Statistical models to improve ACC claims approval and registration process

August 2018

Version history

Version	Date	Nature of amendment
0.1	21/08/2018	Final draft

Contents

Executive Summary

ACC is improving how claims are registered, assessed and approved

Each year, ACC teams manually review and process over two million claims. Over 96 per cent of these claims are accepted, and around 90% are accepted without any need for further information or investigation, because they clearly fit within the Accident Compensation Act 2001 (The Act) criteria.

From September 2018, ACC will use a new system to help identify and approve these straightforward claims for injuries that are clearly caused by accidents. Claims that aren't straightforward or require specialist knowledge will be referred to a staff member, who will review and assess these claims. The business requirements for this system are listed in Appendix 1.

Clients will benefit, as many claims will soon be accepted as they are lodged, and, where ACC has their contact details, they will receive a text message notifying them what's happening with their claim. This is a faster outcome than currently, where claim decisions are confirmed via letters in the post. This will free up time for staff to focus on more complex and sensitive claims, which will continue to be reviewed manually, in the same way they are now.

Claims can only be accepted or referred for manual review by an ACC staff member, they cannot be declined by the new system.

The models informing this new system were developed in partnership with Nicholson Consulting, who are experts in statistical modelling.

Current Process

Right now, it can take some time for injured New Zealanders to find out what's happening with their claim, as ACC staff manually process each claim once it is lodged by a provider (e.g. a doctor, physiotherapist or health practitioner).

Clients currently receive updates on what's happening with their claim via letters in the post. This can take several days.

New process

The new system will identify characteristics of a claim that are relevant to whether a claim will be accepted. Straightforward claims - where the information provided shows that an injury was caused by an accident - will be fast-tracked and immediately accepted. Complex or sensitive claims will be referred for review by an ACC staff member, as all claims are now.

The system will notify clients that their claim has been registered, meaning many clients will know what's happening with their claim more quickly than before.

The system won't be used to decline any claims; it identifies and categorises claims based on the information available.

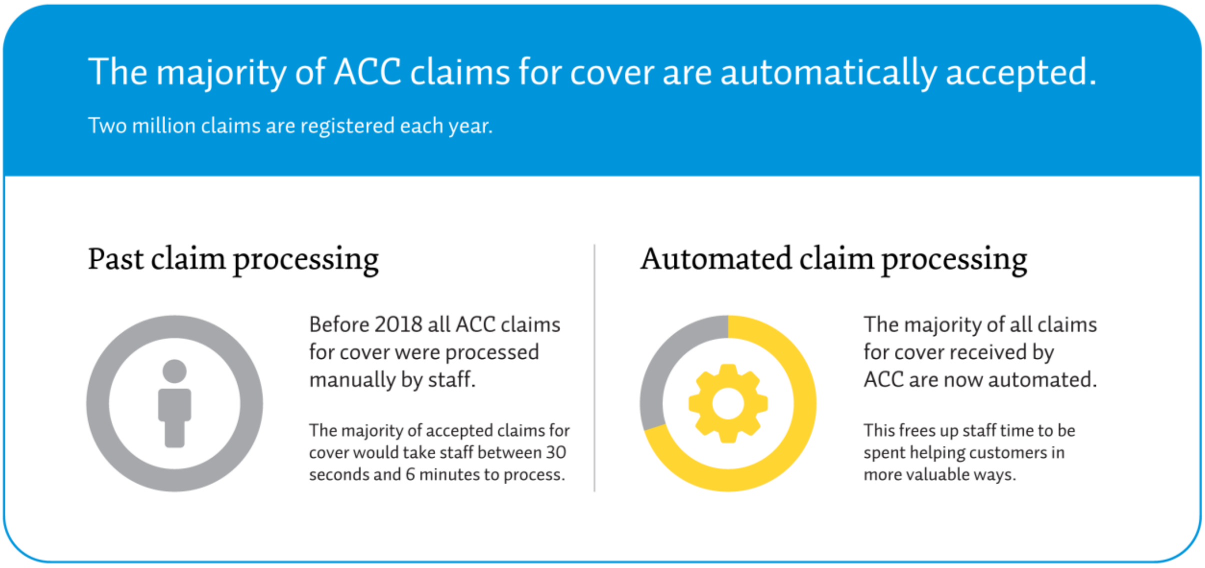
How does it work?

The new system can complete several tasks that are currently done manually, using a combination of business rules and predictive modelling.

For instance, it can take information that's written on a claim form and categorise the claim, defining things like the cause of the injury, if a sport was involved, and what activity was going on just prior to the accident.

The second thing it can do is check for key information or 'flags' on the claim form to identify what type of claim you've made, and whether it's a special claim type such as a dental claim or a claim for treatment injury that needs to be handled by a particular team in ACC.

Lastly, it can look at the information on a claim form and determine whether the claim can be automatically accepted, or whether it needs to be referred to a staff member for manual processing, as all claims are processed now. It cannot decline claims.



Statistical models are used to approve some claims and populate information about claims

The new ACC system uses statistical models and a rules engine to automate much of its current, manual, registration cover process.

This report describes the statistical models involved in this process:

- Cover Decision Service
- Accident Description Service

Cover Decision Service

The Cover Decision Service (CDS) identifies claims that relate to a straightforward injury, that can be accepted without manual review. Claims that do not qualify for automatic approval are referred to staff for processing and cover decision. The model does not decline any claim.

This decision is informed by the scores produced by the statistical models.

