Algorithm Information Request
Data & Digital
Interim Health New Zealand
8 July 2022

Algorithm Name	Risk Score for COVID-19 Call Prioritisation V2.4

Version table

Date	Version number	Changes made	Version author
June 15	1.0	Initial update	Rachel Owens, Pieta Brown, Jon Herries
June 29	2.1	Incorporating feedback from the Algorithm Governance Group	Jon Herries
July 8	2.3	Incorporating feedback from the Algorithm Governance Group	Jon Herries, Kevin Ross, Pieta Brown
July 20	2.4	Incorporating feedback from the Algorithm Governance Group	Jon Herries, Kevin Ross, Pieta Brown

Purpose

This is an adapted form of the Te Pokapū Hātepe o Aotearoa, New Zealand Algorithm Hub - Algorithm Information Request provided by Precision Driven Health. It should be read in conjunction with the Privacy Impact Assessment for the COVID Risk Tool 2.0.

This document is intended to provide a written summary of the information that might be of interest to those who will use the COVID Risk Tool 2.0 and those who may be affected by it and a summary of considerations made and discussions undertaken to understand these. It identifies the key risks and benefits of the tool and also identifies areas which could be explored to make this tool better.

Executive Summary

Precision Driven Health, for the COVID-19 Care in the Community team and the Data and Digital Directorate at the Ministry of Health have developed version 2.0 of a population-based risk scoring tool that will be part of the COVID-19 digital platform when a positive case is identified.

The COVID–19 risk of hospitalisation call prioritisation model V2.0 (V2.0 model) is an extension of the COVID–19 risk of hospitalisation call prioritisation model V1 (V1 model) already in operation. Both models predict risk of hospitalisation for people who have tested positive for COVID-19. They have been developed to be used as a risk stratification tool to determine contact priority at local Care Coordination Hubs, at times where a COVID-19 positive person has not used the self service portal and there is little clinical information otherwise available for that person.

The model risk score is calculated using a mathematical formula that takes age, gender as recorded in the NHI, ethnicity, deprivation, vaccination status, number of hospitalisations and number of medications as inputs. Being older, unvaccinated or only having a single vaccination dose, higher deprivation, and having Māori or Pasifika ethnicity, a higher number of hospitalisations over 24 months and a higher number of prescribed medications all increase someone's estimated risk of being hospitalised after infection with COVID-19.

The V2.0 model is better at identifying people at higher risk of hospitalisation overall than the V1 model. The model also had good performance across all population subgroups, with some differences.

With respect to te Tiriti o Waitangi and the principle of active protection, prioritising calls to those identified by the V2.0 model as being most at risk of hospitalisation will achieve greater equity of outcomes for Māori as the model uses predictors (including higher deprivation, and higher hospitalisations) where Māori are over represented.

The model will be deployed and maintained on the publicly available Te Pokapū Hātepe o Aotearoa, New Zealand Algorithm Hub and will be refreshed on a regular basis (initially monthly due to availability of data and effort to refresh) and model iterations will be refined over coming months as the evaluation and data influence model outcomes and drift.

Appropriate use of the model will depend on how the use of the model is communicated to the intended users (the Care Coordination Hubs), it is intended that this is clear about the intended use case.

We consider that there is no material increase in privacy risk in the deployment of the V2.0 model. While the model uses additional information that is already held by the health system, this information is used only in calculating the score, and isn't made

available with the score. Further we consider that the V2.0 model will reduce harm through better targeting of higher risk people for priority contact at times when the available resource for calling and house visits is exceeded by the demand of a number of cases that have not engaged in the online self-assessment process.

1. Model purpose, description and intended use

1.1. What is the rationale for this model?

The COVID–19 risk of hospitalisation call prioritisation model V2.0 (V2.0 model) is an extension of the COVID–19 risk of hospitalisation call prioritisation model V1 (V1 model, public facing documentation for the V1 model can be found here). Both models predict risk of hospitalisation for people who have tested positive for COVID-19. They have been developed to be used as a risk stratification tool to determine contact priority at local Care Coordination Hubs (CCHs), at times where a COVID-19 positive person has not used the self service portal and there is little clinical information otherwise available for that person.

The Risk Score has a particular use case when the available resource for calling and house visits is exceeded by the demand of a number of cases that have not engaged in the online self-assessment process. The risk score will aid the prioritisation of the people for urgency of making initial contact.

The rationale behind the model is that as infections rise, people at most risk of hospitalisation from COVID-19 will be more likely to receive calls in a timely way so that necessary support can be provided to them.

The V2.0 is proposed to replace the V1 model given its significantly improved performance.

1.2. What current practice does this model support or replace?

The model will support the determination of contact priority at local Hubs, replacing the V1 model. Calling people according to the date and time of reported result (essentially, at random) would mean the people who need support the most are likely to wait longer for contact than if they could be prioritised. The V1 model currently supports local hubs to prioritise people. The V2.0 is more performant so should reduce even further the time between reporting a result and receiving a call for those who most need support.

1.3. Who are the intended end users? Please name an example and describe how this adds value to their role.

The intended end users are people who call COVID-19 positive people from local CCHs. Scores from this model will help these users target those who have not used the self service portal, the users will have the risk score that will identify those who are most at risk from deterioration from COVID-19 for

contact.

Further the risk score tool will be published on a public facing website along with the Privacy Impact Assessment and this document.

1.4. What are the inputs and outputs of the model?

Inputs are:

- Age (capped at 59)
- Age difference to 59 (underpinned at 0)
- Biological sex (Male, Not male)
- Ethnicity (Māori, Pasifika, Other)
- Deprivation decile bin (low, medium, high)
- Vaccination doses (0-3)
- Number of hospital admissions in last 24 months, with two-month lag (capped at 15)
- Number meds prescribed July 2021 - January 2022 (capped at 28)

The output is a risk score per person who has tested positive for COVID-19. Call priority can be ranked by the ordering of risk scores from high to low.

