BMJ Open Sport & Exercise Medicine

Kumanu Tāngata: the aftermatch study – protocol to examine the health outcomes of high-level male rugby union players using linked administrative data

Stephanie D'Souza ,[1,2] Barry J Milne,[1,2] Chao Li,[1] Francesca Anns,[1,3] Andrew Gardner,[4] Thomas Lumley,[5] Susan M B Morton,[6] Ian R Murphy,[7] Evert Verhagen ,[8] Craig Wright,[9] Ken Quarrie[10]

To cite: D'Souza S, Milne BJ, Li C, *et al*. Kumanu Tāngata: the aftermatch study – protocol to examine the health outcomes of high-level male rugby union players using linked administrative data. *BMJ Open Sport & Exercise Medicine* 2024;**10**:e001795. doi:10.1136/bmjsem-2023-001795

Accepted 16 January 2024

For numbered affiliations see end of article.

Correspondence to
Dr Stephanie D'Souza;
s.dsouza@auckland.ac.nz

ABSTRACT

There is increasing interest in the potential long-term outcomes of participation in contact and collision sports, driven by evidence of higher rates of neurodegenerative diseases among former athletes. Recent research has capitalised on large-scale administrative health data to examine health outcomes in contact sport athletes. However, there is limited research on outcomes associated with participation in rugby union, a contact sport with a relatively high incidence of head trauma and musculoskeletal injuries. Additionally, there is scope to investigate a greater range of health outcomes using large, population-based administrative data. The Kumanu Tāngata project is a retrospective cohort study that will use linked information from the New Zealand Rugby Register and health records within a comprehensive deidentified whole-population administrative research database known as the Integrated Data Infrastructure. First-class male rugby union players (N=13 227) will be compared with a general population comparison group (N=2 438 484; weighting will be applied due to demographic differences) on a range of mortality and morbidity outcomes (neurodegenerative diseases, musculoskeletal conditions, chronic physical conditions, mental health outcomes). A range of player-specific variables will also be investigated as risk factors. Analyses will consist primarily of Cox proportional hazards models. Ethics approval for the study has been granted by the Auckland Health Research Ethics Committee (Ref. AH23203). Primary research dissemination will be via peer-reviewed journal articles.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Previous research has highlighted a higher risk of neurodegenerative diseases in contact sport athletes, though there have been limited insights into rugby union and a narrow focus on specific health outcomes.

WHAT THIS STUDY ADDS

⇒ By using New Zealand's Integrated Data Infrastructure, this protocol proposes a longitudinal examination of a comprehensive range of health outcomes (mortality, neurodegenerative diseases, musculoskeletal conditions, chronic physical conditions, mental health outcomes) in high-level male rugby players, comparing them to a general population cohort. The large sample size increases the study's statistical power to detect effects more effectively than previous research in the field.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Findings from a large-scale, retrospective study of this nature could influence future health policies and athlete care strategies in contact sports, particularly in mitigating health risks associated with rugby union.

INTRODUCTION

The long-term health effects of participating in physical contact or collision sports ('contact sports'), particularly various forms of football, have become a key focus in sports health research. When compared with those taking part in non-contact sports, participants in contact sports tend to report a high incidence of head trauma, such as concussions.[1] The significance of this issue is highlighted by the identification of traumatic brain injuries (TBI) as a potentially modifiable risk factor for dementia.[2] In addition, postmortem examinations have revealed evidence of chronic traumatic encephalopathy neuropathological change (CTE-NC) in the brains of former athletes,[3 4] with associations observed between CTE-NC and repetitive head impacts in contact sports.[5] Although questions remain about the clinical significance of CTE-NC,[6] these findings have led to increasing attention, both in scientific publications[6–8] and in books and media reports,[9 10] of the potential

long-term impacts of mild TBI and repetitive head acceleration events resulting from exposure to contact sports.

Consequently, much of the research on health outcomes linked to participation in contact sports has focused on neurodegenerative diseases. Recent systematic reviews and meta-analyses have indicated an increased risk of neurodegenerative diseases and mortality attributed to neurodegenerative diseases, particularly in professional-level contact sport athletes.[6–8 11] In studies using large-scale administrative data, higher mortality from neurodegenerative diseases has been reported in professional association football players[12–15] and former National Football League athletes[16 17] when compared with the general population.

There have also been concerns regarding the mental health outcomes of contact sport athletes, partly due to neuropsychiatric presentations such as depression and suicide that are believed to be associated with CTE-NC.[18–20] However, emerging literature using linked administrative health records has challenged this idea. In cohorts of Finnish contact sport athletes[21] and professional Scottish[18] and elite Swedish[22 23] football players, athletes were shown to have no increased risk, and possibly even a reduced risk, of mood and substance use disorders as well as suicide relative to the general population.

It is also important to acknowledge the many health-related benefits linked to participation in physical activity, including improved cardiovascular health and increased longevity.[24 25] A meta-analysis on studies of former elite athletes reported increased longevity and a reduced incidence of cardiovascular disease (CVD) and cancer mortality in athletes relative to the general population, though this study did not examine contact sport athletes exclusively.[25] Studies reported above that demonstrated an increased risk of neurodegenerative disease mortality in professional athletes have also indicated lower all-cause mortality in players when compared with the general population.[12–16] Therefore, this research suggests that contact sports may increase an athlete's risk of neurodegenerative disease mortality but may also protect against all-cause mortality and other chronic physical health conditions such as CVD.

