Development and initial application of a harmonised multi-jurisdiction work injury compensation database

Michael Di Donato (iD) **, Luke R. Sheehan, Shannon Gray,
Ross Iles, Caryn van Vreden and Alex Collie**

Abstract

Objectives: Workers' compensation schemes provide funding for wage replacement and healthcare for injured and ill workers. In Australia, workers' compensation schemes operate independently in different jurisdictions, making comparisons of health service use challenging. We sought to develop and deploy a new database of health service and income support data, harmonising data from multiple Australian workers' compensation jurisdictions.

Methods: We worked with workers' compensation authorities from six Australian jurisdictions to combine claims, healthcare, medicines and wage replacement data for a sample of compensated workers with claims for musculoskeletal conditions. We designed a structured relational database and developed a bespoke health services coding scheme to harmonise data across jurisdictions.

Results: The Multi-Jurisdiction Workers' Compensation Database contains four data sets: claims, services, medicines and wage replacement. The claims data set contains 158,946 claims for low back pain (49.6%), limb fracture (23.8%) and non-specific limb conditions (26.7%). The services data set contains a total of 4.2 million cleaned and harmonised services including doctors (29.9%), physical therapists (56.3%), psychological therapists (2.8%), diagnostic procedures (5.5%) and examinations and assessments (5.6%). The medicines data set contains 524,380 medicine dispenses, with 208,504 (39.8%) dispenses for opioid analgesics.

Conclusions: The development of this database presents potential opportunities to gain a greater understanding of health service use in the Australian workers' compensation sector, to measure the impact of policy change on health services and to provide a method for further data harmonisation. Future efforts could seek to conduct linkage with other data sources.

Introduction

Work incapacity occurs when an illness, injury or other health condition reduces the ability of a worker to participate in employment. In some cases, this is referred to as work disability.[1] The scale and impact of work incapacity is evident from multiple international data sources.[2–4] For example, there were approximately 2.8 million non-fatal

Healthy Working Lives Research Group, School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia

Corresponding author:
Michael Di Donato, Healthy Working Lives Research Group, School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne, Victoria, Australia.
Email: michael.didonato@monash.edu

workplace injuries and illnesses in the USA in 2019, equating to 2.8 cases per 100 full-time equivalent workers.[5] Work-related musculoskeletal disorders accounted for 239,345 time-loss years in Canada between 2004 and 2013.[6] The International Labour Organization estimates that there are 340 million occupational accidents and a further 160 million workers affected by industrial illness every year.[7] In Australia, the setting for this study, over half a million workers experienced a traumatic work-related injury in 2014,[8] while more than 780,000 received income support for a period of work incapacity during the 2015–2016 Australian financial year.[9] Recent estimates indicate that work injury costs approximately AUD\$ 61.8 billion per annum or 4.1% of Australian gross domestic product.[9]

In many nations, individuals experiencing an episode of work incapacity that is due to the circumstances of work can seek financial support from an income support system for the period of incapacity. These are known as cause-based systems, with financial benefits paid on the premise that work contributed to loss of earning capacity.[4] Caused-based systems are more commonly known as workers' compensation systems and are in operation in nations including Australia, the USA, Canada, Hong Kong, Korea, Japan and Italy.[4,9]

Workers' compensation systems typically fund healthcare, rehabilitation and other services in addition to income support.[4,9] These funds are made available in order to enhance the functional capacity of incapacitated workers, to support recovery from injury and illness, to improve incapacitated workers' ability to engage in paid employment and to provide long-term care for workers with permanent disability. Payments for healthcare and services constitute approximately one quarter of total workers' compensation annual expenditure, or more than \$2.38 billion in 2019–2020, among Australian workers' compensation schemes.[10]

Common data standards enable accurate comparisons that utilise multiple data sources. For example, comparisons of workers' compensation claims in different Australian jurisdictions reporting on injury and illness, occupation and industry are routinely published, as national standards are available for all these variables.[11–13] However, there are currently no data standards or accepted systems or indicators for monitoring healthcare service provision in Australian workers' compensation schemes. While several schemes utilise the Australian public healthcare scheme item codes (i.e. the Medicare Benefits Schedule (MBS)), most also fund services not covered under this scheme, and therefore do not have an MBS item code. In such instances, workers' compensation regulators rely on internally generated service item codes, making comparisons between jurisdictions challenging. Agreed standards for healthcare services data are critical for understanding the prevalence, nature, timing, continuity and quality of healthcare provision to incapacitated workers.

Evidence from other jurisdictions (e.g. USA) indicates that certain procedures, services and medicines may be harmful in some circumstances and may prolong disability. For example, for workers with low back pain, spinal surgery (e.g. decompression or fusion) is associated with prolonged disability duration,[14] diagnostic imaging often triggers a 'cascade' of additional and unnecessary treatments[15–19] and certain medicines such as opioids lead to harmful side effects that outweigh any benefits.[20–23] However, analyses of these health services in compensated workers have not been readily performed in Australia. There are also currently no data across jurisdictions to enable us to determine the extent to which these potentially low value services are being provided to workers.

Despite healthcare provision and income support being the two main tools available to Australian workers' compensation schemes to support workers during their periods of incapacity, we currently have no data (beyond that available to individual insurers or regulators) to understand the prevalence, impact, continuity and quality of health services across jurisdictions. Existing standards of data collection in workers' compensation have provided valuable understanding so far: the national data set for compensation-based statistics and the national return to work survey are both used for monitoring injury and illness trends and research.[24] However, this existing infrastructure still does not provide comprehensive information about health service use in compensated workers.

The objective of this study was to develop and deploy a new database of health service and income support data for research purposes, encompassing data from multiple Australian workers' compensation jurisdictions. We believe that insights generated from such a database will support knowledge generation and policy and practice development. It will also enable an understanding of service and benefit delivery at a national level, and grant greater statistical power that allows for more analyses of more complex relationships between variables. In this paper we describe the methods and process used in the development of this new database.

Methods

Setting

There are 11 workers' compensation schemes in Australia: one per state and territory, and three national.[9,25] These schemes are regulated by government organisations, with claims variously managed and underwritten by government and private organisations. Workers' compensation schemes provide coverage to approximately 93% of the Australian labour force.[25,26] Further information on the structure and coverage of workers' compensation in Australia is available elsewhere.[9,25] These are cause-based systems, meaning that a worker is only eligible for payments for benefits and

services from workers' compensation if they are injured in the course of employment.[25] Income support benefits are typically paid for two to two-and-a-half years, depending on jurisdiction, before workers must provide evidence of significant incapacity preventing a return to work.[25] These benefits are typically paid in fortnightly instalments directly to the injured worker, and cover a percentage of the worker's pre-incapacity weekly wage up to a maximum cap. The percentage and the cap vary between the Australian jurisdictions.

Workers' compensation schemes also fund any healthcare or rehabilitation services or medicines deemed reasonable and necessary for worker recovery in addition to income support.[9,25] The availability of funding for healthcare often extends beyond the end of income support. Workers may still access healthcare outside the workers' compensation scheme via other funding sources. The Australian government operates the MBS and Pharmaceutical Benefit Scheme (PBS): national health insurance systems with subsidised healthcare regulated and funded by the national government and operated through private and local government entities.[27–29] Healthcare may also be funded by the worker from out-of-pocket expenses or private health insurance. Workers' compensation healthcare funding provides full coverage of healthcare costs (i.e. avoiding gap payments), allows injured workers faster access to healthcare (i.e. avoiding wait times common in publicly funded non-emergency healthcare) and greater flexibility in choice of healthcare provider. In some Australian jurisdictions, workers or their employer are responsible for funding the initial healthcare costs in an episode of work incapacity (typically about the first ~AUD$ 800) before the compensation scheme funds further expenditure.[30] Further detailed information about Australian workers' compensation schemes is published elsewhere.[9,25]

Sample

We chose to focus on conditions that are common within Australian workers' compensation systems to establish the first iteration of the database; specifically, low back pain, limb fracture and non-specific limb conditions. Low back pain is the leading cause of disability globally and occurs predominantly in people of working age.[31,32] There are consistent and accepted treatment and management options for low back pain,[33] providing a potential for comparing actual service delivery with 'best practice'. The two other conditions, limb fracture and non-specific limb conditions, were selected as comparator musculoskeletal conditions. Limb fractures have a traumatic mechanism, and the return of functional capacity usually progresses on a reasonably consistent trajectory, and non-specific limb conditions typically involve damage to similar types of structures as low back pain (i.e. soft

tissue). Included conditions were defined using Type of Occurrence Classification System nature of injury codes (see Supplemental Appendix 1). Claims with these conditions occurring between 1 July 2010 and 30 June 2015 (i.e. five Australian financial years) were included. This sample selection criterion has been used in previously published studies.[34]

Project development

We developed the project proposal in mid-2018. As a multidisciplinary team including clinician specialists, allied health professionals and policy analysts, the proposal was developed with a wide array of input. We also invited feedback from industry experts and researchers with experience in similar projects conducted in Canada.[35] The proposal included a list of desired claim- and service-level variables that we determined necessary to complete our analyses.