1.5. Who developed this model, where and when?

This model was developed by Precision Driven Health (PDH) for the COVID-19 Care in the Community team and the Data and Digital Directorate at the Ministry of Health (MoH) in May 2022. The V2.0 model was developed in a MoH cloud-based Databricks development environment. Stage 1 data preprocessing was undertaken by Deloitte on behalf of Data & Digital.

Additional questions

1.6. Describe, and quantify where possible, the benefits of this model to New Zealand.

It is expected that those who do not use the self-service portal when reporting results are more likely to be at risk, due to either an unwillingness to participate in or an inability to participate in the health response. Access is commonly an issue of equity in the health system and is often a predictor of need, therefore it is important that higher risk people in this group are contacted quickly.

Timely identification of needs will enable appropriate support to be provided to higher risk people (as it is for higher risk people who use the self-service

portal), to improve their outcomes and reduce load on the health system that may arise from unmet needs at a time when resources are under strain.

The model is designed to allow callers to rank those in the list to call by their score, with higher scores being at the top and lower further down. Alternatives to this are random ranking or ranking alphabetically or by the order in which the case comes in.

1.7. Is data pre-processing and post-processing required?

Data needs to be transformed into the form specified in model documentation. This is expected to be straightforward, as the model was developed from the original data as it is collected and stored.

For Application Programming Interface (API) users, data pre-processing depends on how raw data is stored. Data for a person should be as up to date as possible.

Post-processing will be minimal and guided by documentation.

1.8. Will this still be relevant beyond the immediate COVID-19 crisis (or is a short-term solution more appropriate)?

Later model iterations may serve different use-cases (such as a general model for risk of hospitalisation) as the information used in this is commonly used by hospitals around the world for identifying those who are eligible for admission prevention initiatives, however model version 2.0 is only appropriate for call prioritisation for people who have tested positive for COVID-19. In the short term the focus will be on helping hubs access and use the tool effectively.

1.9. What version of the model is proposed to be used?

Model version 2.0

2. Model development

2.1. Describe the methodology used and provide evidence that it was appropriate

A set of data profile visualisations were created to provide insights of feature value distribution and their association with the target.

A logistic regression model was chosen for its transparency and appropriateness for a ranking use case. L2 regularisation was included to mitigate overfitting in model training.

Model development started with a set of base features that were agreed via clinical consultation, which includes age, number of hospitalisations, number of medications, biological sex, GP enrolment, deprivation decile, ethnicity and vaccination status.

In a base model, numeric features were used as is and categorical features were binary encoded. Other features were iteratively added into this base model, and additional feature engineering/modelling techniques were applied in iterations.

For each feature or technique, if it improved the model performance, we kept it in later iterations. If there is only marginal influence on the model performance, we applied a principle of simplicity (e.g., keeping binned deprivation decile instead of the raw decile, dropping the additional vaccination dose instead of keeping it). The main reason for this is to preserve explainability for people who will be affected by the score and those using the score.

Feature engineering/Modelling techniques tested:

- Capping numeric features
- Scale transformation of numeric features
- Normalisation of numeric features
- Deriving new features from existing feature (linear/polynomial transformation)
- Interaction terms
- Sample weight (apply weights inversely proportional to class frequencies)

In the final stage of model development, a drop-off experiment over all the chosen categorical features was done to further explore the possibility of simplifying the model with its cross-validated performance retained.

The model includes the following modules as a pipeline:

- An estimator that scales each numeric feature individually in its given

range on the training set and returns values between zero and one.
- An estimator that transforms each categorical feature into one-hot encoded values (either zero or one).
- An estimator that concatenates all the transformed results in the previous estimators.
- A final estimator which is a logistic regression model with L2 (Ridge) regularisation.

2.2. What data was used to develop and test this model and is this appropriate for the intended use?

Data sources available were:

- The National Minimum Data Set (NMDS), which records hospitalisations for discharged patients. This was an extract from January 2019 (based on discharge dates) to March 2022 (there is also incomplete data for later dates). The NMDS data set was necessary to measure the outcome (is hospitalised), linked to the persons unique NHI (national hospital index number).. This data was also used to produce two-year hospitalisation counts per person. Clinical guidance (the Care in the Community Clinical Leaders) indicated that hospitalisation count would be an appropriate indicator for clinical risk. This information is collected and synthesized from the care of the individual with their consent and then stored.
- New Zealand ePrescription Service (NZePS) data, which contains prescribed medicines per patient at the national level with high coverage of prescriptions (we understand this is around 97%). This was a six-month extract from July 2021 to January 2022. This data set was used to produce counts of medications per person. Clinical guidance (as above) indicated that a polypharmacy count would be an appropriate proxy indicator for clinical risk. This information is collected and synthesized from the care of the individual with their consent and then stored.
- The Operational Data Set which contains national information about vaccinations, COVID-19 infections, and demographic data. A GP enrolment flag was added to this data set based on whether a person's NHI exists in the National Enrolment Service (NES). Age, biological sex, ethnicity, deprivation index and vaccination status were all included in the model. International research and guidelines support age (higher risk for older people), biological sex (higher risk for males) and vaccination status (higher risk for the unvaccinated) being included in the model. Clinical guidance indicated that ethnicity and deprivation index could be proxy indicators for clinical risk. We note this is in the context of comorbidities not being defined in the model.

We have included the deprivation index, which is a common indicator used in the health system to account for a variety of social and economic factors that are known and proven to have an impact on health outcomes. It is a measure that is standardized (deciles) and aggregated to a domicile/neighbourhood level.

The same data sources, updated at appropriate intervals, will be inputs to risk score calculations when the model is in operation, therefore the data the model is built on is the most appropriate for the intended use. The data that has been used has been summarized into high-level features, this minimizes the amount of data used but also the potential accuracy. More complex use of the data will need to be weighed against the risk of using it (e.g. privacy and social license) along with the effort to do this. An example of this may be the use of medicines data. At present the use of a count of medicines a person takes could be improved by understanding which medicines and their doses influenced the likelihood of admission. This is likely to require significant effort, and the need to do this is likely to be limited.

