From Decisions to Insights: Using GenAI to Decode Trends in NSW PIC Cases on Whole Personal Impairment

Prepared by Carmen Burraston and Francis Beens
with special thanks to Kelly Chu

Presented to the Actuaries Institute
2025 Injury and Disability Schemes Seminar
16-18 November 2025

Actuaries Institute
ABN 69 000 423 656
Level 34, Australia Square, 264 George Street, Sydney NSW 2000
P +61 (0) 2 9239 6100 | actuaries.asn.au

Abstract

Whole Person Impairment (WPI) is a key 'gateway' to benefits across multiple injury compensation schemes in Australia. In New South Wales, both the Workers Compensation (Workers) and Compulsory Third Party (CTP) schemes rely on WPI assessments to determine eligibility for entitlements. In the Workers scheme, WPI determines access to long-term statutory benefits and the ability to make a damages claim. In the CTP scheme, WPI determines access to non-economic loss in damages claims.

WPI is assessed in NSW using detailed guidelines and different rules apply depending on whether the injury is physical or psychological. Disputes frequently arise over WPI assessments and these disputes will often be resolved in the Personal Injury Commission (PIC).

This paper investigates how GenAI can be used to extract structured information about WPI assessments from the unstructured text of published decisions handed down by the PIC. We test the nature of the information that can be reliably extracted from decisions. We then analyse the resulting dataset to reveal patterns and trends in decision-making separately for physical and psychological injuries.

We also address the practical challenges of applying GenAI to legal text, including model reliability, transparency and the need for human guidance and oversight. The findings illustrate the potential of using alternative sources of information (beyond what is often accessible from insurer claims datasets) to enhance our understanding of important aspects of WPI, offering actuaries, executives and policymakers a novel tool for monitoring the overall health of schemes.

Table of contents

1 Introduction

1.1 Purpose of the paper

1.1.1 What we are trying to achieve

The GenAI characterised by ChatGPT and similar models have been easily accessible and useable by the public since November 2022. GenAI has significant potential to assist in extracting information from large volumes of text.

Our paper tests how we can use these Large Language Models (LLMs) to do this in a specialised field that is relevant to actuarial practice – decisions handed down by the NSW Personal Injuries Commission (PIC) relating to medical questions, including Whole Person Impairment (WPI) assessments. The PIC is the dispute resolution body for Motor Injury (CTP) and Workers' Compensation (Workers) disputes in NSW and therefore has the potential to provide a rich source of information. The PIC publishes aggregated statistical information in relation to disputes, which is useful on its own. Our paper demonstrates how we can extract information from PIC decisions to uncover interesting observations and trends relating to dispute resolution.

This paper is intended as a practical demonstration of what can be achieved using GenAI in a specialised context. It does not attempt to provide a comprehensive analysis of all PIC decisions, nor does it cover every aspect of CTP and Workers disputes. Instead, we focus on a specific and important segment—finalised appeals relating to medical decisions—to illustrate the potential of this approach and highlight opportunities for future work.

1.1.2 Summary and conclusions

Our work demonstrates that GenAI can play a valuable role in transforming unstructured legal text into structured datasets that support meaningful analysis. While the technology is not flawless and requires careful prompt design, iterative refinement and human oversight, the efficiency gains are undeniable. What previously would have taken months of manual effort can now be achieved in hours, at a fraction of the cost. Beyond the immediate application to PIC decisions and WPI assessments, the approach offers broader potential for insurers, actuaries and policymakers: enriching existing datasets, uncovering new insights and enabling more nimble scheme monitoring and targeted claims management. As with any actuarial tool, success depends on applying professional judgment, understanding limitations and incorporating uncertainty into the analysis. Used appropriately, GenAI can become a powerful complement to existing data sources, helping us to better understand injury compensation schemes and identify opportunities for improvement.

1.2 Overview of paper structure

This paper is set out as follows:

- Section 2 sets out background information on the types of medical disputes we are analysing (primarily those to do with WPI) and the PIC
- Section 3 sets out how AI can be used to extract data from text, discusses structured prompting and how to reduce GenAI hallucinations and inaccuracies
- Section 4 sets out how we went about this project, including sourcing the data, interacting with the LLMs, the process of developing prompts (questions) and how we tested the results
- Section 5 then shows the results of our analysis and insights
- Section 6 provides our conclusion and considers professional implications and other applications

2 Background

2.1 Whole Person Impairment

2.1.1 What is Whole Person Impairment?

WPI is a percentage amount that expresses the extent to which an injury affects a person's overall health and ability to function. It is used as a threshold or a gateway to access certain types of compensation across many different Australian compensation schemes.

Each scheme has its own rules for how WPI is to be assessed.

2.1.2 How is WPI assessed in NSW for physical and psychological injuries?

Our paper focuses only on NSW. In NSW, WPI is relevant to both the Workers and CTP schemes.

Workers

The WPI assessment process for Workers is set out in the guidelines published by SIRA in the "NSW workers compensation guidelines for evaluation of permanent impairment" (Workers Guidelines). The Workers Guidelines adopt the 5th edition of the American Medical Association's Guide to the Evaluation of Permanent Impairment for most physical injuries (with modifications and clarifications) and sets out its own rules for the assessment of primary psychological (or psychiatric) injuries. The degree of impairment for psychological injuries is assessed using the Psychiatric Impairment Rating Scale (PIRS) set out in the Workers Guidelines.

CTP

The WPI assessment process for CTP is set out in the Motor Accident Guidelines (CTP Guidelines), also published by SIRA. The CTP Guidelines adopt the 4th edition of the American Medical Association's Guide for physical injuries (with modifications and clarifications) and uses the PIRS for psychological injuries.

Assessors

Under both schemes the degree of WPI is assessed by a trained medical assessor. They must be a medical specialist (for example, a psychiatrist, neurosurgeon, or orthopaedic surgeon), who has suitable qualifications, training and experience relevant to the body system being assessed and also completed training in using the relevant guidelines for the body system being assessed.

SIRA maintains a list of personal impairment assessors.

Different states will have different approaches to assessing WPI.

2.1.3 Thresholds and 'gateways' in NSW

In NSW both the Workers and CTP schemes rely on WPI assessments to determine eligibility for key entitlements. In the Workers scheme, WPI determines access to long-term statutory benefits and the ability to make a damages claim. In the CTP scheme, WPI determines access to non-economic loss in damages claims.

The following table sets out how WPI is used for key thresholds in the Workers scheme:

WPI	Allows entitlement to benefits
0%-10%	Income replacement benefits continue to 260 weeks of benefit and medical benefits continue for two years after income benefits cease. Income replacement benefits are subject to "Work Capacity" testing at various points in time. No entitlement to permanent impairment benefits, no entitlement to lodge a common law claim.
11%	Entitlement to permanent impairment benefits (physical injuries only). Income replacement benefits continue to 260 weeks of benefit and medical benefits can continue for five years after income replacement benefits cease.
15%	Entitlement to permanent impairment benefits for psychological injuries. Entitlement to lodge a common law claim. Income replacement benefits continue to 260 weeks of benefit and medical benefits can continue for five years after income replacement benefits cease.
21%	Income replacement benefits can continue to retirement age plus one year, but still subject to "Work Capacity" assessments. Medical benefits can continue for life. Entitlement to lodge a common law claim, which replaces income replacement benefits and ceases access to medical benefits. Entitlement to permanent impairment benefits.
31%	Same as for WPI 21%+, however, work capacity assessments not required.

The following table sets out how WPI is used for key thresholds in the CTP scheme:

WPI	Allows entitlement to benefits
0-10%	While a claimant can make a common law claim, it is limited to economic loss. There is no entitlement to non-economic loss. Income replacement benefits limited to 156 weeks (as long as a common law claim is pending).
11%	Entitlement to claim non-economic loss (pain and suffering, or general damages) as part of a common law claim. Income replacement benefits extend to 260 weeks (as long as a common law claim is pending).

2.1.4 Relevance to actuarial valuations

The thresholds are important, as exceeding particular thresholds brings an entitlement to more extensive compensation and support. Insurance liabilities for insurers providing cover in these schemes (which includes both public and private sector insurers and self-insurers) can be significantly affected by the number of claims that exceed particular thresholds. Understanding any trends in claims experience that relate to WPI thresholds and WPI assessments, is therefore one important element of actuarial valuations for insurers operating in these schemes.

2.2 The Personal Injury Commission

2.2.1 Role of the Personal Injury Commission and types of disputes

The PIC is the dispute resolution body for many Workers and CTP disputes in NSW. It began hearing disputes in 2021 and replaced previous specialist tribunals that had heard disputes in the two schemes.

The PIC deals with a range of different types of disputes and has a variety of different methods for dealing with them. The Commission's annual report provides a good summary. Only some are relevant for this paper (discussed below), but all of the different pathways include:[1]

- Workers Compensation division:
  - Expedited assessments (up to 12 weeks benefit)
  - Legal disputes (benefits exceeding 12 weeks)
  - Medical disputes
  - Work injury damages disputes (these relate to common law claims in NSW)
  - Appeals
- Motor Accidents (CTP) division:
  - Merit reviews (including appeals)
  - Miscellaneous claims assessment
  - Damages assessments and exemptions from assessments
  - Damages settlement approvals
  - Medical disputes, reviews and further assessments
  - These can arise under the current (2017) or the previous (1999) schemes

The focus of our paper is on WPI, which largely falls within the medical dispute (and related appeal/review) pathways.

Most disputes are lodged by a claimant (86% in CTP and 97% in Workers). Nearly all disputes lodged by a claimant are legally represented (94% in CTP and over 99% in Workers; note that in NSW the Independent Legal Assistance and Review Service provides access to legal advice and representation for injured workers).[2]

The PIC, across both Workers and CTP, finalised the following number of disputes by year:

- 2021/22: 13,700

[1] Personal Injury Commission Annual Review 2023/24, https://www.pi.nsw.gov.au/resources/annual-review, pages 25-29
[2] Personal Injury Commission Annual Review 2023/24, https://www.pi.nsw.gov.au/resources/annual-review, page 50.

- 2022/23: 17,100
- 2023/24: 18,400

Medical disputes, however, are only a subset of all of these.

2.2.2 Nature of medical disputes and decisions

The PIC Annual Review contains the following summary of medical disputes and related appeals:

Workers' Compensation: Medical disputes – Medical disputes in respect of the degree of permanent impairment resulting from an injury are usually referred to a Commission-appointed medical assessor for assessment. In some instances, where there is a liability dispute regarding the injury, a claim may be referred to a member for conciliation and/or determination. Medical disputes in respect of past and future medical expenses are usually referred to a member for conciliation and/or determination.

Workers' Compensation: Appeals – A party to a dispute may lodge an appeal… [The Annual Review initially describes non-medical appeals and then continues]… A party may also appeal against a medical assessment of permanent impairment. If the President's delegate is satisfied, on the face of the application and submissions, that a ground of appeal has been made out, the matter is referred for determination to a medical appeal panel, consisting of a member and two medical assessors.

Motor Accidents: Medical disputes – Medical disputes include whether the degree of permanent impairment resulting from an injury caused by the motor accident is over 10% or whether the treatment provided or to be provided is reasonable and necessary and related to the injuries caused by the accident. Disputes under this scheme also arise in relation to whether an injury is a 'threshold injury'. Such disputes are determined by a medical assessor. A binding certificate is issued to the parties.

