COVID-19 Risk Score Tool
Release 2

Privacy Impact Assessment

August 2022

Document creation and management

Version 2 Document Approval

Name/Title	Sign-off date
Approved by Group Manager, Care in the Community	9 September 2022
Approved by Risk & Privacy Manager, Ministry of Health	4 August 2022

This Privacy Impact Assessment ("the Assessment") is the second PIA on the Risk Score Tool that will form part of the digital response platform when a positive COVID-19 case is identified. This Assessment concerns the second iteration of the tool.

This document will be made publicly available on the Ministry of Health website.

Disclaimer

This updated Assessment has been prepared to assist the Ministry of Health ("the Ministry") to review the use of Ministry-held information for the purposes of improving the accuracy of a risk calculation tool already developed to support triaging for a managed pathway of care those who may be at greater risk of hospitalisation from contracting COVID-19, and the privacy safeguards that are required to manage those purposes. It is not necessary or appropriate to focus on every possible privacy risk (such as the specific details of how security will be applied) but rather the focus is on the most critical points of the Risk Score Tool.

Every effort has been made to ensure that the information contained in this report is reliable and up to date. No inspection of the Risk Score Tool operation or its solution software has taken place as part of this assessment, and any performance representations are as reported to the author.

This Assessment is intended to be a 'work in progress' and may be amended from time to time as circumstances change or new information is proposed to be collected and used.

Contents

Glossary

Term	Meaning
Care in the Community	The Care in the Community model is based on enabling people to be cared for in their home, when it is safe to do so, when they or a member of their household are considered to have COVID-19. The model is flexible, nationally supported, regionally coordinated and locally led, in order to meet the needs of local populations and effectively allocate system resources especially in a time of uncertainty when parts of the local health system may well become non-functional for short term as well.
Case	A person who is considered to have COVID-19
CCCM	**Covid Clinical Care Module**, a shared coordinating clinical record solution in the Border Clinical Management System (BCMS) to nationally support the Care in the Community requirements of individuals who are required to self-isolate as cases, and their household contacts. BCMS was originally created to manage the clinical component of Managed Isolation and Quarantine processes.
CPIR	This commonly refers to the **Covid Population Register** which is a campaign platform which commonly has the purpose of managing communication to the community. In this context it includes the information held in common across the COVID response which is provided through the integration of services such as vaccination, border management and contact tracing.
CIR	**COVID Immunisation Register**, which holds all COVID vaccination records
HIPC	Health Information Privacy Code 2020
NCTS	**National Contact Tracing System**, which enables accurate and timely information on all COVID-19 cases and contacts to be recorded and allows all regions of New Zealand to work together when required.
NHI	**National Health Index** number – this is the unique identifier that is assigned to every person who uses health and disability support services in New Zealand.
PHU	Public Health Unit
CCH	**Care Coordination Hub** set up in the different regions to coordinate and oversee all active management COVID Cases in the community for that region. Activities include Case investigation, assignment for clinical assessment and management, welfare referral/management and overall coordination to ensure all people under care have been appropriately cared for during the period of care. These can be public health units and clinical or welfare manaaki hubs, and in some regions they are combined, and in others they operate separately (for example 'Case investigation' could be managed separately from the clinical and welfare component).
Reach Aotearoa	A Ministry of Health-contracted national provider responsible for contacting Māori and Pacific Island people over 35 (outside Auckland) within 12 hours, and anyone over 65 or who is not enrolled with a GP within 24 hours, unless they have completed the online self-service form.
AWS	**Amazon World Service** secure data platform which hosts the Ministry's CPIR database in Sydney, Australia.

NZePS	**New Zealand e-Prescription Service** which provides a secure messaging channel for prescribing and dispensing systems to exchange prescription information electronically and holds patient medication history information.
NMDS	**National Minimum Dataset** which holds hospitalisation discharge information.
Deprivation Status	The **New Zealand Index of Deprivation** is a small-area-based index providing a measure of neighbourhood deprivation, by looking at the comparative socioeconomic positions of small areas and assigning them decile numbers from 1 (least deprived) to 10 (most deprived). The index is based on 9 socioeconomic variables from the Census. https://ehinz.ac.nz/indicators/population-vulnerability/socioeconomic-deprivation-profile/

Section One – Background and Overview

Background

1. COVID-19 is a serious threat to the safety of New Zealanders, and its impact on both the health of individuals and the capacity of the health system to cope continues to be significant. The health system is struggling to catch up on 2 years of deferred care caused by the pandemic, and this is in turn increasing pressure on primary, emergency, and secondary care. Patients are presenting in greater numbers, are more unwell, and are at later stages of illness than ever before. The ongoing outbreak also continues to impact staffing levels due to illness, isolation rules, and attrition (eg burnout and emigration).

2. New Zealand is currently in Phase Three of the Omicron response[1], in which a key aim is to maintain our national hospital capacity by slowing spread and supporting those positive Cases who can to isolate and recover in their own homes. The Care in the Community Framework has been established to support a regionally coordinated, locally-led approach to managing COVID-19 Cases and their whanau to ensure they receive the support they need to isolate and recover safely.

3. The majority of Omicron Cases experience a mild to moderate illness and are able to safely self-isolate with minimal or no clinical assistance. However, there remain a number who are vulnerable to more serious illness and death, and these people need to be identified as early as possible to ensure they are supported safely where possible and quickly escalated for clinical care if required. The health system therefore needs to be able to rapidly identify those at risk when they test positive and prioritise them for timely personal contact to ensure they receive appropriate assistance.

4. Using clinical and demographic factors which are known to impact the risk of hospitalisation, the Ministry of Health's Data & Digital Directorate and COVID-19 Care in the Community teams developed a population-based Risk tool derived from a model built by Waitematā District Health Board for assessing the risk of hospital admission for COVID patients in the Northern Region. The Waitematā model was developed using Cases from the Delta-strain outbreak and data available to the Northern Region from admissions, community services, and primary care, and was simplified to reflect the data available nationally to the Ministry of Health. Version 1 of the tool was implemented as part of a suite of digital and assisted pathways to identify those at risk and manage the high numbers of COVID-19 Cases, and is held in a secure, Data & Digital-controlled Salesforce platform. It is used to support primary care and COVID Community Hubs in Case contact decision making only, and is not used in support, or in place, of clinical assessment.

5. In Version 1 of the tool, the calculation is made from age, ethnicity, and vaccination information held in the Covid Population Register (CPIR) which is a campaign platform for managing communication to the community. In this context CPIR includes the information held in common across the COVID response which is provided through the integration of services such as vaccination, border management, and contact tracing. The score is used to support decision-making to prioritise contact for those Cases who:

   - do not respond to the initial automated text outreach from the National Contact Tracing Solution (NCTS) system to indicate that they are positive for COVID-19;

---

[1] https://www.health.govt.nz/covid-19-novel-coronavirus/covid-19-response-planning/omicron-community-what-means-you#:~:text=We%20are%20currently%20in%20phase,effective%20weapon%20against%20the%20virus

- do not have activity on their file in this time to indicate they have been assessed and/or contacted by their own or another health provider; or

- for whom no cell phone number is available to send a text to.

6. The tool calculates a risk score between 0 and 1 that identifies whether a person is at higher risk of hospitalisation[2]. The score is calculated and added to case files when they are created in the National Contact Tracing System (NCTS) and the Covid Clinical Care Module (CCCM).

7. Once added to an individual's CCCM file, the score is displayed in the header of the case page alongside the Acuity rating, which is a clinical calculation added by a health provider after a clinical assessment. In both NCTS and CCCM it appears on the dashboard, which displays a patient per row with each row containing key status, upcoming activity, and other summarised information. The dashboard is used for COVID case allocation and monitoring by the regional Care Coordination Hubs (CCHs). Telehealth teams (clinical) also have access to this in order to support people who are in the "supported self management pathway" and do not have digital resources.

8. CCHs run a daily report in either NCTS or CCCM (depending on which system they prefer to use) to identify those who have not responded or been assessed/contacted within 24 hours. The report ranks identified Cases according to their Risk Score to enable prioritisation for contact of those identified as at higher risk.

9. In response to the winter surge of Cases and to ease pressure on CCHs, Reach Aotearoa, under the direction of the National Investigation and Tracing Centre (NITC), also began using the Risk Score tool in late July 2022 to support prioritisation of Cases for contact.

10. The risk score is only available to clinical users who can see the dashboards in CCCM. Clinicians managing COVID care do not have access to the dashboard when they manage patients individually and work with their patient's full record. However, where a practice has large numbers of people under their care, they have found the need for the dashboard overview and workflow capability and the Regional Hubs have permitted access to a Facility level dashboard. A monitoring and audit programme is being implemented to oversee access via this expanded access option (see CCCM PIA).

11. An assessment carried out under the Algorithm Charter (when the tool was initially intended to support clinical assessment) scored the risk of unintended adverse outcomes for individuals as low probability/low impact. This remains the same with Version 2 but, with an improved model, further reduced risk of adverse outcomes and increased use of the tool are likely.

12. The Ministry's Māori directorate has been consulted about the tool and approves of its use as a way to identify high risk people, including Māori, for follow up. The Ministry also worked with Whanau HQ/Northern Region Health Coordination Centre (NRHCC) which was involved in developing the tool it was based on, and has consulted with DHBs and CCHs.

13. Subsequent iterations to improve the algorithm's accuracy are planned, and Version 2 is the first of these. They will include ingesting information from other databases and linked via NHI numbers as the tool is further developed and adjusted to improve its accuracy.

14. Use cases of future iterations of the tool could include to facilitate identification and prioritisation for clinical assessment for interventions such as prescribing therapeutics. Use for

---

[2] This is everyone in the Health Services Utilisation (HSU) database, which includes all individuals who have used the New Zealand health system in the last two years, and the National Enrolment Service (NES) which holds information from everyone who has enrolled with a general practitioner.

any purpose other than supporting contact decision making will be subject to appropriate review and clinical approval, and additional privacy assessment.

15. Version 2 involves no change to the purpose of prioritising Cases for contact based on risk of hospitalisation. The changes being made are to improve the accuracy of the calculation by including further indicators of increased likelihood of hospitalisation. These include underlying conditions, hospitalisation within the last 2 years, biological sex, and social deprivation. The new calculation also updates the definition of fully vaccinated from two doses to two plus a booster, and has been trained on Omicron data from the current outbreak.

16. Version 2 further differs from Version 1 in that it makes the calculation on creation of a Case in NCTS, where Version 1 made and held the calculation population-wide in advance. The Ministry notes that if a future surge results in undue burden being placed on the systems involved in calculating on an as-required basis, it may return to calculating in advance. This will be subject to further privacy review and updating of this PIA.

17. Future use cases of the tool may include health sector planning and readiness purposes at a population level to understand the impact of an outbreak across regions and practices. Individuals would not be identified in the outputs of this use.

18. The Office of the Privacy Commissioner and the Government Chief Privacy Officer have been consulted and provided comments on a draft Privacy Impact Assessment for Version 1 of this tool. Their comments have been considered by the Ministry and incorporated as appropriate.

19. As additional data sets are layered onto subsequent versions of the model, this Privacy Impact Assessment will be reviewed and updated.

20. It is also noted that as of July 1, 2022, ownership of the tool has been transferred, with the Care in the Community programme, from the Ministry of Health to the National Public Health Service (NPHS) in Health New Zealand (HNZ).