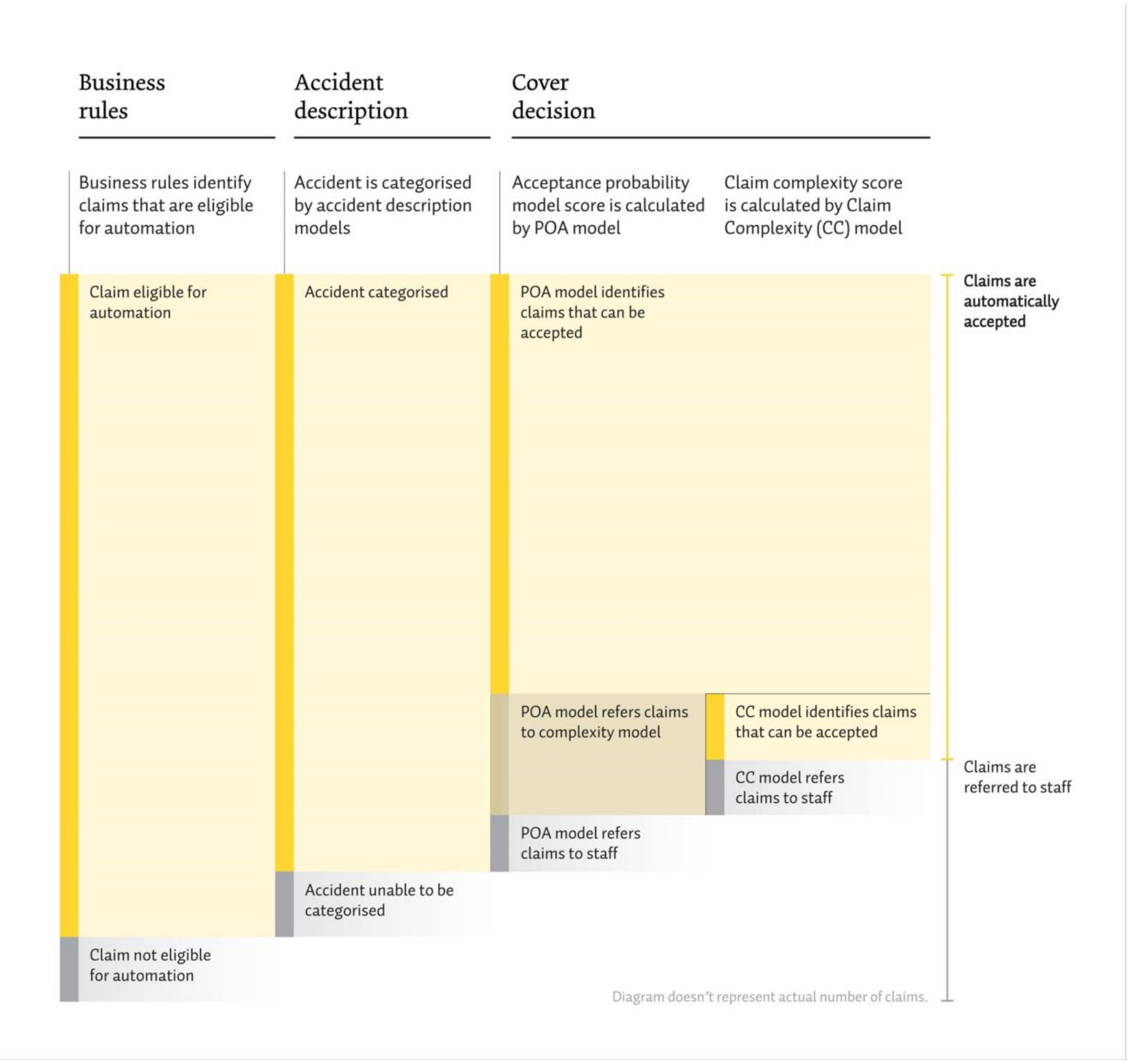
Accident Description Service

The Accident Description Service (ADS) searches the free text in the ACC45 claim form, looking for key words that could help categorise the type of accident being claimed for, e.g. “rugby accident”, or “fall”.

It uses statistical models to auto-populate certain data fields that are used for injury prevention, monitoring and reporting purposes as well as by the Actuarial team. The ADS information is not used as part of claims approval.

Where there is no statistically likely result for a field, the content will be referred for manual population by a staff member. As these fields are not used to determine whether a claim is approved, this will not delay a cover decision.

Figure 1. The models supporting automated services

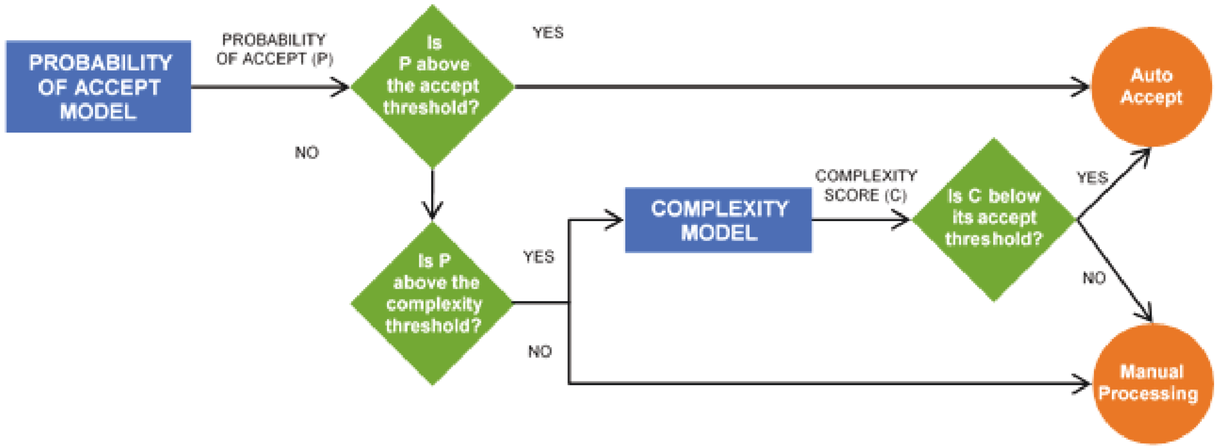


Cover Decision Service

Overview of the models

CDS uses two statistical models that work in tandem to calculate the probability of acceptance and case complexity. For any claim, the cover decision service can only 'auto-accept' or 'hold for manual review'.

Figure 2. Process chart for cover decision service models



The **Probability of Accept** (POA) model predicts the likelihood that a claim would be approved, based on historical data. It analyses variables such as injury diagnosis, client age and lodgement delay using the information from the ACC45 claim form.

A score (expressed as a percentage) is produced for each diagnosis on the claim, and if the score exceeds the set threshold the claim will be automatically accepted

The **Case Complexity** model predicts the complexity of a claim. This model has been built using historic treatment and entitlement cost data. Like the POA model, factors such as diagnosis, age and earner status contribute to the score. A score out of 100 is produced for each diagnosis on the claim.

The case complexity model automatically accepts low cost claims that don't quite meet the POA threshold. Claims of this nature do not warrant referral for manual processing as this offers no value for the client or for ACC.

Business rules are also used to specify certain claim types that are always held for manual review as they need to referred to specialist ACC teams for manual handling. This includes accidental death, or gradual process injury. Over time, such claims may be incorporated into the CDS. Regardless of the model outcome, there are business rules that may apply to flag a claim needs manual handling due to requirements for further information or support services.

Thresholds control the volume of claims that are automatically accepted

The probability and complexity scores result in either an 'auto-accept' or 'held – referred for assessment' decision through the application of thresholds.

When setting thresholds, ACC balances the volumes automatically accepted at each threshold, and the impact of incorrectly accepting a claim. More detail about thresholds and associated measured is set out in table 9.

Claims that score just below the POA threshold are referred to the complexity model. The complexity model will automatically accept low complexity claims on the basis that these are low cost claims that would not benefit from further consideration. More complex claims are referred to manual processing.

Claims with a POA score well below the threshold will be referred to manual processing, regardless of complexity.

Business rules control how thresholds are applied

The models have been built in a way that allows automatic approval of a large volume of claims. The model scores do not determine auto-acceptance; results are fed back to the rules engine for a final decision.

Where there is more than one diagnosis on the claim, business rules determine the overall model result for the claim. Currently these are configured so that if any diagnosis falls below the threshold, the claim will be referred to manual processing. Business rules can therefore apply to override a model recommendation to auto-accept.

For example, a claim may include diagnoses of ‘alcohol poisoning’ and ‘ankle sprain’ - the model would most likely recommend that the ankle sprain is accepted but refer ‘alcohol poisoning’ to manual handling. In this case the whole claim would be sent for manual handling.

Models were built using anonymised historical data

ACC has used data from 12 million previous, anonymised claims to build its models.

The cover decision models were built using the latest cover decision, and the cover decision reason. From a legal perspective, cover is continually assessed throughout the life of the claim and using the latest cover decision rather than the registration decision allows the models to be built using the actual cover outcome, rather than what the initial decision may have been.

It is recognised that initial diagnoses on a claim form can change because of further information, which can affect whether a claim should be covered. ACC is comfortable that an initial approval decision can be revisited if further information becomes available during the life of a claim.

Accredited Employer (AE) claims, duplicates, contributing insurer claims and claims with invalid diagnoses were excluded from the modelling dataset because ACC data shows these as ‘declined’, even though they may appear to be valid claims (in the case of duplicates) or may be accepted by other parties (e.g. AE claims).

Over-sampling improves modelled results

ACC accepts approximately 96% of claims registered each year. To counter the high acceptance rate, oversampling[1] has been used to ensure sufficient data points are available to establish the factors relevant to a decision to decline cover. Oversampling is a standard technique used to ensure that the model doesn’t ignore rare outcomes. Oversampling the data produced the proportions shown in the figure below.

---

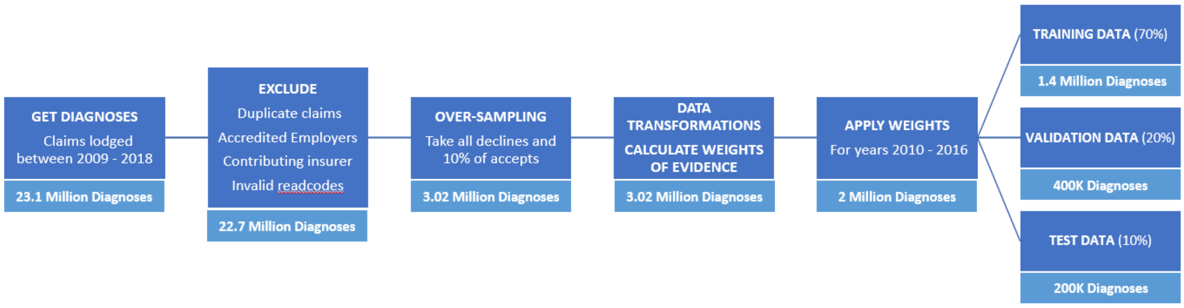
[1] Oversampling involves selecting more samples from one class than another, to force the model to account for the rarest outcome.