Longitudinal studies on outcomes in team-based contact sports have largely been conducted in cohorts of elite or professional associations or American football players, and questions remain about whether results can be extended to other contact sports. One sport in which we may expect to see long-term health consequences is rugby union, a popular contact sport with a relatively high incidence of head trauma.[1 26] A 2018 meta-analysis of 47 studies on head injuries in amateur and professional contact sport athletes reported that rugby had the highest concussion incidence (3.89 concussions per 1000 hours) when compared with association football, ice hockey and American football.[1] A more recent meta-analysis of studies published from 2012 to 2020 focusing only on elite men's rugby union players reported an even higher match concussion incidence rate of 12 per 1000 hours.[26] The physically demanding nature of the sport also results in a relatively high incidence of a range of other injuries among players relative to participants in non-contact team sports, with musculoskeletal injuries being one of the most prevalent injury types reported at a professional level.[27]

Results from a single longitudinal study using administrative data to investigate mortality outcomes in former elite Scottish international rugby players[28] were similar to those reported with other contact sports. Specifically, the study by Russell *et al* observed that while all-cause mortality was similar between the players and a general population comparison group—in fact, rugby players were less likely to die before the age of 70—the risk of dying with a neurodegenerative disease was higher among rugby players, as was the risk of developing a neurodegenerative disease. A much earlier study on the longevity of representatives of New Zealand's (NZ) national male rugby team, the All Blacks, found no difference in life expectancy between rugby players and the general population.[29]

Cross-sectional research on the health outcomes of rugby union players has reported that higher proportions of former elite rugby players have osteoarthritis.[30 31] In another cross-sectional NZ study, amateur and elite rugby players were more likely to engage in hazardous drinking of alcohol than former non-contact sport athletes, but few differences were observed between the groups in terms of current self-reported mood, other substance use and general physical and psychological health.[32] A systematic review published in 2018 concluded that the evidence regarding self-reported concussion, mood and cognitive outcomes among former elite rugby players internationally is equivocal.[33]

To understand the link between participation in contact sports and long-term health outcomes, there is a need for large-scale longitudinal research. While several studies have used administrative data to examine health outcomes in contact sport athletes relative to the general population, many of these studies lack a sample size large enough to examine health outcomes with low prevalence (eg, specific neurodegenerative diseases). The current study aims to address this gap by using a whole population data resource from NZ—the Integrated Data Infrastructure (IDI)[34 35]—to investigate a wider range of health outcomes with considerably larger sample sizes than the studies conducted to date.

The IDI is a population-based linked administrative database maintained by the national statistical agency of NZ, Statistics NZ, which is made available to research under strict access conditions designed to maintain the confidentiality of the data.[34] The IDI contains deidentified data on individuals derived from various governmental sectors and other sources, including comprehensive health data from Manatū Hauora, the NZ Ministry of Health and data on male first-class rugby players from the NZ Rugby Register. Compiled by NZ rugby historian Clive Akers, the NZ Rugby Register contains a list of NZs

who played rugby at a first-class (provincial or higher) level and were active between 1870 and 2015.[36] Details for each player include date of birth, the number of first-class matches played for NZ teams and the years in which the individual was active in play at a first-class level.

To address gaps in knowledge regarding the health impacts of playing high-level rugby, our team of researchers will use data from the NZ Rugby Register within the IDI to develop the Kumanu Tāngata project. The project will investigate the development of health outcomes in representative male rugby players relative to a population-based comparison group. Outcomes will include all-cause and cause-specific mortality, neurodegenerative disease, mental health diagnoses, musculoskeletal outcomes and other chronic physical health conditions such as CVD. This selection was based on previously cited studies that have observed an association between participation in contact sports and these outcomes. The project name 'Kumanu Tāngata' was graciously gifted to our project by Māori Kaumātua (respectable indigenous NZ elder), Luke Crawford, a Māori advisor for NZ Rugby. This name embodies the concept of combining multiple data threads to care for players, aligning with our study's goal to explore what lies beyond a professional sporting career.

METHODS AND ANALYSIS

The Integrated Data Infrastructure

As a whole-population, deidentified research database, the IDI contains a central concordance table known as the IDI 'spine', which comprises an ever-resident NZ population based on tax records (from 1999), birth records (from 1920) and long-term visas (from 1997).[37] Data tables supplied by government and non-government agencies are probabilistically linked to the spine by Statistics NZ.[38]

Cohort identification

Rugby players

Name, sex and date of birth information for individuals in the NZ Rugby Register were used by Statistics NZ to enable linkage to the IDI spine. Following linkage, data from the Rugby Register were stripped of identifying information before access was made available to approved researchers. A total of n=16 101 records for male, first-class players active between 1950 and 2000 were submitted to Statistics NZ for linkage to the IDI by NZ Rugby, and accurate linkage was achieved for 13 800 (85.7%) players.[39]

Additional exclusion criteria were applied to derive an analysis cohort for the proposed study, as described in figure 1. First, individuals born before 1920 (based on year of birth from the Rugby Register), the earliest year containing reliable date of birth information in the IDI, and any players who ceased their playing career before 1950 but were provided in error were excluded. Second, to exclude cases with possible linkage errors, we excluded players who were classified as female in the IDI's 'Personal Details' table (the primary demographics table) or who had birth years that were discrepant by greater than 1 year between the Personal Details table and the Rugby Register. Third, we only included players aged at least 18 years in their final year playing first-class rugby (based on final-year active data in the Rugby Register). Finally, all individuals who were deceased (based on the date of death from Personal Details) or left the country (based on the IDI's 'Person Overseas Spell' table) before their 30th birthday were excluded to allow for most players to have finished their high-level rugby career while also providing time for the development of most chronic health issues. The final rugby cohort consisted of 13 227 players (82.2% of the original register), all born between 1920 and 1984.

