ABSTRACT

Objectives Clinical risk calculators (CRCs), such as NZRisk, are used daily by clinicians to guide clinical decisions and explain individual risk to patients. The utility and robustness of these tools depends on the methods used to create the underlying mathematical model, as well as the stability of that model in relation to changing clinical practice and patient populations over time. The later should be checked by temporal validation using external data. Few if any of the clinical prediction models in current clinical use have published temporal validation. Here, we use a large external dataset to temporally validate NZRisk; a perioperative risk prediction model used in the New Zealand population.

Methods A sample of 1,976,362 adult non-cardiac surgical procedures collected over 15 years from the New Zealand Ministry of Health National Minimum Dataset, was used to temporally validate NZRisk. We divided the dataset into 15 single year cohorts and compared 13 of these to our NZRisk model (2 years used for the model building were excluded). We compared the area under the curve (AUC) value, calibration slope and intercept for each single year cohort, to the same values produced by the data used to create NZRisk, by fitting a random effects meta-regression with each year cohort acting as a separate study point. In addition, we used two-sided t-tests to compare each measure across the cohorts.

Results The AUC values for the 30-day NZRisk model applied to our single year cohorts ranged from 0.918 to 0.940 (NZRisk AUC was 0.921). There were eight statistically different AUC values for the following years 2007–2009, 2016 and 2018–2021. The intercept values ranged from −0.004 to 0.007 and 7 years had statistically significant different intercepts during leave-one-out t-tests; 2007–2010, 2012, 2018 and 2021. The slope values ranged from 0.72 to 1.12 and 6 years had statistically significant different slopes during leave-one-out t-tests; 2010, 2011, 2017, 2018 and 2019–2021. The random effects meta-regression upheld our results related to AUC (0.54 (95% CI 0.40 to 0.99), F 67.57 (95% CI 40.67 to 88.50), Cochran’s Q<0.001) and slope (r 0.14 (95% CI 0.01 to 0.23), r 98.61 (95% CI 97.31 to 99.50), Cochran’s Q<0.001) between year difference.

Conclusion The NZRisk model shows differences in AUC and slope but not intercept values over time. The biggest differences were in the calibration slope. The models maintained excellent discrimination over time as shown by the AUC values. These findings suggest we update our model in the next 5 years. To our knowledge, this is the first temporal validation of a CRC in current use.

BACKGROUND

Clinical risk calculators (CRCs) are commonly used to augment clinical decisions. Their role in clinical decision making has increased over the last decade.1 Most medical and surgical specialties now have one or more CRCs, using them to predict the occurrence of medical events or procedure related risk. One of the main aims of CRCs is estimation of individualised risk.2 As such, clinical prediction models have an important clinical impact, guiding clinicians and governing patient choices. Guidelines to ensure rigour in CRC creation have been published.3

CRCs are required to be internally and externally valid to ensure generalisability and clinical usefulness.15 Further, models should also be temporarily validated to ensure that changes in population demographics and clinical practice over time, do not affect model stability and utility. Most models undergo internal validation prior to publication and popularisation of their use in clinical practice. In contrast external validation occurs less frequently and can sometimes demonstrate severe degradation of model
performance in populations other than the derivation population. Often models that have been externally validated, have been reported to perform less favourably than in the original study.\(^3\) It is difficult to ascertain whether this poorer performance is due to model instability in small datasets, selection bias due to different casemix, different geographical populations, and healthcare systems, or changing populations and medical practices over time, or a combination of these factors.\(^5\)

The importance of temporally validating a prediction model relates to the fact that patient demographics, medical characteristics, surgical techniques and surgical risks, all evolve over time.\(^6\) Indeed, we have not been able to identify any CRC in perioperative medicine that has been formally temporally validated and published.

NZRisk is a non-cardiac surgical risk tool internally and externally validated in a 2-year, full coverage national dataset.\(^7\) It is derived from a New Zealand (NZ) non-cardiac surgical casemix from a relatively stable national population. NZRisk has been developed to be a simple and easy to use risk calculator based on the NZ population. It has been compiled using data from over 270 000 patients aged 18 or over. It includes eight risk factors that can be entered below to give a 30-day, 1-year and 2-year estimate of mortality. Each risk factor is associated with mortality and improves the performance of the calculator. There are few similar large, national, databases. This presents an ideal data environment in which to assess the temporal stability of NZRisk, with minimum influence of selection bias, differing casemix, differing geography and healthcare systems and differing incidence of outcomes.