Figure 3. Proportion of automatic approvals before and after oversampling

Decision	Original proportions	Oversampled proportions
Accept	96%	85%
Declined	4%	15%

Once the data was oversampled a data partition was created. For the CDS models this used a 70:20:10 split to create a holdout datasets, meaning that of the two million data points 70% are used to build the model, 20% are used to test the models and optimise the parameters, and 10% are retained for a final unbiased assessment of model performance.

Figure 4. Creating the dataset



Logistic regression models best meet requirements

Various model solution options were considered for delivery of the business requirements. Logistic regression models[2] were identified as best meeting the key requirements:

Transparency

Regression models enable the creation of plain language rules to identify which variables increase or decrease the probability of acceptance or complexity. This makes it possible to see how inputs have influenced the scoring and resulting decision to automatically approve or hold a claim.

Flexibility

The logistic regression models use a scoring system that allows ACC to apply thresholds for decision making. This gives the ability to control and monitor the claims flow, and adjust the thresholds as necessary. The modelling approach also allows ACC to adjust model inputs in response to external or policy changes.

Variables can be added and removed, which gives ACC the flexibility it needs to quickly change the models if new input data becomes available or data sources change. For example, there could be changes in how an injury diagnosis is recorded, which would require a change in model parameters.

[2] A logistic model (or logit model) is a statistical model applied to situations where there are two possible outcomes. Input variables are used to estimate the probability of each possible outcome. In this case, the two possible outcomes are that a claim is “accepted” or “help for manual processing”.

Deployment

Logistic regression models can use a series of look-up tables for each input variable. These tables supply a parameter value for each variable in the model which are then added together to get a final score. By using lookup tables, the models can be deployed efficiently into different systems.

See Appendix 2 for further discussion of the model options considered and the solution rationale.

The models went through a quality assurance and testing process

The code that builds the POA and Case Complexity models went through a quality assurance and testing process that involved:

- the code being reviewed by another member of the team.
- sense checking of the data and results at each stage to identify irregularities.
- following individual diagnoses through the data processing stage to ensure the dataset is being built up correctly.

As well as these tests, both models were applied to both a validation and test dataset to get an unbiased measure of the accuracy of the model.

Deloitte was also commissioned to undertake certain test procedures on the model development, model validation, and data sourcing processes. They also re-performed a subset of calculations undertaken by the Model and compared the results with the existing model validation testing. All issues raised have been resolved.

Ongoing monitoring and review activity will ensure ongoing accuracy of the models. This is described in more detail below in the section

Ongoing monitoring and review activity will ensure ongoing accuracy of the models

Several inputs will be used to monitor the performance of the statistical models to ensure they continue to produce reliable outputs, and adapt to any external changes. These include:

- Operational reports will provide detailed information about volumes automatically approved, including within specific groups.
- Exception reports will identify where models do not perform as intended.
- Detailed information about the scores generated for each claim, and how each score was reached.

This information will be used for:

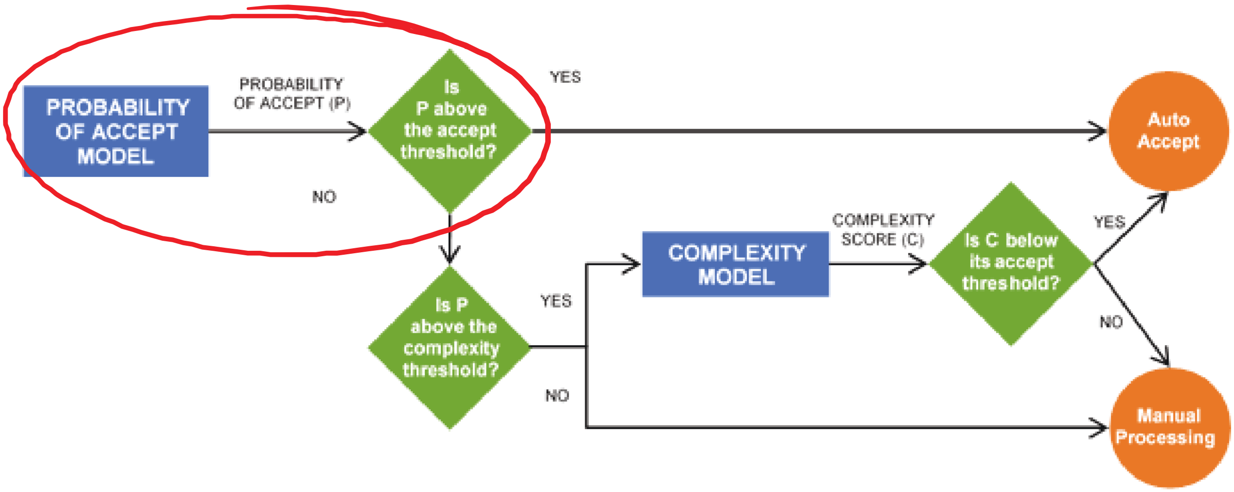
- Checking for 'concept drift[3]' by monitoring trends/changes in model scoring and the underlying variables. This will show where small changes in the data are having an impact on the models and the outcomes of the claims.

[3] Concept drift is where the statistical properties of a modelled target variable changes over time in unforeseen ways. This causes problems because the predictions produced by a statistical model become less accurate over time.

- Checking modelled scores for unexpected changes over time, and monitoring trends in underlying variables used in the models.
- Analysing a sample of claims to confirm whether the models are producing expected outcomes.

Each of these activities form the basis for daily, quarterly and annual review processes. In addition, a governance process guides and approves changes to model parameters in response to model drift, or external factors such as policy or legislative change.

Probability of Accept (POA) model



The POA model was designed to replicate the function of staff accepting claims for cover following ACC's existing business process. The model was developed by examining the existing decision inputs, to determine how they could be used in modelling. These inputs were:

- Information available on the ACC45 claim form:
  - The client details (name, address, DOB, NHI etc.).
  - The date and location where the accident occurred.
  - A description of the accident.
  - The injury diagnoses recorded by the provider in the form of Read codes, ICD 10 or ICD 9[4] codes.
  - Information on whether the accident happened at work and the client's employer details (to determine if the claim is an accredited employer claim).
- 'Hold' criteria provided to the registration team:
  - Particular diagnoses (e.g. identifying diagnoses where further information is needed)
  - Certain text in accident descriptions (e.g. indicating specialist claim types).
  - Instances where providers have diagnosed outside their scope/competency.
  - Large delays between when the ACC45 was signed and when the accident happened.
  - Duplicate claims.

[4] These codes are standardised ways to describe injury diagnoses.

Business subject matter experts were consulted on which factors to include (for example the lodging provider) and inputs that must be excluded from a human rights or privacy perspective (for example client gender and ethnicity). A short list of model inputs or 'variables' was created.

The historic claims data highlighted which variables were most predictive of acceptance, and these were validated with the business and tested from legal and technical perspectives (see Appendix 3: Model validation and internal decision-making).

Variables included

The POA model contains the variables that were proven to be relevant, predictive and ethically sound. The variables are described below.