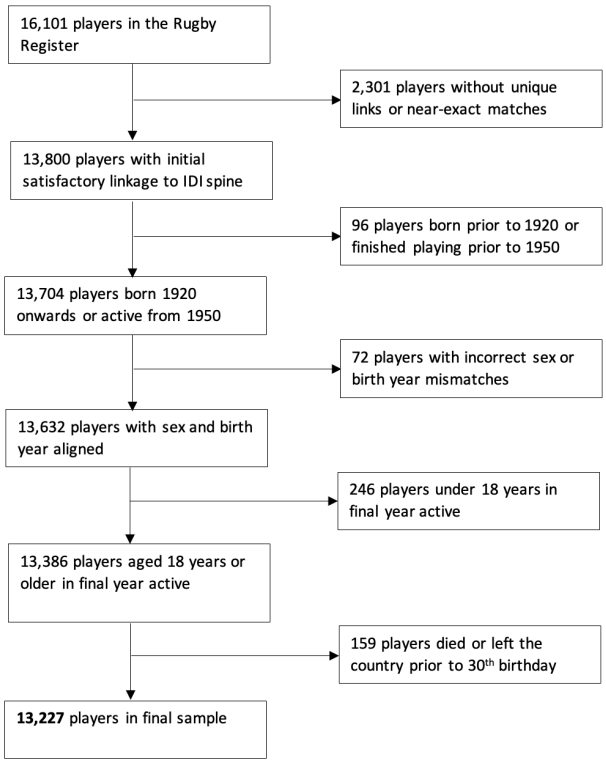

Figure 1 Flow chart demonstrating exclusion criteria applied to players. All counts have been random rounded to base 3, per the reporting requirements of Statistics New Zealand.

Descriptive statistics on variables from the Rugby Register for the original 16 101 players provided to Statistics NZ and the final cohort of 13 227 are provided in table 1. The distributions of total years active, number of career matches, player position and level of play were similar across the original and final cohorts.

General population

To align with players, the general population comparison group comprised all males born between 1920 and 1984 who were part of NZ's ever-resident population (ie, in the IDI spine), were alive and in the country at the age of 30,

Table 1 Characteristics of rugby players in the original cohort and the final cohort

	Original cohort (N=16 101) n (%)	Final cohort (N=13 227) n (%)
Total years active		
1	5922 (36.8)	4428 (33.5)
2–3	3141 (19.5)	2646 (20.0)
4–6	3069 (19.1)	2670 (20.2)
7–10	2625 (16.3)	2334 (17.6)
>10	1344 (8.3)	1149 (8.7)
Total no of career matches		
1–2	4392 (27.3)	3228 (24.4)
3–10	4377 (27.2)	3579 (27.1)
11–25	2889 (17.9)	2472 (18.7)
26–50	1983 (12.3)	1761 (13.3)
51–100	1515 (9.4)	1338 (10.1)
>100	777 (4.8)	714 (5.4)
Missing	168 (1.0)	135 (1.0)
Player position		
Back*	75 (0.5)	57 (0.4)
Centre	513 (3.2)	399 (3.0)
First 5th–8th	438 (2.7)	372 (2.8)
5th–8th	1047 (6.5)	891 (6.7)
Flanker	1356 (8.4)	1095 (8.3)
Forward*	564 (3.5)	456 (3.4)
Fullback	933 (5.8)	795 (6.0)
Halfback	1134 (7.0)	942 (7.1)
Hooker	831 (5.2)	690 (5.2)
Lock	1548 (9.6)	1272 (9.6)
Loose forward	429 (2.7)	378 (2.9)
Midfield	357 (2.2)	315 (2.4)
No 8	408 (2.5)	321 (2.4)
Prop	1473 (9.1)	1164 (8.8)
Second 5th–8th	324 (2.0)	264 (2.0)
Three-quarter	483 (3.0)	408 (3.1)
Wing	2211 (13.7)	1785 (13.5)
Missing	1983 (12.3)	1626 (12.3)
Level of rugby played		
International/ professional	822 (5.1)	744 (5.6)
Amateur/provincial	15 282 (94.9)	12 483 (94.4)

*For some players, positions were not specified beyond forward or back.

and who were not in the 'rugby players' group (n=2 438 484).

Because the ever-resident population in the IDI includes all those who have arrived in NZ on a long-term visa since 1997 and those who have paid tax in NZ since 1999, it is more likely to include individuals who are younger, more ethnically diverse and born overseas. This is demonstrated in table 2 (middle column). Players were more likely to be of European and Māori ethnicity and less likely to identify with Pacific, Asian, and Middle Eastern, Latin American and African (MELAA) ethnicities than the comparison group. Players also had a greater proportion in older birth cohorts than younger birth cohorts and were predominantly NZ-born.

Mortality outcomes vary by ethnicity and migrant status in NZ.[40] Additionally, older individuals will have an increased risk and shorter time for certain health outcomes, such as mortality and neurodegenerative diseases. Therefore, it is important to adjust for the differential age, ethnic and overseas-born composition of the two groups. To control for demographic differences between players and the comparison group, we used direct standardisation to weight the non-players to match that of the players (the standardised population) by total response ethnicity, birth year (5-year bands from 1920–1924 to 1980–1984) and overseas-born status, using the psmatch2 command in Stata (V.16.0). This achieves a one-to-many match between players and non-players for all ethnicity by birth year by overseas-born strata.

Weighted descriptives for non-players demonstrate that after weighting, the sociodemographic composition among players and the comparison group is almost identical (table 2, rightmost column).

Outcome variables and data sources

Mortality outcomes

We will investigate both all-cause and cause-specific mortality. Cause-specific mortality will include death due to neurodegenerative disease (any, Alzheimer's disease, other dementias, Parkinson's disease, motor neuron disease), CVD, cancer, suicide and substance use-related deaths (including alcohol-related deaths). Two sources from the IDI will be used to determine mortality information: the Personal Details table and the Ministry of Health Mortality Collection. The date of death in the Personal Details table is specifically sourced from both the DIA Death Registrations and the Ministry of Health Mortality Collection, with dates reliably available for full calendar years from 1920.

