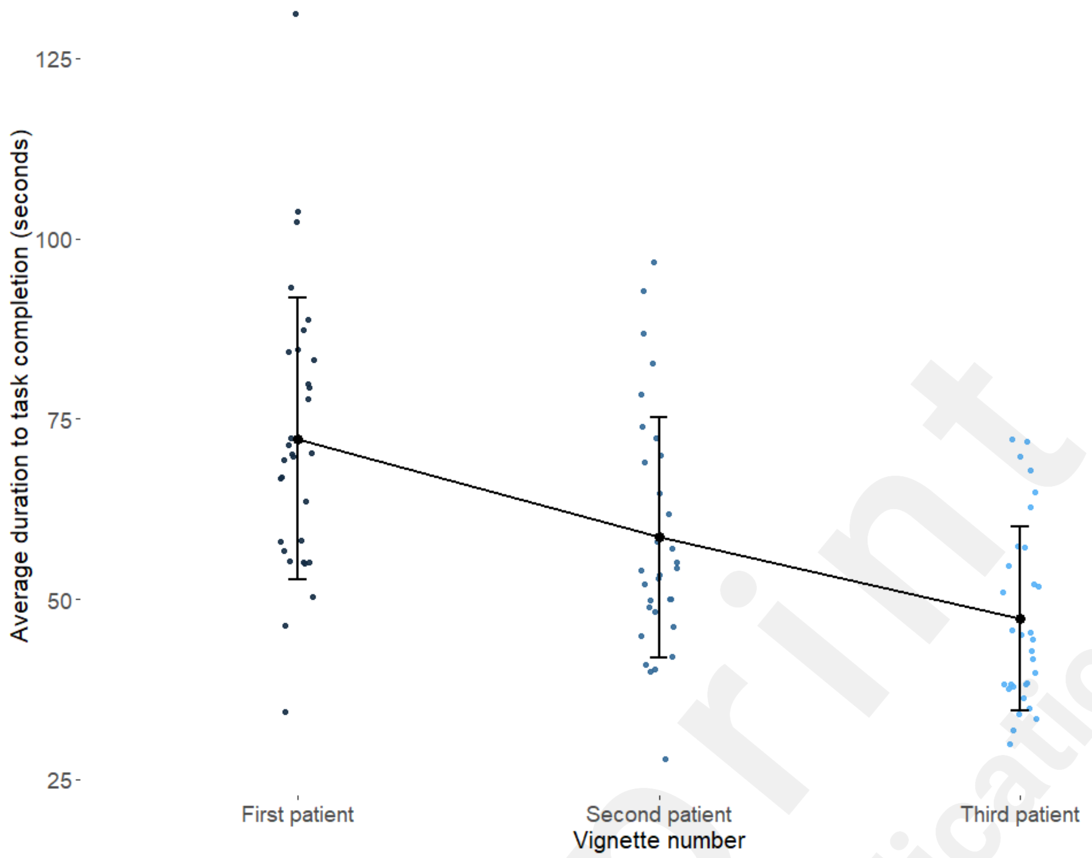



**Figure 2** – Efficiency of clinicians in use of the Artificial Neural Network COVID-19 Prognostication Application.

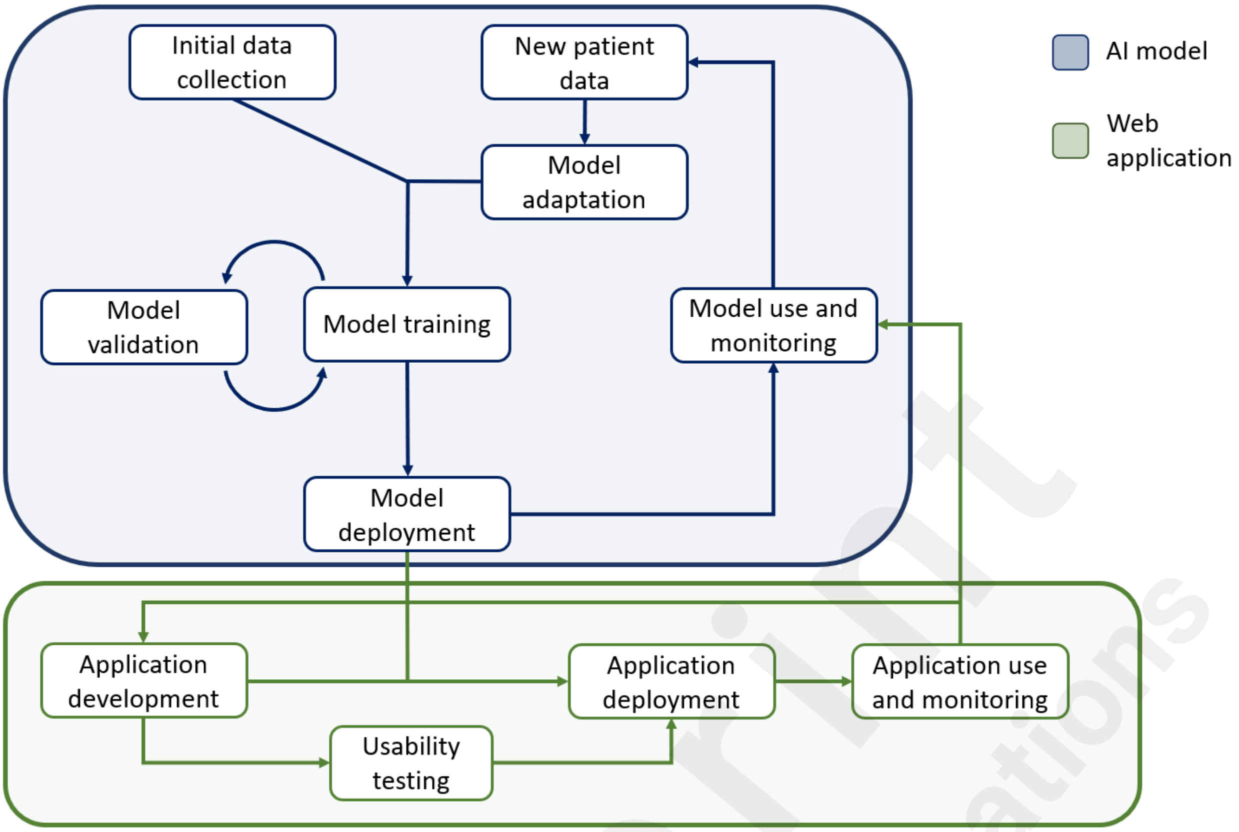



**Figure 3** - Proposed artificial neural network and web application translational workflow, including model training, validation, and adaptation, as well as application development, testing, and deployment.