2.3. How are model accuracy and performance measured, and what has been achieved?

Model performance was measured via 5-fold cross validation. The specific use case is for call ranking, meaning AUROC is the most appropriate measure for performance, however classification metrics were also produced. A classification threshold was set at the 90th percentile of the predicted scores over the training set.

Overall model performance is outlined in the table below. Model documentation also describes metrics by demographic groupings (age, gender, ethnicity) and vaccination status. Performance by population subgroup can be found in the Appendix.

	Performance over training set	**Performance over test set**
AUROC	0.790±0.002	0.789±0.008
Accuracy	0.899±0.000	0.898±0.001
Sensitivity (Recall)	0.468±0.004	0.469±0.015
Specificity	0.909±0.000	0.909±0.001

	Performance over training set	Performance over test set
Precision	0.108±0.001	0.108±0.004
Neg Pred Value	0.986±0.000	0.986±0.000
F1 score	0.175±0.001	0.176±0.006
Balanced Accuracy	0.688±0.002	0.689±0.008

2.4. Describe the constraints and limitations of the model

- **Missing variables:** The model has been developed based on available data at the national level. Comorbidities/long term conditions have not been defined and included and polypharmacy and number of hospitalisations are being used as proxy measures of clinical risk. International guidelines and research suggest other risk factors for unfavourable outcomes for COVID-19 patients. Any absence of important predictors will compromise the ability to assess independent effects of variables in the model.
- **Disability:** The development effort has considered 'disability' a number of times in terms of how this might be a factor in the likely admission of a person for COVID. Despite this being identified, there are a number of barriers to the inclusion of this type of information. These include:
    - The fact that disability is an umbrella term. This means there are a number of people who have a disability that would not affect their risk of admission and some who would. Specificity is important as we risk labelling a number of people as vulnerable to admission who might not be and failing to identify those who are.
    - We lack a comprehensive understanding and documentation of disability at an individual level. The information available to the health system about disability is effectively limited to whether people receive a service related to their disability, and in this there is not good documentation about what that disability is in terms of vulnerability to COVID.
- **NMDS lag:** Lags in the NMDS data set affects both the outcome (incomplete data) and also the window of analysis for hospitalisation count which is shifted back two months to account for the expected lag in scoring cases with the model once it is in operation.
- **Discharge data:** NMDS hospitalisation data relates to discharges. Any person who is included in our evaluation data who was discharged after the latest reporting dates for their DHB in the NMDS extract will not be

flagged as hospitalised as their hospitalisation will not be visible (due to data processing lags). A key limitation in this data that could be described is the difficulty separating whether the admission was for COVID or with COVID. The lag in available coded discharge data due to COVID means this is likely to only be known in hindsight if at all (there is a multi-factorial consideration not evident in admission coding).

- **NZePS:** The NZePS data set is only a 6 month extract. A 12 month extract would be preferable to capture seasonal effects in prescribing (noting that the seasonal effect would have been more muted in the past two years while influenza was suppressed due to border closure). NZePS data does not have full population coverage (estimated to be 97%). It is possible that those not covered by this data set skew to the more vulnerable. Should that be the case, those people will have their risk underestimated by the model as their number of medications will be set to zero. The count of medications derived from NZePS is a proxy for complexity and does not reflect the reason for prescribing of these medications. There is the potential to refine the logic applied to the NZePS data in future model iterations.
- **Cohort definition:** The inclusion dates for the cohort (Omicron cases) cover only a 47 day period. Outcomes are likely to be affected by the total volume of COVID-19 cases and strain on the system at a point in time.
- **Local risk scores:** A range of local and national risk scores are being used by local CCHs. The use of these scores for prioritisation is expected to have some positive impact on people's outcomes. This means it is difficult to separate how good the score is at predicting risk vs. the impact of using the score for prioritisation of clinical assessment and intervention which lowers risk of hospitalisation.
- **Risk scores with more inputs:** Some regions, such as the Northern Region, are using locally developed risk scores that take more information about a person's health into account and use of the score for prioritisation may be expected to have a larger positive impact.
- **Missing cases:** We expect that a large number of true positive COVID cases will be missing where people have not tested or not reported RAT results. Where these people were hospitalised, we understand that a positive test result would be backfilled in their record. We are therefore likely to be missing a large cohort of people who tested positive and were not hospitalised. This potentially skews the cohort towards higher risk people.
- **Access to care:** Testing and hospitalisation reflect access to healthcare and don't provide a complete picture of need for healthcare. This is evident in the difference between the enrolled population, the census and the Health Service User (HSU) datasets used in the pandemic.
- **Deaths:** Deaths in the community were not included due to data availability.

- **Outcome:** There are no specialty exclusions for who is hospitalised. Nor is there a distinction between being hospitalised with or of COVID-19. This may mean hospitalisation rates in lower risk categories are higher than expected for groups that have usually higher overall hospitalisation rates than the prevalence.
- **Suitability of data used for modelling:** while the model will help to adjust for certain variables, the input data may not be representative of COVID-19 patients in the future, therefore the model may not generalise well for future cases, especially where the dominant variant changes, or therapeutics are in wider use.
- **The count of medications** is a simple proxy for complexity and can be refined to better reflect long-term conditions.
- **NZDep (Deprivation Index):** The NZ Deprivation index reflects socio-economic deprivation at the level of an area, not a person, and will be imperfect as a proxy for individual health needs. It is derived from variables collected in the NZ Census.

Additional questions

2.5. What performance should be expected for the intended use of this model in New Zealand?

This model has been trained on recent New Zealand data (as described) and we would expect similar performance to what we found in validation, assuming no changes to outcomes relating factors such as a change in dominant COVID-19 variant or the use of therapeutics. It is intended to monitor and review the model to understand how and if the model performance changes.

3. Model deployment and management

3.1. How do you anticipate the model being deployed and used?

We expect the primary use of this model to be through the national Covid Care in the Community programme and the local CCHs. The population will have risk scores calculated using an API to Te Pokapū Hātepe o Aotearoa, New Zealand Algorithm Hub. The risk scores will be stored in Ministry COVID response systems along with case records and be made available to local Care in the Community Hubs, to be used for call prioritisation.