Workers' compensation authorities from five Australian states and the Commonwealth authority, representing more than 80% of the national labour force, were invited to participate: New South Wales, Victoria, Queensland, South Australia, Western Australia and Comcare (the Commonwealth authority). All invited data custodians agreed to participate, with data from all custodians except New South Wales included in the first iteration of the database. As this was the first iteration of the database requiring substantial methodological development, we chose to invite only the workers' compensation authorities from larger states and territories. In the future versions of the database, we aim to include workers' compensation authorities from all Australian states and territories.

Once data custodians agreed to participate, we met with data and information specialists from each organisation to discuss the data formats and availability of information. Most collected variables were common across jurisdictions; however, there was variation in how some variables were defined. For example, date variables were often identified differently between databases despite being analogous: the date that liability for a claim was accepted by an insurer could be identified as 'date accepted', 'date determined' or 'liability commencement date'. In cases where data dictionaries were maintained, we worked with data custodians to select specific variables to match our requirements. In other cases, we asked data custodians to recommend which variable they thought was most suitable for our requirements.

Code used to develop this database was written in both R 3.5 and STATA, which were executed in Microsoft Visual Studio, RStudio and STATA 15.[36–39] A number of R packages were required to complete the necessary steps referenced throughout this paper, including the following: *dplyr, tidyr, readr, tibble, stringr, ggplot2, lubridate, naniar, knitr, broom* and *haven*.[40–50] A complete reference list of all R packages utilised in the database development is available in Supplemental Appendix 2.

Governance and ethics

We developed and entered data sharing agreements with each data custodian that (1) sought to ensure the privacy, confidentiality and security of the data; (2) specified the ways in which data would be received, stored, used, analysed and published; and (3) established protocols for the review and publication of study findings and the involvement of data custodians in the project. A standard review period was agreed upon between researchers and all data custodians. Researchers would provide any output material (e.g. papers, conference submissions or reports) to data custodians for a 20-business day review period prior to public dissemination. The review period allowed data custodians to comment on findings and analytic interpretation of the data but did not include the option to veto the publication of any findings. We received ethical approval from the Monash University Human Research Ethics Committee (Project ID 17267, November 2018).

Database structure, design and requirements

A worker with an accepted claim is unlikely to receive only a single service, medicine or wage replacement payment. The database was therefore designed with a one-to-many relationship: there are single records for each claim, and many records for each of the services, medicines and wage replacement payments they received. Each of the records, whether single or many per claim, was designated a unique claim identifier by each data custodian. Service, medicine and wage replacement records can therefore be linked to each claim to which they belong. New harmonised identifiers were applied in this database to avoid overlap between data custodians' previously isolated individual databases (e.g. two jurisdictions might both have claim ID 1234). The new identifier was known as the multi-jurisdiction identifier (MJid). This relationship meant that the database consists of four data sets linked via the MJid: the claims data set (one record per claim), services data set (many records per claim), medicines data set (many records per claim) and wage replacement data set (many records per claim). The services, medicines and wage replacement data sets were compiled separately to one another as the variables were too dissimilar. The relationship between these data sets is demonstrated in an example in Figure 1.

Data transfer, storage and privacy controls

Data were transferred from each participating data custodian to the researchers via several secure file transfer protocol services for storage in the Monash University secure research environment. This environment is restricted and can only be accessed by authorised researchers. As per ethics approval, data were de-identified by data custodians prior to transfer. Once received and stored, the common
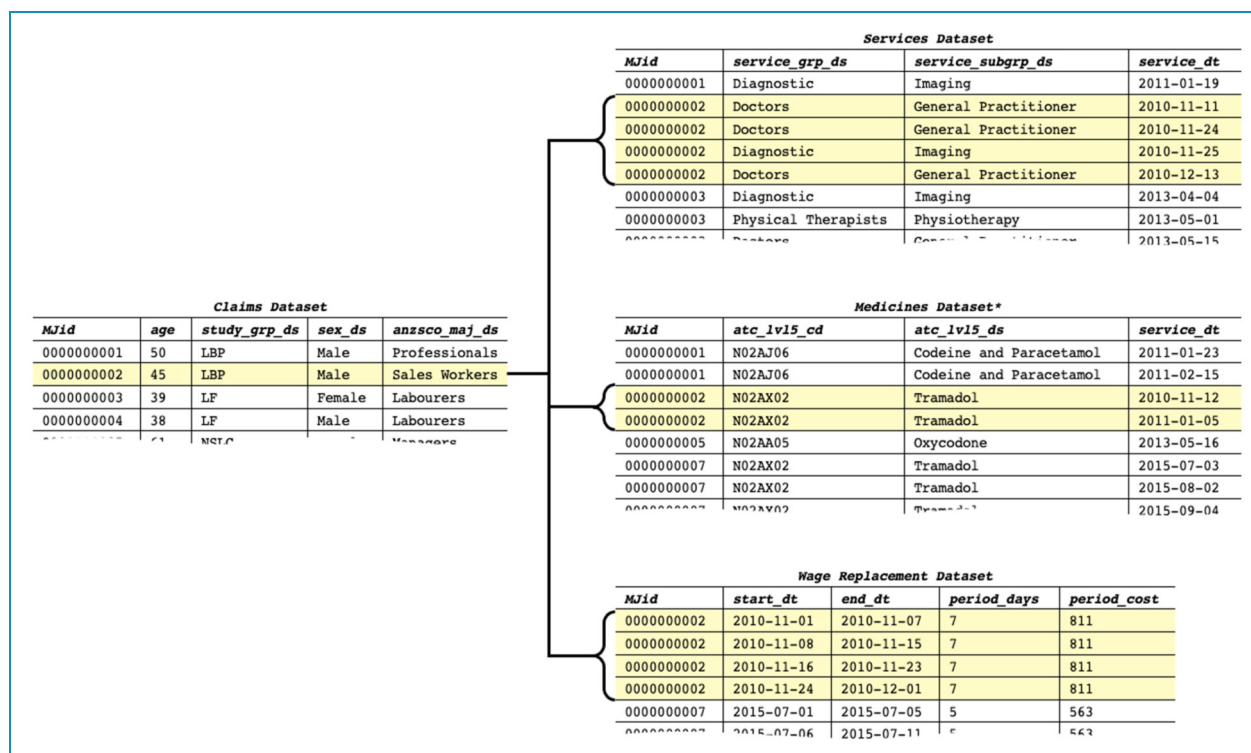

Figure 1. Example of database structure.

unique identifier (i.e. MJid) was generated and applied to claims and services data sets. This identifier serves as the key to link each of the data sets within the database.

Cleaning and quality assurance

Data were first checked for internal consistency. This quality assurance step refers to whether the values in the data were accurately aligned with a given variable. Errors in internal consistency often occurred where data were stored in delimited flat files (e.g. comma-separated value files). For example, an error in value separation when parsing data into the R or STATA environment could result in an industry code being stored in an occupation variable. Count statistics were then generated to identify any unique service items. Any unique services were cross-referenced against a data set of names[51] and manually checked to ensure that no identifying information was erroneously stored within the services data set.

Conceptually common variables were formatted differently between data custodians. This was likely due to each data custodian's data input, storage and handling infrastructure, as well as the format in which the data was transferred to the research team. For example, where data custodians transferred data as flat files, date variables were often parsed as strings, but data transferred in database formats (e.g. SAS) meant date variables retained their date/time properties. Conversion was required to change dates stored as strings to date variables.[52]

The array of service item codes used by each jurisdiction often included a mix of internally generated codes and MBS codes. While MBS codes are numeric only, some internally generated codes included letter prefixes. Depending on file formats, this often resulted in service item codes being stored as strings, and numeric codes frequently having leading zeroes. For example, MBS item 23 (professional attendance by a general practitioner) was frequently stored as '000023' to account for internally generated service item codes.[53] A string trimming loop was implemented to remove leading zeroes from service item codes.