Scope of Assessment

21. This PIA has been prepared to assess the potential impacts on privacy to Aotearoa New Zealand's population of an update to the automated COVID-19 risk score calculation which was initially based on three health and demographic points of information the Ministry holds about them, for use in prioritising for contact those people who are likely to require assistance on confirmation of infection with COVID-19. This PIA covers the second iteration of this tool.

22. An online pathway for Cases who are digitally enabled provides for them to complete a self-assessment Contact Tracing Form following text notification that they are COVID positive, and completion rates for this as at mid-June 2022 were tracking at approximately 70%. The form gathers information around co-morbidities, other health conditions, and risk factors such as living alone, that provide more health information to guide pathway decision-making. The processes supporting decision making and management of these Cases is outside the scope of this PIA but may be considered in future reviews if this information is included in later iterations of the tool. No information from the Form will be included in Version 2, and this PIA does not cover the Form other than as the information submitted through it (or not) impacts use of the risk score calculation in decision making to contact Cases. A separate PIA has been completed for the Form.

23. This PIA does not consider other automated processes in the contact tracing system, such as SMS text notifications and alerts, that support the wider response to community transmission and assistance to individuals and households in self-isolation, the systems that provide data to generate the score, or changes to the systems that will ingest the score such as CCCM. Separate PIAs have been completed for these applications.

24. Assisted channels have been developed in parallel with this tool and its supporting applications, such as the COVID-19 Contact Tracing Form, and assessment of these is outside the scope of this PIA.

Use of local population-based Risk calculations by regions

25. A number of regions, including the Metro Auckland region which developed the model this tool is based on, are using risk calculations based on C0VID-19 and other health information collected and/or held about their own populations to assist with case management decision making. These are outside the scope of this PIA.

Information Collected and User Information Flows

Assessing healthcare needs

Cases will self-manage by default but can escalate to active clinical management if indicated by their self-assessment, their primary healthcare provider's assessment, or contact coordinated by their care coordination hub
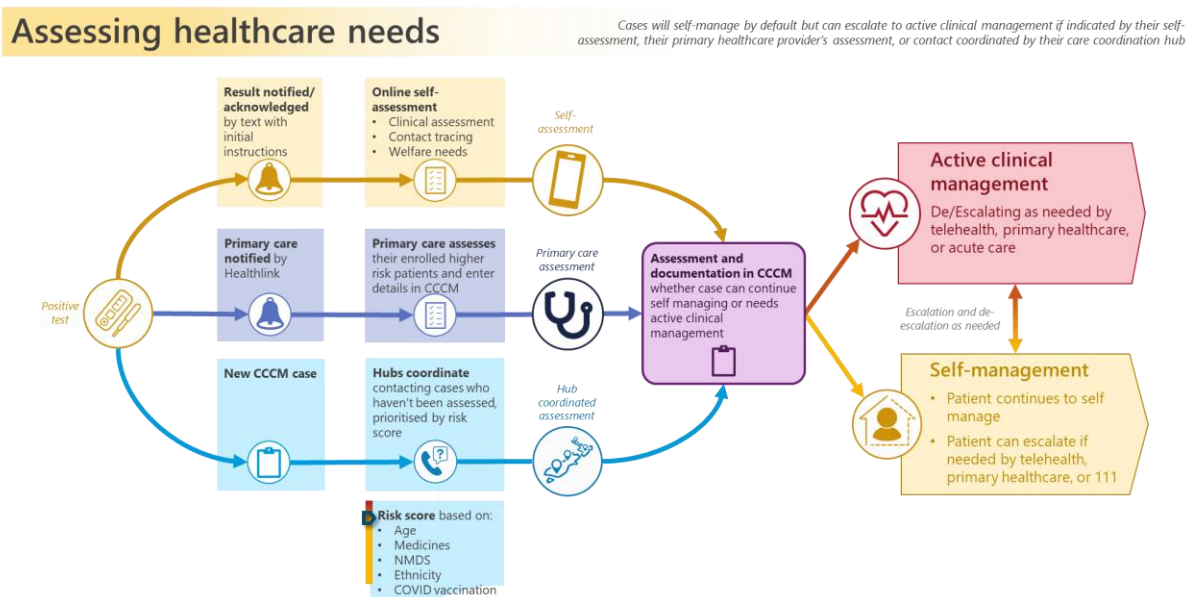
Section Two - Privacy Analysis

The potential privacy impacts resulting from this project are analysed below. The analysis has been completed, against the 13 rules of the Health Information Privacy Code 2020.

The Ministry has conducted its analysis under the Health Information Privacy Code as the information is about Consumers and their health services. Under clause 4(1)(e) it is considered that this is information about an '*individual which is collected before or in the course of, and incidental to, the provision of any health service or disability service to that individual*'.

Purpose of collection (Rule 1)

26. The Ministry already holds the hospitalisation and prescription information to be added to the calculation in Version 2 of the tool and collects this information to support management of individuals' health. Subsets of these datasets will be extracted monthly to enable calculation of an individual's score where they are recorded as a Case, for the directly related purpose of supporting the management of individuals' health in the event they are at risk of hospitalisation on contracting COVID. This data will also be used to improve the accuracy of the tool, in a form that will not identify individuals.

27. The information in use for Version 1 was collected to support manual contact tracing decision making. As daily Case numbers at the beginning of the Omicron outbreak outstripped the ability of Contact Tracers to manually follow up all Cases and contacts, the response moved from both contact tracing and supporting Cases to isolate, to support for isolating only. The tool was developed to apply statistical weightings to these data points to provide automated support to the Case contact decision-making process.

Source of information (Rule 2)

Collection from a source other than the individual

28. Version 1 of the tool uses information collected from the Covid Population Register (CPIR), which is a part of the platform for managing the COVID response. In this context CPIR includes the information held in common across the COVID response which is provided through the integration of services such as vaccination, border management, and contact tracing. In Version 1, the calculation is made using age, vaccination, and ethnicity information. Use of these factors is based on the known increased risk of hospitalisation from COVID associated with greater age, whether the person has been vaccinated and, for some populations, ethnicity. To date, Māori and Pacific Island people have suffered disproportionately worse outcomes from COVID infection and, due to historic and ongoing inequities, are also more likely to have underlying conditions or other inequity-related risk factors than other ethnicities.

29. Age and ethnicity information held in CPIR is sourced from the Health Service User database (HSU), and vaccination status from the Covid Immunisation Register (CIR). Age and vaccination status information is held for 100% of individuals in the database, and ethnicity information for 99.14%.

30. Version 2 will expand the variables used in the calculation to include information collected from:

31. The New Zealand e-Prescription Service (NZePS), as a proxy for underlying conditions based on medicines prescribed;

32. the National Minimum Data Set (NMDS), which holds hospitalisation discharge information; and

33. the Deprivation Index, a small-area-based index providing a measure of neighbourhood deprivation. This information will be matched with individuals' domicile code information to inform the calculation.

34. It will also include biological sex, as there is some increased risk if an individual is not male. This information is collected from the National Health Index (NHI).

35. The information will be collected from a source other than the individual under Rule 2(2)I(iii), that the Ministry believes on reasonable grounds that compliance with the requirement to collect the information from the individual would prejudice the health or safety of any individual.

36. The tool is used to support decision-making for the portion of the population aged 18 and over that does not respond within 12 hours to the initial text outreach from the national contact tracing solution (NCTS) system to indicate that they are positive for COVID-19. Cases for whom no contact details to enable digital outreach are held are immediately prioritised for follow up. These people are contacted via telephone, if their contact number is held, or by a community health worker visiting their home.

37. In parallel with the Risk Score tool, demographic information is also being used to support faster contact pathways for at-risk populations. On creation of a case for them in NCTS, Māori in Auckland are immediately referred to the Māori Regional Coordination Hub (MRCH), and Pacific Island people to the Pacifica Regional Coordination Hub (PRCH), for contact. Outside Auckland, the Ministry-contracted provider Reach Aotearoa is actively reaching out to Cases who have not completed the online form and are of Māori/Pacific Island ethnicity over 35, or in Decile 9/10 on the New Zealand Index of Deprivation[3], within the first 12 hours. Everyone else is expected to be contacted within 24 hours, unless they complete the form within this time and are assessed from the information provided as not at risk. Those who are not contacted will have been provided with information via the SMS and COVID Health Hub link on how to self manage and how to escalate their care if they have any concerns. GPs also receive notification via CCCM where one of their patients is a Case, and assess whether the patient is potentially at high risk. If so, and the case has not been assigned to another provider, they will contact the patient for a clinical assessment to determine a management plan. Under Phase 3 of Omicron, however, agencies have struggled to meet these timeframes, and it is known that some Cases have not been able to be contacted within the at-risk period of their illness. This tool is intended to assist with ensuring that those who are more likely to be at risk are not 'lost'. A review of Version 1 in operation completed by Precision Driven Health has identified areas where it is used and confirmed that Version 1 performed somewhat better than random guessing (Appendix 1). In comparison with Version 2, including where this first model was retrained on Omicron data, they found improved performance with Version 2 (Appendix 2).

Manner of collection (Rules 3 and 4)

38. The Ministry considers that the manner of collection from sources other than the individual (that is, from health datasets held by the Ministry) is lawful and, in the circumstances of a serious threat to public health and safety, and the health and safety of individuals, does not intrude to

---

[3] The New Zealand Index of Deprivation is a small-area-based index providing a measure of neighbourhood deprivation, by looking at the comparative socioeconomic positions of small areas and assigning them decile numbers from 1 (least deprived) to 10 (most deprived). The index is based on 9 socioeconomic variables from the Census. https://ehinz.ac.nz/indicators/population-vulnerability/socioeconomic-deprivation-profile/

an unreasonable extent upon the personal affairs of the individual concerned. The information will be collected from, and remain within, the Ministry's secure systems.

39. External communications to the health sector and general public have been developed to explain and clarify what the tool is and what it does. This communication is included in existing channels such as health key messages and sector webinars. The Ministry does not intend to actively promote the tool to the public as it is not a public facing tool, but information on how it works, how it fits into the wider care in the community strategy, and the role it plays in ensuring clinical and welfare support can be targeted to where it is needed the most, is available on the Ministry website together with this PIA. FAQs are being developed to support both health sector and public comms and engagement.

40. Information about the algorithm has been published on the Ministry's website in accordance with Rule 3 of the HIPC, and the transparency requirements of the Algorithm Charter. The tool will operate in the Ministry's secure environment on a Salesforce platform held in AWS servers in Australia, and the algorithm calculation has been made publicly available on Te Pokapū Hātepe o Aotearoa New Zealand Algorithm Hub[4].

Security (Rule 5)

41. All identifying information is held and handled within the Ministry's secure AWS systems in Australia, and carries the security classification of Medical IN-CONFIDENCE.

42. The risk score is calculated in CIPR then viewed and linked to a case record when a person is recorded as positive in NCTS and a case is created for them in CCCM. Access to CCCM includes non-Ministry users who are part of the public health response, and is subject to user controls and auditing. Non-Ministry users are:

> CCHs – regional care coordination hubs which run a daily NCTS or CCCM report of Cases who have not responded to digital outreach or had activity on their file to indicate contact by another provider, within 24 hours, and triage them for contact with the aid of the Risk Score.