The aim of this study is to determine the temporal stability of the predictive performance of NZRisk. The secondary aim is to determine the appropriate time period for model updating. Our hypothesis is that the population demographics (eight variables) are not going to significantly change over time and likely surgical care will not have changed sufficiently over the 15 years to influence stability of NZRisk. This study has been conducted and reported according to the TRIPOD statement for reporting multivariable prediction studies.\(^3\)

To our knowledge, this is the first example of temporal validation using a dataset, which includes near complete population capture over a 15-year period.

**METHODS**

**Patient and public involvement**

Patients and the public were not involved in the design, conduct or reporting of this research.

**Data**

Data from the NZ Ministry of Health National Minimum Dataset (NMDS)\(^8\) were obtained for participants having surgery between 1 January 2007 and 31 December 31 2021 inclusive. Mortality data were obtained from the death registry at the Ministry of Health and merged using a unique National Health Identifier code. The current version of the NZ Ministry of Health NMDS was introduced in 1999 and has continuously recorded data since then(dataset).\(^8\)

**Participants**

Analysis was performed for all adult patients aged over 18 years who underwent a non-cardiac surgical procedure in public or private hospitals in NZ. Participants were subjects in the NMDS who had been assigned a surgical procedure code and an anaesthetic code. Participants who only had local anaesthesia as an anaesthetic code were excluded. Patients who had multiple procedures during an episode of care had the most complex procedure analysed. Demographic data comparing the cohorts, reflective of covariates included in NZRisk are shown in online supplemental table 1.

**Study outcome**

The primary outcome was 30-day perioperative mortality as defined by a date of death recorded in the NMDS within the time period following the surgical procedure. Outcome assessment was blinded by the routine reporting of the outcomes and variables to the NMDS by outcome assessors unrelated to this study. Secondary outcomes were 1-year and 2-year postoperative mortality as these were secondary outcomes in the original NZRisk validation.

**Variables**

The exposure was a surgical procedure as listed in the NMDS. All quantitative variables were handled as categorical variables, except age which was handled as a continuous variable.

**Missing data and bias**

Date of death is a mandatory report in NZ. We assume all mortality outcomes have been reported and there are no missing outcome data. The NMDS is a national registry of all inpatient events reported to the NZ Ministry of Health. It captures 99% of all surgical procedures performed in NZ. The nearly full coverage within this national dataset minimises selection bias. Our data were missing ASA for 1 612 036 patients (22.7% of eligible patients). During the derivation of NZRisk, it was assumed that patients without ASA had data missing at random, after investigating our data we have come to the same conclusion and thus conducted a complete-case analysis.

**Sample size**

The sample size of 1 976 362 was derived from all adult patients having non-cardiac surgical procedures in NZ in the 15-year dataset. A 15-year sampling period was chosen as the NMDS had nearly full coverage for surgical procedures in NZ over this time. In addition, NZRisk was internally validated using a 2-year dataset from 2013 to 2014. The 15-year dataset allowed external validation in eight sequential 1-year cohorts. Some authors have determined that external validation requires 10 endpoints per covariate, while others have recommended more than
200 endpoints in the dataset.\textsuperscript{9} The smallest of the 1-year cohorts has N=104 569 and mean 1-month mortality of 0.88\% giving a total of 920 endpoints, providing sufficient events for estimates of model predictive performance.

**Statistical analysis**

**Comparison of model performance metrics**

In order to compare the discrimination of the NZRisk model over time, we calculated the NZRISK score for each patient with the required data using the 30-day model. Subsequently, the pROC package,\textsuperscript{10} V.1.18.0, was used to calculate the AUC value for the patients in each single year cohort. The data were broken down into single year cohorts for each year between 2007 and 2021, excluding 2013 and 2014 as those years data were used to build the original model. The ‘set mean’ AUC values are generated by grouping all single year cohorts (excluding 2013/2014) along with the NZRisk values from the original paper and then calculating the average AUC value of that set. Two-sided t-tests were conducted on each cohorts AUC value against the rest of the values as a set, after removing the year being tested, with p values adjusted using the Bonferroni correction.