There are many weightings for each variable, and the relative contribution of each to a final score varies. The relative importance of variables in the POA model is as follows:

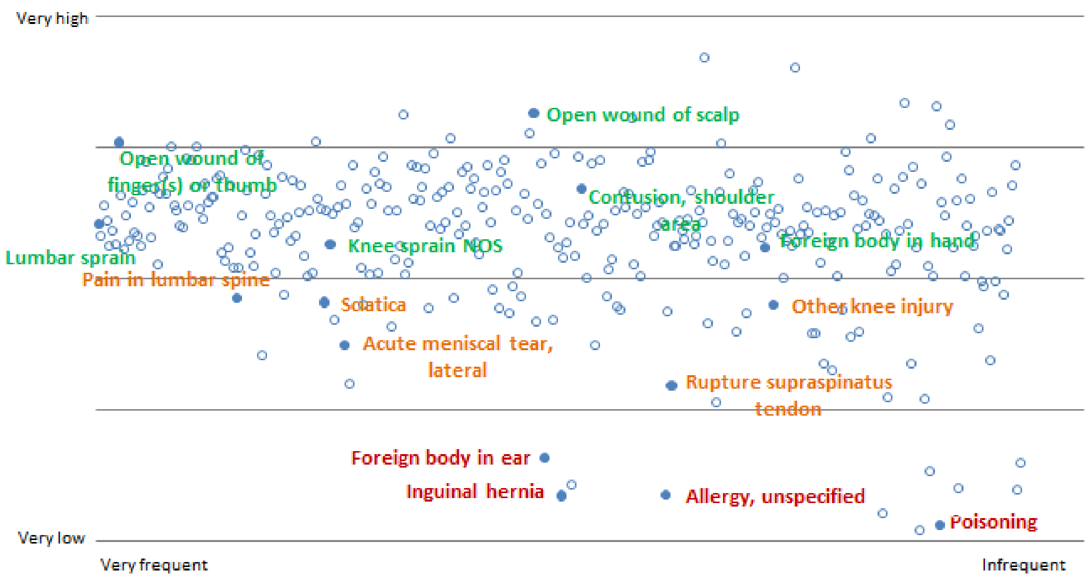
- *Claim diagnosis* is always the most important variable, contributing much more to the final score than all the other variables combined.
- *Client age* is the least influential.
- *Provider history*, *days since last decline*, *lodgement delay*, *accident location* and *key terms* contribute to scoring in a way that is roughly equal.

Diagnosis

The injury diagnosis is the most critical component in the cover decision. On the claim form, it is recorded as either a Read code, an ICD9 or an ICD10 code. Most injuries will be covered, however certain diagnoses require investigation into the circumstances and cause of injury to establish that they meet the requirements of ACC cover.

To develop the variable scoring, the history of cover decisions by diagnosis was analysed. The figure below shows a subset of diagnoses that have been plotted against acceptance rates and frequency. Some diagnoses are labelled to illustrate where there is a consistent pattern of high or low acceptance for certain injury types.

Figure 5. Acceptance rate for selected diagnoses



Each of the over 80,000 available diagnoses has a parameter estimate that contributes to the final score. For common diagnoses, this was calculated using historical data. For rare diagnoses where there is insufficient data to establish a reliable estimate, the overall acceptance rate was adjusted depending on the number of observations.

If a diagnosis is *very* rare, the parameter estimate is set to a large negative number to ensure such claims will always be held for a staff member to determine cover manually.

Most claims feature one diagnosis, but there can be up to 10 diagnoses recorded on the claim form. The model calculates a score for each diagnosis.

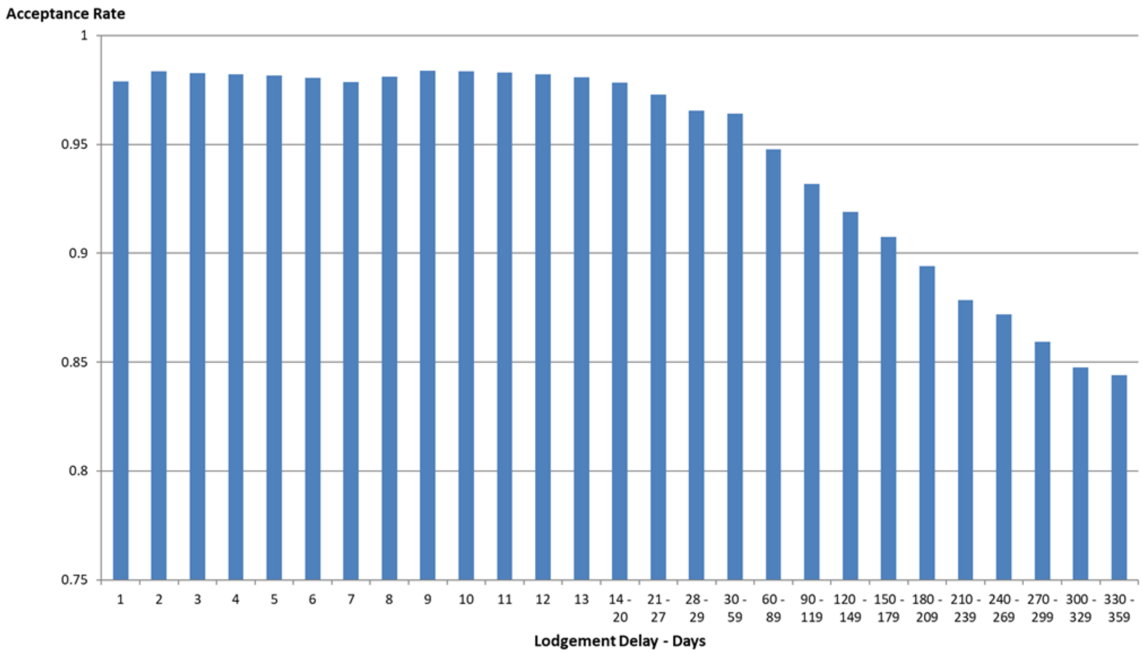
Weightings for the diagnosis variable will be refreshed regularly to ensure the model remains accurate.

Lodgement delay

The Act sets out restrictions around cover if there is a significant delay in lodging the claim. Where this occurs, the model ensures these claims are referred to specialist staff for cover assessment.

The available data on length of lodgement delay and cover decision outcome was reviewed to determine how this variable should influence scoring. The figure below shows claims submitted soon after the injury are more likely to be accepted.

Figure 6. Acceptance rate by lodgement delay



The model calculates the impact by days of delay from 1-360 days, and a set weighting is applied for any delay of more than 360 days.

The model is trimmed at 360 days to remove the effect of providers or clients completing the claim form with the correct accident day and month, but the incorrect year. Any claim with a delay of a year or more will be referred for manual processing. This also ensures claim types that are often associated with long delays (such as sensitive claims) are held for manual consideration.

Client age

Analysis of historical data shows patterns of claim acceptance are influenced by age. Age is only mildly predictive and is weighted in the model accordingly. This is because factors that determine variance in acceptance by age are also picked up by other variables in the model (e.g. key terms).

People in the very young (1-14 years) and very old (70-99 years) age groups have consistently high rates of acceptance and people in their 30s to late 60s share a generally high rate of acceptance.

Those in the 18-24 years age band have the lowest historical rate of acceptance. Data shows that of the declined claims in this age bracket, many had diagnoses related to self-harm, alcohol or poisoning. These are not generally covered under the Act.

In the model, the POA is highest for children under 10. It then decreases in the 10 to 20-year age range to reach a minimum of 20 years old. After that the probability of claim acceptance increases slowly with age.

Accident description key terms

Injury 'key terms' are words that appear in claim forms which indicate that the claim may not qualify for cover, or requires further investigation.

Such key terms have previously been used in the cover assessment process. In developing the model, historic claim data was analysed to validate the terms were associated with low acceptance rates.

The model identifies these terms by searching through the free text within the Accident Description and Injury Comments fields of the ACC45. If any of the key words are identified this negatively influences the Probability of Acceptance calculation, thereby making it more likely that the claim will be held for manual processing.