> Reach Aotearoa – a national provider responsible for contacting Māori and Pacific Island people over 35 (outside Auckland) within 12 hours, and people over 65 or who are not enrolled with a GP within 24 hours. Reach Aotearoa began using the score from late July 2022 as Cases increased in the winter surge.

> GPs – though the score is available to them, it is expected GPs will generally rely on their knowledge of their own patients, and their patient records, when making decisions about who to prioritise for contact.

> Hospitals – staff providing care can access patient files of individuals whose records are set to Active Management. They do not actively use the score as it is not provided to support clinical decision making.

> DHBs – District Health Boards data analysts have access for reporting purposes.

43. Version 2 ingests more information about individuals than Version 1. While this information will not be disclosed outside the Ministry's secure systems, or used for any purpose other than the risk score calculation, as a dataset of information that has not previously been linked it presents a higher security risk than the information collated for Version 1.

---

[4] https://algorithmhub.co.nz/

44. This risk is mitigated by using joined data sets when building the Risk of Hospitalisation Model, and then immediately deleting the joined data from all parts of the system once the calculations are made. This process is regularly re-run (monthly) to ensure accuracy of the model with more current data sets.

45. The NMDS and NZePS data subsets will be extracted initially, held separately, and then refreshed monthly from their parent databases. They will be ingested separately by the model tools as required.

46. To support the calculation of a Risk Score for an individual who is identified as a Covid Case, the data will be held within CPIR and kept up-to-date. Operational systems will only ever see the scores, linked to individuals by their NHI number, and these will be retained within the systems and data warehouse. This means the only output is the method, the date the score was calculated, and the score itself where a file is created for an individual in NCTS. Systems and processes will record relevant metadata should recalculation or an audit or reconstruction of the score be required.

47. Prior to each substantive release, the Project will be subject to Ministry security review processes.

Access and Correction (Rule 7)

48. Individuals are able to request access to, and correction of, their information in accordance with the Ministry's standard channels and as permitted under the Health Information Privacy Code 2020. This includes access to audit log information for records held in NCTS and CCCM.

49. As noted above, the tool is iterative and will be updated as more data are reviewed and deemed appropriate to strengthen the model's predictive capacity. This may mean that people's scores will be updated over time. The Ministry is ensuring that as newer scores are generated, the original and previous scores will still be available in the database. Further work is intended to make calculated scores available as a viewable "history" of scores.

Accuracy and verification of information (Rule 8)

50. In its current (Version 1) state, the model has significant limitations due to the limited data that the formula is based on, and the key risk is the accuracy with which it can identify those people able to safely undertake self-management. These limitations are due to the model being powered by data from the Delta outbreak when, in addition to the variant being less infectious than Omicron, vaccination coverage was lower and boosters were not yet available. Further limitations were the availability of other information nationally that could also be used to identify those at risk.

51. An external peer review of Version 1 of the tool undertaken by Precision Driven Health to validate the statistical methods showed that the tool was not appropriate for clinical decision making, but would be appropriate to assist with prioritising calls to people who do not respond to the text notification within a determined timeframe. It recommended that if the tool were deployed that "significant and appropriate" protections should be put in place to mitigate the likelihood that people assessed as lower risk by the model may still experience poor outcomes. It also noted the authors believed that the accuracy of the tool could be improved by including comorbidity or other information as explanatory variables.

52. The Ministry considered the question of when a tool is good enough to use, or good enough to be of value in the circumstances (when the benefits outweigh the risks, and/or the risks can be sufficiently mitigated). Despite its limitations, it considered that using the tool in a limited scope, to prioritise people likely to be at greater risk of hospitalisation for contact, and where other outreach has either not occurred for any reason (for example, no mobile contact details are

held) or has not been responded to, is an appropriate use. On 23 February 2022 the Director General of Health authorised deployment of Version 1, with mitigations in place, to support this use case. Version 1 was deployed in mid-March 2022.

53. The Algorithm Governance Group, an advisory and oversight body formed after the tool went live, subsequently recommended it should be turned off on the grounds that the "quality of the model" was unknown. The Care in the Community programme considered the Group's recommendation but, with approval from the Care in the Community Leadership team, decided to keep the tool running and complete an evaluation of its use and efficacy.

54. A Precision Driven Health evaluation of Version 1 in operation, based on early results and with some limitations (eg all hospitalisations are included with no specialty exclusions), found that overall the model did help identify people more likely to be hospitalised in the Omicron outbreak, and that across all groups, and with varying performance, it performed better than random selection. Performance for the age group 60 and over was very good (see Appendix 1).

55. It is expected that additional data and further mathematical considerations will materially improve the quality of the tool. The Ministry has limited data holdings on specific comorbidities (for example, cancer), but does not have access to patient-level information about all conditions that may pose an elevated risk to a patient testing positive for COVID. For Version 2 it is therefore relying on the 'proxy' information of prescription data (unique medications and number prescribed in the 6 month period July 2021- January 2022 available for this data) and hospitalisation information (number of discharges after more than 12 hours in the preceding two years) to infer the presence of any condition that may place a Case at higher risk. This use has some limitations, including those noted below in paragraph 53, but is, on balance, considered the most broadly accurate and comprehensive means of identifying likely risk. Its use in the calculation has been carried out with close clinical advice and oversight.

56. To maintain maximum possible currency, the NZePS and NMDS data will be extracted monthly to enable calculation of the scores using the most up-to-date information, though the Ministry notes that there are some timeliness limitations on the data available. The NZePS has been taken from a one-off 6 month data extract, which is all that is currently available, and NMDS from the preceding 2-year period which, due to NMDS data lags, may be up to three months out of date by the time the case is identified (ie admissions in the preceding period up to three months may not be included).

57. Including further, clinically relevant information in the calculation, and training the tool on more recent and relevant data that consists of Cases that have been infected with the Omicron variant, is expected to increase the accuracy of the calculation, and thus compliance with Rule 8. However, both ongoing evaluation of the tool's use in practice and further statistical analysis as more Omicron outbreak data is collected will be critical to determining whether the additional information materially increases the calculation's accuracy.

Retention, Use, and Disclosure (Rules 9 – 12)

Retention

58. Under rule 9 of the Health Information Privacy Code, health information may not be retained for longer than is required for the purposes for which the information may lawfully be used. The information used in the first iteration of the calculation is already held in CPIR and subject to the retention period specified for that system. The score will be sent to, and held in, NCTS and CCCM, and subject to their retention schedules. Patient information will remain accessible in CCCM to service providers for six weeks after they are recorded as recovered, to support any follow-up care that may be required, and will then be archived and inaccessible to clinical users.

59. As noted above, the tool is iterative and will be updated as more datasets are reviewed and deemed appropriate to strengthen the model's predictive capacity. This may mean that information from other databases will be copied to, and held in, CPIR, and that people's scores will be updated over time. As the model is upgraded the team will communicate the change in scores and possible impacts to the Ministry and the health sector, and the Ministry intends to ensure that as newer scores are generated, the original and previous scores will continue to be available.

60. The NMDS and NzePS data subsets extracted and then refreshed monthly from their parent databases will be held separately within CPIR (on the Ministry's secure AWS Platform) for use in calculation of the score. These data sets are also stored for used within the Model Tools when rebuilding the model, and are deleted from it when they are no longer required for this purpose.

61. This is a new use of the additional health datasets that the tool will ingest information from. The Ministry is using it for this purpose under Rule 10(1)(d)(i) and (ii) that the use of the information is necessary to prevent or lessen a serious threat to public health or safety, or the life or health of the individual concerned or another individual. This is because the Ministry's clinical advisors consider that COVID-19 continues to pose a serious risk to public health and the life or health of individuals. The variants currently circulating are highly infectious and the disease can have serious outcomes for many individuals. The health system continues to be stretched by the ongoing impact of more than two years of deferred care caused by the pandemic, and this impact continues to be compounded by ongoing staffing pressures resulting from illness, COVID isolation requirements, and attrition. Targeting those at most risk from COVID-19 is critical to reducing preventable hospital admissions, preventing serious outcomes for individuals, and reducing pressure on a system that is struggling to both recover and manage high loads of complex and serious illness.

62. The Ministry also notes that while the context of this use continues at present to be to prevent or lessen a serious threat to public health or the life or health of any individual, the information held in the NZePS and NMDS is collected to support management of the health of the individuals' concerned. The Ministry therefore considers that use to calculate their risk of hospitalisation if they contract COVID also falls under Rule 10(1)(b), that the purpose for which the information is to be used is directly related to the purpose for which the information was obtained.

63. The information will be used for the purpose of calculating the risk score for an individual with COVID, and improving the accuracy of the risk score tool in identifying those at higher risk of hospitalisation on becoming infected with COVID. It will be used in identifying form within the tool to model risk to the individual for purposes of prioritising them, as necessary, for assistance, and will be deleted from the tool once the calculation has been made.

64. In non-identifying form, the risk scores may be used for response planning at a population level.

Use of the Risk Score

65. Version 1 has been integrated into the community assessment and follow up workflows, and a report downloaded daily by CCHs (and, from July 2022, Reach Aotearoa) ranks Cases in priority order for contact. The tool provides a risk score between 0 and 1 that identifies whether a person is at higher risk of hospitalisation and likely to be in need of active care management, or at lower risk and likely to be able to manage their COVID-19 infection through the self-service pathway. The Version 1 scope of use is limited to providing an initial triaging for contact of people who do not respond to the initial automated outreach, or for whom contact details to enable this are not held, and there is no change to this use in Version 2. This version is expected to increase the accuracy of the tool in identifying those Cases at higher risk of hospitalisation,

but will not increase it sufficiently to support clinical or other decision making such as prioritisation for COVID therapeutics.

66. A health or Local Hub provider reviews each referral made via the tool prior to contacting the individual. Health providers include the person's own GP where they are enrolled with one, and it is expected these will access and review their patient's records rather than rely on the prioritisation assessment. Where the Case is not enrolled with a provider, a Public Health Unit (PHU) or contracted PHU provider assesses them. The score is only used where no other clinical information about the Case is available.

67. Use of algorithms for decision-making about individuals is sensitive and, as a government agency and signatory to Aotearoa New Zealand's Algorithm Charter, the Ministry is committed to ensuring New Zealanders can have confidence in how it uses algorithms. The requirements of the Algorithm Charter include transparency and accountability to ensure the public can trust and support the government to use these tools in appropriate ways. Since Version 1 of the tool went live, the Ministry has established an Algorithm Governance Group, which includes privacy representation, to maintain oversight of the tool's development and use, and ensure transparency and compliance with the Charter.

68. Until the end of June 2022 the Health System Preparedness Program Steering Group was the governing body that the Covid Care in the Community Team reported to, and which assumed responsibility for this work. As of 1 July 2022, this responsibility has been assumed by the National Public Health Service.

69. Use of Ministry-held data is governed by the Data and Information Governance Group, and it has approved use of NMDS and NZePS data for this purpose.

70. The information used in Version 1 is health and demographic information (age, ethnicity, and vaccination status). The calculation is made and held in CPIR, and added to the individual's NCTS and CCCM Cases when they are created. With Version 2, the calculation will be made when a Case is created in NCTS.