Approximately 5% of data were missing in the received data sets. This was typically confined to specific variables. Missing data were often defined in non-standard ways, such as 'NA-', 'Null' or '9999'. These instances were identified and converted to accurate standard missing values.[54] We did not make attempts to impute missing data in this first version of the database, instead leaving this decision to researchers conducting the analyses.

All variables were assessed for logical values. This was usually required for continuous variables where the range of the variable was in question. For example, it was unlikely that a worker would earn AUD$ 999,999 per week pre-injury, and that instead this represented missing or unknown pre-injury weekly earnings. The same logic checks were required for date variables. For example, a date of injury in decades into the future of the study period was unlikely to be a realistic data entry. Illogical values were converted to missings.

Data harmonisation

Data harmonisation involved the combining of variables from different data custodians in such a way that consistent and reliable meaning can be derived from a single new variable. For example, some data custodians recorded worker sex as 'M', 'F' and 'U', whereas others utilised '0', '1' and '9'. There were also frequent differences in date arrangement, with data custodians reporting in day-month-year, month-day-year or year-month-day. It was important to identify what each variable's values represented and arrive at a harmonised version. In cases where data custodians used internally generated variables harmonisation, this was more challenging. For example, employer size was recorded as a continuous variable (i.e. number of employees) or as a categorical variable (e.g. 'small', 'medium' or 'large'). In such cases, data dictionaries and information were assessed, and a number of possibilities were discussed with the research team and data custodians to arrive either at an assumption for that variable or a decision not to include it in the final database.

Services data were harmonised in tranches, based on priorities identified by the research team. General Practitioner services were prioritised as they are a key service in the Australian workers' compensation process.[30] Diagnostic imaging, physical therapy services (e.g. physiotherapy, chiropractic and osteopathy) and psychological therapy services (e.g. psychology, counselling and rehabilitation counselling) are all common services that were also selected based on their relevance to best practice care for low back pain.

Harmonising each tranche of service data followed a standard methodology (see Figure 2). A list of unique services was extracted from each data custodian's services data set. These lists also included the total frequency of each service. A list included a service item code (where applicable) and a service item description (usually a short string of up to approximately 100 characters). A service item code was either a numeric MBS item or an internally generated alphanumeric string. An attempt was firstly made to attach the MBS to each list of unique services by the item code. Where services were matched to the MBS, the research team collectively decided upon eligible services based on MBS descriptions. For example, when examining the tranche of diagnostic imaging services, any service in MBS Category 5 – Diagnostic Imaging Services was considered eligible. The MBS was not exclusively used to identify an eligible service but did assist investigators where a data custodian used an MBS item number or description.

A service definition was created by the investigator team for the service to be harmonised. Two investigators then independently screened each service item for eligibility.
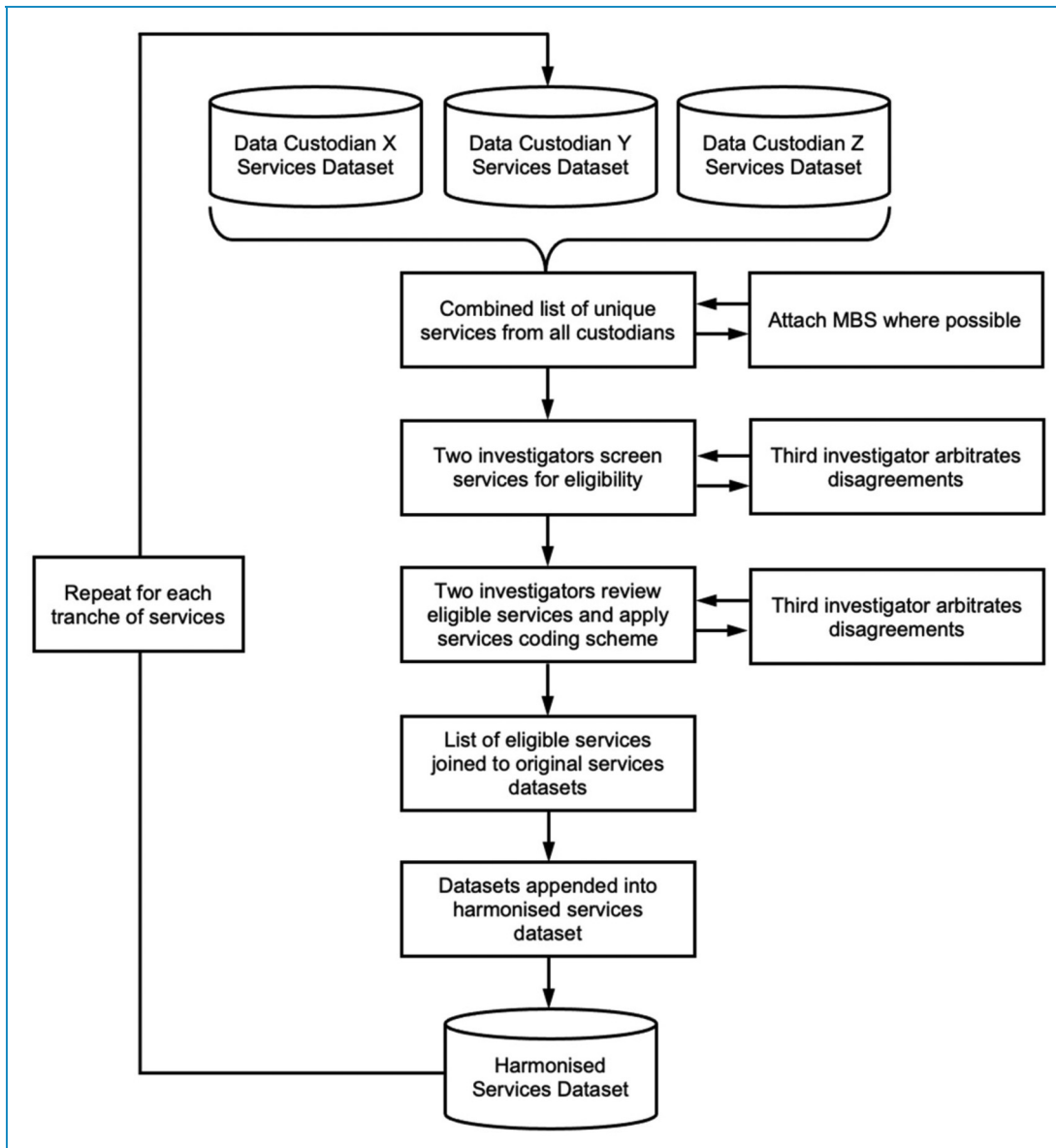
Figure 2. Services data harmonisation process.

Disagreements were arbitrated by a third investigator. After a round of eligibility assessment had been completed, investigators then met to finalise how this eligible tranche of services should be categorised. For example, once general practice (GP) services had been identified, the investigator team then had to decide on definitions for GP services that were 'patient interactions' or 'report writing/ other' (see Supplemental Appendix 3 for examples). Part of the harmonisation process was to develop a bespoke services coding scheme that could retain sufficient services detail from all data custodians. This is described in the following section. Two investigators independently reviewed eligible services and applied the services coding scheme. A third investigator again arbitrated disagreements.

Following screening for eligibility and the application of the services coding scheme, a final list of harmonised service items from all data custodians was available. The list of eligible health services was joined to the original service data sets by one or both of the service item code or description (whichever was available). The filtered and categorised data from each data custodian was then checked and refined to ensure that the same variables were available and commonly identified in each. For example, if service cost was not available in one data custodian's data set, the variable was added with all data as missing. These data sets were then appended to create the final harmonised services data set. This harmonisation methodology will continue to be used in harmonising the remainder of the database.

Medicines data were also prioritised for harmonisation. There is clear guidance for the use of opioids for low back pain: judicious use at low doses and for a short duration if a trial of non-steroidal anti-inflammatory medicines is contra-indicated or unsuccessful.[55–59] Like diagnostic imaging, this provided an evidence-based comparison for any findings. A similar harmonisation methodology to services was utilised. In this instance, unique lists of all medicines were created from the two data custodians who had medicines data (Victoria and South Australia). Investigators then assessed each medicine description to manually identify the appropriate Anatomical Therapeutic Chemical (ATC) code.[60] The remaining ATC structure (i.e. levels one through four) was attached to this data later. Based on the research objectives, opioids were also identified within the data. A new variable was generated to identify a medicine as either not an opioid or a strong or weak opioid. This opioid classification is available in Supplemental Appendix 5 and has been published elsewhere.[61,62]

Finally, wage replacement data were harmonised using relatively common variables. Data were provided from all data custodians in a similar format that included a wage replacement payment start date, payment end date and payment amount. These variables were appended to a single wage replacement data set. No modifications were made to overlapping periods or illogical time periods at this point.