The Ministry of Health Mortality Collection has date and cause of death information for deaths from 1988 to 2018. The lag in data availability for these records is due to the timeframe required for the coronial process regarding the cause of death.[41] For deaths before July 1999, the cause can be determined using codes from the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). For deaths occurring from July 1999 onwards, the cause of death was coded using the ICD and Related Health Problems, 10th Revision, Australian Modification (ICD-10-AM).

All-cause mortality will be identified for our analyses using recorded deaths within the personal details table

Table 2 Sociodemographic characteristics of rugby players and general population males in the study cohort

	Players (N=13 227) n (%)	Unweighted general population males (N=2 438 484) n (%)	Weighted general population males (N=2 438 484) n (%)
Ethnicity			
European	10 425 (78.8)	1 466 574 (60.1)	1 921 917 (78.8)
Māori	2424 (18.3)	231 780 (9.5)	446 880 (18.3)
Pasifika	483 (3.7)	122 685 (5.0)	89 229 (3.7)
Asian	39 (0.3)	73 371 (3.0)	7374 (0.3)
MELAA	102 (0.8)	40 917 (1.7)	18 117 (0.7)
Other	267 (2.0)	48 324 (2.0)	40 737 (1.7)
NZ-born status			
NZ born	12 564 (95.0)	1 435 932 (58.9)	2 316 441 (95.0)
Overseas born	660 (5.0)	1 002 552 (41.1)	122 043 (5.0)
Birth year			
1920–1924	444 (3.4)	97 848 (4.0)	81 672 (3.3)
1925–1929	1029 (7.8)	101 796 (4.2)	189 885 (7.8)
1930–1934	1356 (10.3)	104 571 (4.3)	250 356 (10.3)
1935–1939	1206 (9.1)	109 809 (4.5)	221 964 (9.1)
1940–1944	1227 (9.3)	141 642 (5.8)	226 392 (9.3)
1945–1949	1323 (10.0)	177 291 (7.3)	243 906 (10.0)
1950–1954	1185 (9.0)	202 284 (8.3)	218 277 (9.0)
1955–1959	1119 (8.5)	238 455 (9.8)	205 926 (8.4)
1960–1964	1209 (9.1)	299 589 (12.3)	223 071 (9.1)
1965–1969	1152 (8.7)	276 978 (11.4)	212 565 (8.7)
1970–1974	1146 (8.7)	300 720 (12.3)	210 903 (8.6)
1975–1979	753 (5.7)	229 638 (9.4)	139 191 (5.7)
1980–1984	78 (0.6)	157 866 (6.5)	14 379 (0.6)

All counts have been random rounded to base 3, per the reporting requirements of Statistics NZ. 'Total response' ethnicity is reported, which allows individuals to identify with more than one ethnic identity, which is common in NZ.[55]
MELAA, Middle Eastern, Latin American or African; NZ, New Zealand.

to allow for a wider observation window for investigating our outcome. Cause-specific mortality will be identified using recorded deaths within the mortality collection, but only events from 1988 will be coded. All contributing diagnosis codes (not just primary diagnosis) will be used to determine the cause of death to increase case identification. Due to data availability, study cohort members who died or left the country before 1988 (when mortality collection records were available) will be excluded from cause-specific mortality analyses.

Morbidity outcomes

Morbidity outcomes in the current study will include neurodegenerative diseases (any, Alzheimer's disease, other dementias, Parkinson's disease, motor neuron disease), musculoskeletal conditions, chronic conditions (CVD, cancer, diabetes) and mental health outcomes (any, substance use disorders, psychotic disorders, mood disorders, neurotic disorders, physiological-disturbance disorders, personality disorders, behavioural disorders, unspecified). The primary sources for identifying morbidity outcomes will include the National Minimum Dataset (NMDS) and the mortality collection. The NMDS contains information on all publicly funded hospital admissions in NZ for the full calendar year from 1988. Like the mortality collection, the NMDS used the ICD-9-CM for admissions before July 1999 and the ICD-10-AM for admissions after this date. All contributing diagnosis codes (not just primary diagnosis) will be used to determine morbidity outcomes. As NMDS data are only available from 1988, analyses using hospitalisation records will exclude study cohort members who died or left the country before this date.

Where appropriate, community pharmaceutical dispensing data from the pharmaceutical collection will also be used to identify cases for morbidity outcomes. These data contain information for all

government-subsidised pharmaceuticals dispensed, with reliable data available for full calendar years from 2007.[42] Pharmaceutical codes such as chemical and formulation IDs can identify specific types of dispensed medications. Pharmaceutical data will be used for case identification only when a particular type of medication is used exclusively or predominantly to treat an outcome of interest, consistent with previous studies using IDI data to identify specific health outcomes.[43–47] However, as it is impossible to determine why an individual received a particular medication, we cannot rule out the possibility that a medication was dispensed for a condition separate from our outcome of interest. Pharmaceutical data are also unavailable from 1988 to 2006, while hospitalisation records are available. As such, all analyses will be conducted including and excluding pharmaceuticals, to determine if associations still hold.

Planned statistical analysis

Full population analyses

Weighted Cox proportional models estimating HRs will be conducted for each mortality and morbidity outcome, comparing players to the general population comparison group. Proportional hazard assumption tests and an evaluation of competing risks will be conducted.

Power calculations were conducted to estimate the effect size (HR estimate) able to be detected with 80% power, given the sample size and estimated prevalences of exposures and outcomes. For full population analyses comparing players to non-players, we can detect an HR of 1.05 for all-cause mortality.[i] HRs can be detected for cause-specific mortality ranging from 1.08 for the most common outcome (cardiovascular-related death; estimated prevalence 9.6%[ii]) to 1.34 for the rarest outcome (suicide; 0.7%[iii]). For morbidity outcomes, we can detect HRs ranging from 1.07 for the most common outcome (diabetes; 14.5%[iv]) to 1.19 for the rarest outcome (neurodegenerative diseases; 2.0%[v]).