71. Ethnicity (Māori, Pacific, and other) is included in this calculation because Māori and Pacific Island people have to date been impacted by COVID at a higher rate than other ethnicities, and has been retained in Version 2 as Precision Driven Health's review found it improved the overall performance of the model in a more than minor way. The Ministry also has responsibilities under Te Tiriti o Waitangi to achieve equity and improve outcomes for Māori. Early identification of people at higher risk, including by reason of the statistically poor outcomes indicated by ethnicity, will enable them to be quickly directed for contact and support through their local Care Coordination Hub (CCH) or, where available, a culturally appropriate provider. Where these providers are not able to reach them within 24 hours, they are prioritised via their score in the daily report run by their regional CCH.

Disclosure

Internal disclosure of Ministry-held information for use in the Risk Score calculation

72. This is a new disclosure of information from the health datasets that the tool will ingest information from. The Ministry is disclosing it for this purpose under Rule 11(c) that, as it is to be disclosed to support management of the individual's health where they test positive for COVID, the disclosure of the information is one of the purposes in connection with which the information was obtained.

73. The disclosure and calculation will occur within the Ministry's secure AWS systems, with no disclosure of any component data points out of these systems.

Disclosure of Risk Score calculation output

74. The tool sends an algorithmic score between 0 and 1 to NCTS and CCCM, and this is available to authorised users in the Case record and a dashboard to support the patient contact triage decision where no other clinical information is available. The score goes to both systems to ensure that CCHs using CCCM have access to it when needed.

Disclosure outside New Zealand (Rule 12)

75. There is no expectation of any disclosure of information outside New Zealand (otherwise than for safe custody or processing in compliance with s11 of the Privacy Act 2020, due to the hosting sites located in Australia).

Unique identifiers (Rule 13)

76. Information is sourced and linked using NHI numbers. This use is consistent with the purposes for which these are assigned.

Section Three – Privacy Risk Assessment and Controls

Privacy Risk Table

Risk Reference Number	Privacy Risk Description	Raw Risk Rating Consequence / Likelihood	Existing Controls	Current Risk Rating Consequence / Likelihood	Planned Controls	Target Risk Rating Consequence / Likelihood	Rationale for Target Risk Rating
R01	**Source:** The prioritisation calculation is not necessary to the public health response **Risk:** the collection and use of personal information is unlawful **Effect:** Reputational damage; trust and confidence in the health sector undermined	High (18) Significant / Possible	**PIA03** Clinical approval for limited use **PIA01** Ministry Algorithm Governance **PIA02** Risk Score Tool Governance **PIA04** Limitations on use	Medium (9) Moderate / Unlikely	**PIA12** Ongoing evaluation and review	Medium (9) Moderate / Unlikely	Where no other information is available to support contact decision making, the score enables those who are statistically more likely to be at risk to be prioritised.
R02	**Source:** That a person was identified as being at risk is retained on their clinical records **Risk:** That the individual was calculated as at risk is	High (18) Significant / Possible	**PIA10** Limitations on disclosure **PIA11** Training **PIA06** Privacy Statement, algorithm transparency, and	Medium (6) Moderate / Rare		Medium (6) Moderate / Rare	Controls can mitigate the likelihood; consequence if realised can be reduced through clear communications about the purpose and limitations of the score.

Risk Reference Number	Privacy Risk Description	Raw Risk Rating Consequence / Likelihood	Existing Controls	Current Risk Rating Consequence / Likelihood	Planned Controls	Target Risk Rating Consequence / Likelihood	Rationale for Target Risk Rating
	disclosed to another party, such as an insurer **Effect:** Reputational damage; trust and confidence in the health sector undermined		public communications plan **PIA07** Sector Communications plan				

Risk Reference Number	Privacy Risk Description	Raw Risk Rating Consequence / Likelihood	Existing Controls	Current Risk Rating Consequence / Likelihood	Planned Controls	Target Risk Rating Consequence / Likelihood	Rationale for Target Risk Rating
R03	**Source:** Individuals are concerned that their health information is being used for a purpose other than that for which it was collected **Risk:** Individuals complain about breach of privacy **Effect:** Reputational damage; trust and confidence in the health sector undermined	High (17) Moderate / Likely	**PIA01** Ministry Algorithm Governance **PIA02** Risk Score Tool Governance **PIA03** Clinical approval for use **PIA04** Limitations on use **PIA06** Privacy Statement, algorithm transparency, and communications plan **PIA07** Sector Communications plan	Medium (5) Minor / Unlikely		Medium (5) Minor / Unlikely	The Ministry considers the use is lawful under under Rules 10(1)(d)(i) and (ii) and 10(1)(b). The controls in place support ongoing review, transparency, and governance of this use.

Risk Reference Number	Privacy Risk Description	Raw Risk Rating Consequence / Likelihood	Existing Controls	Current Risk Rating Consequence / Likelihood	Planned Controls	Target Risk Rating Consequence / Likelihood	Rationale for Target Risk Rating
R04	**Source:** An unauthorised party accesses, alters, uses, and/or discloses personal information extracted from parent databases and held separately for use in the calculation **Risk:** Breach of sensitive health information **Effect:** Reputational damage; trust and confidence in the health sector undermined	High (18) Significant / Possible	**PIA08** Security Review **PIA09** No disclosure outside Ministry systems	Medium (10) Significant / Rare	**PIA13** Limited retention of source data **PIA14** Separation of source data	Medium (10) Significant / Rare	The impact if this information were breached cannot be mitigated, but security measures can reduce the likelihood. The additional measures to reduce likelihood do not reduce the risk rating as more information is involved than in Version 1 and these are being implemented to protect that information.
R05	**Source:** An unauthorised party accesses, alters, uses, and/or discloses the scores held in the tool **Risk:** Breach of sensitive, identifiable health information **Effect:** Reputational damage; trust and	High (18) Significant / Possible	**PIA08** Security Review	Medium (10) Significant / Rare	**PIA13** Limited retention of source data	Medium (10) Significant / Rare	The tool will hold risk scores linked to NHI numbers only. The impact if this information were breached cannot be mitigated, but security measures can reduce the likelihood.

Risk Reference Number	Privacy Risk Description	Raw Risk Rating Consequence / Likelihood	Existing Controls	Current Risk Rating Consequence / Likelihood	Planned Controls	Target Risk Rating Consequence / Likelihood	Rationale for Target Risk Rating
	confidence in the health sector undermined						
R06	**Source:** a person is at risk due to factors other than those used in the calculation, or to limitations in the source data **Risk:** a person at higher risk is not identified as such by the score, and suffers adverse consequences due to not being prioritised for contact **Effect:** Reputational damage; trust and confidence in the health sector undermined	High (18) Significant / Possible	**PIA05** Additional collection from individual **PIA11** Training **PIA07** Sector Communications plan	High (14) Significant / Unlikely		Medium (10) Significant / Rare	Including further, clinically relevant information in Version 2 of the calculation, and training it on more recent and relevant data that consists of Cases that have been infected with the Omicron variant, is expected to increase the accuracy of the calculation and therefore further reduce this risk. All other controls remain in place.
R07	**Source:** Datasets of health information about individuals are combined in a new way	High (18) Significant / Possible	N/A		**PIA13** Limited retention of source data	Medium (6) Moderate / Rare	The NMDS and NZePS data will not be combined anywhere outside the tool, and will be immediately deleted from

Risk Reference Number	Privacy Risk Description	Raw Risk Rating Consequence / Likelihood	Existing Controls	Current Risk Rating Consequence / Likelihood	Planned Controls	Target Risk Rating Consequence / Likelihood	Rationale for Target Risk Rating
	Risk: A breach results in more information being disclosed than was previously possible**Effect:** Reputational damage; trust and confidence in the health sector undermined				**PIA08** Security Review **PIA09** No disclosure outside Ministry systems **PIA14** Separation of source data		within the tool once the calculation is made.

Privacy Control Table

Control Reference Number	Control Name	Control Description	Status	Control Owner
PIA01	Te Whatu Ora Algorithm Governance Group	A body with Privacy representation will be put in place to advise the implementation, use of, and any changes to this tool and any others the Ministry implements. This body is to ensure that risk is assessed alongside benefit before use, that risk of unintended consequences can be monitored throughout the lifetime of an algorithm, and that there is oversight, accountability, clear change processes, and adherence to the Algorithm Charter.	Implemented	General Manager, Emerging Health Technology & Innovation – Data & Digital Te Whatu Ora Health NZ
PIA02	Risk Score Tool Governance	The National Public Health Service is the governing body that the Covid Care in the Community Team reports to, and which assumes responsibility for this project.	In progress	COVID-19 Care in the Community Clinical lead and Risk Score Business Lead COVID-19 Care in the Community National Public Health Service
PIA03	Clinical approval for limited use	The purpose for using the tool has been considered and endorsed by clinical experts as appropriate for use in prioritising unresponsive or unreached Cases for contact only, in conjunction with parallel processes, in the circumstances. This supports the lawful use under Rule 10(1)(d)(i) and (ii) that the use of the information is necessary to prevent or lessen a serious threat to public health or safety, or the life or health of	Implemented	Clinical Lead, Covid Care in the Community Team COVID-19 Care in the Community, National Public Health Service

Control Reference Number	Control Name	Control Description	Status	Control Owner
		the individual concerned or another individual.		
PIA04	Limitations on use	Tool in its first iteration is to be used to support patient contact triage only and is not to be used in any circumstances as an aid to, or substitute for, clinical assessment.	Implemented	COVID Care in the Community, Risk Score Business Lead, National Public Health Service
PIA05	Additional collection from individual	Online self-reporting form will collect additional information from the individual	Implemented	Programme Manager, Data and Digital, Te Whatu Ora
PIA06	Privacy Statement, algorithm transparency, and communications plan	Clear privacy statement on website, including how to request access and correction; plain language information about the algorithm and what the score is used for; and publication of this PIA	In progress	Manager External Communications Te Whatu Ora
		Update Algorithm Charter information on website to include Risk Score tool	In progress	General Manager, Emerging Health Technology & Innovation – Data & Digital Te Whatu Ora Health NZ
		Publication of algorithm's mathematical formula on the government's Algorithm Hub, in accordance with transparency requirements under the Algorithm Charter	Implemented	General Manager, Emerging Health Technology & Innovation – Data & Digital Te Whatu Ora Health NZ
		Clear communications strategy to inform the public about how their information is being used and protected	In progress	Manager External Communications Office of the Director General
PIA07	Sector communications plan	Clear communications to health sector about the purpose, intended use, and limitations of the tool	In progress	Manager External Communications Office of the Director General

Control Reference Number	Control Name	Control Description	Status	Control Owner
PIA08	Security Review	The standard security Te Whatu Ora Health NZ security review processes will be completed to Authority to Operate level prior to go live. If any risks are identified they will be mitigated or eliminated prior to go live.	Completed	IT Security Manager, Te Whatu Ora Health NZ
PIA09	No disclosure outside Te Whatu Ora systems	All identifying information will remain within Te Whatu Ora systems which operate in a secure AWS environment.	Implemented	Programme Manager, Data and Digital, Te Whatu Ora
PIA10	Limitation on disclosure	The risk score is not to be included in records saved from CCCM to GPs' PMS files	Implemented	Programme Manager, Data and Digital, Te Whatu Ora
PIA11	Training	Training for non-clinical users in use of score	In progress	Change Management Lead, COVID Care in the Community, Data & Digital, Resourcing & Commercial
PIA12	Ongoing evaluation and review	Periodic evaluation to ensure validity and accuracy.	In progress	General Manager, Emerging Health Technology & Innovation – Data & Digital Te Whatu Ora Health NZ
PIA13	Limited retention of source data	No source data will be retained in the tool once the calculation is made.	Implemented	General Manager, Emerging Health Technology & Innovation – Data & Digital Te Whatu Ora Health NZ
PIA14	Separation of source data	The NMDS and NzePS data subsets extracted and then refreshed monthly from their parent databases to update the	Implemented	General Manager, Emerging Health Technology &

Control Reference Number	Control Name	Control Description	Status	Control Owner
		scores will be held separately within HNZ's AWS secure systems		Innovation – Data & Digital Te Whatu Ora Health NZ

Appendix One – Precision Driven Health Evaluation of Version 1 of the Risk Score for Call Prioritisation

Evaluating the Risk Score for Call Prioritisation (National V1)

COVID–19 risk of hospitalisation model v1

Document version 1.0

Wednesday 4 May, 2022

Notes

This document contains early results and is only intended for initial review of the evaluation process.