Service item coding

In order to capture the totality of services funded by workers' compensation schemes, it was necessary to develop a bespoke coding scheme that covered services with codes in the MBS, PBS and services not covered by those coding schemes.

This coding schema was designed to be applicable to all services potentially funded by workers' compensation. In designing this schema, we sought to strike a balance between having services data detailed enough to not lose valuable insight during analysis, but not so detailed that harmonisation between data custodians was not possible. An example of this balance is demonstrated in Figure 3. With complete harmonisation, we could just identify a general practitioner service as a GP service, but this would result in information loss: in this instance, what the GP service involved. At the other end of scale, preserving a great deal of information about a GP service means that harmonisation is not possible. This is where we sought to strike a balance in harmonising data and preserving service information.

We developed the services coding schema iteratively. The first step was to plan an overall structure. This was conducted collaboratively, and changes and alternations were made through several group discussions with the entire research team. Once we had decided on a preliminary structure we proceeded to harmonise tranches of services data, as described in the previous section. Throughout this process, we continued discussions to strike a balance between services data that were too basic or too detailed. Modifications to a given section of the coding schema were made once the given tranche of services data had been filtered for eligibility. An operational definition of each service was decided upon by the research team prior to the final service item assignment and harmonisation. An example of one of these operational definitions is available in Supplemental Appendix 3.

Like the MBS, our coding schema contains four levels of categorisation: service category, group, subgroup and item. Each level has a code and a description. The code is hierarchical, in that levels above a service can be determined from the code. An example of how GP services fit within this schema is provided in Figure 4. The full structure of the coding schema is available in Supplemental Appendix 4.

It was not necessary to develop a service coding schema for the medicines data set. The content and type of
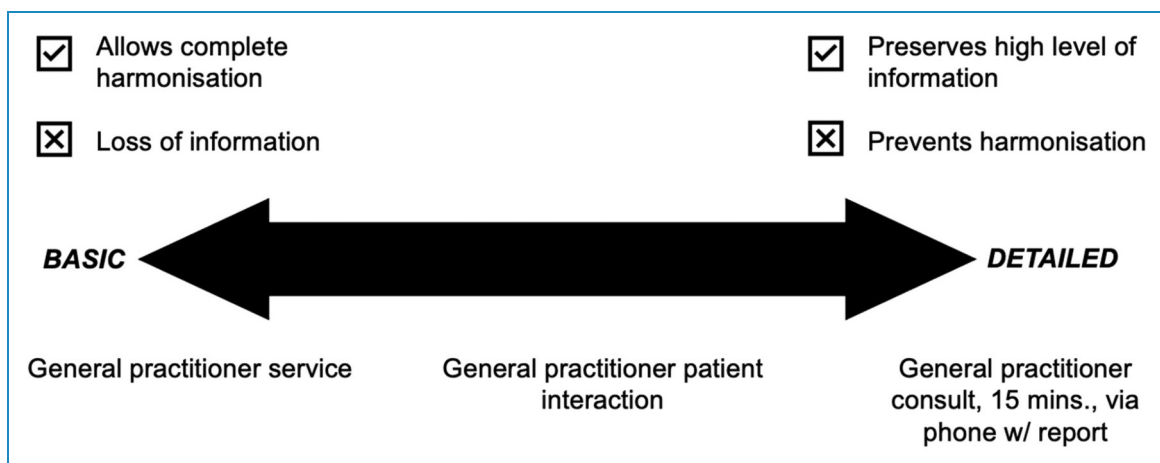

Figure 3. Example of balancing services data detail.

medicine, unlike a service description, is not open to interpretation. As described in the previous section, we therefore assigned all medicines to the ATC coding scheme based on primary chemical.

Other derived and calculated variables

Several additional variables were derived through either mapping to existing standards or calculated from other variables. Australian Statistical Geography Standard and Socioeconomic Indexes for Areas data were linked to the claims data set via worker residential postcode.[63,64] This practice has been performed before in several other research papers[65,66] and provides important covariates for analysis including the Accessibility/Remoteness Index of Australia and Index of Relative Socioeconomic Advantage and Disadvantage.

Wage replacement aggregation

As described in the previous section, wage replacement data were harmonised in the wage replacement data set. Total cumulative wage replacement duration was derived from individual wage replacement payments and linked to the claims data set via the unique identifier. Like services and medicines data, wage replacement payments were stored as many records per claim, with one record per payment period per worker. The difference in days between the start and end dates of a payment period for each worker was calculated. To standardise payment period duration between data custodians, business weeks (i.e. 5 days) were standardised to calendar weeks (i.e. 7 days). Overlapping periods were removed or merged to ensure that individual days were not double counted. The sum of paid days for each worker was then calculated and divided by seven (i.e. a calendar week). This provided the variable wage replacement duration, measured in paid calendar weeks.
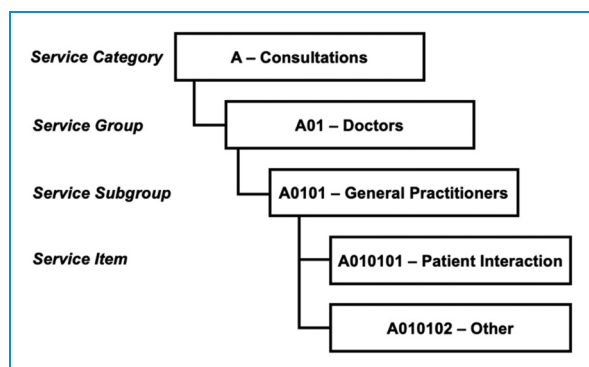

Figure 4. Services coding schema for general practice (GP) services.

Results

Database structure and volume

The Multi-Jurisdiction Workers' Compensation Database is a structured relational database comprised of four data sets: the claims data set (one record per worker), services data set (multiple records per worker), medicines data set (multiple records per worker for two jurisdictions only) and wage replacement data set (multiple records per worker).

In this first iteration of the database, we have included a total of 158,946 claims: 78,796 (49.6%) for low back pain, 37,753 (23.8%) for limb fracture and 42,397 (26.7%) for non-specific limb conditions (Figure 5). The majority of claims (63.6%, $N = 101,086$) involve wage replacement payments for periods of work incapacity (i.e. time loss).

The services data set contains a total of 4.2 million services (see Figure 6). Services from doctors (29.9%) and physiotherapists (56.3%) were the most common in the data.

The medicines data set (see Figure 7) contained medicines from two of the five included jurisdictions: Victoria and South Australia. This data set includes a total of 524,380 medicines, 39.8% ($N = 208,504$) of which are opioid analgesics.

Basic insights

In this section, we demonstrate three examples of basic insights made possible by this new Multi-Jurisdiction Workers' Compensation Database. This brief analysis is conducted using a sub-sample of the complete database. There are a number of notable policy differences between workers' compensation schemes in Australia that should be considered when making inter-jurisdictional comparisons or conducting analyses. The background to these differences can be extensive and is documented elsewhere.[67] For this reason, the following figures represent a sample including claims for any condition (i.e. low back pain, limb fracture and non-specific limb conditions) from Victoria, Queensland, South Australia, Western Australia and Comcare lodged between 01 July 2011 and 30 June 2015. Services data for these claims have been filtered to 30 days prior to 730 days after the claim acceptance date. The 30 days prior accommodates the possibility of services being retrospectively funded after a claim was accepted. A follow-up period of 730 days, or two years, is based on the typical period for which healthcare services are funded by these workers' compensation schemes.

Figure 8 presents the prevalence of service categories among the sample of claims. This is likely reflective of the conditions in the sample, with physical therapy services (77%) and diagnostic imaging services both highly prevalent (59%) relative to psychological therapy services (7%).