Motor Accidents: Medical reviews – Reviews are available if it is shown that the medical assessment is incorrect in a material aspect. If a delegate of the President is satisfied that the review application can proceed, the matter will be referred to a medical review panel constituted by two medical assessors and one member who will conduct a new assessment. Unlike a medical appeal in the Workers Compensation Division, the review is not limited to only that aspect of the assessment, which is alleged to be incorrect, rather it is a new assessment of all matters with which the medical assessment is concerned. A new certificate will be issued which will either confirm the certificate of assessment of the single medical assessor or revoke that certificate.

In summary, across both schemes, disputes around the degree of WPI are initially referred to a PIC appointed medical assessor. Some other medical related disputes may be referred to a medical assessor or to a general member.

Determinations regarding the degree of WPI can be appealed (Workers) or reviewed (CTP). In both cases a panel consisting of one member, and two medical assessors will hear and determine the appeal.

2.2.3 Volume of medical disputes and appeals/reviews

The following table sets out the volume of disputes relevant to medical issues, as well as appeals. The data is sourced from PIC Annual Reviews and SIRA's Open Data scheme reporting.

	CTP	WC	Total
1: Total claims for CTP and WC 2023/24	13,474	114,111	127,585
2: Medical disputes finalised at PIC 2023/24	4,126	5,191	9,317
3: Panel Reviews / Appeals finalised 2023/24 (total)	971	519	1,490
3.a: Determined by a panel	479	319	798
3.b: Dismissed, settled, withdrawn	492	200	692
(3.a.) divided by (2.)	12%	6%	9%

Item 1 is the total number of claims lodged in the CTP and Workers scheme in 2023/24. The CTP figures are the average of the 2023 and 2024 calendar years.[3] In total there were over 127,000 claims lodged in the year.

Item 2 is then our estimate of the number of medical disputes finalised at the PIC in 2023/24. This is the number of *finalised* disputes and is shown against the number of *all claims lodged* (item 1), which is an inconsistent basis, but is shown simply for the rough scale of disputation.

We have estimated the number of medical disputes based on the figures provided in the PIC Annual Review. In CTP this includes Medical Assessments, while in Workers this includes disputes around WPI and medical expenses. In Workers each dispute can encompass multiple issues; we have scaled medical issues down to the total number of disputes.

Item 3 then shows the number of *medical appeals* finalised in 2023/24. In the CTP scheme these are heard by a Medical Review Panel, and in the Workers scheme these are heard by Medical Appeal Panel. In total there were 1,490 reviews/appeals that were finalised, of which around 800 ended with a determination. The remainder were dismissed, settled, or withdrawn.

The last line shows the number of reviews/appeals that ended with a determination (798) as a proportion of all medical related disputes that were finalised (9,317). Across both schemes, around 9% of medical disputes will end up going to review/appeal and then be finalised through to determination.

The reviews/appeals that end with a determination are **published** on Austlii. We use GenAI to read these published determinations, many of which will relate to WPI assessments, to identify trends.

Importantly, the published decisions are only around 9% of the total number of medical disputes. They are by no means a comprehensive set of decisions that capture information on every WPI dispute, let alone every WPI assessment (some of which will not be disputed).

2.2.4 What is published?

Like the decisions of many courts and administrative tribunals, certain PIC determinations are published on Austlii (www.austlii.edu.au). The following table shows the different PIC series published on Austlii and the average number of decisions per year.

[3] Claim numbers were sourced from SIRA's Open Data portals for CTP and Workers: https://www.sira.nsw.gov.au/open-data-schemes

	Decisions published on Austlii (average per year)
Personal Injury Commission	700
Personal Injury Commission - Presidential Decisions	85
Medical Appeal Panel and Review Panel	850
Merit Review (CTP scheme only)	45
Merit Review Panel	<5

The Medical Appeal Panel and Review Panel (highlighted) has an average of 850 determinations published each year. This is quite similar to the number of appeals/reviews finalised each year that result in a determination (and, therefore, are published).

The following table shows for the Medical Appeal Panel and Review Panel determinations alone, the number by year and the split between CTP and Workers.

Table 2.1 – Number of PIC decisions

Decision	Number of decisions		
Year (calendar)	Workers	CTP	Total
2021	232	10	242
2022	271	158	429
2023	334	355	689
2024	430	433	863
2025 [1]	305	384	689
Total to date	1,572	1,340	2,912

[1] Decisions to October 2025

These are the decisions that we are using GenAI to help us 'read'. In total, there are around 3,000 published decisions from 2021 through to October 2025.

2.3 Other sources of data on claims and WPI

2.3.1 SIRA's Open Data

The State Insurance Regulatory Authority (SIRA) publishes statistical information about the CTP and Workers schemes. In the Workers scheme, there is a specific lump sum statutory entitlement based on the degree of permanent impairment an injured worker has suffered. This is known as a "Section 66" payment, after the relevant section of the *Workers Compensation Act 1987*.[4]

The total volume of section 66 payments made provides a very high level overview of trends in claims receiving WPI assessments that lead to an entitlement to a s 66 payment. The following table sets this out for the entire Workers system. This information has come from SIRA's Open Data portal.[5]

[4] There are also payments known as "Section 67" payments that are made for pain and suffering. These payments are only available to emergency services workers (Police, Fire, Ambulance). They were abolished by the 2012 amendments to the *Workers Compensation Act* but the 2012 amendments did not and still do not, apply to emergency services workers.

[5] Available online at: https://www.sira.nsw.gov.au/open-data/payments-data

Table 2.2 – Volume of s 66 payments, source SIRA

	2022/23	2023/24	2024/25
	$m	$m	$m
Psychological injuries	66	88	95
Physical injuries	214	230	226
Total	281	318	321

There is a clear increase in payments made to injured workers with primary psychological injuries. Payments made to other workers increased to 2023/24 and were then slightly lower in 2024/25. Note that these are total scheme-wide payments (including self-insurers) and do not necessarily reflect the experience of any individual insurer or self-insurer.

As far as we are aware, there is no publicly available information on the *number* of claims seeking or finalising a WPI assessment.

2.3.2 Privately available data

Insurers will obviously have detailed data on their claims, including which claims have been paid a benefit under section 66. This information is not publicly available.

In addition to any payment or numerical data, insurers will also have voluminous information on each claim, in the form of case notes and medical reports. This information is rich in text, but often unstructured. Historically, it has been a challenge to extract information in a cost-efficient way from this unstructured text. If an insurer wanted to investigate a particular topic or potential trend, they would typically undertake a claims file review, having people manually review a sample of cases and record information about each case. They would then extrapolate what they learnt from the case file review.

3 AI and Legal Text Analysis

3.1 Overview of AI capabilities

What do we mean when we say AI?

In our context, we are referring to Generative AI, a subset of AI, that uses Large Language Models (LLMs) to create new output in response to a user's prompt.[6] LLMs are trained on massive amounts of data and return answers based on the most probable outcomes given the context the models derived from the prompt.

In this paper, we are using LLMs to ingest pages of legal text and output information in a structured database format. Beyond capturing the identifying variables of each case, we are prompting the models to determine the roles of participants, the outcomes of the legal process and the presence or absence of certain features that may not be uniformly present or named in the text.

We are relying on these models to do this reasonably *efficiently* and *accurately*.

The use of AI discussed in this paper is not intended to be pioneering, in fact it is decidedly quotidian. The novelty of this paper is in using GenAI to synthesise presently unstructured text-based data to determine if there is any systemic insight to be gained from a cohesive analysis of distinct and unrelated legal decision points. This is the kind of rich data source that *should* tell us something about scheme trends, but in the past has been uneconomic to put together retrospectively.

There already exists GenAI products in the legal field broadly aimed at seeking efficiencies in the way law firms deliver advice and provide these legal services. These models, such as Thomson Reuters "CoCounsel",[7] purport to be an 'agentic' AI model that acts like a legal junior that can summarise documents, 'do' legal research and prepare scaffolds for advice. Real world applications of GenAI in an insurance context are also already in train. Suncorp has announced to the market its "aspiration of automating the end-to-end claims process with agentic AI," and has also trialled an AI CTP claims assessor.[8] The ability of GenAI to leverage insights from unstructured text is upon us. What we outline in this paper is an approach that can be easily repointed to and adapted for any text.

To misquote James Joyce: GenAI has given us the tools to "arrange things in such a way that they become easy to survey and to judge".

3.2 Importance of structured prompting

The results from LLMs are improved by structured prompting. The research and language of 'prompt engineering' is emerging concurrent with practical users developing and learning their own methods alongside ever-improving LLMs.

Anthropic's publicly accessible GenAI model, Claude, identifies, in reply to the importance of structured prompts, the following key principles:[9]

[6] Response from Gemini 2.5. Prompt: "What is Generative AI". Sources identified, IBM, Amazon etc.

[7] https://www.thomsonreuters.com/en/cocounsel

[8] JP Morgan update October 2025.

[9] Response from Claude to the prompt: What is the importance of structured prompting and how best to set out prompts for more accurate results.

- **Be specific and detailed:** instructions could be given across several sentences. Think more broadly here about what common pitfalls, alternative interpretations could arise from your query.
- **Provide context:** audience, goal, role, intended outcome.
- **Use examples:** this can also help with the format and style of the output. Examples could be both positive (do this) and negative (do not do this).
- **Specify formats and constraints:** provide length of output, type of output (string, array, date, number etc.) format of output (DDMMYY) and any constraints (i.e. what is not possible as an outcome).
- **Break complex tasks into steps:** this could give better control over the accuracy of outcomes.
- **Request step by step reasoning:** for complex problems, asking to think through this step by step or show reasoning, encourage more accurate, 'thoughtful' responses.
- **Use clear delimiters:** for longer prompts, use delimiters ("---" or "**") to separate sections of the prompt, combined with the labels for "instruction", "context", "example" etc. This helps the model to parse the prompt more efficiently and accurately.

Essentially, the task of developing prompts is complex and can feel, at times, unnatural. A reader, used to discerning context and inferring meaning from vague and varied text, can parse and evaluate a document overcoming inconsistent presentation and other shortcomings. GenAI, by contrast, can be literal, overconfident, agreeable and inventive if not provided with sufficiently detailed **context**, **instruction** and **constraints**.

The following are three examples of prompts that were strengthened to manage unintended interpretations from the GenAI:

Example 1

To identify if a secondary psychological injury is present in a decision, requires first defining that a secondary psychological condition is a) related to the physical injury that is the subject of the dispute and b) is not a previous primary psychological condition.

If the secondary psychological condition is not part of the dispute and the intention is still to capture its presence, then this eventuality also needs to be reflected in the prompt.

Example 2

To identify if an injured person has had surgery, requires first defining that the surgery must be related to the injuries relevant to the dispute and not be a surgery that a person has had at some point in their life (often times a decision can include a comprehensive medical history of the person, with much not relevant to the dispute).

Example 3

To list the body parts involved in the injury that is the subject of the decision requires specification that even if the panel determines there is no permanent impairment following these claimed for injuries, that the claimed for injuries are still identified and returned. It also required an instruction to avoid repetition (as sometimes injuries can be repeated in the text of the decision with slightly different descriptions).

3.3 Risks of hallucination and how to mitigate them

"Gemini can make mistakes, so double-check it".

"Claude can make mistakes. Please double-check response".

"ChatGPT can make mistakes. Check important info."

These appear at the bottom of the Gemini, Claude and ChatGPT response pages.