All results are based on assumptions made with respect to data definitions.

All hospitalisations are included with no specialty exclusions.

Data for deaths in the community was neither identified nor available for this evaluation.

Fully vaccinated is defined as "had two doses of vaccine". We note that there is at least one single dose vaccine (Janssen). Single dose vaccines have not been accounted for (therefore one dose = not fully vaccinated). We will require a more complete definition of "fully vaccinated" to account for these cases.[5]

Evaluation summary

The evaluated model was originally developed using Delta data, and validating against this cohort is a useful review of the model's performance against the outcomes it was trained over.

The model is being applied to Omicron cases through the Simplified Risk Score for Call Prioritisation. Evaluating the model for the Omicron cohort tells us about the model performance in the context where it is being used.

This evaluation is part of a roadmap of development and improvement, each time we develop a model, that model is evaluated and that evaluation informs the next iteration.

Data quality

The cohort for the Omicron analysis is defined as those with a positive covid diagnosis between 23 January to 14 February 2022 inclusive. The plot below shows count of unique hospital events by week/DHB in the hospitalisations data. The number of unique hospital events found drops off considerably after February 21. Given lag to hospitalisation, we assume some hospitalisations are missing in the data for this cohort.

---

[5] The names of vaccines found in the CPIR data set (first dose) are included in the Appendix
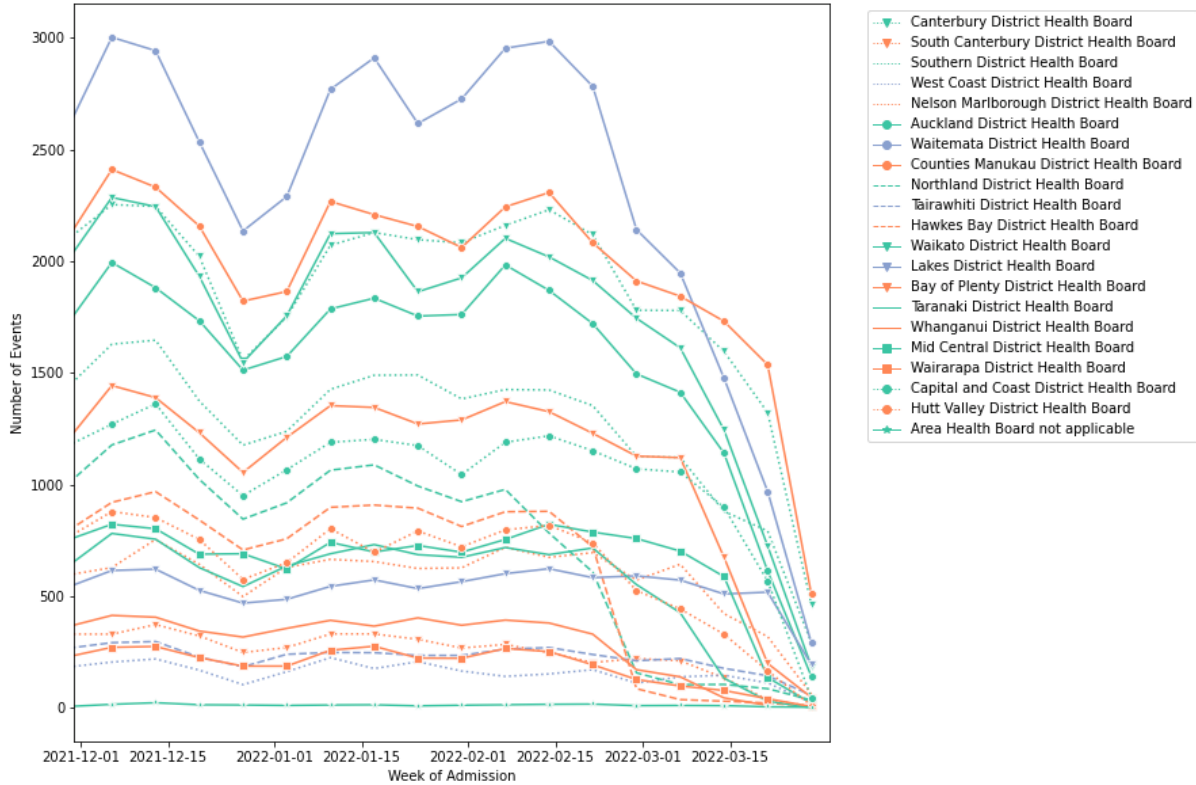
The Omicron cohort (23 January - 14 February 2022)

Calling by risk score vs calling by random order

When used for call prioritisation for cases where little information is known about the patient, the model does a better job at capturing higher risk cases than calling patients in the order in which they are presented (which would be random).

For instance, we can look at the scenario where the tool is used to call the 10% of cases which have the highest scores, and compare rates of hospitalisation for this group against a randomly ordered list. Overall and across all groups, with varying performance, the model identifies more people who require hospitalisation than the baseline prevalence (i.e. the number identified had those people been called at random).

| Group | Total number of COVID positive cases (total number of hospitalised cases) | Calling by risk score ranking | | Calling at random | |
		Number in top 10% by risk score	% of the top risk group hospitalised (Number of hospitalised cases)	Number in randomly selected same number of COVID positive cases (% of the group in overall cases)	Group specific prevalence (Number of hospitalised cases)

Overall	3,151 (109)	316	12.3% (39)	316	3.5% (6)
Māori	324 (19)	35	11.4% (4)	32 (10.3%)	5.9% (2)
Over 60s	272 (29)	198	14.1%(28)	27 (8.6%)	10.7% (3)
Unvaccinated	133 (19)	98	16.3% (16)	13 (4.2%)	14.3% (2)

Highlights

- The model helps us identify higher risk people who are more likely to be hospitalised in the Omicron cohort.

- Overall, the model assigns higher risk to Māori and Pasifika than to other ethnicities. However, there are higher rates of hospitalisation across all risk levels for Māori and Pasifika cases compared to other ethnicities. There are two possible explanations for this. One is that the increase in risk for Māori and Pasifika may not be sufficiently large. Another consideration is that Māori and Pasifika experience higher hospitalisation rates overall regardless of COVID-19 status. It is possible that higher hospitalisation rates reflect this, noting that there is no distinction in the outcome between hospitalisation with and of COVID-19, and no exclusion of specialties that are unlikely to be COVID-19 related (e.g. gynaecology).

- Model performance for the unvaccinated (AUC-ROC 0.698) was just below the commonly used threshold for average performance (AUC-ROC 0.7).

- Model performance for the age group 60 and over was very good.

Evaluation measures

The simplified risk model is evaluated with:

- Rates of hospitalisation across predicted risk levels,

- AUC-ROC

- Classification metrics (e.g. accuracy, sensitivity, specificity).

All results are at the end of the document. The summary below focuses on rates of hospitalisation by risk level and AUC-ROC which both measure how well the simplified risk model ranks cases according to risk of hospitalisation (noting that the model is used for call ranking).[6]

Interpreting rates of hospitalisation across predicted risk levels

If the hospitalisation rate at higher risk levels is greater than the prevalence of hospitalisation overall, then we are more likely to call the 'right' people using the ranking, compared to random calling.

Interpreting AUC-ROC

AUC-ROC is a measure that indicates how good the model is at ranking people based on risk (how likely any two cases are correctly ordered). 0.5 indicates that the model is no better than random at ranking

---

[6] The classification metrics measure the model according to how it would perform as a classifier as opposed to a ranking tool (hospitalised vs not hospitalised, given a risk threshold which for this evaluation was set to 0.15).

the cases in order of risk and 1 indicates a perfect model.[7] For the purposes of this document, AUC-ROC is defined as follows: 0.5-0.54 no better than random; 0.55 - 0.59 a little better than random; 0.6-0.64 better than random; 0.65-0.69 below average; 0.7-0.74 average; 0.75- 0.79 good; 0.8+ very good.

Note that "average" refers to the commonly used threshold for average performance (AUC-ROC 0.7) and that "below average" does not necessarily imply substandard performance in the context.

Variable definitions

Variables for the model are outlined in the table below. Included in the evaluation are people who have tested positive for COVID-19 between 30 November 2021 and 14 February 2022. Excluded from the evaluation are those who have COVID-19 status "Under investigation".

Variable	Definition
Tested positive for COVID-19	Between 30 November 2021 and 14 February 2022.
Outcome (Flag if hospitalised or death)	Include a case in the outcome if hospitalised up to 28 days after and 2 days before a positive test for COVID-19. No specialty exclusions. Deaths in the community are not included in the outcome due to data availability at the time of this evaluation. There is no distinction between hospitalisation with and of COVID-19.
Age	Age in years. Exclude under 18.
Ethnicity	Level 1 ethnic codes for prioritised ethnicity: ● "Māori" (2) ● "Pacific Peoples" (3) ● "Other" (1,4-9)
Vaccination status	Prior to testing positive: ● Not vaccinated - No dose dates present ● One dose - One dose date present, no two or three dose dates present ● Fully vaccinated - Two or three dose dates present.

Overall results

● We note that AUC-ROC indicates how well the model ranks cases across all risk levels (in other words AUC-ROC measures how well the model ranks everyone across both high risk and low risk cases). In practice, the model is used for call prioritisation of the high risk cases. Therefore, while some of the AUC-ROC measures are low, if we look at just the hospitalisation rates at the

---

[7] The statistical literature generally describes an AUC-ROC between 0.6 and 0.7 as 'poor' or 'average' and 0.7 - 0.8 as acceptable (for example, Hosmer & Lemeshow (2013). Applied logistic regression. p.177).

higher risk level, we can see that across all groups the model performs better than random selection.

- As expected, hospitalisation rates increase as risk level increases for both the prospective Delta (30.11.21 - 22.01.22) and Omicron cohorts. This means that when the risk score is used for call prioritisation, the probability of calling the people who would be hospitalised is better than random selection.

- Rates of hospitalisation are affected by the measurement of the denominator (total covid cases), especially for the Omicron cohort. It is expected that actual rates of hospitalisation would be lower across all groups than measured here due to undetected cases.

- The hospitalisation rate for the Omicron cohort is lower than for the Delta cohort (3.5% cf 8.3%).

- The model had average overall performance in prospective evaluation for both cohorts combined (30.11.21-14.02.22) (0.715 AUC-ROC).