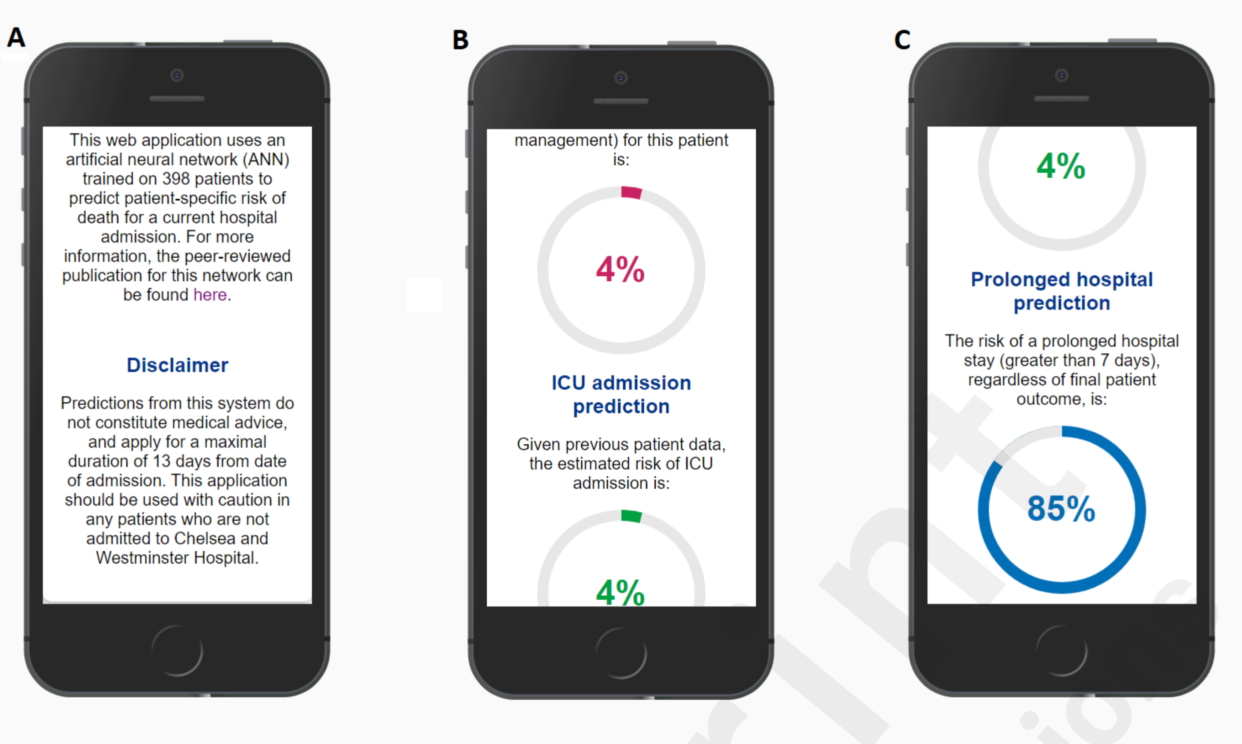
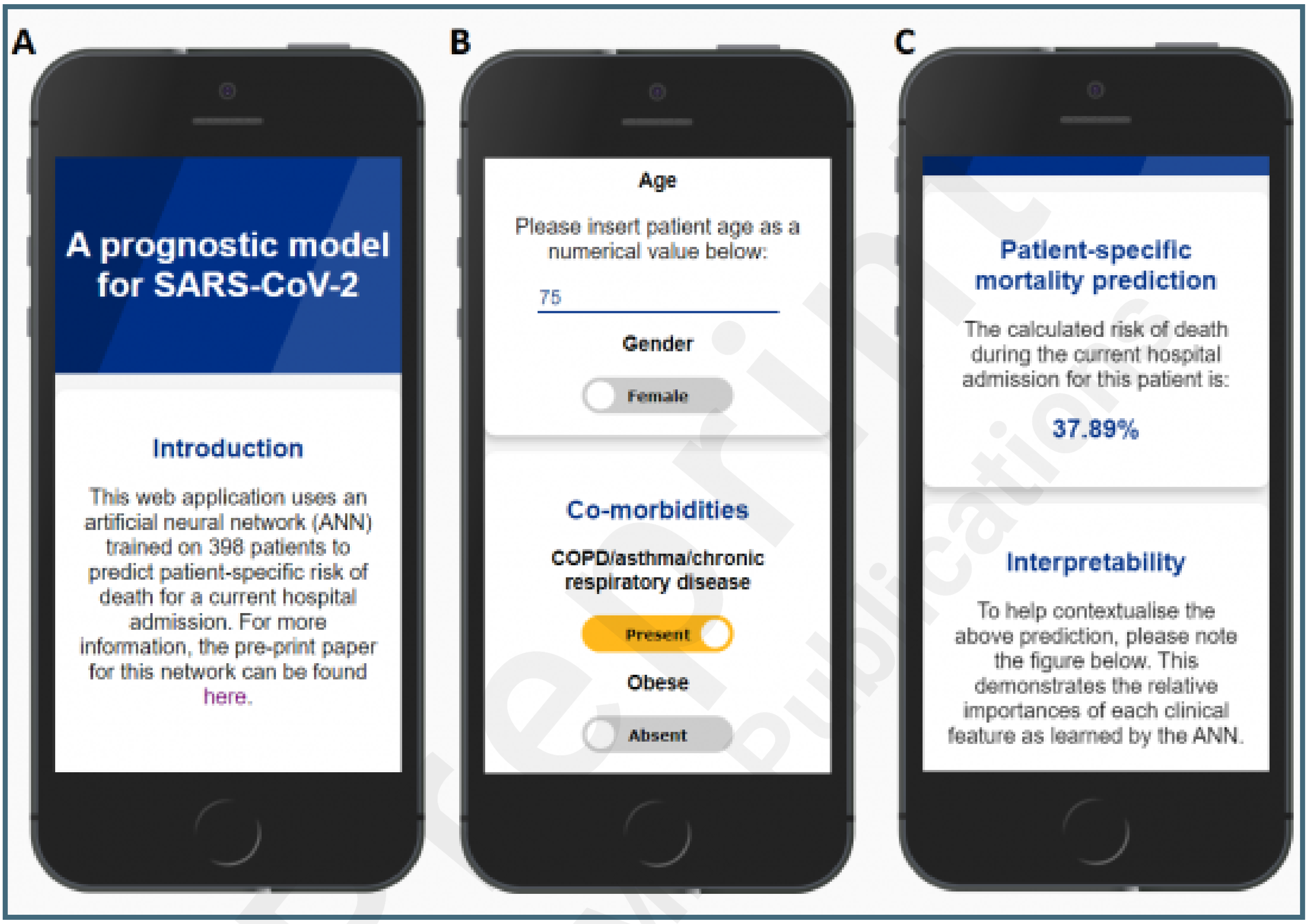



**Figure 4** - Screenshots of the matured Artificial Neural Network COVID-19 Prognostication Application. **A** shows the introductory screen an added disclaimer for use. **B** and **C** show a portion of the results screen. The mortality, intensive care unit and prolonged hospital stay prediction are presented as a human-readable percentages and colour coded, to reflect retraining of the underlying algorithm.