GenAI models by their own admission prioritise efficiency and coherence when delivering their responses and this can come at the expense of accuracy and reliability.[10] Put simply, GenAI models can present confidently as fact, something it may have made up.

Hallucinations can be factual (fabricating people, events, data and sources) and reason-based (illogical reasoning)

3.3.1 Hallucination example

To illustrate this risk, we were able to simply and quickly engender a hallucination from Claude. Our initial two prompts were to ask for the number of separate judgements handed down in a recent High Court decision (outside its training data window of January 2025, though Claude states it can do more recent internet searches when required).

[10] Response from Gemini 2.5

tell me the majority justices in the high court decision of Farmer v Minister for Home Affairs [2025] HCA 38 (15 October 2025)

I'll search for information about that specific High Court case.

Farmer v Minister Home Affairs 2025 HCA 38 — 10 results

Based on the search results, the High Court unanimously held that s 501(6)(d)(iv) of the Migration Act 1958 (Cth) does not infringe the implied freedom of political communication (Federal Court of Australia) in *Farmer v Minister for Home Affairs* [2025] HCA 38.

The Full Court bench consisted of: Gageler CJ, Gordon, Edelman, Steward, Gleeson, Jagot, and Beech-Jones JJ (Hcourt).

Since the decision was unanimous, all seven justices comprised the majority—there were no dissenting opinions in this case.

how many separate judgements were there and who were the authors of each

I'll search for more detailed information about the separate judgments in this case.

"Farmer v Minister for Home Affairs" HCA 38 2025 separate judgments — 3 results

Let me fetch the full judgment PDF to see the structure of the separate judgments.

Failed to fetch https://www.hcourt.gov.au/sites/default/files/eresources/2025-10-15/HCA/Far...

Farmer Minister Home Affairs HCA 38 judgment Gageler Gordon Edelman authors — 10 results

Based on the search results, there was one joint judgment in *Farmer v Minister for Home Affairs* [2025] HCA 38 delivered by Gageler CJ, Gordon, Edelman, Steward, Gleeson, Jagot and Beech-Jones JJ (High Court of Australia).

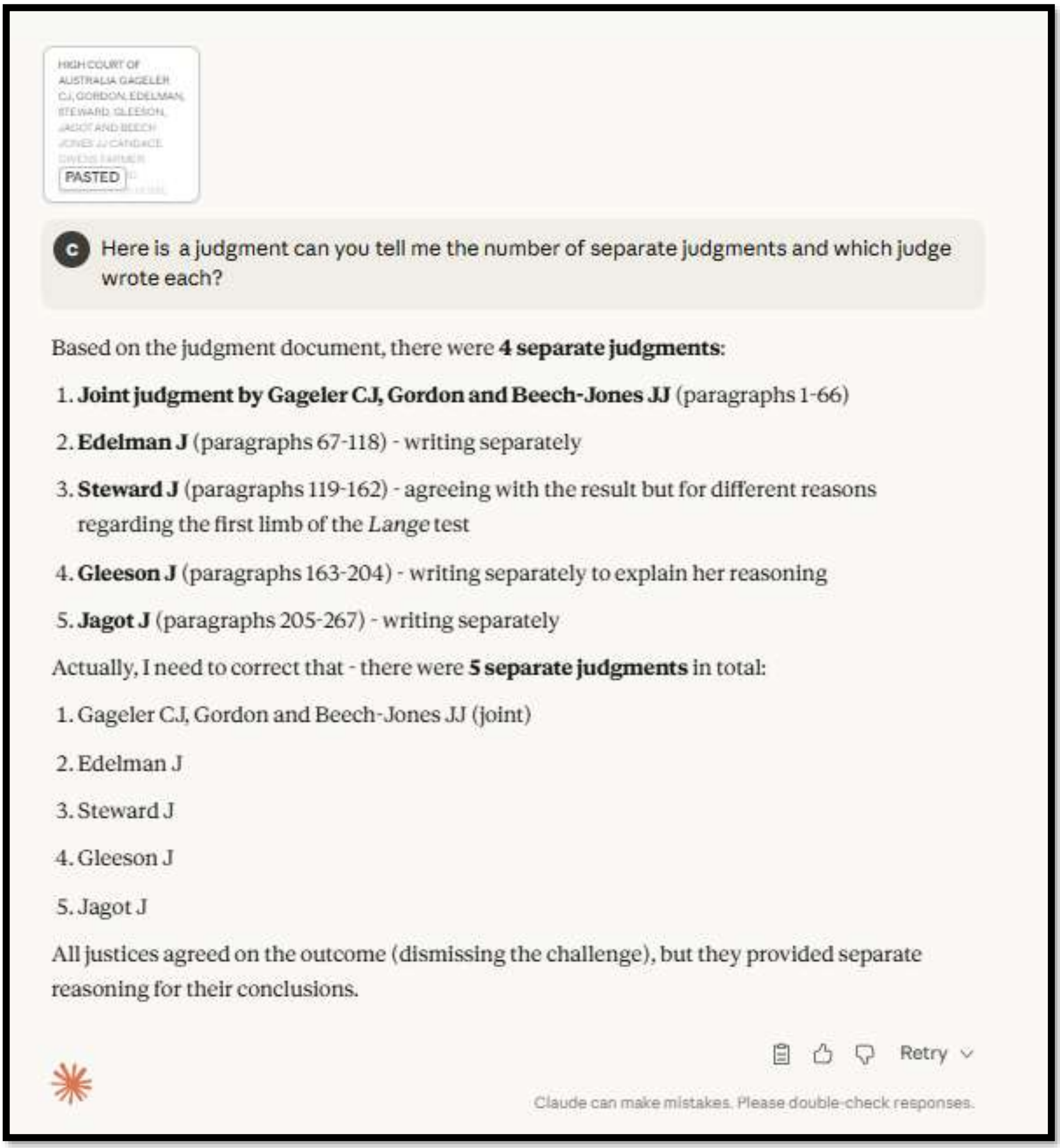
The case was decided unanimously with a single joint judgment rather than multiple separate judgments. All seven justices agreed on both the reasoning and outcome, finding that section 501(6)(d)(iv) of the Migration Act did not infringe the implied freedom of political communication and that the Minister's decision was valid.

So to answer your question directly:

- **Number of separate judgments:** 1 (a single joint judgment)
- **Authors:** All seven justices jointly - Gageler CJ, Gordon, Edelman, Steward, Gleeson, Jagot and Beech-Jones JJ

Claude very confidently asserts that all seven High Court justices agreed on the outcome, which is true, and that they wrote "a single joint judgement rather than multiple separate judgements" which is completely false.

We then uploaded a copy of the High Court judgement we had mentioned to Claude.



HIGH COURT OF AUSTRALIA GAGELER CJ, GORDON, EDELMAN, STEWARD, GLEESON, JAGOT AND BEECH-JONES JJ CANDACE

PASTED

Here is a judgment can you tell me the number of separate judgments and which judge wrote each?

Based on the judgment document, there were **4 separate judgments:**

1. **Joint judgment by Gageler CJ, Gordon and Beech-Jones JJ** (paragraphs 1-66)
2. **Edelman J** (paragraphs 67-118) - writing separately
3. **Steward J** (paragraphs 119-162) - agreeing with the result but for different reasons regarding the first limb of the *Lange* test
4. **Gleeson J** (paragraphs 163-204) - writing separately to explain her reasoning
5. **Jagot J** (paragraphs 205-267) - writing separately

Actually, I need to correct that - there were **5 separate judgments** in total:

1. Gageler CJ, Gordon and Beech-Jones JJ (joint)
2. Edelman J
3. Steward J
4. Gleeson J
5. Jagot J

All justices agreed on the outcome (dismissing the challenge), but they provided separate reasoning for their conclusions.

Retry

Claude can make mistakes. Please double-check responses.

Apart from an initial counting mishap, when we fed the judgement to Claude it correctly identified that there were 5 concurring judgments and correctly identified the justices.

We have found this across Gemini, Claude and ChatGPT – if we ask it to search the web for a specific judgment it will often hallucinate aspects of the judgment. This seems to be the case even where models search the web for information. But when we upload the judgment in a document directly to the model and then ask questions about the document, we get more accurate responses.

3.3.2 Mitigation

We have found that hallucinations can be mitigated by:

- providing structured prompts that lay out clear instructions, context, examples, constraints and the desired format of any response,
- iteratively testing the outputs from the prompts and refining,
- asking GenAI models to review the quality and clarity of prompts,
- if relevant and possible, using a "Retrieval Augmented Generation" (RAG) to augment the prompt, and
- on a smaller scale than RAG methods – providing the actual text that you would like the GenAI to parse.

Given the pace of development, we have found it worthwhile to test across a number of GenAI models and compare outputs.

Iterative testing

If the prompt is intended to be applied to numerous datapoints, a user could test the output on a limited dataset and refine as necessary.

In our case, iterative testing helped to refine the majority of prompts used in the development of our NSWPICMP database, including, identifying the presence of secondary psychological conditions, the number of surgeries, the party successful, the degree successful, the number of body parts affected and the presence of consequent injuries and what they were.

Iterative testing was most useful in realising the 'errors' or different interpretations that could arise from weak or vague prompts.

Asking for GenAI's input

Asking GenAI for feedback on the clarity and structure of a prompt is a further refinement available. This was done individually for a number of prompts.

We had mixed results when asking GenAI models for feedback on prompts. In one instance, it oversimplified the prompt (by removing the different naming conventions for parties between Workers and CTP decisions) and in other prompts, it 'forgot' the format of the decision text (pdf) and so added extraneous formatting detail for the output (html). These were easily resolved issues but indicate that any output needs to be reviewed. If a user has a fair understanding of how to construct a prompt and a sense of the range of possible outcomes from their text, then there is little additional to be gained from asking for further input from the GenAI that cannot be informed by testing the actual output of the prompts.

RAG methods

RAG allows GenAI models to improve the initial prompt by first searching domain and context-specific databases to 'augment' the initial prompt with additional context and so return a more accurate answer that is built on the domain-specific database and the LLM's own training data.

In this paper we have not relied on RAG methods but by providing the 'source text' (the NSWPICMP decisions) and asking the GenAI model direct questions in response to the text our approach can be considered 'RAG-lite'.

4 Methodology

4.1 Data Collection

As noted previously, we are using the published determinations from the PIC Medical Appeal Panel and Medical Review Panel. We are using all decisions available from Austlii, as shown in the table:

Table 4.1 – Number of PIC decisions

Decision Year (calendar)	Number of decisions Workers	CTP	Total
2021	232	10	242
2022	271	158	429
2023	334	355	689
2024	430	433	863
2025 [1]	305	384	689
Total to date	1,572	1,340	2,912

[1] Decisions to October 2025

These decisions are all publicly available. They are text based, available either as a web page (HTML), or as an RTF or PDF document. Note, there were a handful of decisions that errored in extraction and we have not included these (less than 20 decisions).