A similar approach was taken for the model calibration slope and intercept statistics, each year was tested against the set of all other values with a two-sided t-test to look for statistically significant differences.

In terms of determining relevant values, a priori, we deemed that since our original model had a c=0.92, within the bounds of Steyerberg’s excellent risk model, we assumed a drift to c<0.90 would be a failure of temporal calibration.\textsuperscript{11} Further, we deemed a calibration slope of 0.8–1.2 acceptable and beyond that NZRisk would need updating. Finally, in terms of intercept, we deemed a bias of ±0.2\% acceptable and beyond that an indication that the model needs updating.

We also conducted a meta-analysis regression considering each year to be a study within the meta-analysis by fitting first a random effects model to each of the AUC, slope and intercept values and second fitting a mixed-effects model where the study year was treated as a moderator variable. When using study year as a moderator, we assigned year to be year—2007 so as to centre the intercept. For the AUC values, we used 100 × AUC for the meta regression to avoid rounding issues with the analysis outputs. These results were also analysed visually with bubble plots to assess trends over time.

All statistical analyses were carried out in R V.4.0.3\textsuperscript{12} using Rstudio V.1.3.109. We used the forestplot package, V.1.10 for Forest plots.\textsuperscript{13} We manipulated data using data.table, V.1.13.0.14 and the tidy verse package, V.1.3.0.15. We conducted the meta-analysis using the metafor package V.3.8–1.\textsuperscript{16}

**RESULTS**

Data included in this modelling are summarised in the STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) diagram (figure 1). The AUC values for the 30-day NZRISK model applied to our single year cohorts ranged from 0.918 to 0.940 (NZRISK AUC was 0.921). During leave-one-out testing, there were eight unadjusted statistically different AUC values between the 1-year test cohorts; 2007–2009, 2016 and 2018–2021 (please see figures 2 and 3 and table 1). The AUC values also generally increased over time. The intercept values for all cohorts had values very close to zero in all years. The calibration slope values ranged from 0.72 to 1.12 and 7 years had unadjusted statistically significant different slopes during leave-one-out t-tests, 2010, 2011, 2017, 2018, 2019–2021. The slope values generally decreased over time, in contrast to the AUC values. Two sided t-tests on each cohorts AUC value against the rest of the values as a set, after removing the year being tested, and are reported in table 1 with p values adjusted using the Bonferroni correction.

The meta-regression random effects model fitted to each of sets of model performance metrics; AUC, slope and intercept. For the AUC set we saw a τ value of 0.54 (95\% CI 0.40 to 0.99), an I² of 67.57 (95\% CI 40.67 to
Figure 2  Forestplots for each single year cohort when the 30-day NZRISK model is applied to the patient data from that cohort. The top graph shows the AUC values by year, with a vertical line to indicate the overall set mean from all the AUC values. The middle and bottom graphs show the slope and intercept values, respectively. The vertical lines represent the ideal slope and intercept of 1 and 0. Each box is the point estimate from a single year cohort with the associated 95% CI in blue on either side. Each plot also contains the NZRISK value from the original model and the set mean for that measurement. AUC, area under the curve.
Figure 3  Bubble plots for each meta-analysis regression with year as a modifier. From top to bottom, the plots represent the meta-analysis regression for the AUC, slope and intercept values, respectively. Each bubble represents the estimate for that statistic (AUC, slope, intercept) at the associated year between 2007 and 2021, excluding 2013/2014. The trend lines represent a meta-regression mixed-effects model fitted to the data with the associated 95% CI around it. The point size represents the weight of each point within the model. AUC, area under the curve.
Table 1  Shows the AUC, slope and intercept values from each single year cohort of patients assessed with the NZRISK algorithm