Figure 9 highlights the median number of services per claim in those claims that had that service. It is important to consider this figure in the context of the previous: while diagnostic procedures were relatively prevalent (59%), the median
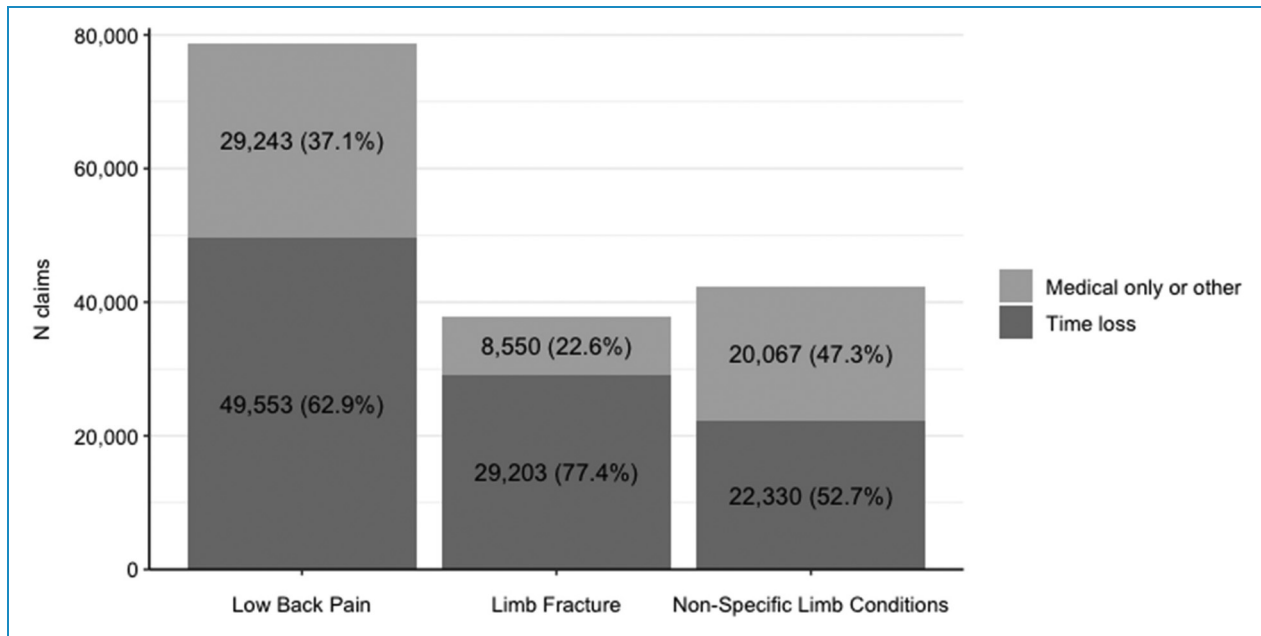
Figure 5. Number of claims in the Multi-Jurisdiction Workers' Compensation Database (claims data set) by claimed condition and claim type.
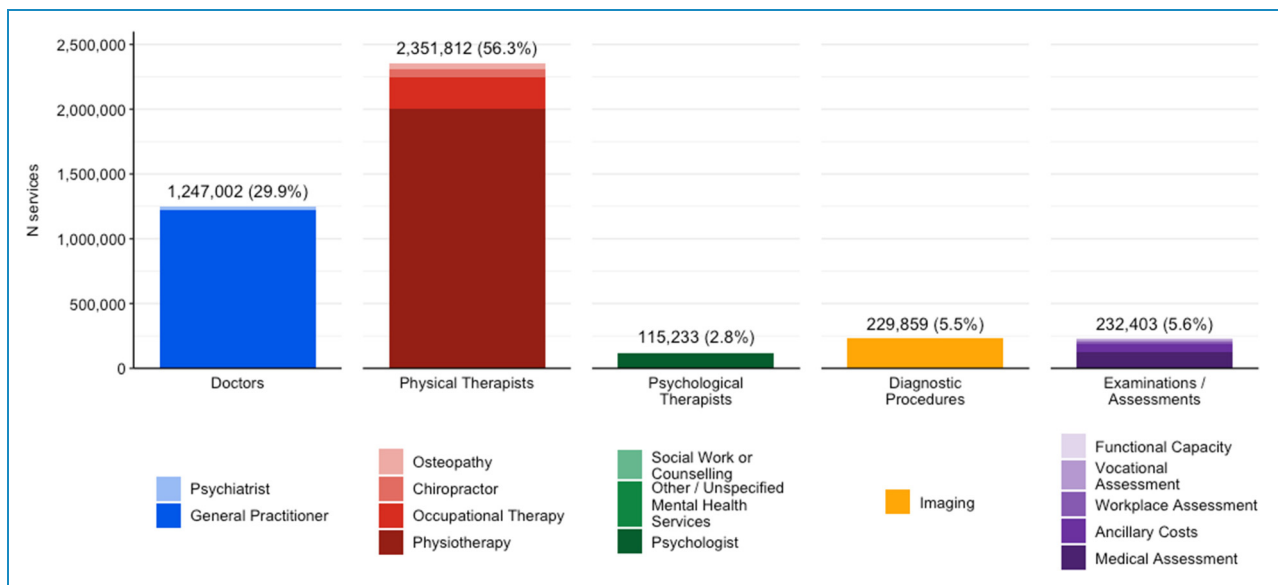

Figure 6. Number of services in the Multi-Jurisdiction Workers' Compensation Database (services data set) by service group and subgroup.

number of services in those that claims that received them is relatively low. In contrast, the small number of claims that received psychological therapy services received a median number of services similar to that of doctors.

Finally, Figure 10 demonstrates the timing of services relative to claim acceptance. This plot shows the relative density (i.e. frequency) of services in the 30 days prior to and 730 days after claim acceptance. Trends are immediately apparent, with the majority of diagnostic procedures

occurring early in a claim likely driven by the acute severity of limb fractures in this sample.

Published analyses

The database has already been utilised in four published academic studies. In our first study, we explored the prevalence of diagnostic spinal imaging and its association with wage replacement duration.[68] We then used group-based

trajectory modelling to identify patterns of opioid dispensing, and the association between these patterns and wage replacement duration.[69] We also assessed the variation in general practitioner services between jurisdictions in an exploratory analysis.[70] Finally, we examined the continuity of care provided by general practitioners and how this was associated with wage replacement duration.[71]

Discussion

This project sought to harmonise claim- and service-level data across multiple Australian workers' compensation jurisdictions. Previous efforts to utilise service-level data in the workers' compensation sector have been specific to single jurisdictions. For example, the Compensation Research Database combined data from the Victorian workers'

Figure 7. Number of medicines and number of opioids in the Multi-Jurisdiction Workers' Compensation Database (medicines data set) by opioid strength.

compensation and Transport Accident Compensation Schemes into a single research database.[72] This database has since been utilised in a multitude of research projects but has only provided insight in a single Australian jurisdiction. In this paper we have described a process for collection, coding and harmonisation of data from multiple Australian jurisdictions. To date, all claims data and general practitioner, physical therapy, and psychological therapy services, diagnostic procedures and medical assessments have been harmonised across five workers' compensation jurisdictions. Further, medicines data have been harmonised across the two available jurisdictions. The development of this database presents a number of future opportunities and benefits for workers' compensation stakeholders including regulators, insurers, healthcare providers, researchers and workers.

The first potential application for the database is to gain a greater understanding of health service use in the Australian workers' compensation sector. This potential has been demonstrated in our recent studies of imaging, opioid analgesics and GP service prevalence and patterns in compensated workers with low back pain.[68–70,73] Future research would be able to analyse the use of different services, and the combination of services. Clustering methods, such as latent class analysis, have previously been used in more specific samples to measure types of service users.[74,75] These same methods could be applied in future research using the database.