All of the determinations are set out in the same format, as shown in the following Workers example:

New South Wales

Personal Injury
Commission

DETERMINATION OF APPEAL PANEL

CITATION:	**Westpac Banking Corporation v Albertsen [2025] NSWPICMP 790**
APPELLANT:	Westpac Banking Corporation
RESPONDENT:	Susie Albertsen
APPEAL PANEL	
MEMBER:	Marshal Douglas
MEDICAL ASSESSOR:	Graham Blom
MEDICAL ASSESSOR:	Michael Hong
DATE OF DECISION:	14 October 2025
CATCHWORDS:	WORKERS COMPENSATION - *Workplace Injury Management and Workers Compensation Act 1998*; review of Medical Assessment Certificate (MAC); whether a report on surveillance the appellant had done of the respondent established the ground for appeal provided in section 327(3)(a); whether the ratings the Medical Assessor (MA) made for the respondent's impairment in social and recreational activities and social functioning were based on a correct history; *Held* – the surveillance report was not additional relevant information and accordingly did not establish the ground for appeal provided in section 327(3)(a); the MA's ratings of the respondent's impairment in social and recreational activities and social functioning were not supported by the evidence; respondent re-examined; MAC revoked.

The CTP determinations also include a short summary of "Determinations Made". Following that, both the Workers and CTP determinations have a body of paragraphs that can run to tens of pages. There is a significant amount of information embedded in the text, but all in a completely 'unstructured' way (from a computational point of view – there is obviously sophisticated linguistic and legal structure!).

4.2 AI Processing

4.2.1 Model design

At a high-level, our model used Python script to download as pdfs decisions published on the Austlii[11] website from the NSWPICMP. A second Python script then ingested these individual pdf decisions into Gemini 2.5, which was then asked to extract information in response to 50 different prompts detailed in an excel-based schema that was also ingested. This process of ingestion and extraction occurred individually for each decision, thereby improving the accuracy of the output - Gemini was asked to 'read' and 'respond' to each specific decision. Additionally, some of the prompts enforced accepted values as the only output available to Gemini, thereby allowing the dataset to be more helpfully structured with categorial variables. The extracted information was then returned in csv format as a dataset with a row record per decision.

Also output were logs, showing the success of each download and the success of the extraction.

For each decision, output separately was a csv file, capturing Gemini's response and against each prompt Gemini's 'accuracy score' and a comment justifying its extraction decision. This made reviewing and understanding the results of individual decisions easier.

We used Finity's ***Identifi*** platform to interact with Gemini at scale. Identifi is a plug-and-play engine that can be added to any ingestion workflow and is specifically designed to take in large amounts of unstructured text, feed it to an LLM with structured queries and return structured information based on user defined queries. It also includes pre-built error checking and logging.

While Identifi allowed us to do this quickly and efficiently at scale, you could achieve the same result by downloading a case as a PDF or RTF file, opening up Gemini in a browser, uploading the PDF or RTF file, typing in prompts and then saving the answers that are returned in a spread sheet. In total, we downloaded and queried Gemini on around 3,000 decisions, so Identifi allowed us to automate (at speed) the input/output interaction with Gemini.

4.2.2 Prompt design and iteration

Developing and refining the prompts was an iterative process. The prompts were iterated and refined over five to ten cases and outputs were reviewed for their reasonableness.

Refinement focused not just on the accuracy of the output but also the preferred format of the outputs, requirements to avoid repetition in the output and the seeming difficulty that Gemini had with enforcing 'yes/no' answers as accepted values.

The prompts were informed by our domain knowledge of NSW accident compensation schemes.

The prompt structure for each prompt was set out as follows:

Field Name: <name of field to be extracted>

DataType: (choice of: number/string/array)

Context:

[11] Australasian Legal Information Institute, https://www.austlii.edu.au/cgi-bin/viewdb/au/cases/nsw/NSWPICMP/

description of field <enter brief sentence on description of field>

Extraction instructions <a few sentences explaining the conditions for extraction, detail any constraints and output format, listing any keywords that may be useful>

Example: <example output>

Classification rules: <additional instructions to classify output from a list of accepted values>

Accepted Values: <defined categorical variables>

The spaces, labelling and formatting of the prompt help GenAI more accurately and quickly parse information.

4.2.3 Extraction of structured data

The following table shows the data fields that were populated for each decision. Some fields are more important than others, some fields are more difficult to ascertain than others and some fields were used as a check on other fields.

Table 4.2 – data fields

Identifying facts of the decision	Dates	Facts about injured person and injury	WPI information
Decision name	Date of decision	Name of injured person	Initial WPI score
Decision citation	Date of injury	Broad injury classification (physical/psychological)	Final WPI score
Name of PIC member	Date of initial medical assessment certificate	Sex of injured person	Deductions to WPI score
Names of Medical Assessors	Date of reissued medical assessment certificate	Brief description of mechanism of injury	PIRS class
Name of appellant	Current age of injured person	Employer (for Workers)	PIRS score
Name of respondent	Age at injury of injured person	Body systems impacted	PIRS domain scores by domain
Appellant role (worker/employer or claimant/insurer)		Number of body systems impacted	Injured person claimed for WPI
Respondent role (worker/employer or claimant/insurer)		Pre-existing injuries flag	Insurer claimed for WPI
Was the decision WPI related		Presence of secondary psychological issues if a primary physical injury	

Party successful		Description of secondary psychological condition - brief sentence to describe how secondary psychological issues present in the person.	
Degree of success (whole/partial)		Had the individual had surgery as a result of their injury	
Brief description of outcome of the decision		Brief description of surgeries	
Capturing of the 'Catchwords' section of each decision		Number of surgeries	
Venue (Appeal Panel or Review Panel)			
Remote/in-person assessment flag			

4.3 Validation and Auditing

4.3.1 Overview of how we checked

Our approach to reviewing the outputs from the GenAI model was to audit a sample of decisions. Seeing as the GenAI is 'reading' through each decision and answering a series of questions, we randomly chose a sample of decisions, read through them ourselves and asked ourselves 15 of the same questions. We then compared our answers against the GenAI produced answers.

Our sample was chosen as follows:

- 14 CTP decisions and 16 Workers decisions
- A roughly similar number of decisions from each year 2021 to 2025 (i.e. 2 or 3 CTP and 2 or 3 Workers decisions from each calendar year).

The 30 decisions represented around 1% of all published decisions.

We spent around 8 hours in total reviewing the 30 decisions.

4.3.2 Comparison of GenAI outputs with human interpretation

The results of our review are shown in the following table:

Field	AI was correct	AI partly correct	AI was wrong	Total	Fully correct
Jurisdiction (CTP or Workers Compensation)	30	0	0	30	100%
Appellant	30	0	0	30	100%
Appellant role (claimant, employer/insurer)	30	0	0	30	100%
Respondent	30	0	0	30	100%
Respondent role (claimant, employer/insurer)	30	0	0	30	100%
Party successful	27	1	2	30	90%
Degree of success (partial, full)	22	5	3	30	73%
Date of injury	29	0	1	30	97%
Nature of injury (physical or psychological)	30	0	0	30	100%
Was secondary psych an issue?	26	2	2	30	87%
Final WPI score on appeal / review	29	0	0	29	100%
Was there a deduction for pre-existing conditions?	28	0	1	29	97%
WPI score on initial assessment	25	3	1	29	86%
Total	366	11	10	387	
Proportion	95%	3%	3%		95%

Note that some questions have fewer responses than others. This is because the questions were not relevant to every single decision. We are also comparing the 'final' answers produced by the GenAI, after we had refined the prompts as noted in section 4.2.

We found our answers agreed with the GenAI answers in most cases. There were a small number of differences. These tended to be because either:

- the information to answer the question was available, but quite difficult to identify. For example, a case disputing the degree of WPI may have referred to the original assessment and noted that the original assessment held the original injuries were soft tissue injuries and likely to improve, however, the decision did not explicitly state the original assessment found 0% WPI - we needed to infer this. The GenAI did not always do this.
- there was no single correct answer to the prompt/question.

Legal decisions are complex, particularly where they also incorporate specialised medical assessments. This led to a few examples where you need to take care with the prompts/questions being asked (some of which we dealt with in prompt design, some of which appear in our audit table, above, as differences to the GenAI response). The following are some further observations from our audit.

Observation 1

We had designed our prompts on the assumption that each decision would have involved one claim and therefore have one accident date. Occasionally, however, a PIC decision around WPI will involve *multiple* claims (made by the same injured worker) at the same time. In this situation, as a human, we treated the most significant claim as the relevant injury and used this to determine the "date of injury", while the GenAI simply took the date of the first injury. In this situation the GenAI was not wrong – the question simply had multiple possible correct answers.

Observation 2

Another example was the "party successful". If, in a Workers case, the original decision found a WPI of 13%. At this WPI, the injured worker is just shy of being able to lodge a common law claim (eligibility for common law requires a WPI of 15% or higher). If the worker appeals the decision and claims that the WPI should be higher, who is “successful” if the Medical Appeal Panel finds:

- the original WPI assessment was defective for a particular reason and must be revoked and re-issued, but,
- the final WPI should be 14%?

On one hand, the injured worker was successful, because they ‘won’ as the original WPI assessment was revoked and reissued at a higher level. On the other hand, the employer/insurer is successful because the final WPI remains below the 15% threshold that allows access to common law. As a result, we also capture the ‘degree of success’, but there is not always an obviously correct answer.

Observation 3

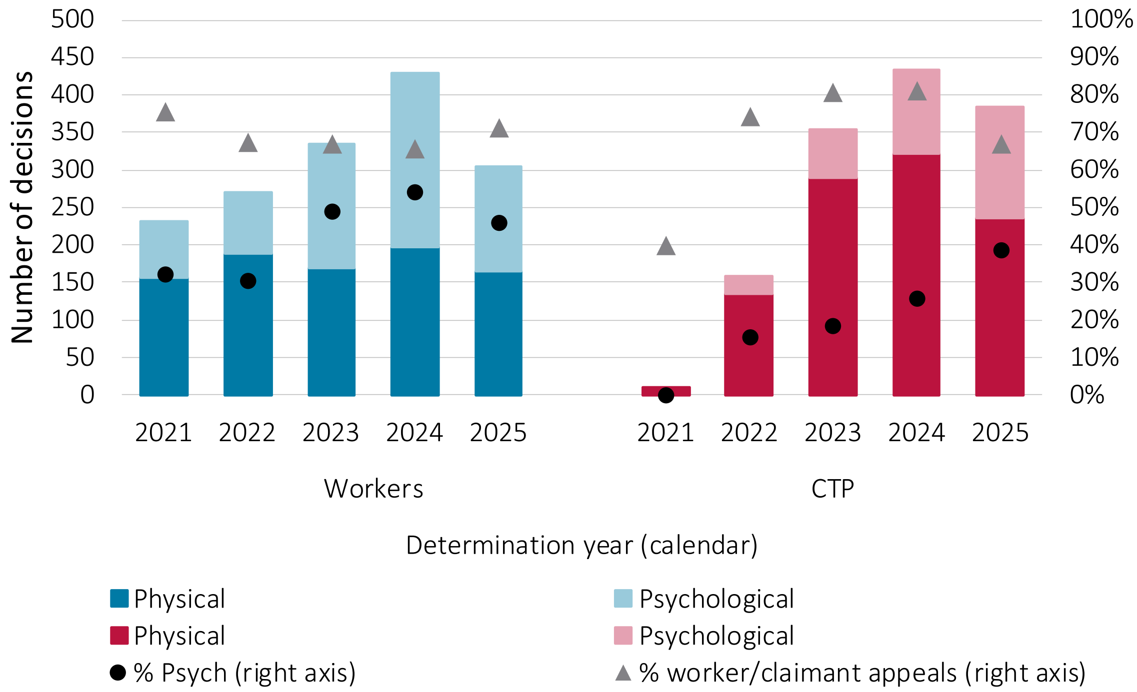
The prompt for the degree of success was (in hindsight) too narrowly defined and relied too much on using the change in WPI as the primary way of differentiating the degree of success (we only allowed the responses to be partially or wholly successful). In our audit, this prompt was the most inaccurate (score of 73%). The prompt gave rise to some confusion in the GenAI model when the dispute did not relate to WPI (and say related to threshold injuries in CTP), resulted in no change to the WPI score. These types of complexities would possibly also lead to a range of outcomes from human reviewers.