A web form will also be available on Te Pokapū Hātepe o Aotearoa, New Zealand Algorithm Hub for the purposes of trying out the model and updating scores for individuals.

3.2. What is the proposed approach to monitoring, maintaining and updating the model?

Monitoring: Performance of the V2.0 model will be evaluated once sufficient data is available for evaluation.
Maintaining: The model will be maintained by the Te Pokapū Hātepe o Aotearoa, New Zealand Algorithm Hub on behalf of the Ministry of Health.
Updating: The model is part of an ongoing process of iteration to improve risk score with more data and new data sources. Future model iterations may be trained in coming months.

3.3. Who is responsible for model performance, and making updates?

MoH is responsible for model performance and has contracted PDH to support this work.

3.4. How should/will use of the model be monitored?

As noted, the model will be evaluated. It is expected that this should occur within three months of implementation (a number of cases need to be admitted and then discharged in order to evaluate the outcome), but monitoring could begin sooner than this. Use of the model will be logged via API calls to Te Pokapū Hātepe o Aotearoa, New Zealand Algorithm Hub. Feedback from those who use the outputs of the model will require engagement with the appropriate CCHs. Feedback from website users can also be submitted from the website via the ‘contact us’ page.

Additional questions

3.5. How will user feedback be incorporated?

The Care in the Community programme is already engaging with CCHs to receive feedback on the model, its use and effectiveness. There is interest from a number of regions in how this has been developed and the increased quality of the model.

User feedback can also be submitted through the website for the Te Pokapū Hātepe o Aotearoa, New Zealand Algorithm Hub Team to evaluate. Feedback will be included in the model where appropriate.

3.6. How will the results, including accuracy, be explained to users?

There is specific information on model performance that will be shared with interest groups (e.g. hub clinical leads) for dissemination to end users, with a practical focus on how the model increases the number of hospitalised people contacted per X number of calls. This is a part of existing change management and education in the COVID Care in the Community response.

3.7. What training and user documentation will be provided?

Training for the use of the rankings produced by the model will be provided to CCHs by the Covid Care in the Community programme. It will be an adaption of the content and approach provided for V1.

No specific training is required for use of the model. The model will rank people for call prioritisation and those rankings will be provided to local CCHs.

3.8. Are specific user groups to be provided with supplemental support?

We do not believe that there are groups which will need extra support. Users who need extra support are likely to be non-medical users who are interested in the model and explore the model on Te Pokapū Hātepe o Aotearoa, New Zealand Algorithm Hub website where use of the model will be documented. We advise that the model is primarily used by healthcare related users.

4. Māori impact

4.1. What are the relevant Māori considerations for the development and use of this model, and how have/are these being addressed? Consider specifically how this work can uphold the principles of Te Tiriti o Waitangi with reference to *participation*, *protection* and *partnership*.

Māori are known to have poorer health outcomes, and therefore will have a disproportionate need for support after infection with COVID-19. This expectation is supported by the model itself where being Māori is a risk factor for hospitalisation after infection.

The use case (which is the same for the V1 model which is currently operational) is to prioritise calls to people infected with COVID-19 who have not used the self-service portal. It is likely that Māori are overrepresented in the group of people who do not use the self-service portal. Given Māori are also more likely to be at risk of hospitalisation after infection with COVID-19, in terms of the *protection* principle, prioritising calls to those most at risk should be of benefit to Māori.

4.2. Describe how appropriate decision making and community engagement has taken place.

There is an active engagement underway with data Iwi leaders group to ensure that Māori interests are furthered by this work. There is also engagement with the Māori Health Directorate in the Ministry of Health about this work.

Specific Maori community engagement has not taken place as part of the development of this algorithm, although there is Māori membership on the Algorithm Governance Group.

5. Equity

5.1. Does the model explore, or is it able to detect, differences in outcome by population subgroup e.g. ethnicity, gender as described above, age? Please include a specific Māori lens in your response.

Population subgroups were explored in the development of this model. Being older, Māori or Pasifika, in a higher deprivation subgroup, and/or not being male increases the risk score.

As noted, Māori are known to have poorer health outcomes. If a person who identifies as Māori has more hospitalisations and is prescribed more medications (whether or not these medications are dispensed), they will also be assigned higher risk by the model. Further, Māori are more likely to experience higher deprivation, which also increases the risk score.

The Māori population skews younger than the population overall, which will have the opposite effect, slightly reducing overall risk in that population.

Male gender is associated with lower risk, given the other factors already included in the model. While international literature suggests males may be at increased risk of poor outcomes from COVID-19, caution should be applied when interpreting a coefficient like this in isolation. Gender is likely to be highly correlated with other factors such as hospitalistions and medications. It is included in V2 due to having a statistically significant effect and improving model performance.

5.2. Has the algorithm been tested for differential accuracy or validity by population subgroups e.g. ethnicity? (Comment with respect to factors such as goodness of fit, performance metrics, treatment of missing data). Please include a specific Māori lens in your response.

Performance of the model is good for all population subgroups analysed, although slightly lower for Māori compared to other ethnicities, not male biological sex compared to male biological sex and is slightly higher for the over 60s compared to other age groups. (The male/not male definition was based on international research indicating that being male was a risk factor for those infected with COVID-19.)

Due to a lack of data, the ability to analyse impacts on disabled people, homeless people and refugees have not been assessed.

Performance metrics by population subgroup are in the Appendix.

The outcome of the model does not discriminate between hospitalisation of and with COVID-19 and it is possible that small differences in performance for subgroups relates to this. For instance, Māori have a higher rate of hospitalisation aside from COVID-19 risk which may not be sufficiently predicted by the inputs to this model. Further, this is a possible reason why being biologically male decreases rather than increases risk.

There is likely to be different reporting of COVID-19 depending on population subgroup. For instance, some subgroups may test and report positive tests at a higher rate than other subgroups. This will skew the certain subgroups towards higher risk people (as well as the cohort overall), and will likely affect performance of the model for those subgroups, however this is speculative and difficult to quantify.