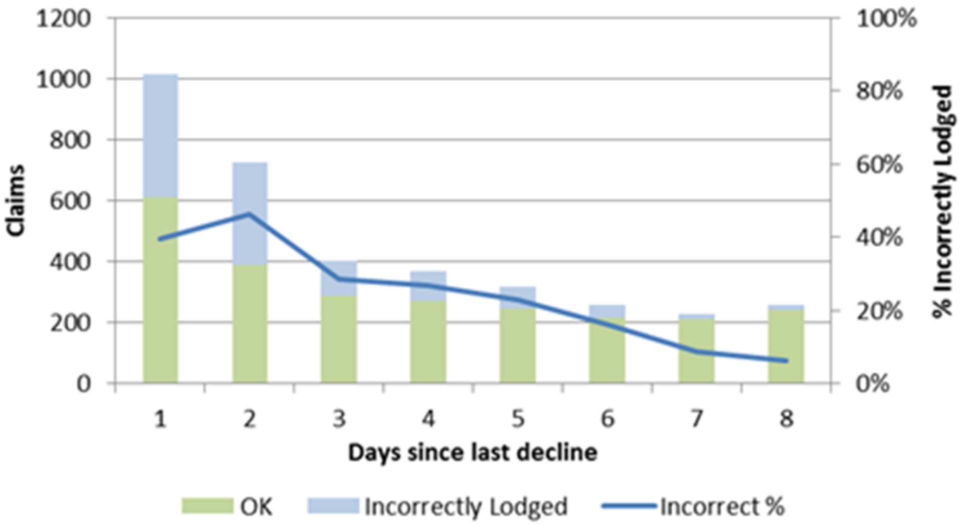
Model key terms will be regularly reassessed and may be added to over time.

Days since last decline

Sometimes claims are submitted more than once (duplicate claims), or resubmitted following an initial decline. To address these scenarios, a check for recent declined claims for the same client is included in the model.

A duplicate claim is most likely where there are fewer than seven days since the last decline, so model parameters adjust the probability of acceptance downwards where a recent decline is identified. Beyond this, decline decisions become less useful in predicting whether a claim should be accepted. The pattern of consequential declines is illustrated below.

Figure 7. Days since last decline



Note that the number of days since the most recent claim was submitted was not identified as a useful indicator of whether a claim should be automatically accepted.

Location

The Act provides cover for accidents in New Zealand, and a small subset of claims outside of New Zealand where certain criteria are met. Most accidents that occur overseas are not covered.

The ACC45 claim form identifies the location of the accident and on registration these are recorded as a physical address, or as a non-specific location such as 'at sea'.

Where the location is in New Zealand there is no weighting applied to the calculation. In general, model parameters will result in any claim for an accident that does not state it occurred in New Zealand, being held for manual investigation.

Provider history

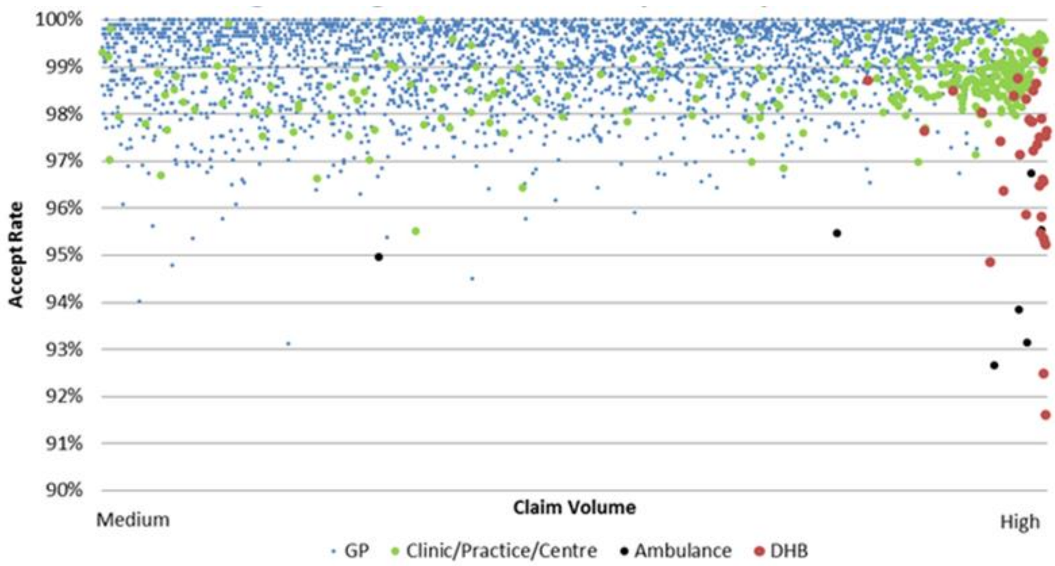
The lodging provider's historical rate of claim acceptance is included as a variable in the POA model to reflect the wide variance in acceptance rates by the ~30,000 providers ACC deals with. Provider acceptance rates are influenced by factors such as:

- Quality of information submitted by providers (e.g. where the person who files the claim is not the person who assessed or treated the client).
- The system the provider uses to input required information on the claim (e.g. a provider working with a Practice Management System has a greater likelihood of complete information).
- The type of injury the provider commonly treats (e.g. a physiotherapist may have greater likelihood of injuries treated receiving cover under the Act than an orthopaedic surgeon).

Analysis of the data shows that the variance is specific to provider, rather than provider types. For example, District Health Boards have a large range in rates of acceptance from high to quite low, and they are consistent in these rates over time.

Figure 8 maps the pattern of lodging provider claim acceptance rates.

Figure 8. Acceptance rates by registering provider



The data on provider acceptance rates will be refreshed on a regular basis, so that this variable remains an accurate predictor and ensures any improvements in provider processes and behaviours are reflected in the model parameters.

Manual changes were made to some parameter estimates

In some instances, the parameter estimates were illogical and a manual adjustment was required. For example, if client age is somehow recorded as negative the claim needs to be manually processed. To achieve this, the parameter estimate for these was manually set to -1234.5 which is a sufficiently low weighting to ensure a 'held' outcome.

The same parameter value was used for any diagnosis with fewer than 20 accepted claims in the past or with fewer than 30 claims in total; i.e. sufficiently rare to require manual processing for cover.

Models have performed well in testing

The accuracy of the POA model in testing was very high; the validation Receiver Operating Characteristic (ROC) score[5] for the model is 95.26%, although the actual rate of accuracy is influenced by where the threshold for auto-acceptance is set.

Testing shows the inaccuracy of the model increases as the threshold is increased, as expected. This is illustrated in the figure below which shows model performance using a test dataset of 10% of claims lodged in the last six months of 2016.

Figure 9. Probability of Accept model example thresholds and the accuracy of the model

Percentage of eligible diagnoses auto-accepted	Auto-approve model score threshold	Number of diagnoses auto-accepted	Number of diagnoses subsequently declined	Accuracy of auto accept
10%	99.91%	12,840	2	99.98%
20%	99.87%	25,680	13	99.95%
30%	99.83%	38,519	24	99.94%
40%	99.78%	51,359	42	99.92%
50%	99.71%	64,199	60	99.91%
60%	99.60%	77,039	87	99.89%
70%	99.43%	89,879	132	99.85%
80%	99.09%	102,718	218	99.79%
90%	97.73%	115,558	409	99.65%

The number of diagnoses subsequently declined is mostly attributable to information received after registration that changes the decision outcome (e.g. the diagnosis changed or a new diagnosis was added). This does not reflect the actual rate of incorrect acceptance by the model, which is only a small percentage of this volume.

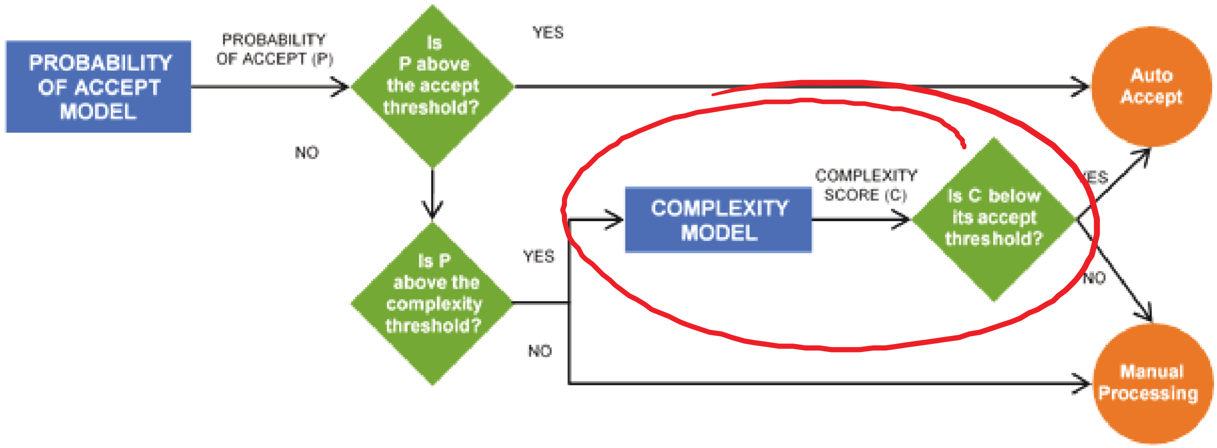
---

[5] The Receiver Operating Characteristic (ROC) is an indicator of the rate of false positives relative to the rate of true positives. It is a useful measure of performance of logistic models that shows how well a model can distinguish between expected outcomes. A score close to 1 indicates a very good performance \.