5 Results and Analysis

5.1 Extracted Data Overview

In this section we highlight some of the core contextual information we were able to extract from NSWPICMP decisions. Figure 5.1 sets out the number of decisions, split by whether the injury was physical or psychological, and who lodged the appeal or review. Nearly all of the appeals deal with assessment of WPI, and in both the Workers and CTP schemes physical and psychological injuries are assessed separately for the purposes of WPI. Where a claimant has both a physical and psychological injury that leads to a WPI assessment, the 'most severe' injury will be used for the purposes of determining entitlement to benefits.

Figure 5.1 – Number of appeal/review decisions by primary injury



We observe the following:

- It took longer for CTP disputes to 'flow through' the appeal/review process, with volumes increasing substantially to 2023.
- Psychological claims (the black circles for each year) have increased across both types of accident compensation. In 2025 they now represent more than 50% of appeals in Workers and almost 40% of CTP.
- Workers/claimants are the most likely to initiate appeals; around 70% in Workers and 80% in CTP. Though there has been some movement by determination year, these proportions are broadly consistent.

Figure 5.2 and Figure 5.3 show the appeal success rates by injury type for CTP and Workers and by which party initiated the appeal.

Figure 5.2 – Appeal success rates by scheme and injury type for worker/claimant initiated appeals

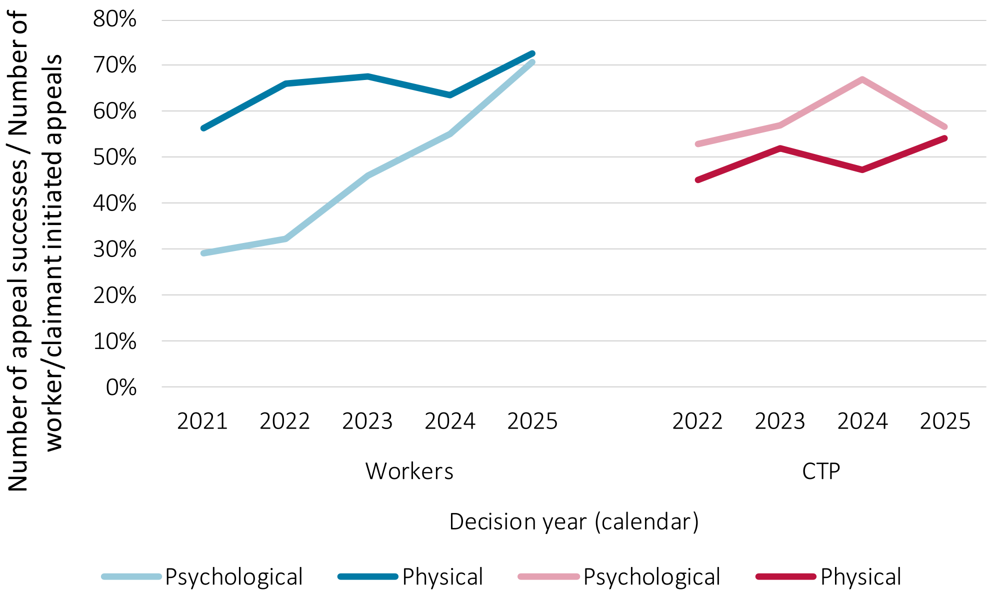
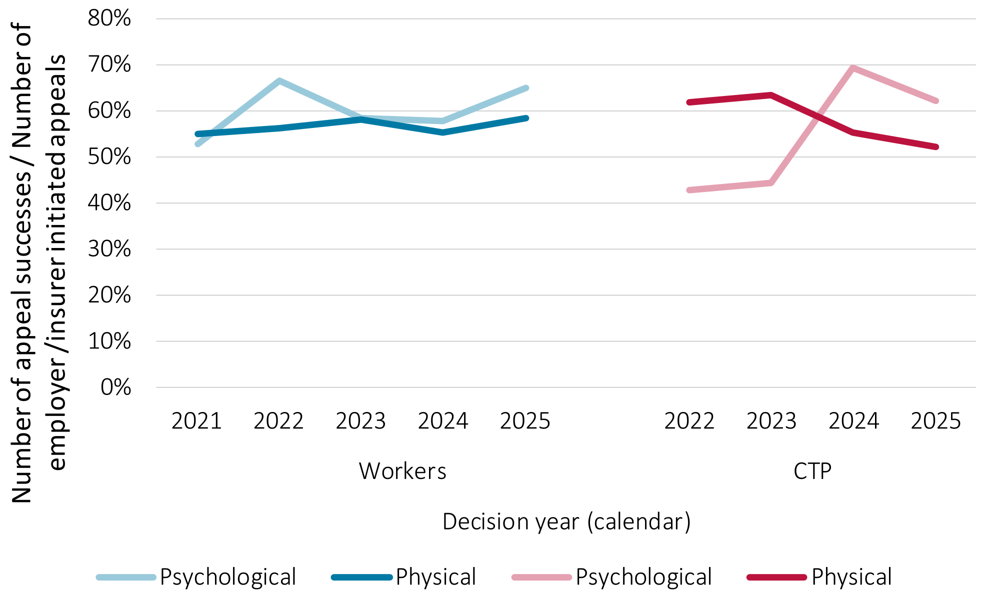


Figure 5.3 – Appeal success rates by scheme and injury type for employer/insurer initiated appeals



We observe the following:

- For worker initiated appeals, there is an upward trend in appeal success rates by decision year for both physical and psychological injuries – though the trend is much stronger in psychological claims.
- Success rates for worker initiated appeals are (now) around 70%, that is workers are successful in 70% of cases where they initiate the appeal.
- For claimant (in CTP) initiated appeals, there is possibly a slight upward trend in appeal success rates for both physical and psychological claims.
- Success rates for claimant initiated appeals are around 55%.

- The success rates for employer/insurer initiated appeals are generally lower than for worker/claimant initiated appeals. The success rates for employer initiated appeals are around 60% and broadly stable. For CTP, the success rates for physical claims have decreased from above 60% to 50% and for psychological claims, have increased to above 60%. Note, the volumes of these appeals are smaller in number – sometimes less than a dozen determinations in a year. Care should be taken in interpreting these trends.

5.2 Insights from PIC Decisions

In this section we offer some examples of the types of 'deep dives' we undertook into our constructed dataset. These are not exhaustive. Interpretation of our observations requires care. This is because the underlying database is a special subset of PIC decisions (reviews/appeals only) and is therefore likely to be subject to certain biases. What can be taken from this section with certainty is the variety of the insights that can be gleaned from the creation of this novel dataset.

5.2.1 The initial and final WPI score

Figure 5.4 and Figure 5.5 shows the distribution of determinations by initial and final WPI (i.e. the WPI disputed as part of the appeal and the WPI determined by the appeal) for Workers and CTP. The first chart shows this information for primary physical injuries and the second chart for primary psychological injuries. The distribution is based on determinations from the 2023 calendar year to October 2025. Note, we have attempted to limit our analysis to the decisions relating to WPI (which is the majority).

Figure 5.4 – Change in WPI distribution by band, physical injuries

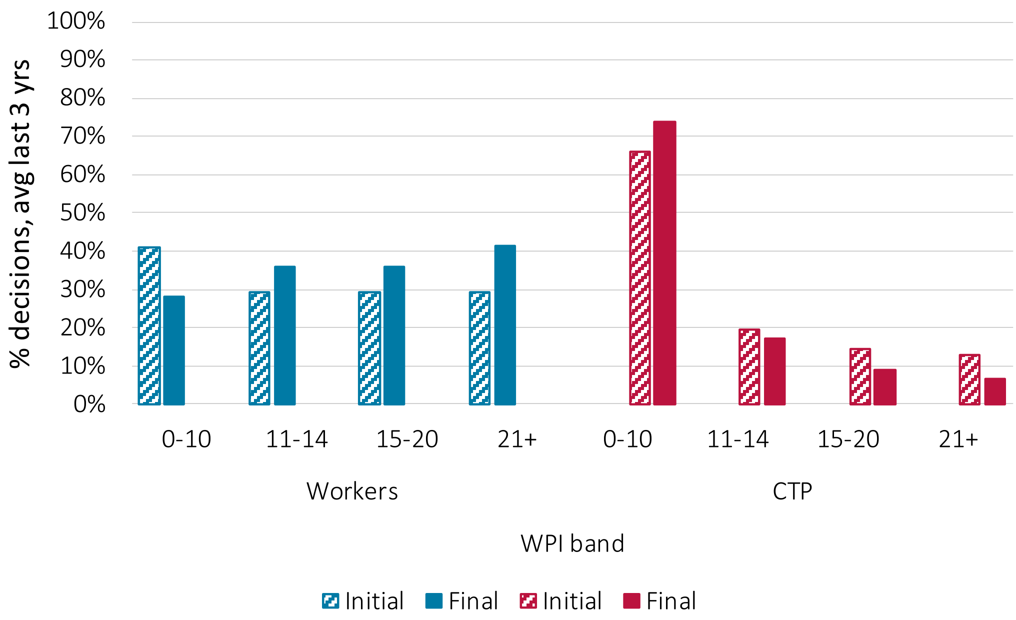
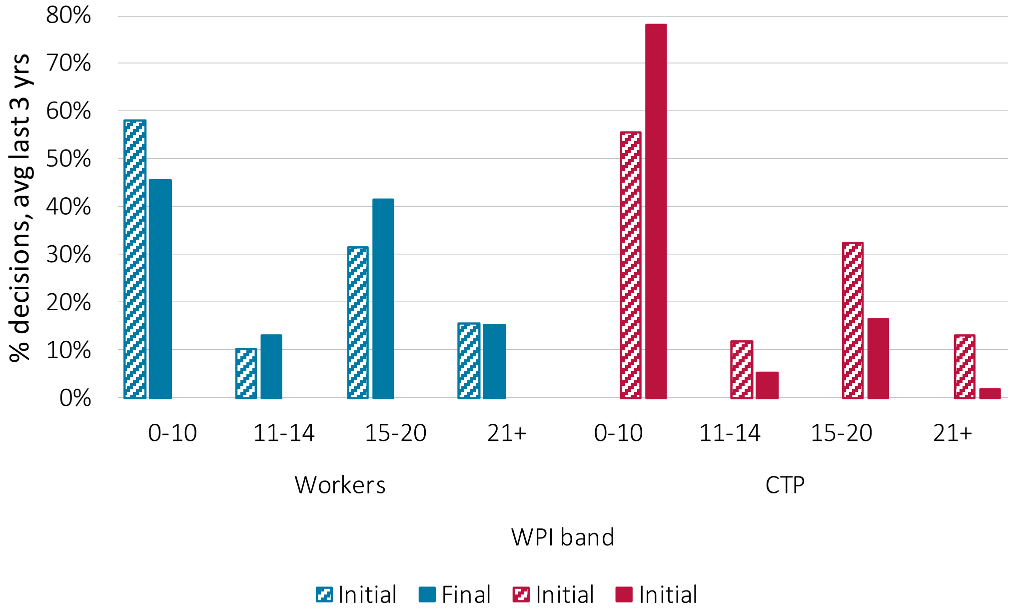