<table>
<thead>
<tr>
<th>Year</th>
<th>AUC (95% CI)</th>
<th>P value (adjusted)</th>
<th>Slope Value</th>
<th>P value (adjusted)</th>
<th>Intercept P value</th>
<th>P value (adjusted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>0.922 (0.915 to 0.930)</td>
<td>0.015 (0.198)</td>
<td>1.027 (0.993–1.061)</td>
<td>0.011 (0.138)</td>
<td>−0.001 (-0.001–0.000)</td>
<td>0.001 (0.011)</td>
</tr>
<tr>
<td>2008</td>
<td>0.923 (0.916 to 0.930)</td>
<td>0.030 (0.394)</td>
<td>1.043 (1.017–1.070)</td>
<td>0.005 (0.059)</td>
<td>−0.001 (-0.001–0.000)</td>
<td>0.025 (1.000)</td>
</tr>
<tr>
<td>2009</td>
<td>0.918 (0.910 to 0.925)</td>
<td>&lt;0.001 (0.001)</td>
<td>1.008 (0.979–1.036)</td>
<td>0.028 (0.367)</td>
<td>0.000 (-0.001–0.000)</td>
<td>&lt;0.001 (&lt;0.001)</td>
</tr>
<tr>
<td>2010</td>
<td>0.925 (0.918 to 0.931)</td>
<td>0.194 (1.000)</td>
<td>1.041 (1.019–1.063)</td>
<td>0.005 (0.066)</td>
<td>0.001 (-0.001–0.000)</td>
<td>0.019 (0.247)</td>
</tr>
<tr>
<td>2011</td>
<td>0.929 (0.922 to 0.936)</td>
<td>0.190 (1.000)</td>
<td>1.079 (1.038–1.120)</td>
<td>0.001 (0.009)</td>
<td>0.001 (-0.002–0.000)</td>
<td>&lt;0.001 (0.004)</td>
</tr>
<tr>
<td>2012</td>
<td>0.924 (0.917 to 0.932)</td>
<td>0.145 (1.000)</td>
<td>0.946 (0.906–0.986)</td>
<td>0.393 (1.000)</td>
<td>−0.001 (-0.002–0.000)</td>
<td>0.238 (1.000)</td>
</tr>
<tr>
<td>2015</td>
<td>0.925 (0.919 to 0.932)</td>
<td>0.361 (1.000)</td>
<td>0.928 (0.894–0.961)</td>
<td>0.689 (1.000)</td>
<td>−0.001 (-0.002–0.000)</td>
<td>0.399 (1.000)</td>
</tr>
<tr>
<td>2016</td>
<td>0.922 (0.915 to 0.929)</td>
<td>0.014 (0.187)</td>
<td>0.851 (0.825–0.876)</td>
<td>0.145 (1.000)</td>
<td>−0.001 (-0.001–0.000)</td>
<td>0.003 (0.038)</td>
</tr>
<tr>
<td>2017</td>
<td>0.927 (0.919 to 0.935)</td>
<td>0.951 (1.000)</td>
<td>0.790 (0.747–0.832)</td>
<td>0.008 (0.098)</td>
<td>−0.001 (-0.002–0.000)</td>
<td>0.232 (1.000)</td>
</tr>
<tr>
<td>2018</td>
<td>0.940 (0.933 to 0.946)</td>
<td>&lt;0.001 (&lt;0.001)</td>
<td>0.761 (0.713–0.808)</td>
<td>0.002 (0.022)</td>
<td>−0.001 (-0.002–0.000)</td>
<td>&lt;0.001 (0.001)</td>
</tr>
<tr>
<td>2019</td>
<td>0.931 (0.924 to 0.939)</td>
<td>0.018 (0.232)</td>
<td>0.735 (0.699–0.770)</td>
<td>&lt;0.001 (0.006)</td>
<td>−0.001 (-0.002–0.000)</td>
<td>0.031 (0.404)</td>
</tr>
<tr>
<td>2020</td>
<td>0.931 (0.924 to 0.938)</td>
<td>0.033 (0.434)</td>
<td>0.717 (0.686–0.747)</td>
<td>&lt;0.001 (0.002)</td>
<td>−0.001 (-0.002–0.000)</td>
<td>0.992 (1.000)</td>
</tr>
<tr>
<td>2021</td>
<td>0.938 (0.932 to 0.944)</td>
<td>&lt;0.001 (&lt;0.001)</td>
<td>0.718 (0.684–0.751)</td>
<td>&lt;0.001 (0.002)</td>
<td>−0.001 (-0.002–0.000)</td>
<td>0.001 (0.019)</td>
</tr>
</tbody>
</table>

NZRISK was built using data from 2013/2014 hence those years are removed. Each value has also been compared with all others as a set using a two-sided t-test and this table includes the associated adjusted and unadjusted p values from those tests.