False positives and false negatives were checked for and reviewed in an audit exercise (a false positive is an auto-accepted claim that should not have been accepted).

The few findings of model error were where there was a lack of external force needed to qualify the claim for cover. In some cases, the force involved was a bodily function that does not meet the requirements for cover under the Act. Thus, terms like ‘yawn’ were added to the model to assist with identifying these scenarios. Further discussion of the review findings can be found in Appendix 3: Review of model outcomes and accuracy.

Complexity model



A primary objective of automated acceptance is to reduce the high volumes of low-cost claims that are referred for manual assessment. The case complexity model supports that goal by automatically approving claims that do not quite meet the Probability of Accept threshold, but are assessed as likely to be low cost. For such claims, the costs of manual assessment exceeds the client and efficiency benefits from automated approval.

The complexity model is not used to identify high cost or complex claims – if a claim meets the POA threshold it will be automatically accepted, regardless of complexity.

The model is based on historic treatment, entitlement and rehabilitation cost data from ACC provider and client payments systems. The model does not include claim costs for external providers such as DHBs as these come under a different funding structure.

The purpose of the model is to identify groups of low complexity claims to automatically approve, rather than produce an accurate estimate of claim cost, so it was not necessary to achieve an accuracy standard as high as the POA model. Several models were trialled that involved adding or removing terms. The best model was chosen by identifying the model with the highest R-squared[6] on the validation dataset.

As most claims are low complexity this gives a skewed distribution in the historic dataset. A log transformation was applied to produce unbiased estimates.

Variables included

Analysis was carried out to determine which variables had a measurable influence on claims cost, which showed some overlap with POA variables, although in each instance the way in which the variable contributes to the model is different.

The *diagnosis* and *provider* inputs are the most important variables, with the remaining variables contributing approximately equally to the score.

[6] An R-square score is a measure of how well data fits a linear model, with scores close to one representing a very good fit.

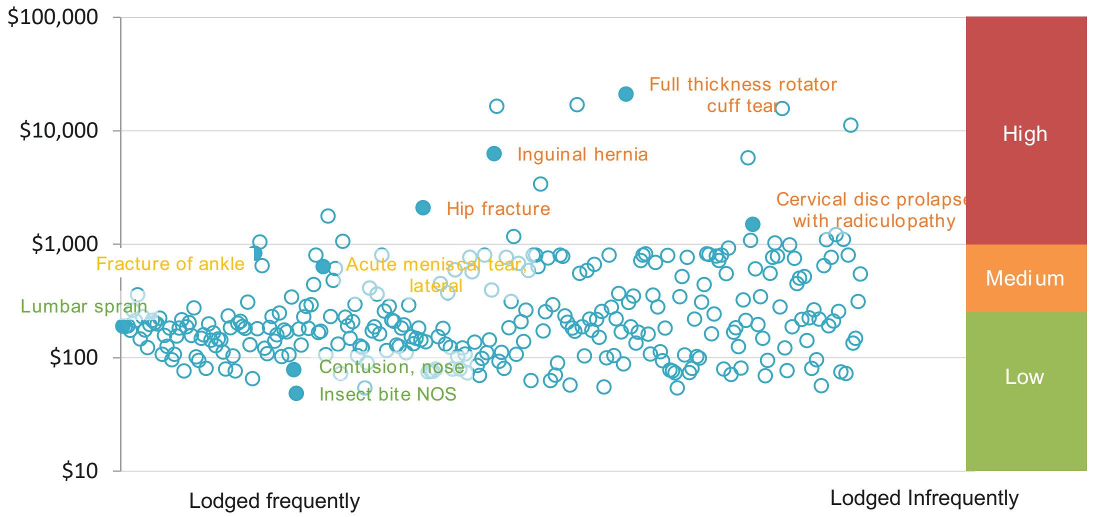
Diagnosis

Diagnosis is critical in predicting how complex a claim is likely to be.

There are factors specific to a given injury diagnosis such as treatment required, associated provider expertise, level of inherent incapacity and average duration for rehabilitation that impact cost.

The chart below shows the relationship between high, medium and low complexity diagnoses (and estimated claim cost) and the frequency of these diagnoses.

Figure 10. Median complexity of a selection of diagnoses using cost as a proxy for complexity



The same approach to diagnosis was used for the Probability of Accept model.

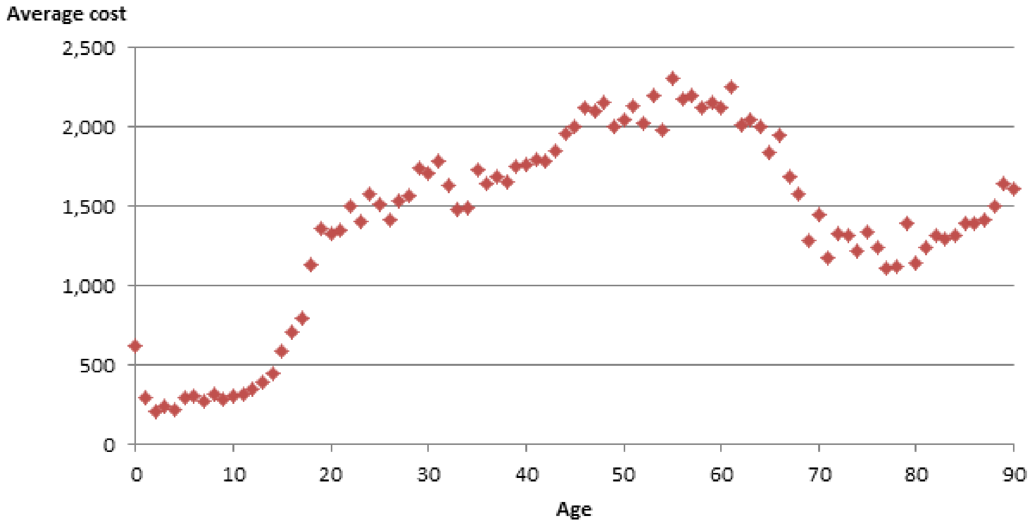
Client age

The data shows that complexity is lowest for those under 20. The complexity model is therefore more likely to automatically approve claims from young children.

The graph below illustrates claim costs by client age. Note that the increase at around 20 years old and decrease between 60 and 70 years old is caused by changes in employment status and therefore eligibility for weekly compensation payment.

Client age is one of the least important variables in the case complexity calculation. This is because part of the variation by age is picked up by other variables in the complexity model, such as diagnosis and key terms.

Figure 11. Average claims costs by age



Incapacity

The incapacity the medical provider specifies on the ACC45 form shows the amount of time a client needs off work due to their injury. A client can either have no incapacity (time off work), be fit for selected work, or fully unfit.

Incapacity is indicated on the ACC45 with either a return to work date or a duration and the units of duration (i.e. days, weeks or months).

Analysis found that where there is incapacity as a result of injury, the count of days increases the cost of the claim in an approximately linear fashion. The weighting for scoring incapacity in the model has been set accordingly.

Provider payment history

This variable considers past payments for claims made by the same lodging provider to identify providers that tend to submit claims with low cost claims. For example, certain types of provider will tend to deal with injuries at the lower end of the spectrum and therefore lodge claims for injuries that attract low cost (and vice versa).

Several models were built to identify the most accurate payment history windows, ranging from three months to eight years. The one with the best R-squared was the eight-year window.

As with the diagnosis variable, there are many possible values for this variable reflecting the number of providers. Variables for very common providers were constructed so that they could have their own parameter estimate that was separate from the single estimate for providers that submit a small number of claims each year.