5.3. Is there a potential for disproportionate benefit or disproportionate harm to one group or another in applying or interpreting the results? How do you propose to mitigate this? Please include a specific Māori lens in your response.

It is unlikely that there will be disproportionate benefit or harm to any group in applying this model. The model allows for differences in risk across population subgroups, and is performant for all groups, improving on the V1 model. There is no human intervention that would bias the rankings produced by the risk score. Further, the use case is specifically intended for quickly reaching more vulnerable people in the population that is infected by COVID-19 - those who do not use the self-service portal, who have clinical indicators of risk and who belong to more vulnerable population subgroups. As noted, for Māori this means they are more likely to be scored more highly and be higher up the call ranking unless young and healthy.

One potential source of harm would be in the case that those not covered by NZePS data set (estimated 3% of the population) skew to the more vulnerable. This includes those who can’t afford or do not collect prescriptions. Should that be the case, those people would have their risk underestimated by the model as their number of medications would be set to zero.

Disability is another factor that isn’t available or model development or operations, and it is possible that having a disability would affect both the risk and vulnerability of a COVID-19 patient. For this reason, we note the exclusion as a factor in training materials, and expect that care providers develop an appropriate response or disabled people independent of their risk score.

6. Legal and risk

6.1. What are the relevant IP rights in the algorithm and who owns them? Please provide any appropriate evidence (e.g. contracts)

This model is joint property of MoH and PDH, and is available for use without restriction in New Zealand.
The model development and evaluation IP is with the developer, PDH, however we note the development and training involve the use of open-source packages.

6.2. Was the model/algorithm trained using any personal information, and was appropriate consent granted?

The model was trained using personal health data in a MoH environment with the consent of MoH, granted through the data governance process. The Ministry of Health Data Governance Group has confirmed that both Rules 10(1)(d)(i) and (ii) and 10(1)(b) provide legitimate grounds for use. Further information is provided in the Privacy Impact Assessment.

6.3. Will the algorithm collect and process any personal information (i.e. information that could identify an individual)?

The algorithm will not collect personally identifiable information, this information already exists. It will and needs to in order to create individual scores, process personally identifiable information. In operation, model inputs provided to the Te Pokapū Hātepe o Aotearoa, New Zealand Algorithm Hub cannot be linked to any individuals.

In model development, NHI was used to link datasets in order for the data to be used in the model (all of this was conducted within a MoH environment).

This is discussed in detail in the Privacy Impact Assessment.

6.4. Is there evidence of social license for this algorithm and the proposed use?

This COVID-19 risk model has been developed and deployed in the context of the COVID-19 pandemic which is a serious threat to public health (Rule 10 (1) (d) of the HIPC). Social license work for risk of hospitalisation models for COVID-19 is ongoing internationally. While some social/cultural licence work has been undertaken in New Zealand around the use of data, on-going work is required, The work done to date suggests New Zealanders have a variety of views around the purpose and value of data sharing and their trust in how their data is used and by whom.

As with the V1 model, the V2 model will be transparent: it is explainable, will be available to inspect on Te Pokapū Hātepe o Aotearoa, New Zealand Algorithm Hub as a public facing web form with documentation, and will be shared with interested groups. There will also be a public-facing webpage from the Ministry of Health on this algorithm. None of these is evidence of social license, however these steps are demonstrative of transparency which is one way to support the growth or maintenance of social license.

As noted, there are a number of New Zealand studies which consider, in a variety of contexts, New Zealanders' views on the purpose and value of data sharing and their trust in how their data is used and by whom.

There are a variety of findings in these studies, some of which support the propositions that: (1) there is a social license for the proposed data use for the Covid scoring/triaging algorithm and; (2) people would expect the Health System to use the data in that way. However, there are also findings which suggest that not all New Zealanders would support the proposed data use because they consider individual consent should be given before sensitive health information, such as Covid data, is used. Some of the findings also suggest that some people identifying as Māori are uncomfortable with their health information being shared with other health professionals involved in their care in other organisations.

Due to a lack of direct evidence of social license for this specific algorithm, the next step for Data & Digital is to complete a review of the current state of social license for algorithms in healthcare and then develop a plan to maintain and grow social license for the use of algorithms (including this one) in healthcare. This may include actively engaging with people, communities and groups using different ways of ascertaining the social/cultural licence (eg, hui, citizens' juries, surveys, consumer forums) to ensure compliance with the Algorithm Charter.

Additional questions

6.5. What permissions, if any, are in place to permit publication and use by others of the algorithm? Please provide any appropriate evidence (e.g.

licence terms)

NA

6.6. If the algorithm contributor does not own the IP, do we have appropriate rights to use the algorithm?

NA

6.7. Does the algorithm utilise open source software or code? If so, is that software subject to permissive or restrictive licence terms? Please provide details of relevant licence terms.

The model was developed in Python 3.8 and runs in a Python 3.7 environment. All packages imported to develop and run the model are open source python packages:
numpy==1.21.6
PyYAML==6.0
Cerberus==1.3.4
mlflow==1.26.1
cloudpickle==1.6.0
pandas==1.2.4
psutil==5.8.0
scikit-learn==0.24.1
typing-extensions==3.7.4.3
pickle5==0.0.12

Licence:
Python - license agreement for 3.7 and 3.8
NumPy - BSD 3-Clause "New" or "Revised" License
PyYAML - MIT license
Cerberus - ISC license
Mlflow - Apache License 2.0
Cloudpickle - license
Pandas - BSD 3-Clause "New" or "Revised" License
Psutil - BSD 3-Clause "New" or "Revised" License
Scikit-learn - BSD 3-Clause "New" or "Revised" License
Typing-extensions - license
Pickle5 - license

6.8. Are there any privacy risks related to use of this algorithm?

The privacy risks associated with the model are discussed in more detail in the Privacy Impact Assessment.

In summary, Personal Health Information was used to create the model, and is used to create scores for individuals. Information that is used to calculate the score is then deleted. This score is stored in systems controlled by the Ministry of Health and is used in the delivery of health services by the community hubs.

7. Ethical considerations

7.1. Please discuss ethical considerations (including the potential for harm) related to the development, use or mis-use of this algorithm. Please include reference to the potential for harm by using this algorithm vs. the status quo.