Rugby-specific analyses

Variables of interest for the rugby-specific analyses will include player position (forward or back; see details of recoding below), exposure to the sport (total numbers of years active and career matches played), and an indicator of the level of rugby played (international/professional level vs amateur representative level). Positions from the Rugby Register recoded as forwards are: forward; loosehead prop; hooker; tight head prop; left lock; right lock; left flanker; right flanker; number eight. Positions recoded as backs are: back; scrum-half; fly half; left wing; left centre; right centre; right wing and full back.

Four Cox proportion hazard analyses will be conducted for each outcome, corresponding to each of the rugby-specific predictor variables stated above. All models will adjust for birth year, dummy-coded ethnicity variables and NZ-born status.

Power calculations were conducted to estimate the effect size (HR estimate) able to be detected with 80% power, given the sample size and estimated prevalences of exposures and outcomes. To demonstrate the range of effect sizes able to be detected, we will report results for our most common exposure (player position, categorised into forward (47.1%) and back (40.6%) and our rarest exposure (level of play, see table 1). For all-cause mortality, we can detect an HR of 1.09 for player position and 1.21 for level of play. When comparing forwards and backs on cause-specific mortality, we can detect an HR of 1.17 for cardiovascular-related death and 1.81 for suicide. When comparing the level of play categories, power calculations place our detectable HRs at 1.39 and 3.40 for cardiovascular-related death and suicide, respectively. For the most common morbidity outcome, diabetes, detectable HRs are 1.14 for player position analyses and 1.31 for level of play. For the rarest morbidity outcome, neurodegenerative diseases, detectable HRs are 1.42 for player position and 2.06 for level of play.

Patient and public involvement

There was no direct patient or public involvement in the development of the Kumanu Tāngata study protocol.

DISCUSSION AND DISSEMINATION

The Kumanu Tāngata project will be the largest study yet undertaken investigating health outcomes in high-level male rugby players. Linkage to the IDI, a world-leading population-based administrative research database, allows for longitudinal follow-up, ascertainment of a range of health outcomes in administrative data and comparison to a general population weighted to match key characteristics of players. The study will provide important insights into the long-term outcomes of playing high-level rugby union and further our understanding of the benefits and risks of playing contact and collision sports at a first-class level.

As the study will employ secondary data analysis, individual consent from cohort members is not required. The study has also undergone independent review by experts in sports injury epidemiology as a part of the funding process. Statistics NZ has approved the use of the IDI for this project (ref MAA2016-21). As part of granting access to the IDI, Statistics NZ reviewed the study to ensure that it meets its 'Five Safes' policy relating to data privacy.[48]

[i] [1]Estimated for mortality by 2018 from New Zealand cohort life tables for males 1921–1981 (https://www.stats.govt.nz/information-releases/new-zealand-cohort-life-tables-march-2020-update/), given the prevalence of birth year in our sample (table 2), and assuming 10% of deaths occur overseas so will be missed.

[ii] [2]Estimated as fraction of total deaths 2011–2018, from https://tewhatuora.shinyapps.io/mortality-web-tool/

[iii] [3]Estimated as 2.4% of all male deaths, see https://mentalhealth.org.nz/suicide-prevention/statistics-on-suicide-in-new-zealand

[iv] [4]Estimated from Virtual Diabetes Register, see https://www.tewhatuora.govt.nz/our-health-system/data-and-statistics/virtual-diabetes-tool/#:~:text=and%20age%20group.-,Key%20findings%20from%20the%202021%20Virtual%20Diabetes%20Register,%2C%2041.7)%20per%201000%20population.

[v] [5]Estimated from[43]

This policy requires that (1) Data can only be accessed by researchers who have undergone confidentiality training by Statistics NZ, signed a declaration of secrecy under the Statistics Act 1975 and signed a contract agreeing to follow the rules and protocols established by Statistics NZ ('safe people'); (2) The project is considered beneficial to the public ('safe project'); (3) Data are accessed within a secure Datalab environment that only approved researchers can access ('safe settings'); (4) Only deidentified data necessary for the project aims are provided to researchers ('safe data') and (5) All data and results produced are confidentialised according to Statistics NZ protocols and have undergone checking by Statistics NZ before being shared outside of the secure Datalab environment ('safe output'). Privacy impact assessments conducted by Statistics NZ have also indicated a low risk of privacy breaches.[49] Statistics NZ has conducted a privacy impact assessment for the inclusion of data from the NZ Rugby Register into the IDI and concluded that benefits outweigh potential risks.[50]

The research team has a history of producing robust, high-impact research using IDI data.[43–46] Therefore, we anticipate this study will produce several manuscripts targeted towards high-impact journals. We foresee at least four peer-reviewed journal articles reporting on the following outcomes: (1) all-cause and cause-specific mortality; (2) morbidity due to neurodegenerative disease; (3) morbidity due to other physical health outcomes (CVD, musculoskeletal conditions, cancer) and (4) morbidity related to mental health conditions. The Strengthening of Observational Studies in Epidemiology (STROBE) guidelines,[51] as well as the STROBE Extension for Sports Injury and Illness Surveillance (STROBE),[52] will be used to guide the dissemination of findings in journal articles. Results will also be shared with key stakeholders and study funders, including NZ Rugby, the NZ Rugby Foundation and World Rugby. We also anticipate sharing preliminary findings at appropriate scientific meetings and conferences.

The current study possesses some limitations. First, we do not have access to some key confounders related to health outcomes, such as weight or body mass index, family health history, socioeconomic status and other lifestyle risk factors (eg, alcohol and drug use). Second, we cannot measure direct exposures, such as head impacts and collisions, that might mediate associations between rugby playing and health outcomes. Third, we do not have information on and are, therefore, unable to accurately adjust for, the sporting history of our comparison group, which will include individuals who participated in lower-level rugby or other contact and collision sports. Fourth, the Rugby Register data in the IDI only contains information on male first-class rugby players. Therefore, our findings may not reflect health outcomes for female rugby players.