Client payment history

Similar to the provider payment history, analysis has shown that the sum of past payments to a client has an influence on the likely cost of a new claim. This variable also reflects other variables in the model, such as earner status.

The client history variable is not as predictive as the provider history variable and has a significantly lower weighting in the model.

Earner status

Weekly compensation is the largest cost category for ACC; having earner status in the model enables a more accurate estimation of the anticipated average costs for a specific earner type.

Earner status is a useful predictor even where there is no incapacity noted on the claim. Frequently lodging providers will not note time off work in the ACC45 because in the first seven days of the claim weekly compensation is not payable. It is in subsequent medical assessments that the question of incapacity is addressed.

Accident description key terms

The accident description often contains keywords that indicate the severity or complexity of the injury beyond what can be established from the diagnosis. Data analysis has identified these 'key terms' impacting claim costs.

For example, the description of an injury as 'mild' or 'minor' helps with identifying low complexity claims.

Accident Description Service

Automation to improve speed and consistency

The Accident Description Service (ADS) automates recording of accident details such as cause, external agent (or 'force') and prior activity. Automating the recording of this information reduces processing time and ensures better consistency within the data. This information is not used as part of the cover decision model.

Automating this process for some claim types is not possible or appropriate, and these will continue to be processed manually, e.g. sensitive claims, dental claims.

The models automatically select the field inputs based on the available information

The ACC45 claim form has a free text field section to describe how the accident happened. As part of registering a claim in the ACC claims management system, the free text in the form is codified into 11 data fields. These fields are used for a range of purposes (excluding cover decisions), including Injury Prevention, setting levy rates and by external organisations such as WorkSafe and Statistics New Zealand, who use this information to produce statistics regarding accidents.

Each field has a pre-set list of input values associated to it. The fields record:

- What was the **cause** of the accident (e.g. slipped).
- What did the person come into **contact** with as they had the accident (e.g. ground/floor).
- What **prior activity** was happening before the accident (e.g. children playing).
- What was the **external agent** that caused the accident (e.g. sharp object).
- If it was a motor vehicle accident, what was the person's position or role at the time (e.g. a passenger) this is referred to as **external agent 1.**
- If it was a motor vehicle accident, what was the thing come into contact with (e.g. a tree) this is referred to as **external agent 2.**
- If the client was **playing sport**, what kind of sport was it (e.g. rugby), was it **organised** (yes or no) and what was the client's **involvement** (e.g. referee).
- What **type of work** does the client ordinarily do (e.g. heavy work).
- Was the person **at work** when the accident happened (yes or no).

For each of the fields, the accident description models search for specific terms matching the input value. They calculate the statistical probability of each value being the correct choice, based on the information in the claim form. If the input with the highest score meets the accuracy threshold, then it is copied to the field. If the threshold is not met, the information will be entered manually.

Developing the Accident Description Service model

Historical data was used to understand correct categorisation

Anonymised data from claims registered between 2010 and 2016 was used to produce roughly 12 million data points.

To minimise the impact of historical variation, data was cleansed to address or remove obvious irregularities in how claims data had been recorded. For example, the standard deviation of each registration field drop down value was calculated and any staff member with a proportion of claims that exceeded 3 standard deviations was removed from the modelling.

Oversampling was used for some rare accident description types

Some registration field values were used for fewer than 2% of claims, for example cycling accidents. To counter this, oversampling was used to ensure sufficient data was available to produce reliable parameter estimates.

Once the data was oversampled, a data partition was created using a 70:30 split so the model could be validated using a holdout dataset.

Text was cleaned so that single terms, bigrams and trigrams [7]could be identified

The free text for the models comes from two fields in the claim form; the *accident description* field and the *injury comments* field. Special characters such as symbols, digits and white space-related characters are removed. The exception to this is the # character which is a known shorthand for the term fracture.

The Accident Description Service model relies on correct spelling of certain terms to determine which registration categories a claim should belong to, and to ensure correct classification of claims. As a high percentage of claim forms have misspellings in accident descriptions, common misspellings and medical terms are corrected using a lookup table.

Stemming[8] was carried out within the spell-checking list. Words were stemmed back to their base verb, and to past tense. A stop list was also created to prevent extraction of common terms that do not improve model performance, such as 'the'. Single terms are extracted, followed by bigrams and then trigrams. The bigrams and trigrams can include stop list words such as ‘and’ and ‘the’ because they provide important context for identifying causes. For instance, 'fell off' is likely to have the cause coded as loss of balance/personal control. However, something liked tripped 'and fell' is less likely to code the cause as loss of balance personal control and instead code it as ‘tripping’.

Common terms that assist in codification are identified

For example, the term "rugby" is a very common term in rugby accidents. There are also less common terms that are highly predictive of rugby injuries such as "ruck" and "maul". These terms are identified and incorporated into the model in a way that allows for rare but highly predictive terms as well as common terms.

Bigrams and trigrams appear as business terms within the model to ensure they are recognised. For example, "all-terrain vehicle" is recognised as “ATV” in the external agent category, rather than "all", “terrain” or “vehicle” as individual terms.

Refining some registration categories

As the model was developed, several opportunities for improvement were identified in relation to the registration field drop downs. Some of these fields were found to be difficult for registration staff to use as there was overlap in the drop downs.

During this process, a dictionary of definitions for each drop down was created. This will help staff follow the same logic as the model when it comes to codifying claims that are sent for manual registration.

---

[7] A pair/group of three consecutive words such as letters, syllables, or words that have a particular meaning when together.

[8] Stemming is the process of reducing inflected words to their word stem, or root form. E.g. ‘falling’ to ‘fall’.

Models have performed well in testing

Most the models have easy to identify terms with most of the ROC validations scores exceeding 0.85.

The median ROC score is 0.93 meaning most of the models are performing very well. Note that the accuracy depends on the cut-off thresholds set. At a threshold of 50% most the registration fields are correctly classified.

Models that performed best were those where categories within fields were specific and distinct, with limited scope for interpretation.

Adjustments to avoid step changes in data fields

An adjustment is used to preserve the current rates of accident registration types so that there is no sudden shift that could impact uses of the data generated from the ADS. These adjustments can be ‘turned off’ individually without impacting the others, and will be removed over time.

The adjustment is run independently across the registration fields: *cause, contact, prior activity* and *external agency*.

After running the models, shifts in categorisation volumes were investigated, which revealed some areas where the model would produce different results. For example, the *cause* field showed the following:

- *Loss of hold* had a large increase, which is most likely caused by the new Subject Matter Expert (SME) definition, which also includes dropping objects in here. This may also be responsible for the large drop in the loss of personal control category.
- *Struck by handheld tool* decreased because anything that breaks the skin now belongs in *puncture*.
- *Flooding overflow of liquid* has had a large change because it now defined as chemical liquid specifically rather than water in general.
- *Oral ingestion of a fungi* had a large increase as these are often scattered across several categories.

These shifts are the result of the model applying a consistent approach that best reflects the definition for each field, and have been used to develop a dictionary of definitions for each field to help staff.

Privacy, human rights and ethical considerations

All stages of the modelling life cycle have been reviewed to ensure that privacy, human rights and ethical considerations have been identified.

The model has been developed to avoid any gender or ethnicity bias and has been independently validated by experts to ensure it is accurate, meets good practice guidelines for use of personal data and is working as intended. It has also been developed with careful consideration of privacy, human rights and ethical issues.

In developing the models and associated material, the 'Principles for the safe and effective use of data and analytics' (May 2018), produced by Statistics New Zealand and the Privacy Commissioner have been taken into account. Research on building trusted analytics published by KPMG (2016) has also been used, which focuses on four anchors of creating trust:

- *Quality* (of data, tools, methods and human capability).
- *Effectiveness* (model process, model accuracy and model utility).
- *Integrity* (regulatory compliance, privacy and ethical use).
- *Resilience* (future proofing, security, governance and monitoring of data and predictive models).