Figure 5.5 – Change in WPI distribution by band, psychological injuries

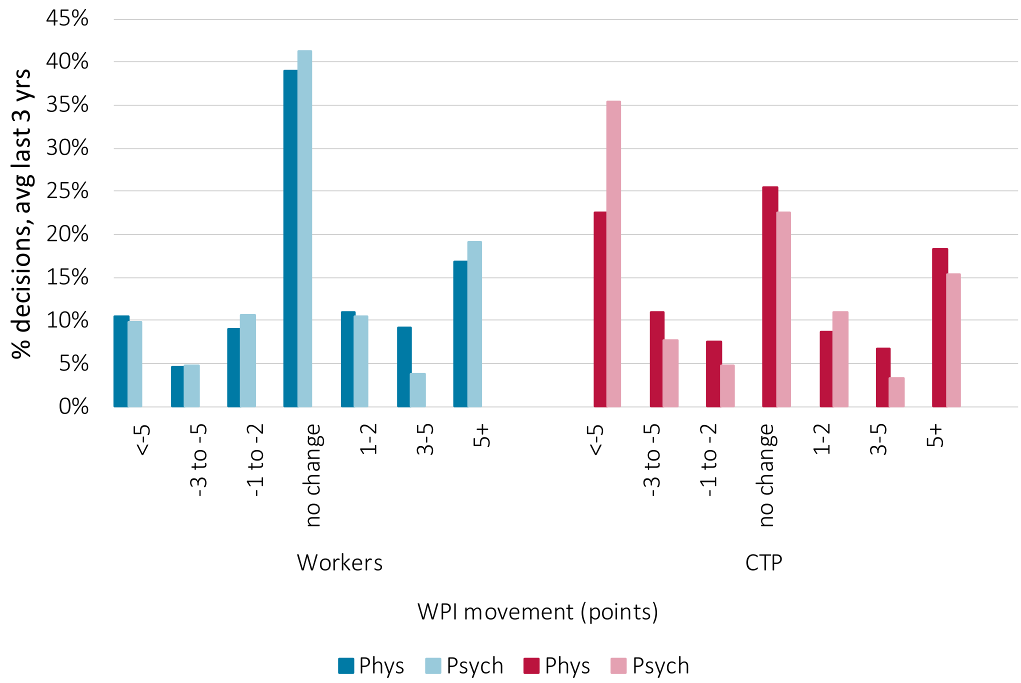


We observe the following:

- For Workers, appeals tend to lead to an upward shift in the distribution of WPIs toward higher WPI scores. This is the case for both physical and psychological claims.
- For CTP, appeals tend to lead to a downward shift in the distribution of WPIs toward lower WPI scores. This is the case for both physical and psychological claims.

Figure 5.6 shows the difference between the final WPI score compared to the initial score. We show this separately by injury type and scheme.

Figure 5.6 – WPI movement before and after appeal by WPI points



We observe the following for Workers:

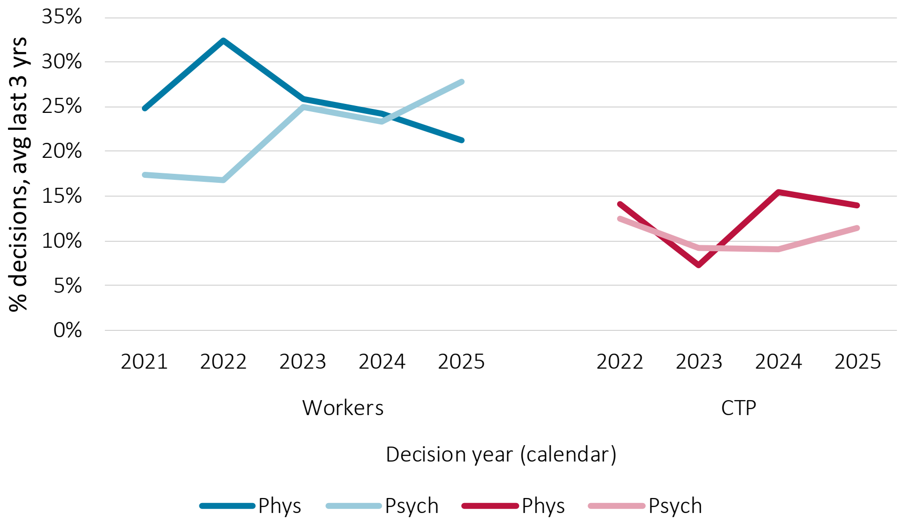
- A high proportion of appeals in Workers result in no change to the WPI score, around 40%.
- The movement in WPI scores is broadly similar for both physical and psychological decisions.
- The movement in WPI scores is slightly more positive (i.e. WPI scores tend to move upward). Though 25% of appeals still do lead to a lower WPI.

We observe the following for CTP:

- There are three 'peaks': around 25%/20% (physical/psychological) of appeals led to no change in the WPI score, 23%/35% a more than a five point decrease in WPI and slightly less than 20%/15% with a more than five point increase in WPI.
- For CTP, psychological injuries show more movement.
- There is greater downward adjustment in WPI scores - almost 50% of appeals lead to a lower WPI score.

Figure 5.7 focuses on key damages thresholds and the appeals that were initially below these thresholds and then as a result of the appeal move above the key threshold. For Workers the key threshold is WPI 15%+ and for CTP it is WPI 11%+.

Figure 5.7 – Determinations that move above key damages thresholds



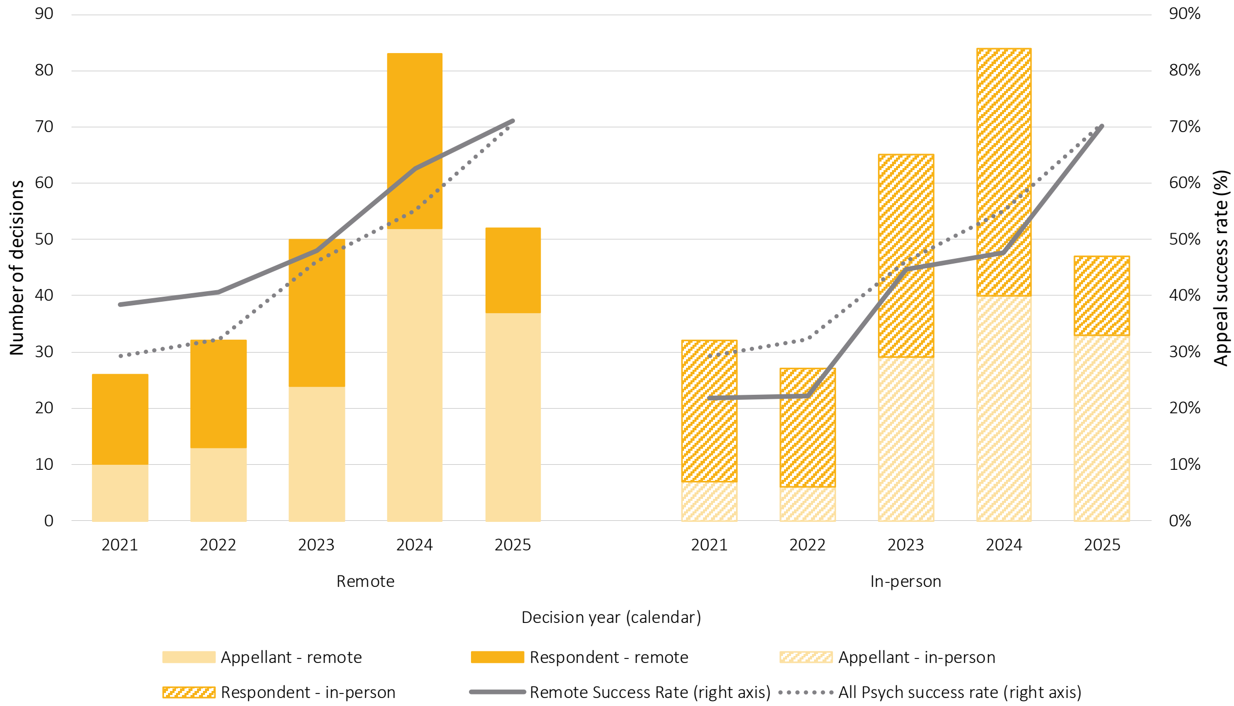
We observe the following:

- For Workers, an upward trend in psychological appeals that move above threshold, from <20% to almost 30%.
- For Workers, a downward trend in physical appeals, from >30% to almost 20%.
- For CTP, trends are less distinct but possibly slightly upward for all injury types.
- Fewer CTP appeals lead to a threshold movement (taking WPI above 11%), compared to >20% for Workers.

5.2.2 Remote assessments of WPI for psychological claims

Figure 5.8 and Figure 5.9 show the number of worker/claimant initiated appeals by the party successful, the venue of WPI assessment (remote or in-person) and the appeal success rates. We compare the appeal success rates to the overall success rate for psychological claims (for worker/claimant initiated appeals). By necessity, nearly all WPI assessments for physical claims occur in person and so are not shown.

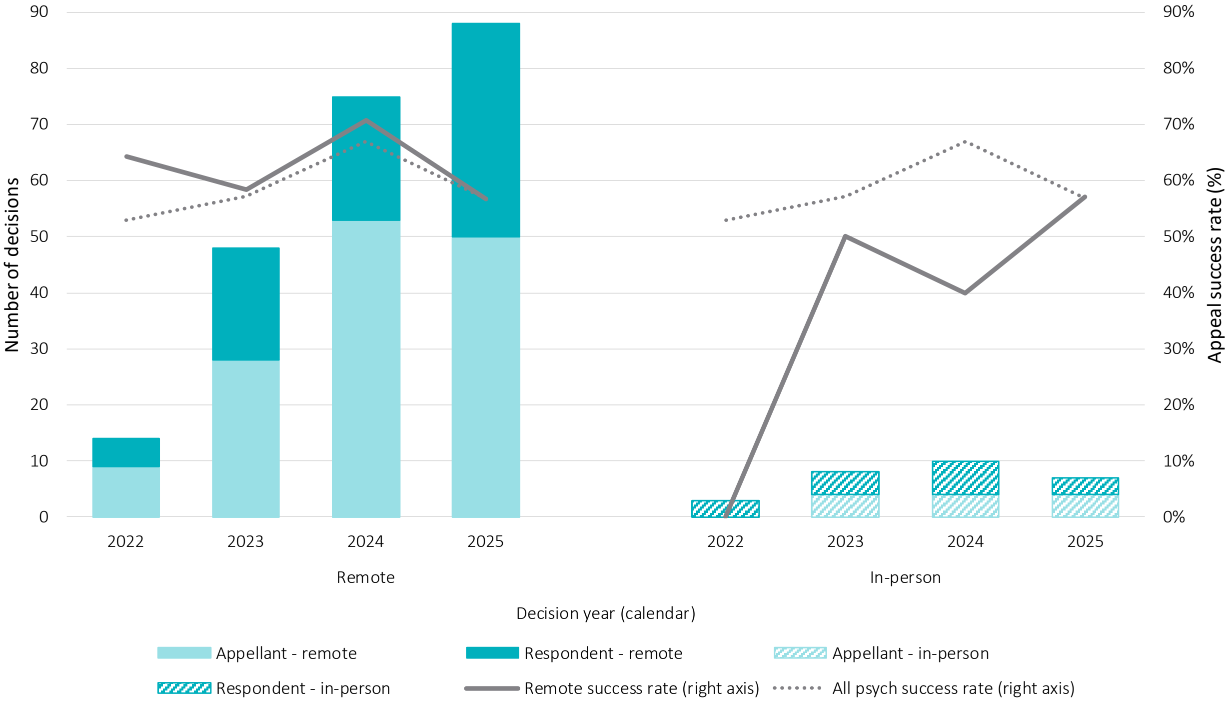
Figure 5.8 – Workers: Appeal success rates for worker initiated appeals for psychological claims by WPI assessment venue



We observe the following:

- The volume of remote and in-person assessments is fairly evenly split.
- Appeal success rates are consistently 5-10% higher for remote WPI assessments compared to in-person.
- For both remote and in-person, the appeal success rates are increasing.