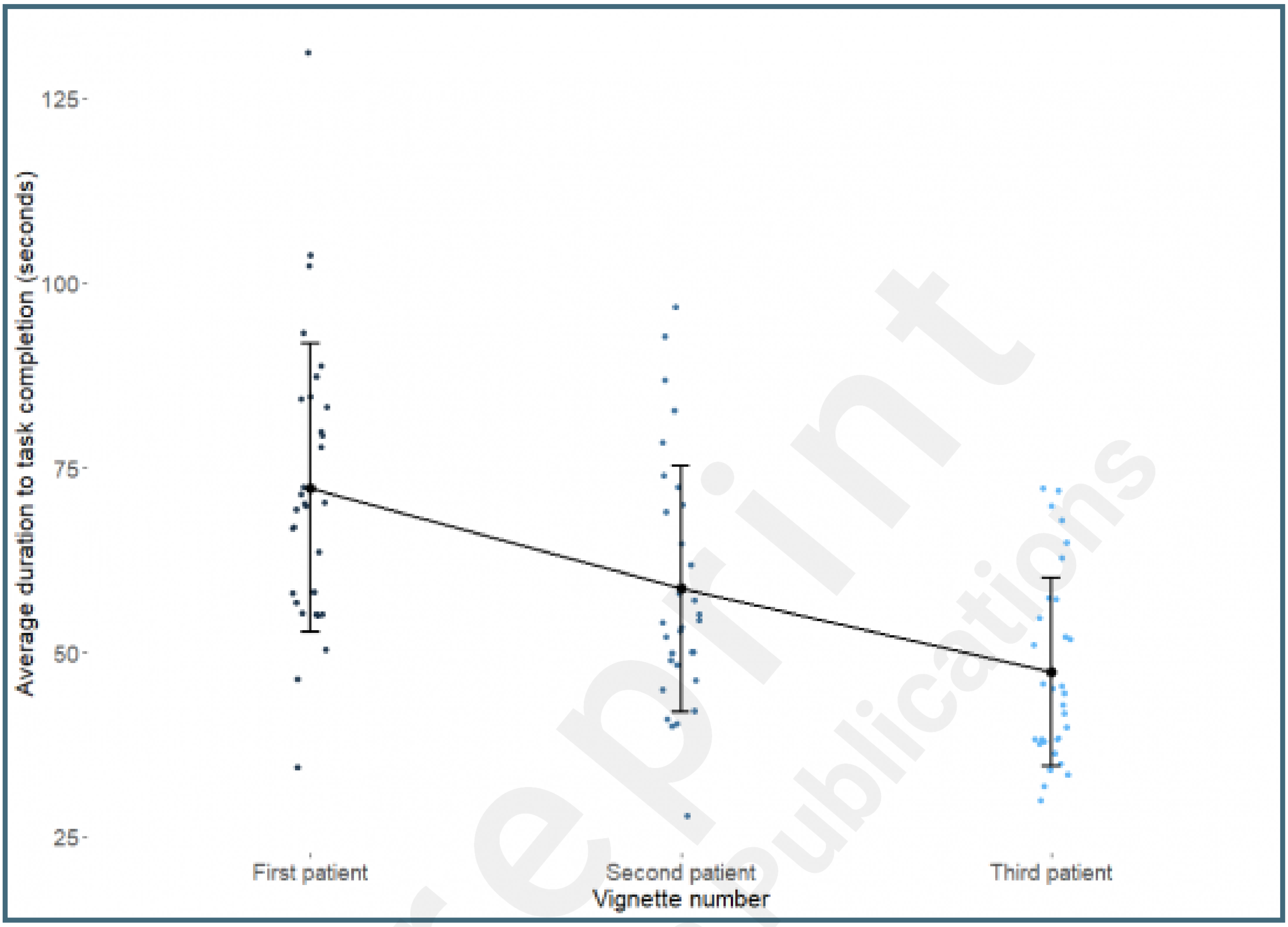
# Supplementary Files

# Figures

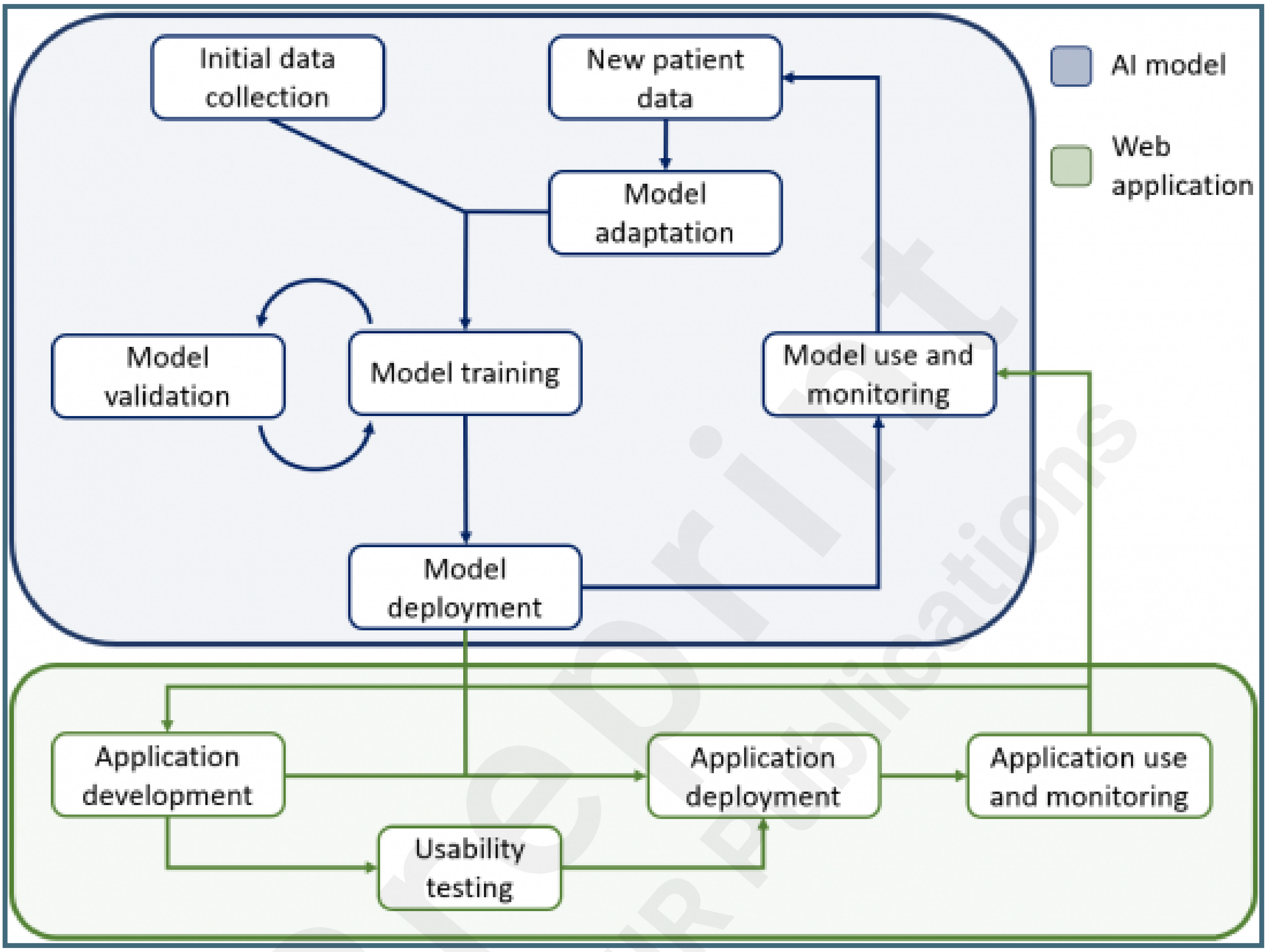
Screenshots of the Initial Artificial Neural Network COVID-19 Prognostication Application. A: shows the introductory screen with a hyperlink to access more data regarding the ANN and its development. B: demonstrates the data input process with examples of numerical and categorical features. Selected categorical features are coloured and labelled. Numerical features have input instructions above the data collection field. C: shows a portion of the results screen. Patient mortality data is presented as a human-readable percentage.



Efficiency of clinicians in use of the Artificial Neural Network COVID-19 Prognostication Application.



Proposed artificial neural network and web application translational workflow, including model training, validation, and adaptation, as well as application development, testing, and deployment.



Screenshots of the matured Artificial Neural Network COVID-19 Prognostication Application. A shows the introductory screen an added disclaimer for use. B and C show a portion of the results screen. The mortality, intensive care unit and prolonged hospital stay prediction are presented as a human-readable percentages and colour coded, to reflect retraining of the underlying algorithm.