There are a number of considerations to be made in using this model.

The first and key consideration in this that the algorithm is to be used only as a risk stratification tool to determine contact priority at local Care Coordination Hubs, at times where a COVID-19 positive person has not used the self-service portal and there is little clinical information otherwise available about that person which enables them to receive appropriate support and treatment for COVID-19. In other words, the use of their information in the algorithm to support their treatment is directly related to the purpose for which the information was obtained (Rule 10 (1)(b)) and it is also necessary to prevent or lessen a serious threat to their life or health (or that of another individual); and public health or public safety (Rule 10 (1)(d)).

The Ministry of Health believes these reasons (for using individual information) provide reasonable grounds for using that information in the absence of individual authorization and it also considers a person would reasonably expect the health system to use information it holds (eg. medical records) to deliver the right care, at the right place, at the right time and therefore this is where consideration of the ethics for this tool begin.

The first and key consideration in this that the subject is seeking diagnosis and treatment for COVID-19. While this couldn't be considered informed consent for the model, it does suggest a person would reasonably expect the health system to use information it holds (eg. Medical records) to deliver the right care, at the right place, at the right time and therefore this is where consideration of the ethics for this tool begin.

Harms and other options have been framed in the broad context of the potential for harm relating to the use of this algorithm and alternatives to and/or reasons for using the model.

These are the potential harms we have identified:

- In some individual cases, a vulnerable person by chance may end up further down the calling list than they would have been under a random calling scenario, especially if they have missing data (e.g. not represented in NZePS) or their vulnerability is not otherwise picked up in the model inputs (eg. missing data relating to disability). This likelihood of this risk is described in the technical documentation.

- There is a limitation in the model regarding disability. This was described in more detail earlier in the document, and relates to the difficulty defining disability as it relates to the admission risk of individuals and the lack of data held by the health system able to be used in the model relating to this term.
- There are some privacy risks associated with this tool being in use, these are described in the Privacy Impact Assessment and are mitigated to a large extent. The Ministry of Health Data Governance Group has determined that using data for this purpose is Rules 10(1)(d)(i) and (ii) and 10(1)(b). The second purpose (that the use is directly related) relies on two precedents. The first is the use of this same information in assessing cardiovascular risk for the population in the national CVD risk assessment programme. Key inputs and key output for that are medicines, demographic details (age, gender as recorded in the NHI, ethnicity) and admissions. The second is that the information used in the model is likely information a clinician would ask in an assessment of admission risk. This model simply uses the information that would be asked for again in an analogue process undertaken by a person asked to identify those who are at risk of poor outcomes from a COVID infection. By using a test of what is reasonable, in general the expectation from the population is likely to be that re-using information is preferable to re-collecting the information (note: this is the entire premise of the Hira programme and Digital Health Strategic Framework).
- Should there be insufficient communication to and oversight of end users, the model may be inappropriately used (outside of the use case) in a way that compromises best care practice. For instance, model scores may be relied upon in situations where clinical judgement should be used instead. This is mitigated through the change management and education provided to hubs for this work.
- There is a social license risk in using this algorithm where the public may not accept the use of their information to predict those at risk of admission due to the linking of data required to do this or a perceived removal of personal choice or autonomy.
- The tool is being used before understanding the outcomes of the use of the tool. It has been rigorously assessed and independently reviewed as being performant across the population and its subgroups. However, given it is a predictive tool certainty about the utility requires the use of the tool to measure the effect of it (beyond a test set of data).

Potential for harm relating to not using the algorithm:

- Compared to random calling (e.g. calling people by order of the reported test result datetime) the V1 evaluation demonstrates that the model is better at identifying higher risk people to call first (and V2 significantly improves on the performance of V1 on the model development data - evaluation data post go-live is currently pending). If the model is not

used and people are instead called randomly, more vulnerable people will wait longer, on average, to be called and may not be called at all in times of limited calling capacity (e.g., at times of infection surge).

- An alternative approach could be considered where a clinician was required to assess all of the people with the information in the model, this would not be practicable (we don't have enough clinicians) and would likely result in significant bias due to the variability in assessment and clinical decision making. This has been documented in the Wai2575 claim and this claim demonstrated bias and racism inherent in clinical service delivery of the health system. It is therefore possible that those who have higher risk scores are more likely to be discriminated against if this score isn't used.
- There is a social license risk in not using this algorithm where the public may expect us to use tools we have available to us (as discussed in question 6.4) and we failed to use it for the benefit of the community.
- A further alternative that could be considered is what is currently happening in some provider hubs at present where calculations of risk are being undertaken. In our conversations with a sample of the Hubs, these haven't been evaluated for quality or accuracy, or been peer-reviewed and it is likely that some do not have rigorous governance and evaluation of their performance. A failure to use V2 of the model would then result in this becoming the prevalent approach. The consequences of this approach could compound the impact - such as biases or poorer risk prediction for certain groups - of models that have not been rigorously developed and evaluated.

7.2. Describe the potential unintended consequences you have identified that could result from use of this model. How will these be mitigated?

- As noted, should there be insufficient communication to and oversight of end users, the model may be inappropriately used (outside of the use case) in a way that compromises best care practice. For instance, model scores may be relied upon in situations where clinical judgement should be used. Clear communication and change management around the release of the model will included when the model is disseminated, including on the public-facing website.
- Risk of hospitalisation may not tell the full story of support needs. The model has been developed to prioritise need for a clinical assessment. This does not reflect the broader picture of individual needs if someone has COVID-19 (e.g. social, childcare, food or other needs) and if the model was over-relied upon for COVID-19 manaaki care it may give undue importance to clinical risk only. Therefore, there may be those at lower risk of hospitalisation that may have higher support needs, e.g.

single parents or those with mental health issues, who will not be prioritised in this model.

Additional questions

7.3. Where applicable, what consents have you obtained from individuals whose information might be used for model training and development? What form did those take? Please provide details.