AUC, area under the curve.
88.50) and the Cochran’s Q test of heterogeneity, was <0.001. For our mixed-effects model of AUC using study year as a moderator, we obtained a τ value of 0.27 (95% CI 0.00 to 0.66), an I^2 of 36.97 (95% CI 0.00 to 77.35). After accounting for study year as a moderating variable, our test for remaining heterogeneity had a non-significant test statistic of 0.1 and year as a moderating variable had a test statistic of <0.001 with an estimate of 0.10 change in the AUC per year. Our model intercept was 92.03 so based on our model, the AUC is estimated to start around 0.92 and increases by about 0.1% per year.

For the slope the τ value was 0.14 (95% CI 0.01 to 0.23), with an I^2 of 98.61 (95% CI 97.31 to 99.50) and Cochran’s Q test of heterogeneity was <0.001. For our mixed-effects model of slope using study year as a moderator, we obtained a τ value of 0.04 (95% CI 0.03 to 0.07), an I^2 of 85.5 (95% CI 70.26 to 95.27). After accounting for study year as a moderating variable, our test for remaining heterogeneity had a test statistic of <0.001 and year as a moderating variable had a test statistic of <0.001 with an estimated change of −0.027 (or a decrease of about 2.7% per year) in slope per year our model had an intercept of 1.08.

For both the fixed effects and random effects models for the intercept values, we saw Q values which did not reach statistical significance and all the τ^2, τ and I^2 estimates were 0.

DISCUSSION

We have carried out a temporal validation and meta-analysis regression of our NZRisk calculator using a database of 1.3 million cases recorded over a 15-year period spanning either side of the NZRisk calculator creation dataset. AUC (discrimination), intercept (bias) and slope (calibration) values have been calculated for each of the 15 single year cohorts and compared with look for pattern and range of variation of these parameters.

When measuring the AUC values, or model discrimination, eight 1-year cohorts were found to be statistically significant in leave-one-out t-testing. These AUC values, however, were all above 0.9 and remained in the range for a model with excellent performance. The values from earlier years tended to be lower and AUC values generally increased over time, indicating that the model is improving in its discrimination. We hypothesise this may reflect a larger decrease in mortality within lower risk cohorts than higher risk ones, making it easier to discriminate between those at high and low risk of perioperative complications or death.

In trying to explain the differences year-on-year in the calibration slope, we believe that the slope may have some sensitivity to small changes in event rate (year-on-year) in the higher risk deciles included in the NZRisk model. A slope value on its own does not measure calibration, it must be taken alongside the intercept. Across our cohorts, the maximum deviation from the ideal intercept of zero is 0.0014 and the mean deviation is 0.001, suggesting that the intercept is close to zero across the cohorts. These intercept values indicate our ideal slope for good calibration would be 1. Our cohort values are either side of 1 with later year cohorts tending to produce slopes <1. This suggests some level of underpredicting of risk as we get closer to the present day and earlier cohorts tending to slightly overpredict risk, with a slope >1. We believe this reflects the overall long-term trend of mortality risk decreasing within the NZ population. A similar effect over time was seen during the development of NZRisk where an earlier calculator, SORT, was applied to NZ data and was found to underpredict.

Our meta-analysis found statistically significant Q values for the AUC and the slope, but no significant changes in intercept values. We, therefore, reject the null hypothesis that there is no difference between AUC or slope values across years. This is consistent with our finding that multiple AUC and slope values are statistically significantly different from the overall set when assessed using t-tests. Consequently, we went on to assess year of surgery as a possible moderating variable to tease out the clinical meaning of this finding.

For the AUC values, after accounting for study year, the heterogeneity between the years was roughly halved (I^2 reduced from 69.6% to 37.1%) and we saw a statistically significant QM value for our moderator (year) but a non-significant result for QE, our residual heterogeneity. A rule of thumb for the interpretation of an I^2 value is that 25%, 50% and 75% represents low, medium and large heterogeneity, respectively. This suggests that after accounting for year, the level of remaining heterogeneity shifted from medium-high to low-medium. We propose that clinical or patient changes over time are likely to be accounting for the differences in AUC across the years. This is consistent with our theory that AUC changes reflect larger decreases in the mortality of lower risk cohorts than higher risk cohorts in the population, making it easier to discriminate between patients likely or unlikely to die within 30 days. While the AUC results showed a statistically significant increase over time, this is unlikely to be clinically important, as the AUC remains over 0.9, maintaining excellent discrimination performance.