Administrative data has previously been used in quasi-experimental studies measuring the impact of policy changes such as benefit eligibility on incapacity duration.[76–78] The database provides a new resource to measure the impact of policy interventions and reforms both inside and outside workers' compensation systems on health service use. For example, the Australian
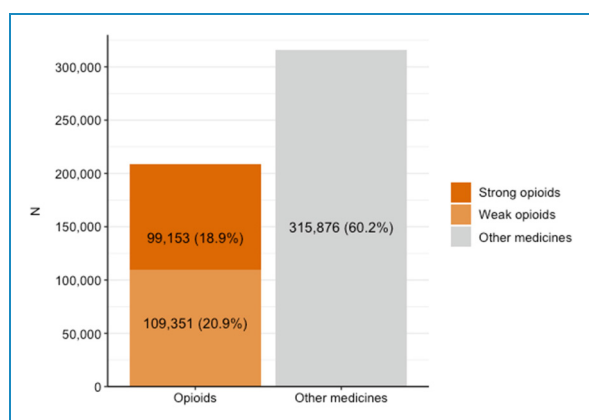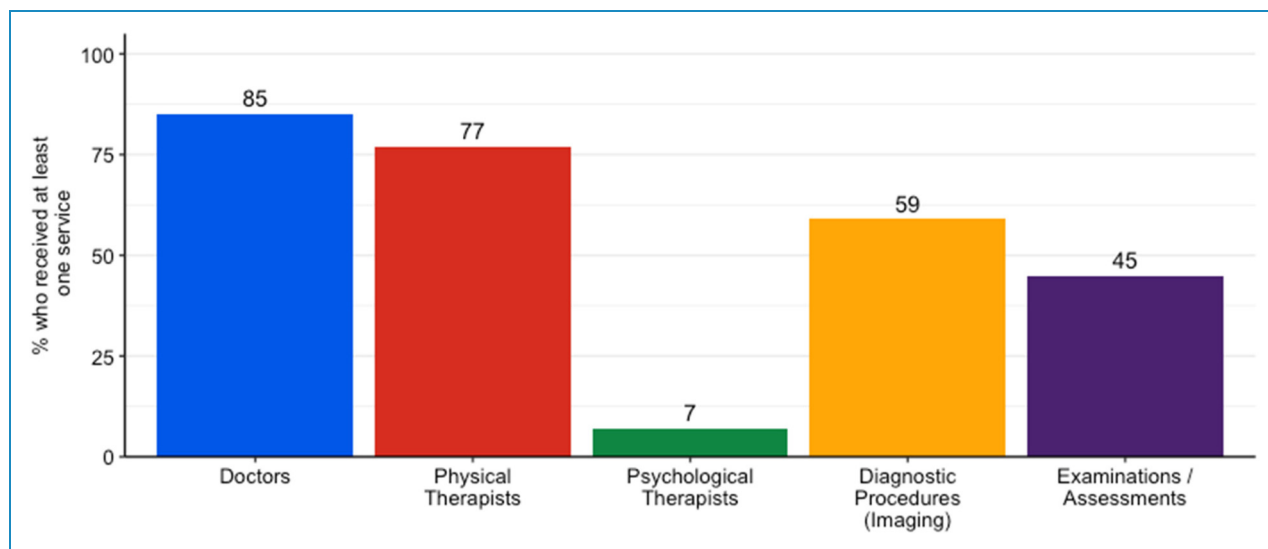

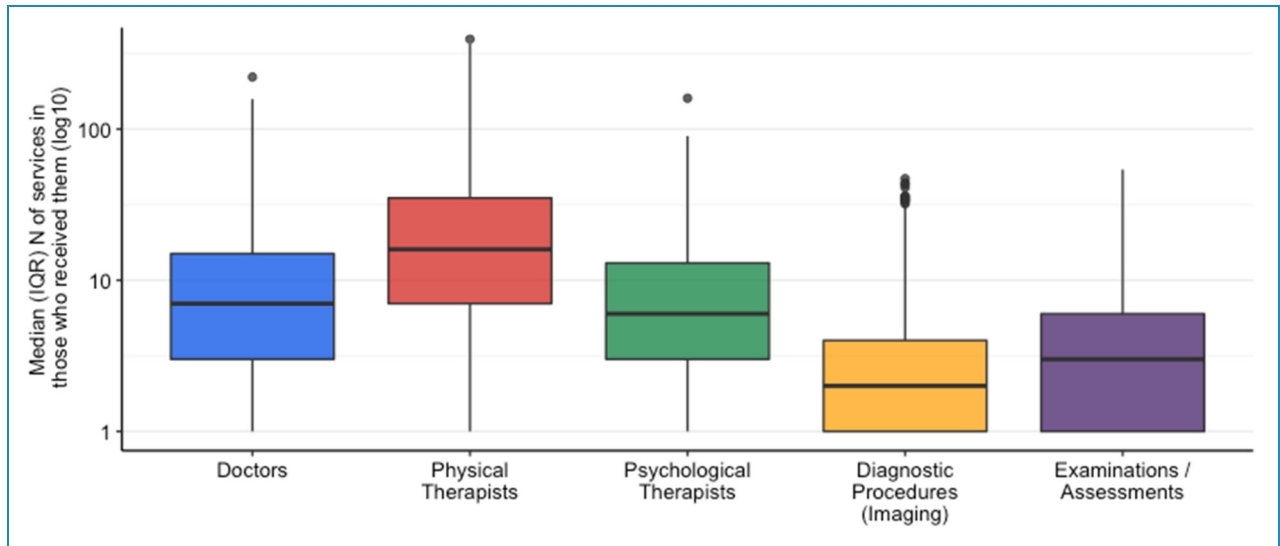**Figure 8.** Prevalence of services.

Figure 9. Median (interquartile range (IQR)) number of services in those who had them.
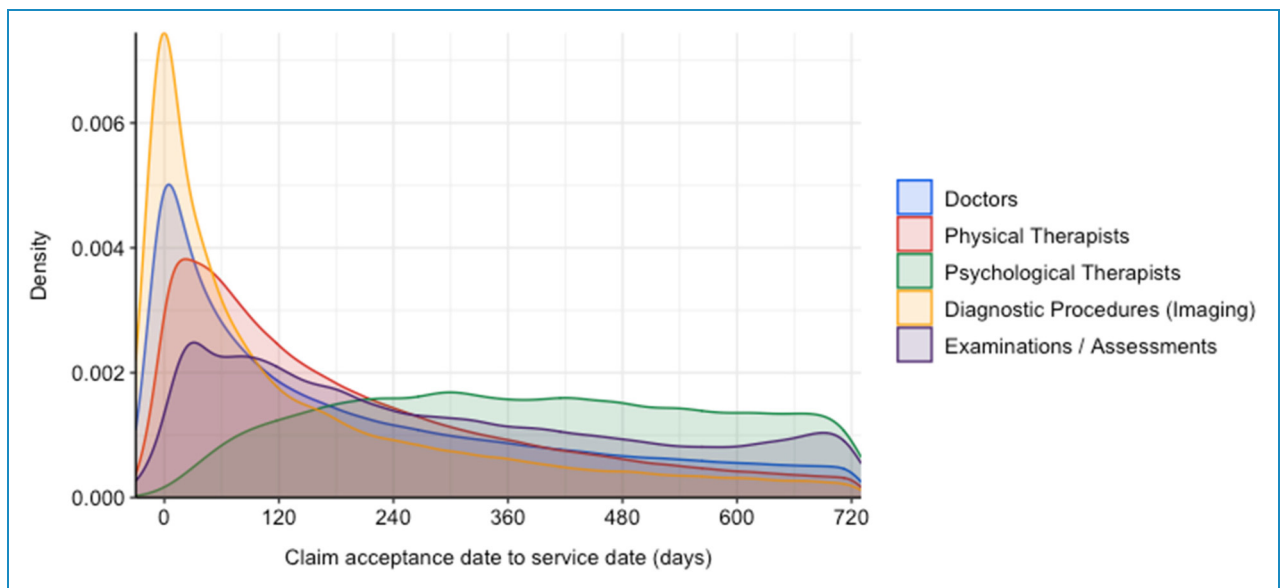

Figure 10. Timing of services relative to claim acceptance date.

Government 'up-scheduled' the commonly used opioid analgesic codeine from an over the counter to prescription-only medicine in February 2018.[79] Based on the findings of opioid dispensing research already conducted using this database,[69] such a policy change may have a substantial impact on opioid dispensing for compensated workers. This is the kind of policy reform that could be better understood using the new database.

This database presents a significant new resource for both research and workers' compensation and healthcare stakeholders. The large-scale coverage of compensated Australian workers and their services data is a new development in the research and compensation sectors. The current version of the database has already been utilised in numerous studies and will continue to be expanded into additional research of compensated musculoskeletal disorder claims in the future. The methods refined in developing the database can also be applied to future efforts to harmonise workers' compensation administrative data, including the expansion of the Multi-Jurisdiction Workers' Compensation Database itself.

It is also important to note several limitations of the database. As a by-product of scheme regulation, administrative data only includes what a scheme was intended to capture. For example, information about a worker's claimed injury or illness is typically only described by the four Type of

Occurrence Classification System variables – injury nature, location, mechanism and agency. While research benefits from this information being both standardised and structured, there is a lack of important information that would benefit any analyses, such as secondary symptoms or conditions and psychosocial factors. The database also only contains data about services funded by the included workers' compensation systems. If a service was not deemed 'reasonable and necessary' by an insurer or claims manager, then it would not have been funded, and thus not captured in the administrative data. Other processes, such as data linkage, would be required to extend coverage to other healthcare schemes like the MBS. Some jurisdictions, due to policy limits, only pay for healthcare and income support up to certain limits, or after certain thresholds have been met. This means that data do not necessarily contain all healthcare services sought by a worker during a period of work incapacity. In particular, this limits insights into service use in short-term claims and the early phase of medium- to longer-term claims. Finally, research using this database will be retrospective. During the lag between the time an event occurs, data are captured, and then available for research may limit the currency of any research findings.