The development of this algorithm has depended on the use of public health information (PHI), linked across multiple data sets. Individual permission was not sought for this work, and the license to do the work was via MoH and associated governance, in the context of a serious threat to public health (Rule 10 HIPC).

Further discussions with the MoH data governance group have identified that the use of this data is also consistent with directly related purpose. This is demonstrated by the use of information that a clinician would alternatively collect directly in the course of an assessment and that this use case (using hospital admission and medicines data to predict outcomes) is not unique and has precedence in a number of hospitals locally and internationally and also nationally through services such as the virtual Diabetes register and the Cardiovascular risk assessment tool. This could be revisited once COVID is no longer a serious threat to public health (Rule 10 HIPC). .

This is discussed further in the Privacy Impact Assessment.

Appendix

Model performance by population subgroups

By age groups

Risk ranking by the model is good for all age groups, and highest for the 60+ age group. As expected, the prevalence is highest for over 60s. The difference in prevalence between the younger cohorts is small. The uptick in hospitalisation post 60 is captured in the model by splitting the age variable into two so that two slopes may be accommodated.

Model performance by age group

	Overall	**18-39**	**40-59**	**60+**
Sample Size	239,604	152,483	67,512	19,609
Prevalence	2.31%	1.87%	2.06%	6.55%
AUROC	0.790	0.769	0.768	0.796
Accuracy	0.899	0.929	0.905	0.637
Sensitivity (Recall)	0.468	0.339	0.423	0.801
Specificity	0.909	0.940	0.915	0.625
Precision	0.108	0.098	0.095	0.130
Neg Pred Value	0.986	0.987	0.987	0.978
F1 score	0.175	0.152	0.155	0.224
Balanced Accuracy	0.688	0.640	0.669	0.713

By ethnicity

Risk ranking by the model is good for all defined ethnicity groupings, but lowest for Māori. Model performance was similarly lowest for Māori in the V1 model. Prevalence for Māori and Pasifika is 57% and 63% higher than the "Other" ethnicity grouping. This may be due to a combination of both increased clinical risk and higher prevalence of hospitalisation regardless of COVID-19 status, noting that the outcome does not distinguish between hospitalisation with and of COVID-19.

Model performance by ethnicity

	Overall	**Other**	**Māori**	**Pasifika**
Sample Size	239,604	144,095	42,428	53,081
Prevalence	2.31%	1.86%	2.92%	3.03%
AUROC	0.790	0.784	0.776	0.790
Accuracy	0.899	0.927	0.856	0.857
Sensitivity (Recall)	0.468	0.400	0.529	0.535
Specificity	0.909	0.937	0.866	0.867
Precision	0.108	0.107	0.106	0.111
Neg Pred Value	0.986	0.988	0.984	0.984
F1 score	0.175	0.168	0.176	0.184
Balanced Accuracy	0.688	0.668	0.697	0.701

By vaccination status

Risk ranking by the model is good for all vaccination groups, and lowest for those not fully vaccinated. As expected, prevalence is highest in the group that is not fully vaccinated and lowest in the boosted group.

Model performance by vaccination status

	Overall	Not fully vaccinated	Fully vaccinated	Boosted
Sample Size	239,604	11,853	122,559	105,192
Prevalence	2.31%	6.55%	2.31%	1.82%
AUROC	0.79	0.772	0.781	0.779
Accuracy	0.899	0.619	0.909	0.917
Sensitivity (Recall)	0.468	0.794	0.419	0.409
Specificity	0.909	0.607	0.921	0.927
Precision	0.108	0.124	0.112	0.094
Neg Pred Value	0.986	0.977	0.985	0.988
F1 score	0.175	0.214	0.176	0.152
Balanced Accuracy	0.688	0.700	0.670	0.668

By gender (as recorded in the NHI)

Risk ranking by the model is very good for Males, and good for other gender categories. Hospitalisation prevalence is a lot lower for Males, as reflected in the coefficients. We can only guess that it is possible that this difference relates to lack of distinction between hospitalisation with and of COVID-19 in the outcome - for instance, all women admitted to hospital for maternity services are captured in the outcome. Lower model performance for “Not male”, may be likewise due to lower risk people being represented in the outcome. Specialty exclusions in the outcome may help to mitigate this.