- The model had higher performance for the Delta cohort compared to the data the model was built on in January 2022, based on Delta data up until 29.11.22 (0.717 cf 0.7 AUC-ROC).

- Performance for the Omicron cohort is below average (0.683 AUC-ROC) and lower than that for the Delta cohort. This was expected as the model was trained on Delta cases.

Performance for groups

- Hospitalisation rates increased with risk level across all sub-groups (age, ethnicity, vaccination status).

- Hospitalisation rates were higher for people who were not fully vaccinated, and for Māori and Pasifika.

Age

- When broken down by risk level, the age trend is less consistent, likely related to group composition and measurement error in the denominator.

- For the Delta cohort, model performance was good for the 40-59 year age group (AUC-ROC 0.781), and below average for 18-39 years (AUC-ROC 0.667) and for over 60s (AUC-ROC 0.691).

- For the Omicron cohort, model performance was no better than random for the 40-59 year age group (AUC-ROC 0.540), better than random for the 18-39 year age group (0.648), and very good for over 60s (AUC-ROC 0.829). We note that the performance of the middle age bracket may be influenced by people in this cohort being hospitalised for reasons other than COVID-19 (there is no distinction between hospitalisation with and of COVID-19 in the outcome, nor is there exclusion of certain specialties).

Ethnicity

- There are higher rates of hospitalisation for low risk Māori and Pasifika cases compared to other ethnicities. There are two possible explanations for this. One is that the increase in risk for Māori and Pasifika may not be sufficiently large. Another consideration is that Māori and Pasifika experience higher hospitalisation rates overall regardless of COVID-19 status. It is possible that higher hospitalisation rates at lower risk levels reflect this, noting that there is no distinction in the outcome between hospitalisation with and of COVID-19, and no exclusion of specialties that are unlikely to be COVID-19 related (e.g. gynaecology).

- For the Delta cohort, performance was good for Pasifika (AUC-ROC 0.799), average for Māori (AUC-ROC 0.719) and below average for other ethnic groups (AUC-ROC 0.673)

- For the Omicron cohort, performance was better than random for Māori (AUC-ROC 0.609) and below average for Pasifika (AUC-ROC 0.672), and other ethnic groups (AUC-ROC 0.688)

Vaccination status

- For the Delta cohort, performance was a little better than random for the fully vaccinated (AUC-ROC 0.598), below average for not fully unvaccinated8 (AUC-ROC 0.689) and average for the unvaccinated (AUC-ROC 0.710)

- For the Omicron cohort, performance was better than random for the fully vaccinated (AUC-ROC 0.638) and below average for not fully vaccinated (AUC-ROC 0.677) and unvaccinated (AUC-ROC 0.698).

Background

The COVID–19 risk of hospitalisation model v1, also known as the "simplified risk model" is available to support local Hubs with call prioritisation.   This model predicts risk of hospitalisation for people who have tested positive for COVID-19. The model was trained on Northern Region COVID-19 data from the start of the Delta outbreak until late November 2021 (people who tested positive for COVID-19 19/08/2021 - 29/11/2021). The model was developed to be used as a risk stratification tool to determine contact priority at times where a covid positive person has not used the self service portal and there is little clinical information otherwise available for that person.

The simplified risk model has three predictor variables - Age, Ethnicity, and Vaccination Status. These variables were chosen for their relevance to clinical deterioration and their ready availability at the national level. The intention has been to build on this model over time by incorporating comorbidity variables. Due to time and data constraints, this model was not prospectively evaluated before being rolled out.

The Ministry of Health would like to now prospectively evaluate this model to assess its performance. Hospitalisation, Age, Ethnicity and vaccination status data will be used to score people who tested positive for COVID-19 from 30 November 2021 up to 14 February 2022[9].

Evaluation Approach

Evaluation cohorts

Cohort 1 - People aged over 18 who tested positive for COVID-19 between November 30 2021 and January 22, 2022 (Delta Cases)

Cohort 2 - People aged over 18 who tested positive for COVID-19 between January 23, 2022 (Omicron and Delta) and 14 February, 2022

Cohorts will be further stratified by age, ethnicity and vaccination status.

The Risk Score was available from 11 March 2022

Outcome measures

Primary outcome measure - hospitalisation defined as hospitalisation up to 28 days after a positive test result and 48 hours before a positive test result.

Secondary outcome measures - None proposed

---

[8] 'Not fully vaccinated' combines those who are unvaccinated and those who have had one dose.

[9] We understand NMDS hospitalisation data is currently available up until the end of February 2022. This data covers discharges only. The cohort of people who tested positive for COVID-19 after 14 February has had insufficient time to have been both hospitalised and discharged by 28 February and is therefore excluded.

Evaluation questions

What is the observed rate of hospitalisation within groups identified by the Risk Score as being at low (risk score <0.1), medium (risk score >=0.1 and risk score <0.2) or high risk (risk score >=0.2) of hospitalisation?

How does this vary by age, ethnicity and vaccination status? (depending on sample sizes available in the data)

How does this vary between the Delta and Omicron variants?

How does this vary over time (alongside total reported cases)?

How does this vary by DHB region? (if applicable) This can be reviewed alongside qualitative feedback on where and how the Risk Score is being used.

Evaluation metrics

Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and F1 Score (a balanced measure of precision and recall).

Limitations

Outcomes are likely to be affected by the total volume of COVID-19 cases and strain on the system at a point in time. This has varied during the course of the Omicron outbreak and is not accounted for in this approach although we propose to review the risk score over time.

A range of local and national risk scores are being used by local Hubs. 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.

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.

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 evaluation cohort towards higher risk people.

Testing and hospitalisation reflect access to healthcare and don't provide a complete picture of need for healthcare. This evaluation approach may miss individuals who needed/continue to need greater care.

The window for hospitalisation has been defined as 28 days after a positive test. We note that we plan to include people who tested positive up to two weeks before the expected maximum date of the NMDS data. We may not have the full hospitalisation window for some patients who tested positive between 31 January 2022 and 14 February 2022 (due to data processing lags).

NMDS hospitalisation data relates to discharges. Any person who is included in our evaluation data who was discharged after 28 February may not be flagged as hospitalised as their hospitalisation may not be visible (due to data processing lags).

Deaths in the community will not be included due to data availability at the time of the evaluation.

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.

Breakdown of the Omicron mixed cohort

By age

Hospitalisation (Number of cases)	Age	Percentage in the hospitalisation group
Not hospitalised (3,042)	18-39yr	63.9%
	40-59yr	28.1%
	60yr+	8.0%
Hospitalised (109)	18-39yr	48.6%
	40-59yr	24.8%
	60yr+	26.6%

By ethnicity

Hospitalisation (Number of cases)	Ethnicity	Percentage in the hospitalisation group
Not hospitalised (3,042)	Māori	10.0%
	Pasifika	38.6%
	Other	51.3%
Hospitalised (109)	Māori	17.4%
	Pasifika	49.5%
	Other	33.0%

By vaccination status

Hospitalisation (Number of cases)	Vaccination Status	Percentage in the hospitalisation group
Not hospitalised (3,042)	Fully vaccinated	95.2%
	One dose	1.1%

	Unvaccinated	3.7%
	Fully vaccinated	79.8%
Hospitalised (109)	One dose	2.8%
	Unvaccinated	17.4%

Detailed Early Evaluation Results

Risk score distribution
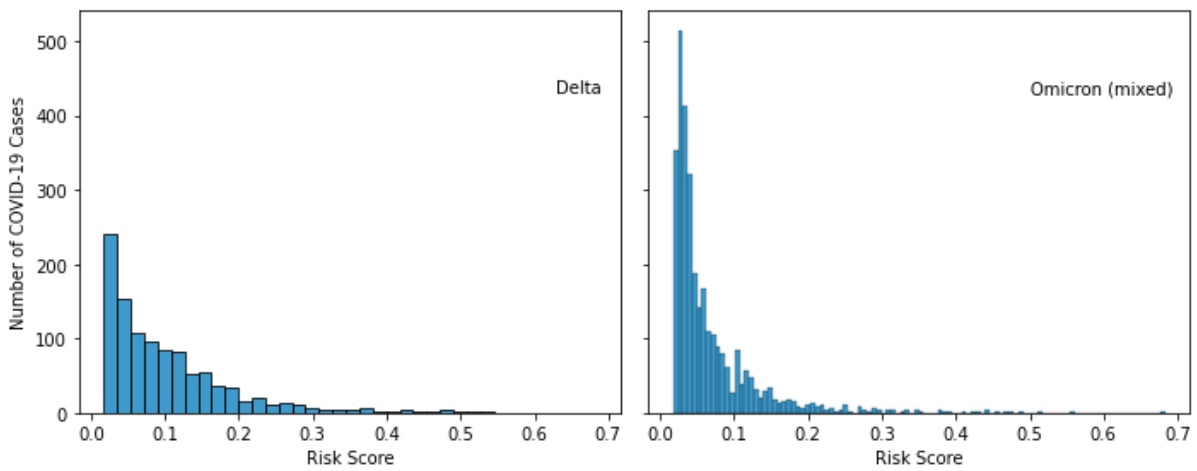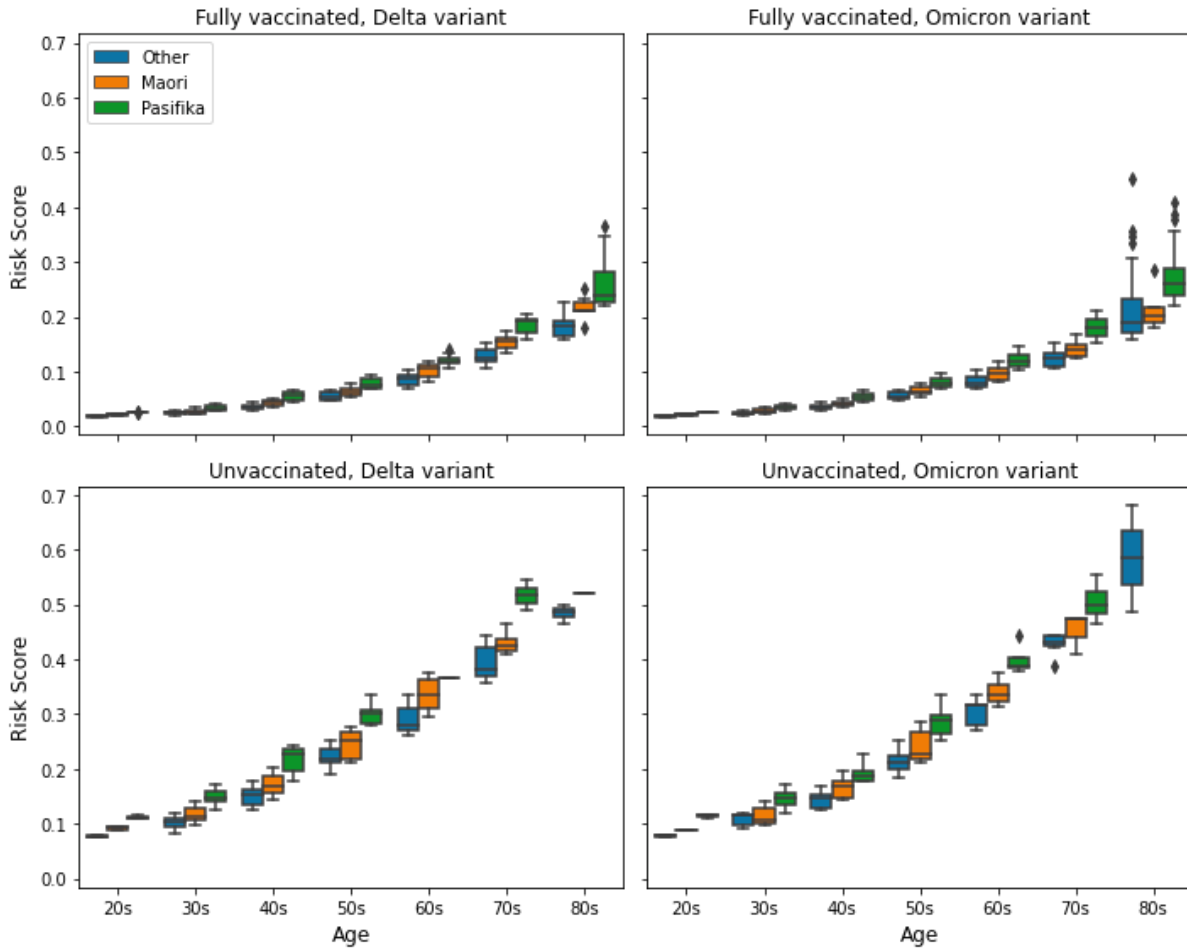
Risk levels are defined as:

Risk score	Risk level
<0.1	LOW
>= 0.1 and < 0.2	MEDIUM
>= 0.2	HIGH

Overall rates

Risk Level	% Hospitalised (Sample size)	
	Delta	Omicron (mixed)
OVERALL	8.3% (1,056)	3.5% (3,151)
LOW	4.0% (645)	2.3% (2,571)
MEDIUM	10.7% (298)	4.7% (448)
HIGH	26.6% (113)	21.2% (132)

Interpretation

Used as a ranking tool, the risk score means we would be more likely to call people who would be later hospitalised than if we called people at random.

For the Omicron cases, had we called the 132 cases in the highest risk group, we would have reached ~28 people (21.2%) who were eventually hospitalised.

By contrast, had we called 132 people at random, we would have reached ~5 people (based on prevalence of 3.5%) who were eventually hospitalised. This is 18% of the number successfully identified above using the risk score.

By age

Note that at some risk levels across both cohorts, the hospitalisation rate for the 18-39 year age group is higher than for older age groups. This may be due to missing cases from the denominator where people with COVID-19 have either not been tested, or not reported a test. Small sample size is also a factor. Group composition within each age band, with respect to vaccination status and ethnicity, may also be a contributing factor.

Cohort	Risk Level	% Hospitalised (Sample size)		
		18 - 39yr	**40 - 59yr**	**60yr+**
Delta	OVERALL	7.1% (619)	7.7% (324)	16.8% (113)
	LOW	4.8% (441)	2.5% (204)	NA[10]
	MEDIUM	12.4% (169)	7.4% (54)	9.3% (75)
	HIGH	22.2% (9)	24.2% (66)	31.6% (38)
Omicron (mixed)	OVERALL	2.6% (1,997)	3.1% (882)	10.7% (272)
	LOW	2.3% (1,926)	2.5% (645)	NA
	MEDIUM	11.6% (69)	3.0% (199)	3.9% (180)
	HIGH	50.0% (2)	13.2% (38)	23.9% (92)

By ethnicity

Cohort	Risk Level	% Hospitalised (Sample size)		
		Pasifika	**Maori**	**Other**
Delta	OVERALL	10.8% (157)	10.2% (335)	6.6% (564)
	LOW	2.4% (85)	5.1% (137)	4.0% (423)
	MEDIUM	12.8% (47)	9.0% (144)	12.2% (107)

---

[10] In the model calculation, this age group cannot be assigned a low risk score.

	HIGH	36.0% (25)	25.9% (54)	20.6% (34)
	OVERALL	4.4% (1,229)	5.9% (324)	2.2% (1,598)
	LOW	3.1% (911)	4.6% (262)	1.4% (1,398)
Omicron (mixed)	MEDIUM	4.5% (245)	8.3% (48)	3.9% (155)
	HIGH	20.6% (73)	21.4% (14)	22.2% (45)

By vaccination status

| Cohort	Risk Level	% Hospitalised (Sample size)	
		Fully vaccinated	**Not fully vaccinated**
	OVERALL	4.9% (673)	14.4% (383)
	LOW	3.6% (553)	6.5% (92)
Delta	MEDIUM	7.8% (102)	12.2% (196)
	HIGH	27.8% (18)	26.3% (95)
	OVERALL	2.9% (2,983)	13.1% (168)
	LOW	2.3% (2,536)	5.7% (35)
Omicron (mixed)	MEDIUM	3.5% (372)	10.5% (76)
	HIGH	21.3% (75)	21.1% (57)

By DHB

Because the sample of data available is limited to the first two weeks of the Omicron outbreak, the majority of cases were still in the Auckland region. A by DHB analysis would not be meaningful in the context of this data.

By risk groups

LOW RISK GROUP

MEDIUM RISK GROUP
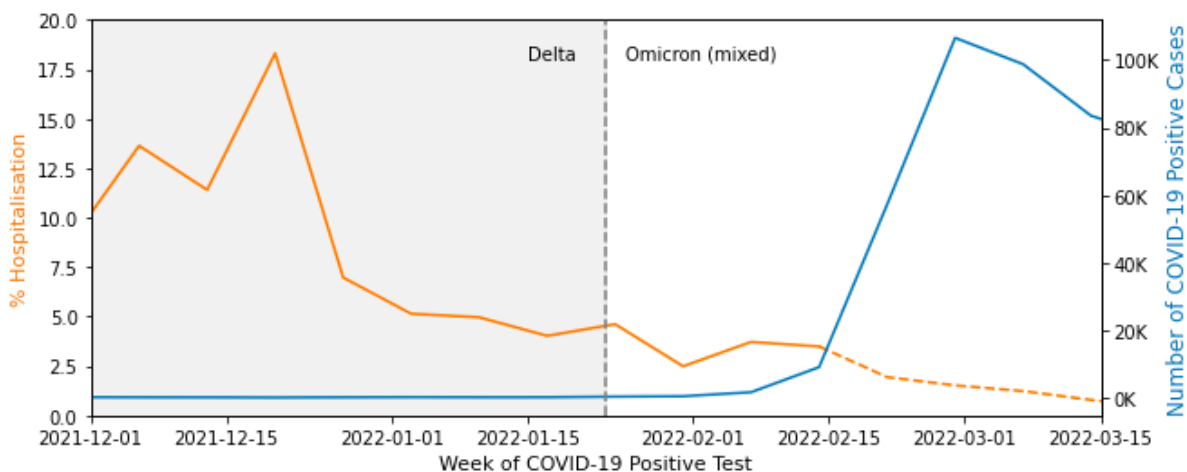
HIGH RISK GROUP
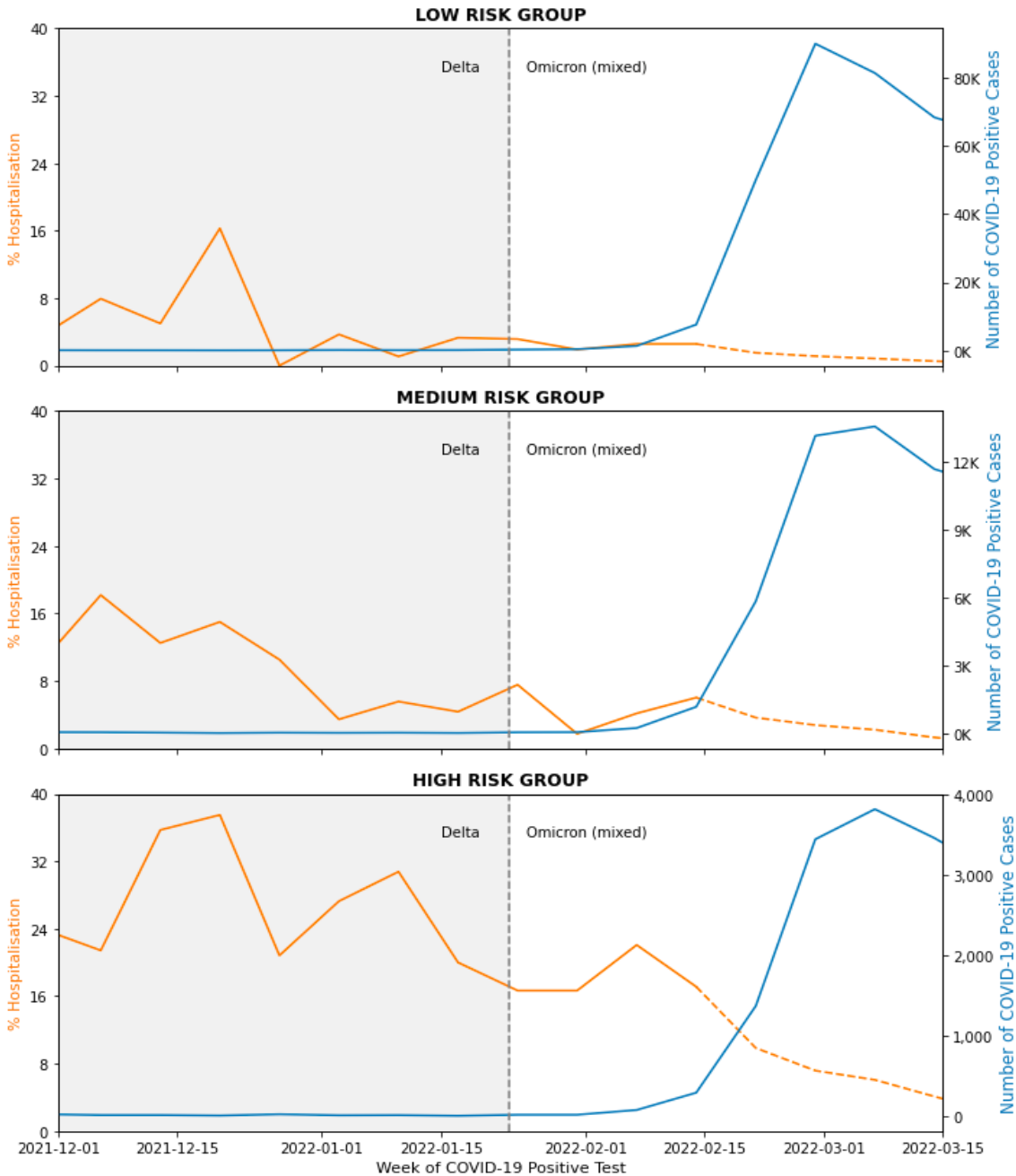
Metrics

Classification metrics (Accuracy, Sensitivity, Specificity, Precision, Negative Predictive Value, F1 Score and Balanced Accuracy) are based on a risk score threshold of 0.15, above which a person is predicted to be hospitalised. The choice of risk score threshold will affect the classification metrics. As the threshold changes, some metrics will improve, others will worsen. The choice of threshold therefore depends on the use case. For the call prioritisation use case, we wish to minimise the number of cases where someone is classified as not hospitalised when they were, in fact, hospitalised (false negative). Therefore, we have tuned this threshold to maximise Negative Predictive Value.

An explanation of the classification metrics is included in the Appendix. An explanation of AUC-ROC is in the Evaluation Measures section.