Next steps

There are a number of future steps to be undertaken with currently available data. First, we will complete the cleaning and harmonisation of the remaining services data, including services such as surgery, exercise physiology and other home modification and social services. Second, we will conduct further analyses including both newly cleaned services and combinations of services. Third, we aim to expand the database to include both additional claimed conditions (e.g. compensable mental health conditions) and a longer time series. Finally, we aim to expand access to future versions of the database. Consistent with information sharing agreements entered into with data custodians, researchers can currently access the data included in the database by either working directly our research team as collaborators or approaching single jurisdictions independently to request the raw data. In the future, we will seek to make de-identified data available through independently managed data request processes such as those that we have established previously with individual states.[72]

The construction of this database has also meant the development of knowledge that may lead to standardised coding and collection of healthcare services data in the future. This would enable more straightforward data aggregation and simpler comparisons: a process that is currently challenging with the 'siloed' and disaggregated nature of healthcare services data in the workers' compensation sector in Australia.

In the future, it may also be possible to attempt linkage of data to other relevant data sets (e.g. Medicare). This would dramatically expand the scope of data and its value but would require the use of identifiable data for either deterministic or probabilistic linkage. While challenging, this is achievable: linkage between workers' compensation data and other data sources has been achieved in at least one Australian jurisdiction by our research team.[80]

ORCID iD: Michael Di Donato ⬚ https://orcid.org/0000-0002-6531-5949

References

1. Anema JR and Loisel P. *Handbook of work disability prevention and management*. New York, NY: Springer, 2013.
2. Di Donato M, Iles R, Lane T, et al. The impact of income support systems on healthcare quality and functional capacity in workers with low back pain: A realist review. *Pain* 2020; 161: 2690–2709.
3. International Labour Organization (ILO). *World social protection report 2017–19 – universal social protection to achieve the sustainable development goals*. Geneva, Switzerland: ILO, 2017.
4. Lippel K and Lötters F. Public insurance systems: A comparison of cause-based and disability-based income support systems. In: Loisel P and Anema JR (eds) *Handbook of work disability: Prevention and management*. New York, NY: Springer, 2013, pp.183–202.
5. Bureau of Labor Statistics (BLS). *Employer-reported workplace injuries and illnesses - 2019*. Washington DC, USA: U.S. Department of Labor, 2021.
6. Macpherson RA, Lane TJ, Collie A, et al. Age, sex, and the changing disability burden of compensated work-related musculoskeletal disorders in Canada and Australia. *BMC Public Health* 2018; 18: 58.
7. International Labour Organization (ILO). The enormous burden of poor working conditions, https://www.ilo.org/moscow/areas-of-work/occupational-safety-and-health/WCMS_249278/lang–en/index.htm (2021, accessed 8 December 2021).
8. Lane TJ, Collie A and Hassani-Mahmooei B. Work-related injury and illness in Australia, 2004 to 2014. What is the incidence of work-related conditions and their impact on time lost from work by state and territory, age, gender and injury type? Report No. 118-0616-R02. Melbourne: ISCRR, 2016.
9. Collie A, Di Donato M and Iles R. Work disability in Australia: An overview of prevalence, expenditure, support systems and services. *J Occup Rehabil* 2019; 29: 526–539.
10. Safe Work Australia (SWA). *Comparative performance monitoring report 23 - Workers' compensation funding*. Canberra, Australia: SWA, 2021.
11. Australian Bureau of Statistics (ABS). *Australian and New Zealand Standard Classification of Occupations, 2013, Version 1.2*. Canberra: Australian Bureau of Statisticshttps://www.abs.gov.au/AUSSTATS/abs@.nsf/allprimarymainfeatures/4AF138F6DB4FFD4BCA2571E200096BAD?opendocument= (2013, accessed 2 July 2019).
12. Australian Bureau of Statistics (ABS). 1292.0 – Australian and New Zealand Standard Industrial Classification (ANZSIC), 2006 (Revision 2.0), https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/1292.0Main+Features12006%20(Revision%202.0)? OpenDocument (2013, accessed 29 June 2022).
13. Australian Safety and Compensation Council (ASCC). *Type of occurrence classification system*. 3rd ed., Revision one. Canberra, Australia: ASCC, 2008.
14. Harris IA, Dantanarayana N and Naylor JM. Spine surgery outcomes in a workers' compensation cohort. *ANZ J Surg* 2012; 82: 625–629.
15. Graves JM, Fulton-Kehoe D, Jarvik JG, et al. Health care utilization and costs associated with adherence to clinical practice guidelines for early magnetic resonance imaging among workers with acute occupational low back pain. *Health Serv Res* 2014; 49: 645–665.
16. Graves JM, Fulton-Kehoe D, Jarvik JG, et al. Impact of an advanced imaging utilization review program on downstream health care utilization and costs for low back pain. *Med Care* 2018; 56: 520–528.
17. Webster BS, Bauer AZ, Choi Y, et al. Iatrogenic consequences of early magnetic resonance imaging in acute, work-related, disabling low back pain. *Spine* 2013; 38: 1939–1946.
18. Webster BS, Choi Y, Bauer AZ, et al. The cascade of medical services and associated longitudinal costs due to nonadherent magnetic resonance imaging for low back pain. *Spine* 2014; 39: 1433–1440.
19. Webster BS and Cifuentes M. Relationship of early magnetic resonance imaging for work-related acute low back pain with disability and medical utilization outcomes. *J Occup Environ Med* 2010; 52: 900–907.
20. Franklin GM, Rahman EA, Turner JA, et al. Opioid use for chronic low back pain: A prospective, population-based study among injured workers in Washington state, 2002-2005. *Clin J Pain* 2009; 25: 743–751.
21. Franklin GM, Stover BD, Turner JA, et al. Disability risk identification study C. Early opioid prescription and subsequent disability among workers with back injuries: The disability risk identification study cohort. *Spine* 2008; 33: 199–204.
22. Gross DP, Stephens B, Bhambhani Y, et al. Opioid prescriptions in Canadian workers' compensation claimants: Prescription trends and associations between early prescription and future recovery. *Spine* 2009; 34: 525–531.
23. Webster BS, Verma SK and Gatchel RJ. Relationship between early opioid prescribing for acute occupational low back pain and disability duration, medical costs, subsequent surgery and late opioid use. *Spine* 2007; 32: 2127–2132.
24. Collie A. *Overview of the COMPARE project*. Melbourne, Australia: Monash University, 2018.
25. Safe Work Australia (SWA). *Comparison of workers' compensation arrangements in Australia and New Zealand 2019*. Canberra, Australia: SWA, 2019.
26. Australian Bureau of Statistics (ABS). 6202.0 – Labour force, Australia, December 2018. https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/6202.0Main+Features1Dec%202018?OpenDocument= (2019, accessed 5 May 2022).
27. Australian Government. Human Services (Medicare) Act 1973. Canberra, Australia: Australian Government, 2020.
28. Australian Government. National Health Act 1953. Canberra, Australia: Australian Government, 2022.

29. Bohm K, Schmid A, Gotze R, et al. Five types of OECD healthcare systems: Empirical results of a deductive classification. *Health Policy* 2013; 113: 258–269.

30. Safe Work Australia (SWA). *Comparison of workers' compensation arrangements in Australia and New Zealand 2018*. Canberra, Australia: SWA, 2018.

31. Hartvigsen J, Hancock MJ, Kongsted A, et al. What low back pain is and why we need to pay attention. *Lancet* 2018; 391: 2356–2367.

32. Wu A, March L, Zheng X, et al. Global low back pain prevalence and years lived with disability from 1990 to 2017: Estimates from the global burden of disease study 2017. *Ann Transl Med* 2020; 8: 99.

33. Oliveira CB, Maher CG, Pinto RZ, et al. Clinical practice guidelines for the management of non-specific low back pain in primary care: An updated overview. *Eur Spine J* 2018; 27: 2791–2803.

34. Di Donato M, Buchbinder R, Iles R, et al. Comparison of compensated low back pain claims experience in Australia with limb fracture and non-specific limb condition claims: A retrospective cohort study. *J Occup Rehabil* 2020; 31: 175–184.

35. Sharpe K, McGrail K, Mustard C, et al. A framework for understanding how variation in health care service delivery affects work disability management. *J Occup Rehabil* 2022; 32: 215–224.

36. Microsoft Corporation. Microsoft visual studio 2017. 2018.

37. RStudio Team. RStudio: Integrated development for R. v1.4.1103. http://www.rstudio.com/ (2022, accessed 29 July 2022).