Figure 5.9 – CTP: Appeal success rates for claimant initiated appeals for psychological claims by WPI assessment venue



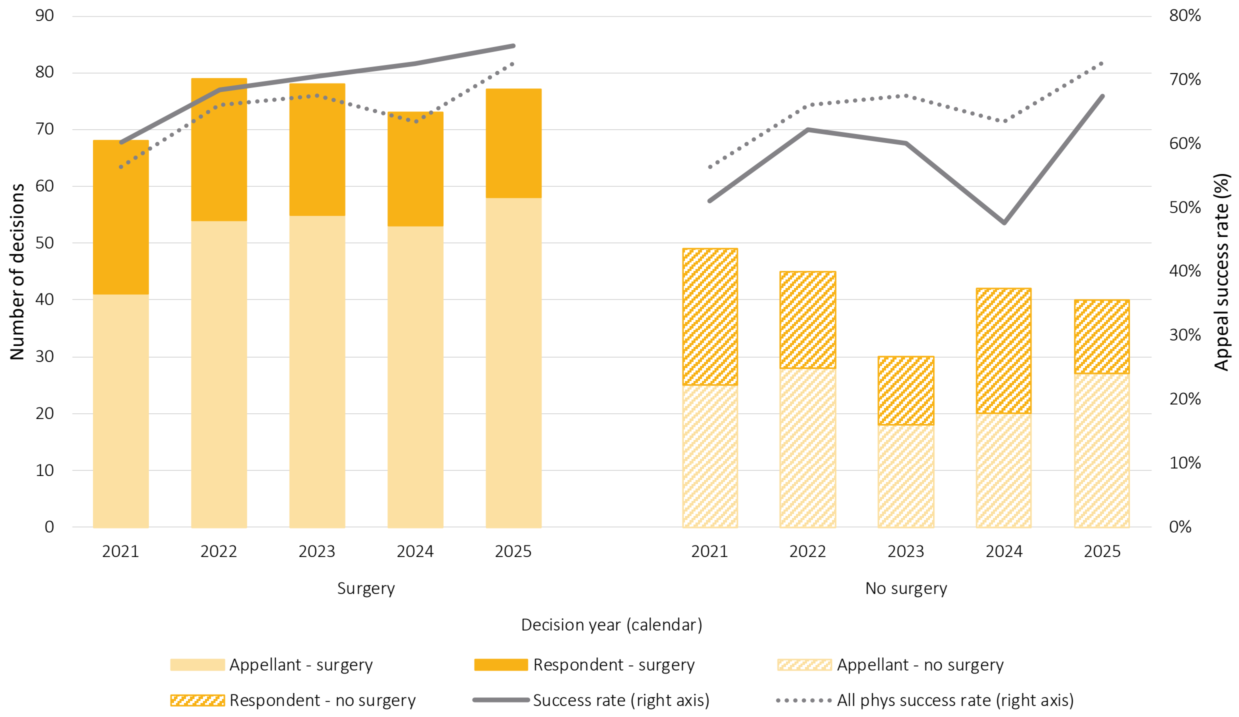
We observe the following:

- Most psychological WPI assessments in CTP appear to be remote.
- Given the small number of in-person assessments (<10 per year), it is difficult to comment on differences in appeal success rates.

5.2.3 The presence of surgeries

Figure 5.10 and Figure 5.11 show the number of worker/claimant initiated appeals by the party successful and compare the appeal success rates by whether the injured person had surgery as a result of their injury. We compare the appeal success rates to the overall success rate for physical claims only.

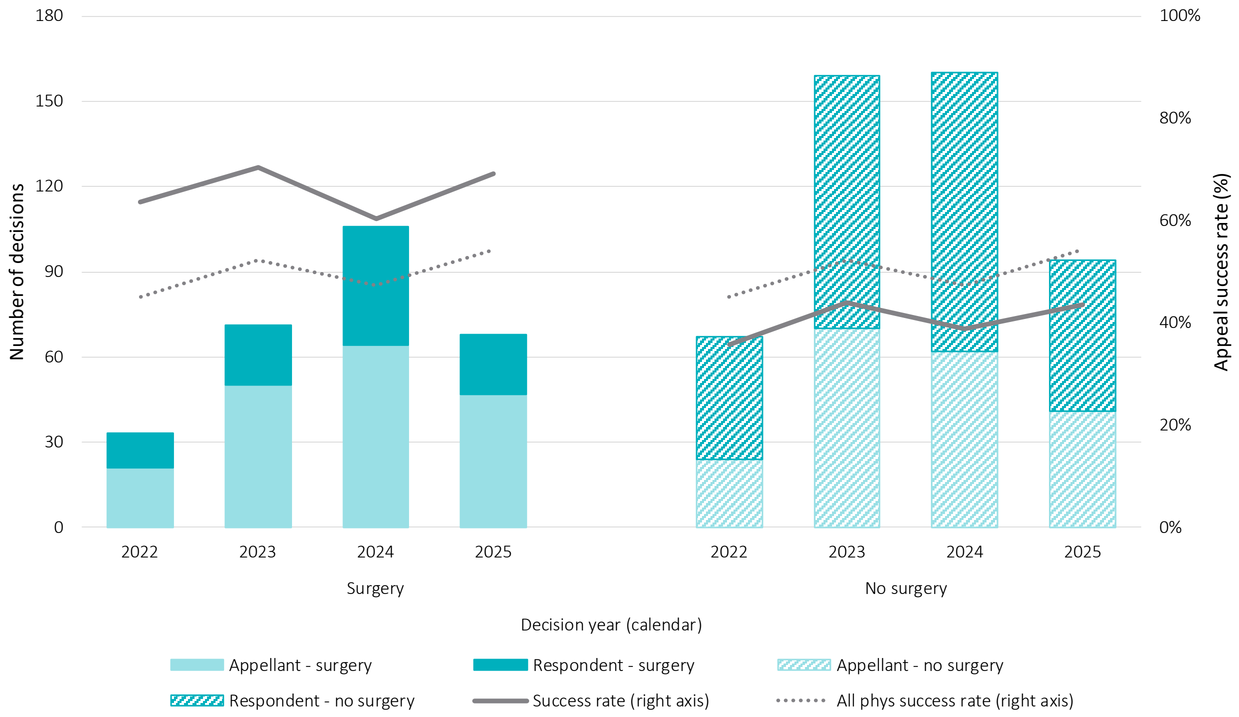
Figure 5.10 – Workers: Success rates for worker initiated appeals by surgery status for physical claims



We observe the following:

- Around 75% of worker initiated appeals with physical injuries have had surgery.
- Success rates for workers with surgery are consistently 2-10% higher than the success rate for physical injury related decisions.
- Success rates where the worker had no surgery are on average more than 10% lower than for injured persons with surgery.
- There is an upward trend in success rates for both decisions that involve injured persons with and without surgery.

Figure 5.11 – CTP: Success rates for claimant initiated appeals by surgery status with physical injuries



We observe the following:

- Around 40% of claimant initiated appeals from 2024 have had surgery. This has increased from 30%.
- Success rates for claimants who have had surgery are consistently 15% higher than the success rate for physical injury related decisions.
- Success rates for where the injured person did not have surgery are on average more than 10% lower than for injured persons with surgery.
- There is possibly a slight upward trend in success rates for both decisions that involve injured persons with and without surgery.

Figure 5.12 and Figure 5.13 show the initial and final WPI distributions for Workers and CTP, separately for whether the injured person did or did not have surgery. The distributions are an average of the last three decision years.

Figure 5.12 – Initial and final WPI distribution for Workers by surgery status for physical injuries

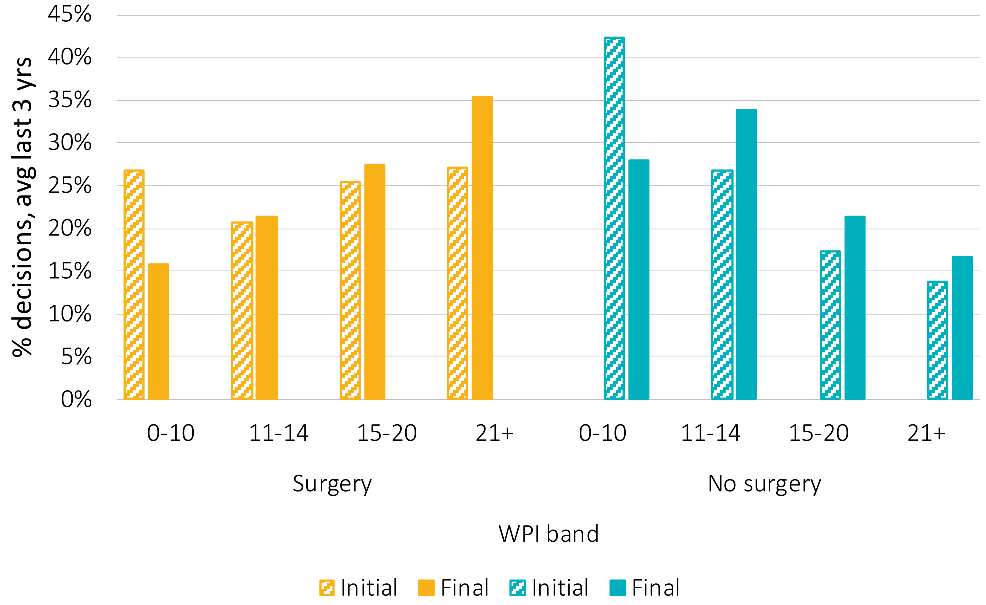
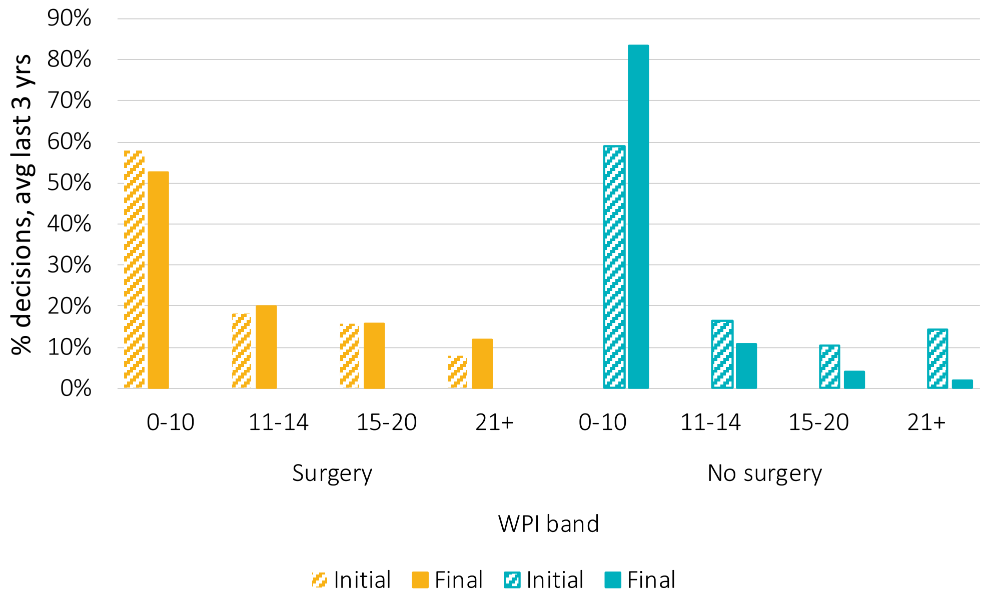


Figure 5.13 – Initial and final WPI distribution for CTP by surgery status for physical injuries



We observe the following:

- The final WPI distribution generally shifts to higher WPI scores for both injured persons with and without surgery. However, for CTP, where there was no surgery, the final WPI distribution tended to shift downward.
- As would be expected, those with surgeries tend to have a higher level of impairment. Though, for Workers, the average level of impairment for appeals is higher than for CTP.