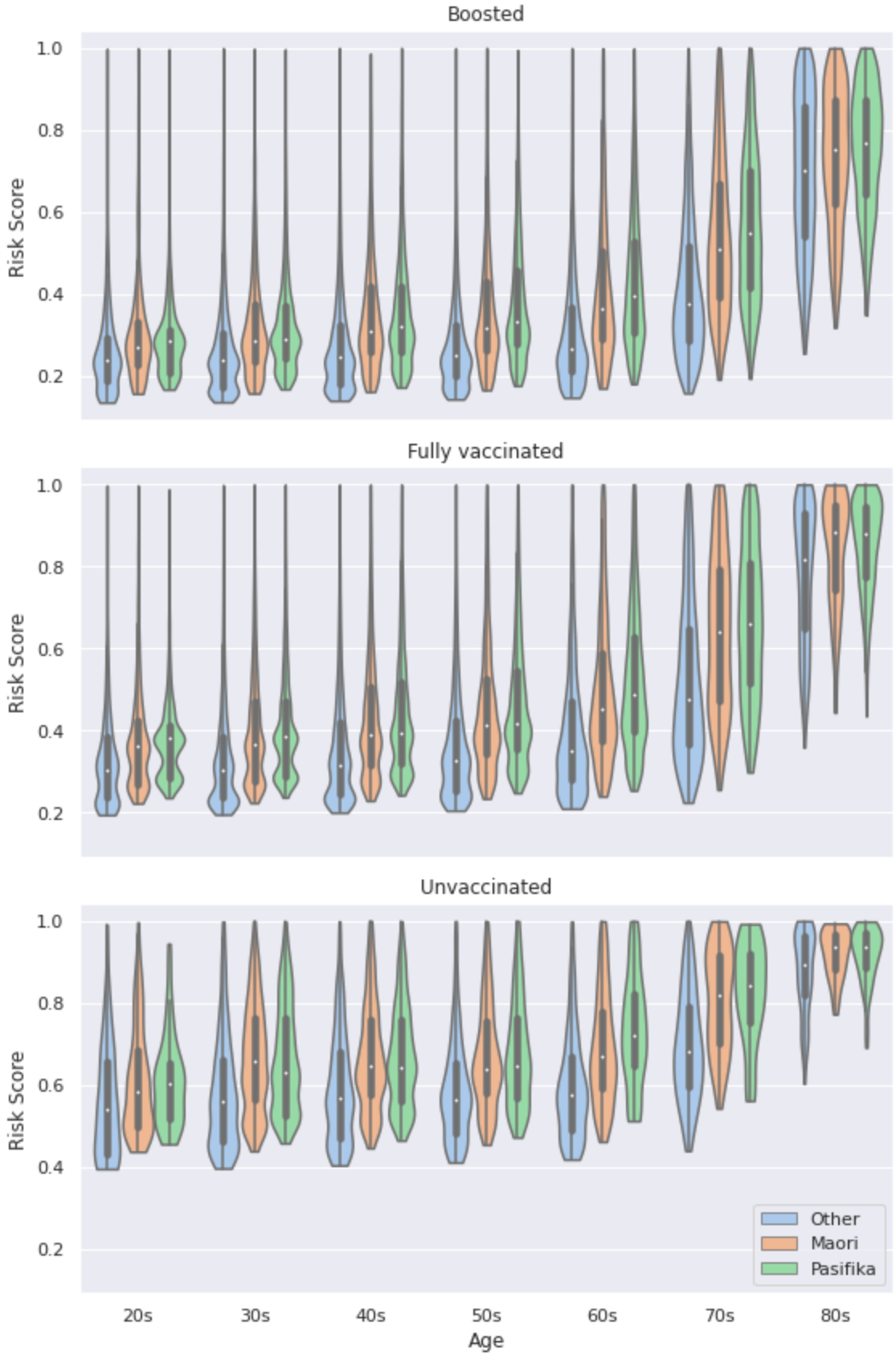
Model performance by gender

	Overall	Not male	Male
Sample Size	239,604	127,448	112,156
Prevalence	2.31%	2.81%	1.73%
AUROC	0.790	0.765	0.801
Accuracy	0.899	0.857	0.946
Sensitivity (Recall)	0.468	0.487	0.432
Specificity	0.909	0.867	0.955
Precision	0.108	0.096	0.145
Neg Pred Value	0.986	0.983	0.990
F1 score	0.175	0.161	0.217
Balanced Accuracy	0.688	0.677	0.694

Predicted Score distribution

Risk score distribution by Age, Ethnicity and Vaccination Status

Risk scores, on average, increase across age groups, and are higher for Māori and Pasifika ethnicities and for the unvaccinated.


Boosted
Fully vaccinated
Unvaccinated
Risk Score
1.0
0.8
0.6
0.4
0.2
Other
Maori
Pasifika
20s
30s
40s
50s
60s
70s
80s
Age

Actual hospitalisation rate

The risk groups are stratified into top 10% highest risk (“High”) and the rest 90% (“Low”), using the model trained with all the data (given no significant performance difference observed over the training set vs. the test set).

Because the high risk threshold is set at the top 10% or risk scores (when the prevalence is 2.3%), and because the model is imperfect, we do expect to see some differences in the hospitalisation rate within the high risk group across demographic groupings which reflect overall prevalence for each grouping.

In general, the significant hospitalisation rate difference in each of the low vs. high risk group pairs indicates that the model is doing well in stratifying cases by their risk level, regardless of the sub-cohort.

Hospitalisation rate by risk level

The hospitalisation rate is significantly higher in the high risk group.

Risk level	% Hospitalised (sample size)
Low	1.4% (215,643)
High	10.8% (23,961)
Overall	2.3% (239,604)

By age groups

In the high risk group, the hospitalisation rate is highest in the 60+ age band.

Hospitalisation rate by age group

Risk Level	% Hospitalised (sample size)		
	18-39	40-59	60+
Low	1.3% (142,604)	1.3% (61,326)	2.2% (11,713)
High	9.8% (9,879)	9.5% (6,186)	13.0% (7,896)
Overall	1.9% (152,483)	2.1% (67,512)	6.5% (19,609)

By ethnicity

The hospitalisation rate is similar across ethnicities in the higher risk groups. The prevalence is higher overall for Māori, and the slightly lower hospitalisation rate in the higher risk group compared to the “Other” ethnicity group likely reflects the slightly lower model performance for Māori.

Hospitalisation rate by ethnicity

	% Hospitalised (sample size)		
Risk Level	**Other**	**Māori**	**Pasifika**
Low	1.2% (134,045)	1.6% (36,245)	1.6% (45,353)
High	10.7% (10,050)	10.6% (6,183)	11.1% (7,728)
Overall	1.9% (144,095)	2.9% (42,428)	3.0% (53,081)

By vaccination status

There are differences in the hospitalisation rate across vaccination status for the high risk group. The highest rate is for the people who have had one dose of the vaccine, although this may reflect variation related to the smaller sample size.

Hospitalisation rate by vaccination status

	% Hospitalised (sample size)			
Risk Level	**Unvaccinated**	**One dose**	**Fully vaccinated**	**Boosted**
Low	2.3% (5,308)	2.3% (1,574)	1.5% (111,922)	1.2% (96,839)
High	12.2% (4,102)	13.1% (869)	11.2% (10,637)	9.4% (8,353)
Overall	6.7% (9,410)	6.1% (2,443)	2.3% (122,559)	1.8% (105,192)

By gender

Within the high risk category, a much higher rate of hospitalisation is being captured for the “Male” group, despite having lower prevalence overall. It would be good to rule out whether a lack of distinction between hospitalisation with and of COVID-19 was driving this difference.

Hospitalisation rate by gender

Risk Level	% Hospitalised (sample size)	
	Not male	**Male**
Low	1.7% (109,286)	1.0% (106,357)
High	9.6% (18,162)	14.5% (5,799)
Overall	2.8% (127,448)	1.7% (112,156)