38. StataCorp. Stata Statisical Software: Release 15. 2017.

39. R Core Team. R: A language and environment for statistical computing. v3.5, https://www.R-project.org (2022, accessed 29 July 2022).

40. Wickham H, Francois R, Henry L, et al. dplyr: A grammar of data manipulation. v1.0.9, https://cran.r-project.org/package=dplyr (2022, accessed 29 July 2022).

41. Wickham H and Girlich M. tidyr: Tidy messy data. v1.2.0. https://cran.r-project.org/package=tidyr (2022, accessed 29 July 2022).

42. Wickham H, Hester J, Francois R, et al. readr: Read rectangular text data. v2.1.2 . https://cran.r-project.org/package=readr (2022, accessed 29 July 2022).

43. Wickham H, Miller E and Smith D. haven: Import and export 'SPSS', 'Stata' and 'SAS' files. v2.5.0. https://cran.r-project.org/package=haven (2022, accessed 29 July 2022).

44. Muller K, Wickham H, Francois R, et al. tibble: Simple data frames. v3.1.8. https://cran.r-project.org/package=tibble (2022, accessed 29 July 2022).

45. Wickham H. stringr: Simple, consistent wrappers for common string operations. v1.4.0. https://cran.r-project.org/package=stringr (2019, accessed 29 July 2022).

46. Wickham H, Chang W, Henry L, et al.. ggplot2: Create elegant data visualisations using the grammar of graphics. v3.3.6. https://cran.r-project.org/package=ggplot2 (2022, accessed 29 July 2022).

47. Spinu V, Grolemund G, Wickham H, et al.. lubridate: Make dealing with dates a little easier. v1.8.0. https://cran.r-project.org/package=lubridate (2021, accessed 29 July 2022).

48. Xie Y. knitr: A general-purpose package for dynamic report generation in R. v1.39, https://cran.r-project.org/package=knitr (2022).

49. Tierney N, Cook D, McBain M, et al. naniar: Data structures, summaries, and visualisations for missing data. v0.6.1. https://cran.r-project.org/package=naniar (2021, accessed 29 July 2022).

50. Robinson D, Hayes A and Couch S. broom: Convert statistical objects into tidy tibbles. v1.0.0. https://cran.r-project.org/package=broom (2022, accessed 29 July 2022).

51. Wickham H. babynames: US baby names 1880-2017. https://CRAN.R-project.org/package=babynames (2021, accessed 29 July 2022).

52. Spinu V, Grolemund G, Wickham H, et al. lubridate: Make dealing with dates a little easier. https://CRAN.R-project.org/package=lubridate (2021, accessed 29 July 2022).

53. Wickham H. stringr: Simple, consistent wrappers for common string operations, https://CRAN.R-project.org/package=stringr (2019, accessed 29 July 2022).

54. Tierney N, Cook D, McBain M, et al.. naniar: Data structures, summaries, and visualisations for missing data. https://CRAN.R-project.org/package=naniar (2021. accessed 29 July 2022).

55. Agency for Clinical Innovation (ACI). *Management of people with acute low back pain*. Chatswood, Australia: ACI, 2016.

56. Foster NE, Anema JR, Cherkin D, et al. Prevention and treatment of low back pain: Evidence, challenges, and promising directions. *Lancet* 2018; 391: 2368–2383.

57. Lee J, Gupta S, Price C, et al. Low back and radicular pain: A pathway for care developed by the British Pain Society. *Br J Anaesth* 2013; 111: 112–120.

58. National Institute for Health and Care Excellence (NICE). *Low back pain and sciatica in over 16s: Assessment and management*. London, UK: NICE, 2016.

59. Qaseem A, Wilt TJ, McLean RM, et al. Noninvasive treatments for acute, subacute, and chronic low back pain: A clinical practice guideline from the American College of Physicians. *Ann Intern Med* 2017; 166: 514–530.

60. World Health Organisation (WHO) Collaborating Centre for Drug Statistics Methodology. ATC/DDD Index 2020. https://www.whocc.no/atc_ddd_index/ (2019, accessed 30 January 2020).

61. Dowell D, Haegerich TM and Chou R. CDC guideline for prescribing opioids for chronic pain. *MMWR Recomm Rep* 2016; 65: 1–49.

62. Mathieson S, Wertheimer G, Maher CG, et al. What proportion of patients with chronic noncancer pain are prescribed an opioid medicine? Systematic review and meta-regression of observational studies. *J Intern Med* 2020; 287: 458–474.

63. Australian Bureau of Statistics (ABS). *1270.0.55.05 – Australian Statistical Geography Standard (ASGS): Volume 5 – remoteness structure, July 2016*. Canberra, Australia: ABS, 2018.

64. Australian Bureau of Statistics (ABS). *Socio-economic indexes for areas (SEIFA)*. Canberra, Australia: ABS, 2018.

65. Collie A, Lane TJ, Hassani-Mahmooei B, et al. Does time off work after injury vary by jurisdiction? A comparative study of eight Australian workers' compensation systems. *BMJ Open* 2016; 6: e010910.

66. Di Donato M, Buchbinder R, Iles R, et al. Comparison of compensated low back pain claims experience in Australia with limb fracture and non-specific limb condition claims: A retrospective cohort study. *J Occup Rehabil* 2021; 31: 175–184.

67. Safe Work Australia (SWA). *Comparison of workers' compensation arrangements in Australia and New Zealand*. Canberra, Australia: SWA, 2021.

68. Di Donato M, Iles R, Buchbinder R, et al. Prevalence, predictors and wage replacement duration associated with diagnostic imaging in Australian workers with accepted claims for low back pain: A retrospective cohort study. *J Occup Rehabil* 2022; 32: 55–63.

69. Di Donato M, Xia T, Iles R, et al. Patterns of opioid dispensing and associated wage replacement duration in workers with accepted claims for low back pain: A retrospective cohort study. *Pain* 2022; 163: e942–e952.

70. Collie A, Sheehan L and Di Donato M. Variation in general practice services provided to Australian workers with low back pain: A cross-jurisdictional comparative study. *J Occup Rehabil* 2022; 32: 203–214.

71. Sheehan LR, Di Donato M, Gray SE, et al. The association between continuity of care with a primary care physician and duration of work disability for low back pain: A retrospective cohort study. *J Occup Environ Med* 2022; 64: e606–e612.

72. Prang K-H, Hassani-Mahmooei B and Collie A. Compensation research database: Population-based injury data for surveillance, linkage and mining. *BMC Res Notes* 2016; 9: 56.

73. Sheehan LR, Di Donato M, Gray SE, et al. The relationship between continuity of care with a primary care physician and duration of work disability for low back pain: A retrospective cohort study. *J Occup Environ Med* 2022; 64: e606–e612.

74. Xia T, Iles R, Newnam S, et al. Driving Health Study Report No 3: Health service use following work-related injury and illness in Australian truck drivers. Melbourne, Australia: Insurance Work and Health Group, Faculty of Medicine Nursing and Health Sciences, Monash University, 2018.

75. Xia T, Iles R, Newnam S, et al. Driving Health Study Report No 4: Use of pharmaceuticals following work-related injury and illness in Australian truck drivers. Melbourne, Australia: Insurance Work and Health Group, Faculty of Medicine Nursing and Health Sciences, Monash University, 2019.

76. Collie A, Lane TJ, Beck D, et al. Evaluating the impact of the 2012 legislative amendments to the NSW workers' compensation scheme. Melbourne, Australia: Insurance Work and Health Group, Monash University, 2018. DOI:10.26180/5c353a4ec498e.

77. Lane TJ, Gray S, Hassani-Mahmooei B, et al. Effectiveness of employer financial incentives in reducing time to report worker injury: An interrupted time series study of two Australian workers' compensation jurisdictions. *BMC Public Health* 2018; 18: 00.

78. Lane TJ, Gray SE, Sheehan L, et al. Increased benefit generosity and the impact on workers' compensation claiming behavior: An interrupted time series study in Victoria, Australia. *J Occup Environ Med* 2019; 61: e82–e90.

79. Australian Government Department of Health Therapeutic Goods Administration. Final decisions and reasons for decisions by delegates of the Secretary to the Department of Health. Canberra, Australia: Australian Government Department of Health, 2017.

80. Lane TJ, Berecki-Gisolf J, Iles R, et al. The impact of long-term workers' compensation benefit cessation on welfare and health service use: Protocol for a longitudinal controlled data linkage study. *Int J Popul Data Sci* 2021; 6: 1419.