For our meta-analysis of slope values, when using study year as a moderating variable the heterogeneity between years decreased but remained high (I^2 decreased from 98.6% to 85.5%). While study year had a statistically significant effect on between group heterogeneity (QM value 111.2 and p<0.001), there continued to be a significant effect on between group heterogeneity (QE 67.5, p<0.001). This suggests that changes to the underlying surgical population over time do not fully account for the decrease in calibration slope. We postulate that the observed residual heterogeneity may be due one or more of the following: (1) overfitting of the original model, (2) a skewing effect of higher risk cohorts on the slope calculation or (3) effects of modifiers we have not accounted for. We believe a calibration slope outside of the range 0.8–1.2 makes a CRC less clinically useful.
CRCs such as NZRISK will always have a delay from the time of development to the time they are used, as models cannot be calculated on live data. NZRISK was developed using data from 2013 to 2014 and first deployed for clinical use in 2018. The temporal validation described in this paper, provides evidence of changes in model accuracy and calibration that occur over time. As the model maintains excellent discrimination, and only slightly decreased calibration we believe the model is still valid for clinical use in 2022. Speaking with clinicians, the model is primarily used to estimate risk and make decisions which are focused on discrimination between the highest and lowest risk patients, making the increase in AUC values over time a positive finding. Based on this, evidence cases which critically depend on accurate calibration should potentially be revisited between 5 and 10 years. Overall, while some differences amid years included in the temporal validation do exist, the chance of these small statistical differences reflecting clinically important variance is low. We have yet to determine the optimal methods for updating the NZRisk model and plan to work on this in the near future.

The main limitation of this work is the relatively short time span either side of model development that we evaluated. This was related to data availability. A larger sample would have given us more robust confirmation of our results. Further, it is possible that a different statistical approach to comparison across time cohorts within our validation dataset, could have modified our findings. There are, however, no published temporal validation examples we could compare to. Many commonly used CRC’s, such as the P-POSSUM and EUROSORE are still in use many years after their development, without evidence of temporal validation. This maybe because of a lack of availability of appropriate external dataset to temporally validate the original model. The main strength of this work is the coverage and completeness of the national dataset we have used to validate NZRisk. Additionally, the robustness and size of the national dataset has allowed us to use data directly comparable to the data used to construct the original model.

To our knowledge, this study is the first attempt at temporal validation of a perioperative CRC. This would suggest there is wide scope for further work in this area. Future work could consider if these findings can be replicated in subsets of patients, particularly high-risk groups. Overall, mortality rates for surgery are low in our cohort (~0.6% 30-day mortality in NZ), but moderate to high in international studies (2.8%) and among other surgical specialities and patient groups in NZ. These groups may show less stability when measured over time, particularly if surgical practice has changed significantly during study periods. We also hypothesise that changes in these higher-risk groups are likely to have an outsized impact on model statistics such as calibration slope and AUC level. Future work evaluating how model performance changes across time or where significant changes in surgical practice have led to a decrease in mortality, such as introduction of laparoscopic bowel surgery, would offer high value for suggestions of changepoint or updated timelines for CRCs to maintain model performance.

In summary, we have presented the temporal validation of a CRC, NZRisk. We have shown the stability of our model in a large external dataset. In view of the importance that such models have come to have on clinical decision making, we would encourage other groups to consider undertaking temporal validation of models they have developed.

Contributors MS, LB, DC and TL generated the hypotheses underpinning this work. LB carried out the statistical analysis with the help of TL. The original NZ Risk model was developed by DC and LB. All authors contributed to the writing and editing of this manuscript. DC acts as the guarantor and accepts full responsibility for the work and/or the conduct of the study, had access to the data, and controlled the decision to publish.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Ethical approval for this study (16/CEN/200) was given by the New Zealand National Health and Disability Ethics Committees.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. Data used are available through the NZ government record system (NMDS).

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Luke Boyle http://orcid.org/0000-0001-6114-1833
Marta Seretty http://orcid.org/0000-0003-0615-1224

REFERENCES