Finally, we are limited to primarily hospitalisation and mortality records, and to some extent, pharmaceutical dispensing, for identifying the health outcomes of interest in our cohort. We cannot access detailed primary health information, where milder cases may be captured. As such, our sample is likely to be biased towards those with more severe impairments in relation to our outcomes of interest. Furthermore, our outcome identification will be based on individuals seeking services or receiving treatment, who may be sociodemographically different from those who do not access these services.[53] This has been observed in previous research where women screened for prenatal depression were socioeconomically and ethnically different from those accessing pharmacological treatment for depression.[54] In addition, we cannot ascertain CTE-NC in these records, an outcome of considerable interest in contact sport research. CTE-NC can only be identified postmortem and attempts to identify it have not been performed as part of routine postmortem examinations. Additionally, diagnostic codes do not exist in the ICD-9-CM or the ICD-10-AM for CTE-NC. These limitations do not negate the notable strengths of the Kumanu Tāngata project but will be acknowledged in the interpretation and dissemination of the study's results.

Author affiliations
[1]COMPASS Research Centre, University of Auckland, Auckland, New Zealand
[2]School of Social Sciences, University of Auckland, Auckland, New Zealand
[3]School of Psychology, University of Auckland, Auckland, New Zealand
[4]Sydney School of Health Sciences, The University of Sydney Faculty of Medicine and Health, Sydney, New South Wales, Australia
[5]Department of Statistics, University of Auckland, Auckland, New Zealand
[6]Research Institute for Innovative Solutions for Well-being and Health (INSIGHT), University of Technology Sydney, Sydney, New South Wales, Australia
[7]Illawarra Shoalhaven Local Health District, Wollongong, New South Wales, Australia
[8]Department of Public and Occupational Health, EMGO, Amsterdam UMC Locatie VUmc, Amsterdam, Netherlands
[9]Social Wellbeing Agency, Wellington, New Zealand
[10]New Zealand Rugby, Wellington, New Zealand

Twitter Evert Verhagen @evertverhagen

Acknowledgements We would like to acknowledge Statistics New Zealand and their staff for loading the New Zealand Rugby Register data to the IDI, for access to the IDI and for their timely review of output data. We would also like to thank the Public Policy Institute at the University of Auckland for access to their Statistics New Zealand Datalab.

Contributors CW and KQ conceived the projec, and digitised the data from the NZ Rugby Register so that it could be loaded into the NZ IDI. KQ, BJM, SD'S, SMBM, CW, AG, IRM, EV and TL contributed to overall project planning and development of the funding application. Analyses were planned by SD'S, BJM, KQ, FA, CL, TL and SMBM. Descriptive statistics were produced by CL, SD'S and BJM. Power calculations were conducted by BJM and TL. SD'S, KQ, BJM and FA drafted the report. All authors contributed to final drafting and editing and accept responsibility for the decision to submit the report for publication. Access to data for analyses in the IDI environment was restricted to CW, BJM, FA, SD'S and CL.

Funding This work was supported by World Rugby and the New Zealand Rugby Foundation (grant number G-2004-00963).

Disclaimer These results are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI) which is carefully managed by Statistics New Zealand. For more information about the IDI please visit https://www.stats.govt.nz/integrated-data/.The funders had no role in study conception or design, data collection, data analysis, data interpretation, or writing of the manuscript. World Rugby obtained an independent, external review of the project plan prior to funding.

Competing interests KQ has been employed by New Zealand Rugby since 2000 and currently occupies the role of Chief Scientist, New Zealand Rugby. He also sits on World Rugby's Scientific Committee and has contributed to various

World Rugby working groups focussed on player welfare issues from 2011 to the time of publication. He has received funding for flights and accommodation from World Rugby and New Zealand Rugby to attend World Rugby Medical Commission Conference meetings. KQ's input into the project was conducted in collaboration with his coauthors, who are independent of New Zealand Rugby and World Rugby, and he declares that at no time did NZ Rugby, nor the funding agencies World Rugby and the New Zealand Rugby Foundation provide advice or input into the data analysis, interpretation of the findings, or the manuscript preparation. A copy of the submitted draft of the project was provided to the funding bodies and New Zealand Rugby at the time of submission. AJG has been a contracted concussion consultant to Rugby Australia since July 2016. He is a member of the World Rugby Concussion Working Group. He has received travel funding or been reimbursed by professional sporting bodies, and commercial organisations for discussing or presenting sport-related concussion research at meetings, scientific conferences, workshops, and symposiums. He has a clinical practice in neuropsychology involving individuals who have sustained sport-related concussion (including current and former athletes). He is a member of the Australian Football League Concussion Scientific Advisory Committee. He is supported by a National Health and Medical Research Council (NHMRC) Investigator Grant. He acknowledges unrestricted philanthropic support from the National Rugby League for research in former elite-level rugby league players.IM was employed by New Zealand Rugby from 2012 to 2019 as the Chief Medical Officer. During this time, he also held various committee appointments with World Rugby focused on player welfare matters.

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

Patient consent for publication Not applicable.

Ethics approval Ethics approval for the study has been granted by the Auckland Health Research Ethics Committee (Ref. AH23203).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. Data from the IDI cannot be shared by the authors. Access to the anonymised data used in this study was provided by Statistics NZ under the security and confidentiality provisions of the Statistics Act 1975. Anyone who wishes to access the data must submit an application through Statistics New Zealand. Requests require a concept paper describing the purpose of data access, ethical approval and provision for secure data access. Further details can be found at www.stats.govt.nz/integrated-data/apply-to-use-microdata-for-research/.