5.2.4 The presence of pre-existing injuries

Figure 5.14 and Figure 5.15 show the number and proportion of decisions with pre-existing injuries, for physical and psychological injuries and scheme.

Figure 5.14 – Number and proportion of Workers decisions with pre-existing injuries

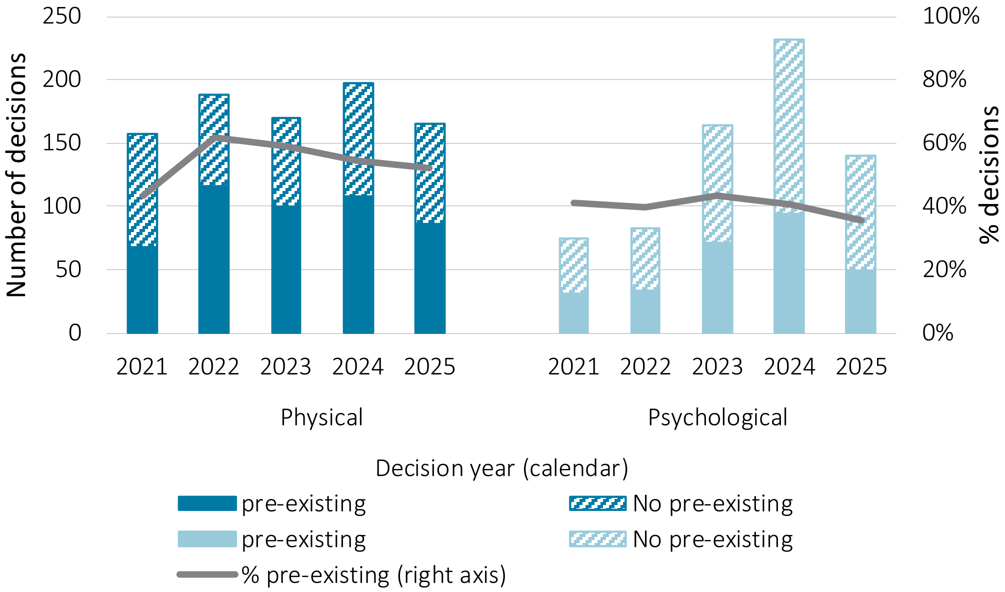
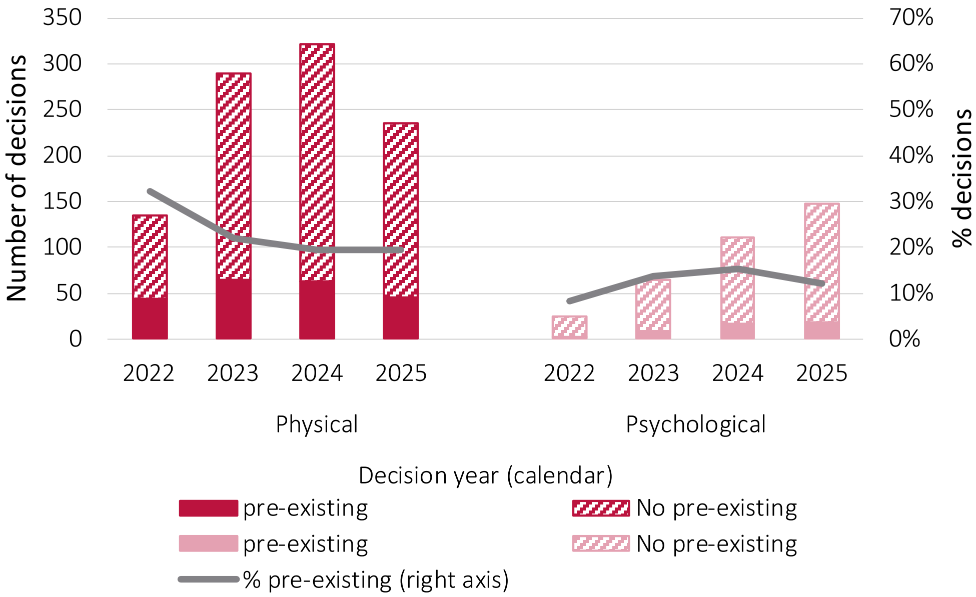


Figure 5.15 – Number and proportion of CTP decisions with pre-existing injuries



We observe the following:

- A higher proportion of pre-existing injuries in Workers compared to CTP.
- A higher proportion of pre-existing physical injuries compared to psychological injuries.

6 Conclusion

6.1 Strengths and weaknesses of GenAI in this context

GenAI is already being used by insurers, for example to assist in summarising information. however, it can also have extensive applications in populating feature rich datasets that can be used for analysis and insight. We were surprised at just how accurate it was, even with complex legal and medical text – although in order to get to that level of accuracy we did need to spend a reasonable amount of time refining the prompts we used.

Does it save time? Absolutely. We used it to extract information from around 3,000 decisions. These decisions range in length from around 10 A4 pages up to 40 or more. They are complex decisions. If we were to manually read through all 3,000 and extract relevant pieces of information it would take us around 1,200 to 1,500 hours (nearly a full working year for one person).

If, at the end of the 1,200 to 1,500 hours you decided you actually wanted to extract a different piece of information, you would need to repeat the entire process again! This is obviously a time and cost intensive approach, which is why in the past we would conduct a sample and then extrapolate from the sample (which has its own limitations).

By using GenAI you can build a feature rich dataset relatively cost-efficient way (it cost around $2,500 in 'tokens' to process all 3,000 decisions with the 50 odd questions we used to extract information) and quickly (in hours – and because processing could be done concurrently it would have been possible to run through all 3,000 decisions even more quickly than hours). And if, at the end of that, you decide you want to query additional information, you can re-run the entire process for the same cost and over the same period of time.

6.2 Need for human oversight and domain expertise

As with any tool you need to use GenAI appropriately. This means that having domain expertise is quite important – including knowing what is *not* available in the text being analysed. In our PIC example we are analysing Medical Appeal and Medical Review decisions – the original (or first instance) decisions are not available. Therefore, no matter how 'clever' or useful GenAI is, we simply cannot get a comprehensive view of the first instance decisions.

The prompts used to extract useful information matter enormously and they often need to be refined through trial and error. We found that testing the outputs against an initial sample of cases was essential to improve accuracy and relevance. The process was inherently iterative – we started with some prompts/questions, reviewed the resulting answers, adjusted the prompts and reviewed again until we were asking the "right" questions. The nature of the decisions made this particularly important as there can be ambiguity in questions that you do not necessarily think through until you review GenAI's answers – which may be technically correct but not helpful.

At the end of the process, an audit of the final dataset is critical. We found it important to make sure that we had not simply trained ourselves on prompt writing on a subset of claims. By auditing the results we could be more confident about overall accuracy.

One last comment is that our final dataset was *not* perfect. There will still be inaccuracies in some of the responses to questions we posed of the GenAI. For our high-level analysis of trends in the PIC Medical Appeal and Medical Review decisions, we were comfortable with the level of accuracy we achieved. For other purposes though, it may be that an even higher level of accuracy is required.

6.3 GenAI and professionalism

The Actuaries Institute has recently updated the guidance material for abiding by the Code of Conduct to incorporate relevant considerations around GenAI. The guidance material notes that Members, amongst other things, will need to consider that the use of GenAI is:

> *"fit for purpose, including understanding and accommodating the limitations of the tools. Members should be aware of the potential for errors or approximations to be introduced, particularly considering the degree of accuracy required for the Service."*

We agree with this – both in terms of striving for better accuracy and also acknowledging that perfection is generally not possible in actuarial work. We rarely (if ever) deal with 'perfect' data – one of the values that actuarial thinking can bring to a problem is acknowledging the limitations in a dataset and incorporating this uncertainty into our analysis and conclusions. Even if a capable human, with legal and medical training, reviewed the 3,000 decisions and entered information relating to them into a database it is likely they would make at least some errors. By applying the actuarial control cycle , we can test, improve and validate that the GenAI is producing useful information, at a sufficiently high level of quality that we can use it for analysis and be confident in the conclusions we are drawing.

We found the following articles helpful, particularly from an actuarial perspective, when thinking about ethics and professionalism in how we were using GenAI:

- Julie Evans and Stuart Rodger, *Is AI Making it Easier or Harder to be a Professional?*, 2 June 2025, https://www.actuaries.asn.au/research-analysis/is-ai-making-it-easier-or-harder-to-be-a-professional
- Dr Fei Huang, *Check Your AI: A Framework For its Use in Actuarial Practice*, 21 August 2025, https://www.actuaries.asn.au/research-analysis/check-your-ai

6.4 Potential for broader application by insurers and actuaries

The approach we tested has clear potential beyond analysing PIC decisions. Insurers and actuaries could use similar techniques to:

- populate databases with structured data and meta-information derived from case notes or other unstructured sources;
- 'back populate' data by creating fields that were never captured in the original claims systems. For example, the case notes around a claim will almost certainly identify the presence of a secondary psychological injury, the length of time over which it developed and potentially what has contributed to causing it. By using GenAI across the case notes this information can be extracted and then added on to existing databases to allow for better analysis and claims management; or
- explore alternative information sources – such as decisions regarding disputes on a claim, or medical reports, or customer interactions, to enrich existing datasets.

This additional information could inform better claims management, with the ultimate goal of supporting injured persons to return to work and especially health to the maximum extent possible.

6.5 Conclusion

Our work demonstrates that GenAI can play a valuable role in transforming unstructured legal text into structured datasets that support meaningful analysis. While the technology is not flawless and requires careful prompt design, iterative refinement and human oversight, the efficiency gains are undeniable. What previously would have taken months of manual effort can now be achieved in hours, at a fraction of the cost. Beyond the immediate application to PIC decisions and WPI assessments, the approach offers broader potential for insurers, actuaries and policymakers: enriching existing datasets, uncovering new insights and enabling more proactive scheme monitoring and claims management. As with any actuarial tool, success depends on applying professional judgment, understanding limitations and incorporating uncertainty into the analysis. Used appropriately, GenAI can become a powerful complement to existing data sources, helping us to better understand injury compensation schemes and identify opportunities for improvement.