Measures have been taken to ensure the data is reliable and secure

Every year around two million claims are processed. Data from 12 million previous, anonymised claims from 2010-2016 was used to inform these models as they were developed offering a wide and reliable base to model from. This also ensures enough time is available for all cover decisions to be made (this can take up to nine months for some claims). These years were chosen as analysis found that data from previous periods was less relevant for predicting outcomes, and closing off the data input at 2016 leaves the 2017 data available for further testing of the model if necessary.

The model data has been anonymised to protect the privacy of clients and providers. Other areas of data consideration such as storage, privacy and security are covered in the Privacy Impact Assessment and Security Risk Assessment.

There was careful consideration of variables in the model

ACC has consent to use all the data which features in the models. When a person signs the ACC45 form they give consent for their claim data to be used to help determine if the claim will be approved.

From the outset, ethnicity and gender were specifically excluded from the models. All model variables were passed through a rigorous internal review and assessment process to determine the appropriateness of their inclusion in terms of privacy and ethics.

The models are highly configurable and variables can be added, removed or adjusted as legislation or process changes.

Testing showed no gender or ethnicity bias

The POA and Complexity models were tested for evidence of discrimination or bias in scoring outcomes, with a focus on gender and ethnicity. These considered the presence of bias in both the historic data series, and modelled results. This is important as bias in historical data could impact the model parameters, and influence the modelled outputs for certain groups. In each case, no material bias was identified.

The assessment was carried out using four methods:

1. **Testing historic data for bias**

Historic data was tested for any differences in the acceptance rate and average complexity of claims manually processed, with a focus on ethnic groups and gender.

Outcome: The table below shows only minor differences in proportion of claims historically accepted by gender and ethnicity. Actual claim costs for each group were converted to a complexity score out of 100, which again showed only minor differences.

Grouping	Percent historically accepted	Average complexity score
European	97.6	26.8
Maori	97.0	25.1
Pacific	97.6	23.4
Asian	97.9	27.2
Other	97.6	26.3
Male	97.6	26.5
Female	97.4	26.4

2. **Testing the model for bias**

The holdout data set was entered into the models to test the percentage of claims automatically accepted and complexity scores produced by the model. Again, results for ethnicity and gender were compared to check for any differences which suggest bias in model outputs.

Outcome: The table below shows only minor differences in the modelled scores for gender and ethnicity.

Grouping	Percent historically accepted	Average complexity score
European	97.7	20.4
Maori	97.5	18.0
Pacific	98.1	17.5
Asian	98.2	19.5
Other	97.1	21.9
Male	97.7	20.1
Female	97.8	19.7

Tests for correlations between variables in the model were completed to identify any relationships between variables, and gender or ethnicity. This test helps identify whether variables included in the model can act as proxies for gender and ethnicity, thereby introducing potential bias.

Outcome: there was no observable correlation between the model variables and gender or ethnicity.

Ethnicity and gender were tested as predictive variables to identify whether they were predictors of probability of acceptance.

Outcome: testing confirmed that neither gender nor ethnicity are statistically relevant predictive variables.

Appendix 1: High level business requirements

The following high level business requirements were identified at the initiation of the project:

- Develop a tool to automate population of fields in the claim registration process.
- Develop a tool to predict the likelihood of claim approval (based on available information), as per the current assessment carried out by staff within the registration and cover decision process. The tool must meet the objectives of enabling automation of cover acceptance and improving consistency in decision-making.
- Develop a tool to predict the likely complexity of a claim, based on anticipated cost of treatment and rehabilitation. The tool must meet the objectives of enabling identification of low complex claims warranting acceptance without manual review.
- Utilise available data and learnings from past cover decisions to support prediction, within the confines of what can be legally and ethically considered.
- Deliver the tools in a way that makes the basis of the predictions available and understandable, to support ACC's disclosure requirements and give confidence in the accuracy of the prediction.
- Utilise in house data analytics expertise (e.g. in the form of SMEs) in the tool development, to enable the solution to be understood, maintained and refined by ACC, reducing dependency on external specialists.
- Deliver the solution in a form enabling straightforward, phased deployment.
- Deliver a solution that enables the addition/removal of specific variables, as business requirements/policy may change over time.

Appendix 2: Overview of solution options

To meet ACC's business requirements the following solution options were considered:

Deliver via development of simple business rules.

For example, if readcode = S572 and description doesn't mention gradual process then accept. This approach is the easiest to implement, understand and maintain but only works if 1 or 2 variables are used. If more variables are required, the rules get too complex. The number of variables requiring consideration for cover decision is more than 5, so this was not considered a suitable solution.

Deliver via a simple model.

1. For example, simple decision tree or regression.
2. One of the key requirements was for a solution that is easy to understand and easy for developers to implement. Simple models are an ideal solution balancing ease of implementation, transparency and accuracy as they allow multiple variables to be considered at the same time while still allowing the user to understand any prediction the model makes. They are also easy to sense check the results with SMEs to ensure the data processing steps don't contain any errors.
3. This option was considered the best fit for the business requirements and development constraints.

Deliver via a complex model such as machine learning or Artificial Intelligence.

4. For example, polynomial regression, neural network or random forest. These models tend to be slightly more accurate than the others and can adapt very quickly to problems that change rapidly. However, they are very hard to interpret which makes sense checking difficult and greatly reduces the transparency of the predictions they make.
5. One of the unique challenges with ACC's data is that the most useful variables also have a very large number of possible values that do not easily roll up into usable groups (for example diagnosis (4000 – 5000) and provider id (approximately 20,000). This makes the interpretation of complex models very difficult. They are also hard to implement in systems other than the one the model is built in.

Appendix 3: Review of model outcomes and accuracy

Approach to assessment

To assess the true failure rate, thirty claims from each decile of POA were sampled where the model suggested these would be accepted but should have been declined. These claims were reviewed by a subject matter expert (SME) to ascertain the true status of these claims.

The results are shown below.

Figure 12. Test results

Threshold quantiles	Accept	Decline	Exclude	Total
0		20	7	27
1		9	13	22
2		14	25	39
3	1	10	13	24
4		11	26	37
5	2	18	19	39
6	1	10	19	30
7		8	20	28
8	1	11	29	41
9	5	8	29	42
Total	**10**	**119**	**200**	**329**

- The threshold quantiles go from the lowest probability of these claims being accepted, 0 to the highest probability of being accepted, 9.
- Accept means that after review the SME decided that these claims should have been accepted.
- Decline means that after review the SME decided that these are correct manual declines. This is our true failure rate.
- Exclude means that after review the SME decided that these claims should have been excluded from the model accuracy assessment as they would be streamed off by business rules prior to the model.

A lot of claims should be excluded from the model assessment because they were declined for good reasons:

Missing mandatory data

The model is correct based on the model input data used but the claim was declined because mandatory information was missing such as a signature on the form or verifying a person who was injured overseas is ordinarily resident. A further situation is when a provider has specified a diagnosis outside their scope of practice, such as an acupuncturist diagnosing a broken bone. In such cases the claim would be declined and the client would have to visit the appropriate provider to be diagnosed. These types of misclassifications can be picked up by upstream data validation or the business rules engine to determine which claims are eligible to have a cover decision made.

Claims that are subsequently withdrawn

The most common scenario is where a claim is accepted, but the provider contacts ACC to correct information on a ACC45 form. In this case, the model was correct based on the information at the time the decision to auto approve was made. Approximately 5% of declined claims are in the withdrawn category.

Cover was later revoked

The injuries on the ACC45 are accepted and at some point during claims management lifecycle it is determined that the client has some other kind of non-accident related condition so the original cover for the entire claim is revoked. In this case, the model is right to auto approve the claim. It is possible that these claims could be dealt with if the cover decision service was triggered each time additional information became available.

Additional diagnosis added and declined

A medical certificate arrives at some point after the original cover is accepted requesting an additional injury, which ACC declines. Currently the cover decision engine is only triggered once per claim so this would be out of scope in the current set up. It is possible that this would be revised at a later date, at which point this type of misclassification would need to be revisited.

Overall findings

- Less than half of the false positives were true failures of the model (declines).
- The true failure rate of the model increases as the Probability of Accept threshold is reduced
- A few claims should have been accepted.
- There were no material areas of model improvement emerging from assessment of the true failures.