Overall

	Overall	Delta	Omicron (mixed)
Sample Size	4,207	1,056	3,151
Prevalence	4.68%	8.33%	3.46%
AUC-ROC	0.715	0.717	0.683
Accuracy	0.879	0.794	0.908
Sensitivity (Recall)	0.411	0.511	0.330
Specificity	0.902	0.819	0.929
Precision	0.171	0.205	0.142
Neg Pred Value	0.969	0.949	0.975
F1 score	0.242	0.292	0.199
Balanced Accuracy	0.657	0.665	0.629

By age

	Overall	Delta			Omicron (mixed)		
		18-39yr	40-59yr	60yr+	18-39yr	40-59yr	60yr+
Sample Size	4,207	619	324	113	1,997	882	272
Prevalence	4.68%	7.11%	7.72%	16.81%	2.65%	3.06%	10.66%
AUC-ROC	0.715	0.667	0.781	0.691	0.648	0.54	0.829
Accuracy	0.879	0.859	0.778	0.478	0.962	0.931	0.438
Sensitivity (Recall)	0.411	0.295	0.720	0.737	0.057	0.185	0.966
Specificity	0.902	0.903	0.783	0.426	0.987	0.954	0.374
Precision	0.171	0.188	0.217	0.206	0.103	0.114	0.156
Neg Pred Value	0.969	0.944	0.971	0.889	0.975	0.974	0.989
F1 score	0.242	0.230	0.333	0.322	0.073	0.141	0.268
Balanced Accuracy	0.657	0.599	0.751	0.581	0.522	0.570	0.670

By ethnicity

	Overall	Delta			Omicron (mixed)		
		Other	Māori	Pasifika	Other	Māori	Pasifika
Sample Size	4,207	564	335	157	1,598	324	1,229
Prevalence	4.68%	6.56%	10.15%	10.83%	2.25%	5.86%	4.39%
AUC-ROC	0.715	0.673	0.719	0.799	0.688	0.609	0.672
Accuracy	0.879	0.851	0.710	0.764	0.942	0.877	0.872
Sensitivity (Recall)	0.411	0.297	0.676	0.647	0.417	0.158	0.333
Specificity	0.902	0.89	0.714	0.779	0.954	0.921	0.897
Precision	0.171	0.159	0.211	0.262	0.172	0.111	0.129
Neg Pred Value	0.969	0.947	0.951	0.948	0.986	0.946	0.967
F1 score	0.242	0.208	0.322	0.373	0.244	0.130	0.187
Balanced Accuracy	0.657	0.594	0.695	0.713	0.685	0.540	0.615

By vaccination status[11]

	Overall	Delta		Omicron (mixed)	
		Fully vaccinated	Not fully vaccinated	Fully vaccinated	Not fully vaccinated
Sample Size	4,207	673	383	2,983	168
Prevalence	4.68%	4.90%	14.36%	2.92%	13.10%
AUC-ROC	0.715	0.598	0.689	0.638	0.677
Accuracy	0.879	0.900	0.606	0.931	0.500
Sensitivity (Recall)	0.411	0.212	0.691	0.253	0.636
Specificity	0.902	0.936	0.591	0.951	0.479
Precision	0.171	0.146	0.221	0.135	0.156
Neg Pred Value	0.969	0.958	0.919	0.977	0.897

[11] The first table compares those who are fully vaccinated and those who are not fully vaccinated (unvaccinated and one dose). The second table compares those who are fully vaccinated and those who are unvaccinated.

F1 score	0.242	0.173	0.335	0.176	0.250
Balanced Accuracy	0.657	0.574	0.641	0.602	0.558

	Overall	Delta		Omicron (mixed)	
		Fully vaccinated	Unvaccinated	Fully vaccinated	Unvaccinated
Sample Size	4,207	673	276	2,983	133
Prevalence	4.68%	4.90%	15.58%	2.92%	14.29%
AUC-ROC	0.715	0.598	0.71	0.638	0.698
Accuracy	0.879	0.900	0.547	0.931	0.414
Sensitivity (Recall)	0.411	0.212	0.814	0.253	0.737
Specificity	0.902	0.936	0.498	0.951	0.36
Precision	0.171	0.146	0.23	0.135	0.161
Neg Pred Value	0.969	0.958	0.935	0.977	0.891
F1 score	0.242	0.173	0.359	0.176	0.264
Balanced Accuracy	0.657	0.574	0.656	0.602	0.548

References

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (Vol. 398). John Wiley & Sons.

Appendix

Risk Model Performance Metrics

These metrics are based on classifying people with estimated risk over the threshold as 'at risk' and those with estimated risk beneath the threshold as 'not at risk'. In practice, there are likely to be multiple risk groups used for specific use cases.

Sensitivity/Recall: What proportion of people with hospitalisation or mortality recorded are classified as at risk?

Specificity: What proportion of people without hospitalisation or mortality recorded are classified as not at risk?

PPV/Precision: Of the people classified as being at risk, what proportion did have hospitalisation or mortality recorded ?

NPV: Of the people classified as being not at risk, what proportion did not have hospitalisation or mortality recorded ?

F1 Score: Harmonic mean of precision and recall (balanced metric to describe precision and recall)

First dose vaccinations in CPIR

Vaccine name (as it appears in CPIR)	Count of first dose
Pfizer BioNTech COVID-19	4,043,725
Paediatric Pfizer	260,427
AstraZeneca	16,309
Moderna	3,555
Covishield	2,677
CoronaVac	2,221
Sinopharm	1,859
Novavax	1,635
Janssen	1,417
Sputnik V	533
Sinopharm Inactivated (Vero Cells)	230
Covaxin	160
ZIFIVAX / ZF2001 / RBD-Dimer	48
Covidecia / Ad5-nCOV	38
KCONVAC / SARS-CoV-2 Vaccine (Vero Cells)	25
EpiVacCorona	21
Sputnik Light	18
COVID-19 Inactivated Vaccine/COVIran Barekat	10
MVC-COV1901	4
COVAX-19/SpikoGen	2
Abdala / CIGB-66	2
Pfizer BioNTech 19	1
Pfizer	1

FAKHRAVAC(MIVAC)	1
TAK-919 (Moderna formulation)	1
KoviVac	1
Sinopharm Inactivated	1
Coronavac	1
EpiVacCorona - N	1

Comparing the performance of V2 vs. V1

- AUROC provides a measure of how good the model is at ranking people by risk. This increases from 0.64 for V1 to 0.79 for V2
- Classification metrics are sensitive to score threshold. **For this comparison**, the risk score at the 90th percentile (top 10% of scores) for each model is used as the threshold for calculating metrics (V1 → V2)*
  - **Sensitivity** (true positive rate): **0.297 → 0.469**
  - **Specificity** (true negative rate): **0.903 → 0.909**
  - **Precision** (correct positive predictions): **0.067 → 0.108**
  - **Negative Predictive Value** (correct negative predictions): **0.982 → 0.986**

*Note this is a different approach than used for the draft V1 evaluation which was based on a fixed risk score threshold of 0.15 risk score.

Comparing the performance of V2 vs. V1

Classification threshold set at 90th percentile for each model on each train split

	V2 train data	V2 test data	Retrained V1	Original V1	Blackbox model
Sample Size	191683±0.4	47921±0.4	47921±0.4	47921±0.4	47921±0.4
Prevalence	0.023±0.000	0.023±0.000	0.023±0.000	0.023±0.000	0.023±0.000
AUROC	0.790±0.002	0.789±0.008	0.652±0.008	0.644±0.009	0.785±0.008
Accuracy	0.899±0.000	0.898±0.001	0.891±0.002	0.889±0.002	0.898±0.001
Sensitivity (Recall)	0.468±0.004	0.469±0.015	0.303±0.010	0.297±0.014	0.457±0.014
Specificity	0.909±0.000	0.909±0.001	0.905±0.002	0.903±0.002	0.909±0.001
Precision	0.108±0.001	0.108±0.004	0.070±0.002	0.067±0.003	0.106±0.003
Neg Pred Value	0.986±0.000	0.986±0.000	0.982±0.000	0.982±0.000	0.986±0.000
F1 score	0.175±0.001	0.176±0.006	0.113±0.004	0.110±0.005	0.172±0.005
Balanced Accuracy	0.688±0.002	0.689±0.008	0.604±0.005	0.600±0.007	0.683±0.007

Performance for population subgroups

- Performance of the model (AUROC) is **good** for all population subgroups*, 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 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 experience 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 (men, for instance, do not use maternity services).

*Performance is quantified in the technical model documentation. AUROC measures how well the model ranks all cases.

Should we complete the Algorithm Charter

The algorithm charter is recommended to be used when there is likely to be a high likelihood many people will suffer an unintended adverse impact. In the context of developing this, it has to be considered that without a prioritisation model, many people will miss out on care who would benefit.

It has been scored in a low probability/low impact space due to:

- Acts as an input to an assessment only
- Training has been provided for this to users to help them understand how it has been created and the use case for it.
- It is a calculated score similar to use cases that are prevalent in healthcare already in many different clinical settings.
- It has moved to the right as the score has improved, the likelihood of use will increase, and therefore the likelihood of harm

Risk matrix

Likelihood

Probable Likely to occur often during standard operations			
Occasional Likely to occur some time during standard operations			
Improbable Unlikely but possible to occur during standard operations			

Impact — **Low** The impact of these decisions is isolated and/or their severity is not serious. — **Moderate** The impact of these decisions reaches a moderate amount of people and/or their severity is moderate. — **High** The impact of these decisions is widespread and/or their severity is serious.

Risk rating

Low The Algorithm Charter could be applied.	**Moderate** The Algorithm Charter should be applied.	**High** The Algorithm Charter must be applied.

Controls associated with the commitments

Commitments	Ways to demonstrate the commitments	How are we addressing these?
Transparency - maintain transparency by clearly explaining how decisions are informed by algorithms.	Plain English documentation of the algorithm	Privacy Impact Assessment Documentation on the algorithm hub website Documentation on the Ministry website
	Making information about the data and processes available (unless a lawful restriction prevents this)	Privacy Impact Assessment Official Information Act Requests
	Publishing information about how data are collected, secured and stored.	Privacy Impact Assessment
Partnership - Deliver clear public benefit through Treaty commitments by:	embedding a Te Ao Māori perspective in the development and use of algorithms consistent with the principles of the Treaty of Waitangi.	Algorithm Information Report Engagement with Maori Health Directorate MoH Planned engagement with Iwi Data Leaders and Maori Health Authority
People - Focus on people by:	identifying and actively engaging with people, communities and groups who have an interest in algorithms, and consulting with those impacted by their use.	Publishing to Algorithm Hub Publishing supporting information on MoH website
Data - Make sure data is fit for purpose by:	understanding its limitations	Independent Technical Review
	identifying and managing bias.	Independent Technical Review
Privacy, ethics, and human rights - Ensure that privacy, ethics and human rights are safeguarded by:	regularly peer reviewing algorithms to assess for unintended consequences and act on this information.	Algorithm Information Report
Human oversight - Retain human oversight by:	nominating a point of contact for public inquiries about algorithms	Governance Group/Publishing to Algorithm Hub
	providing a channel for challenging or appealing of decisions informed by algorithms	Governance Group
	clearly explaining the role of humans in decisions informed by algorithms.	Governance Group