ORCID iDs
Stephanie D'Souza http://orcid.org/0000-0003-1458-2147
Evert Verhagen http://orcid.org/0000-0001-9227-8234

REFERENCES

1 Prien A, Grafe A, Rössler R, *et al*. Epidemiology of Head Injuries Focusing on Concussions in Team Contact Sports: A Systematic Review. *Sports Med* 2018;48:953–69.
2 Livingston G, Huntley J, Sommerlad A, *et al*. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet* 2020;396:413–46.
3 Omalu BI, DeKosky ST, Hamilton RL, *et al*. Chronic traumatic encephalopathy in a national football league player: part II. *Neurosurgery* 2006;59:1086–92;
4 McKee AC, Stern RA, Nowinski CJ, *et al*. The spectrum of disease in chronic traumatic encephalopathy. *Brain* 2013;136:43–64.
5 Daneshvar DH, Nair ES, Baucom ZH, *et al*. Leveraging football accelerometer data to quantify associations between repetitive head impacts and chronic traumatic encephalopathy in males. *Nat Commun* 2023;14:3470.
6 Batty GD, Frank P, Kujala UM, *et al*. Dementia in former amateur and professional contact sports participants: population-based cohort study, systematic review, and meta-analysis. *EClinicalMedicine* 2023;61:102056.
7 Bellomo G, Piscopo P, Corbo M, *et al*. A systematic review on the risk of neurodegenerative diseases and neurocognitive disorders in professional and varsity athletes. *Neurol Sci* 2022;43:6667–91.
8 Iverson GL, Castellani RJ, Cassidy JD, *et al*. Examining later-in-life health risks associated with sport-related concussion and repetitive head impacts: a systematic review of case-control and cohort studies. *Br J Sports Med* 2023;57:810–21.
9 Fainaru-Wada M, Fainaru S. *League of Denial: The NFL, Concussions and the Battle for Truth*. 1st edition. New York: Crown Archetype, 2013.
10 Ward J. 111 N.F.L. brains. all but one had C.T.E. The New York Times; 2017. Available: https://www.nytimes.com/interactive/2017/07/25/sports/football/nfl-cte.html, https://www.nytimes.com/interactive/2017/07/25/sports/football/nfl-cte.html [Accessed 30 Aug 2022].
11 Morales JS, Castillo-García A, Valenzuela PL, *et al*. Mortality from mental disorders and suicide in male professional American football and soccer players: a meta-analysis. *Scand J Med Sci Sports* 2021;31:2241–8. 10.1111/sms.14038 Available: https://onlinelibrary.wiley.com/toc/16000838/31/12
12 Mackay DF, Russell ER, Stewart K, *et al*. Neurodegenerative Disease Mortality among Former Professional Soccer Players. *N Engl J Med* 2019;381:1801–8.
13 Russell ER, Mackay DF, Stewart K, *et al*. Association of Field Position and Career Length With Risk of Neurodegenerative Disease in Male Former Professional Soccer Players. *JAMA Neurol* 2021;78:1057–63.
14 Ueda P, Pasternak B, Lim C-E, *et al*. Neurodegenerative disease among male elite football (soccer) players in Sweden: a cohort study. *Lancet Public Health* 2023;8:e256–65.
15 Orhant E, Carling C, Chapellier J-F, *et al*. A retrospective analysis of all-cause and cause-specific mortality rates in French male professional footballers. *Scand J Med Sci Sports* 2022;32:1389–99. 10.1111/sms.14195 Available: https://onlinelibrary.wiley.com/toc/16000838/32/9
16 Lehman EJ, Hein MJ, Baron SL, *et al*. Neurodegenerative causes of death among retired National Football League players. *Neurology* 2012;79:1970–4.
17 Daneshvar DH, Mez J, Alosco ML, *et al*. Incidence of and mortality from amyotrophic lateral sclerosis in national football league athletes. *JAMA Netw Open* 2021;4:e2138801.
18 Russell ER, McCabe T, Mackay DF, *et al*. Mental health and suicide in former professional soccer players. *J Neurol Neurosurg Psychiatry* 2020;91:1256–60.
19 Omalu BI, Bailes J, Hammers JL, *et al*. Chronic traumatic encephalopathy, suicides and parasuicides in professional American athletes: the role of the forensic pathologist. *Am J Forensic Med Pathol* 2010;31:130–2.
20 Montenigro PH, Baugh CM, Daneshvar DH, *et al*. Clinical subtypes of chronic traumatic encephalopathy: literature review and proposed research diagnostic criteria for traumatic encephalopathy syndrome. *Alzheimers Res Ther* 2014;6:68.
21 Batty GD, Frank P, Kujala UM, *et al*. Suicide and depression in former contact sports participants: population-based cohort study, systematic review, and meta-analysis. *EClinicalMedicine* 2023;60:102026.
22 Kader M, Pasternak B, Lim C-E, *et al*. Depression and anxiety-related disorders and suicide among Swedish male elite football players: a nationwide cohort study. *Br J Sports Med* 2024;58:66–72.
23 Ueda P, Pasternak B, Svanström H, *et al*. Alcohol related disorders among elite male football players in Sweden: nationwide cohort study. *BMJ* 2022;379:e074093.
24 Warburton DER, Bredin SSD. Health benefits of physical activity: a systematic review of current systematic reviews. *Curr Opin Cardiol* 2017;32:541–56.
25 Runacres A, Mackintosh KA, McNarry MA. Health Consequences of an Elite Sporting Career: Long-Term Detriment or Long-Term Gain? A Meta-Analysis of 165,000 Former Athletes. *Sports Med* 2021;51:289–301.
26 Williams S, Robertson C, Starling L, *et al*. Injuries in Elite Men's Rugby Union: An Updated (2012-2020) Meta-Analysis of 11,620 Match and Training Injuries. *Sports Med* 2022;52:1127–40.
27 Williams S, Trewartha G, Kemp S, *et al*. A meta-analysis of injuries in senior men's professional Rugby Union. *Sports Med* 2013;43:1043–55.
28 Russell ER, Mackay DF, Lyall D, *et al*. Neurodegenerative disease risk among former international rugby union players. *J Neurol Neurosurg Psychiatry* 2022;93:1262–8.
29 Beaglehole R, Stewart A. The longevity of international rugby players. *N Z Med J* 1983;96:513–5.
30 Davies MAM, D. Judge A, Delmestri A, *et al*. Health amongst former rugby union players: A cross-sectional study of morbidity and health-related quality of life. *Sci Rep* 2017;7:11786.

31 Hind K, Konerth N, Entwistle I, *et al*. Cumulative Sport-Related Injuries and Longer Term Impact in Retired Male Elite- and Amateur-Level Rugby Code Athletes and Non-contact Athletes: A Retrospective Study. *Sports Med* 2020;50:2051–61.
32 Hume PA, Quarrie KL, Lewis GN, *et al*. NZ-RugbyHealth Study: Self-reported Injury Experience and Current Health of Former Rugby Union and Non-contact Sport Players. *Sports Med* 2022;52:1701–13.
33 Cunningham J, Broglio S, Wilson F. Influence of playing rugby on long-term brain health following retirement: a systematic review and narrative synthesis. *BMJ Open Sport Exerc Med* 2018;4:e000356.
34 Milne BJ, Atkinson J, Blakely T, *et al*. Data resource profile: the New Zealand integrated data infrastructure (IDI). *Int J Epidemiol* 2019;48:677–677e.
35 Statistics New Zealand. Integrated data infrastructure | Stats NZ. Available: https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/ [Accessed 5 Nov 2018].
36 Akers C. New Zealand Rugby Register 1870-2015: The Players, Referees and Administrators in First-class Rugby. *New Zealand Rugby Museum* 2016.
37 Black A. *The IDI prototype spine's creation and coverage*. Wellington, New Zealand: Statistics New Zealand, 2016.
38 Fellegi IP, Sunter AB. A Theory for Record Linkage. *Journal of the American Statistical Association* 1969;64:1183–210.
39 D'Souza S, Li C, Quarrie KL, *et al*. *Linkage review and evaluation of rugby players and referees added to the Integrated Data Infrastructure*. Auckland, New Zealand: COMPASS Research Centre, University of Auckland, 2022.
40 Hajat A, Blakely T, Dayal S, *et al*. Do New Zealand's immigrants have a mortality advantage? Evidence from the New Zealand Census-Mortality Study. *Ethn Health* 2010;15:531–47.
41 Bowden N, Gibb S, Thabrew H, *et al*. Case identification of mental health and related problems in children and young people using the New Zealand Integrated Data Infrastructure. *BMC Med Inform Decis Mak* 2020;20:42.
42 Gibb S, Cunningham R, Telfer K. Mental health and addiction information in the IDI. Virtual Health Information Network; 2019. Available: https://vhin.co.nz/guides/mental-health-and-addiction-information-in-the-idi/ [Accessed 12 Jul 2022].
43 Richmond-Rakerd LS, D'Souza S, Milne BJ, *et al*. Longitudinal associations of mental disorders with dementia: 30-year analysis of 1.7 million New Zealand citizens. *JAMA Psychiatry* 2022;79:333–40.
44 Bowden N, Gibb S, Audas R, *et al*. Association Between High-Need Education-Based Funding and School Suspension Rates for Autistic Students in New Zealand. *JAMA Pediatr* 2022;176:664–71.
45 Richmond-Rakerd LS, D'Souza S, Milne BJ, *et al*. Longitudinal associations of mental disorders with physical diseases and mortality among 2.3 million New Zealand citizens. *JAMA Netw Open* 2021;4:e2033448.
46 Richmond-Rakerd LS, D'Souza S, Andersen SH, *et al*. Clustering of health, crime and social-welfare inequality in 4 million citizens from two nations. *Nat Hum Behav* 2020;4:255–64.
47 Leong KSW, McLay J, Derraik JGB, *et al*. Associations of prenatal and childhood antibiotic exposure with obesity at Age 4 Years. *JAMA Netw Open* 2020;3:e1919681.
48 Desai T, Ritchie F, Welpton R. *Five Safes: designing data access for research*. Bristol, UK: University of the West of England, 2016.
49 Statistics New Zealand. *Integrated Data Infrastructure: Overarching privacy impact assessment*. Wellington, New Zealand: Stats NZ Tatauranga Aotearoa, 2017.
50 Statistics New Zealand. Wellington, New Zealand: Statistics New Zealand; Privacy impact assessment for adding NZ Rugby Register data to the IDI, . 2018Available: https://www.stats.govt.nz/privacy-impact-assessments/privacy-impact-assessment-for-adding-nz-rugby-register-data-to-the-idi
51 von Elm E, Altman DG, Egger M, *et al*. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The Lancet* 2007;370:1453–7.
52 Bahr R, Clarsen B, Derman W, *et al*. International Olympic Committee consensus statement: methods for recording and reporting of epidemiological data on injury and illness in sport 2020 (including STROBE Extension for Sport Injury and Illness Surveillance (STROBE-SIIS)). *Br J Sports Med* 2020;54:372–89.
53 Milne BJ, D'Souza S, Andersen SH, *et al*. Use of population-level administrative data in developmental science. *Annu Rev Dev Psychol* 2022;4:447–68.
54 Svardal CA, Waldie K, Milne B, *et al*. Prevalence of antidepressant use and unmedicated depression in pregnant New Zealand women. *Aust N Z J Psychiatry* 2022;56:489–99.
55 Statistics New Zealand. Census ethnic group summaries. Stats NZ, . 2018Available: https://www.stats.govt.nz/tools/2018-census-ethnic-group-summaries [Accessed 17 Aug 2023].