

# LATENT DRIVERS OF PLAYER RETENTION IN JUNIOR RUGBY

Harini Dissanayake <sup>a,b,d</sup>, Paul J. Bracewell <sup>a,b</sup>, Emma C. Campbell <sup>b</sup>, Holly E. Trowland <sup>b</sup> and Simon Devoy <sup>c</sup>

<sup>a</sup> Victoria University of Wellington

<sup>b</sup> DOT Loves Data

<sup>c</sup> Auckland Rugby Union

<sup>d</sup>Corresponding author: [Harini@dotlovesdata.com](mailto:Harini@dotlovesdata.com)

## Abstract

To help key stakeholders cultivate an environment that fosters long-term participation in rugby, drivers that encourage young athletes to remain in the sport must be identified and understood. This study investigates the latent drivers of engagement in a junior rugby system for better data informed decisions. This study then demonstrates how combining administrative data with dynamic social datasets objectifies biased perceptions to some degree. Administration-level data was collected each annual season across a three-year period (2017-2019) by the Auckland Rugby Union and analysed to identify the predictors of player retention. Players were categorised according to whether they remained in (or departed from) the sport at the end of each playing season. A multivariate logistic regression model with a stepwise AIC variable selection was employed to identify significant independent predictors of player retention. Squad size, rugby sentiment in the media and deprivation were significant contributors to junior rugby player retention. This demonstrates that player retention is not only driven by weight and peer group participation, which has been the main focus of engaging juniors in rugby in the past, there are other social factors associated with churn.

**Keywords:** churn, logistic regression, repeated measures, junior sport

## 1. INTRODUCTION

In recent years safety issues at the junior level (under 13 years of age) has been topical amongst the New Zealand Rugby community. Some provinces, including Auckland have applied weight limits for each tackle grade to minimise injury risk by having players of similar size play against each other. However, these weight-based systems can create an unaccommodating situation for heavier players who are required to play outside their peer group. Campbell, Bracewell, Blackie & Patel (2018) revealed churn due to the impact of weight limits had a statistically significant effect for Auckland junior rugby players across the 2009-2016 seasons. The significance of playing with peers on player retention is thus established and has in recent years played an important consideration in shaping the framework for creating an age-weight grading system that optimises weight limits in New Zealand (Campbell, Patel & Bracewell, 2018). It has been encouraging that this current study (for seasons 2017-2019) revealed that weight limit effects are much less a driver for juniors leaving the sport (only significant for 10 year olds, where 64% are more likely to leave the game). However, this is amid a backdrop of falling player numbers (Campbell *et al.*, 2018) and does not coincide with a decrease in the churn rate. Of concern is the growing level of churn (from 42% in 2012 to 47% in 2019) suggesting that endeavours to retain players in junior sport by minimizing weight limit effects may be a limited approach that needs to be implemented alongside consideration of other drivers of rugby engagement.

While there is wide spread agreement that playing with peers of the same physical and mental age is an important factor for junior player retention (World Rugby, 2013), the aforementioned increase in churn amid efforts to address age-weight issues suggests that this is an incomplete view. As a consequence, we will investigate if there are factors beyond weight limits that are impacting successful engagement with young kiwis in New Zealand's national sport.

## 2. METHODS

### DATA ANALYSIS

The cleaned data contained approximately 8,000 observations (for the seasons 2017-2019) across 10 variables; churn, age, date of birth, weight, peer group indicator (indicating whether a junior is playing outside peer group because their weight falls outside weight band) and additional variables for the social model; squad size, deprivation index, segment, rugby sentiment, player proximity to club). The study adopts a logistic regression analysis to determine the likelihood of player churn driven by weight limit effect (model 1). We then employ a logistic regression to determine the likelihood of player churn driven by social factors (model 2).

## ADMINISTRATIVE DATA

Auckland Rugby Union provided data for all children aged 3 to 18 years who enrolled in Auckland junior competition 2017 to 2019. On the day of enrolment (January 1st for any given year) basic player information, such as Player ID, Weight, Date of Birth, Year of Registration, Player Address and Club were recorded. Hashed Player ID numbers were used to anonymously track player participation across the seasons to determine churn. Players under the age of 7 have been removed from the dataset as they do not play tackle rugby and therefore are not influenced by the age-weight bandings. Whilst players over 13 years old are affected by age-weight bands, as they move in to a different competition governed by college, they have also been removed from the dataset. One of the limitations of the previous studies by (Campbell *et. al.*, 2018, Campbell *et. al.*, 2018) was the absence of player's home address and club details of each player which restricted the ability to combine proprietary data sources from DOT loves data ([www.dotlovesdata.com](http://www.dotlovesdata.com)) and investigate the impact of social factors on player retention. Each player address and the address of the club which they belonged to were geo-coded to calculate club proximity and assign a meshblock. A meshblock is the smallest geographic unit for which statistical data is reported by government agency, Stats NZ. In 2020, there are 53,596 meshblocks to which NZ's population of 5 million is allocated.

## DYNAMIC SOCIAL DATASETS

Geo-coded player addresses enabled merging of external data sets using time (month) and place (meshblock) as the match keys. One such data set, DOT loves data's Dynamic Deprivation Index (DDI) (Ward *et. al.*, 2019) was incorporated into the analysis. The DDI incorporates a combination of massive public and proprietary data sets shaping five dimensions of deprivation: employment, support, income, education and material deprivation. The DDI is a world-first dynamic measure of the monthly changes in socio-economic status of communities where the tool has been three years in the making, due to the scientific rigour, ethical considerations and market testing that underpins the development of this data source for small areas within New Zealand.

Geo-coded player addresses enabled the investigation of the impact of how the media portrayal of a player's community rugby impacts player retention in the sport. We looked at the sentiment of rugby in mainstream digital media in the home communities of players using a geographical focused implementation of the approach described by Bracewell, *et. al.*, 2016. This used a natural language processing tool to describe the commentary regarding rugby events in the media, reporting the sentiment and noise around rugby across the different communities of New Zealand. For the purpose of reporting on current events, DOT loves data has collected an archive of news content from publicly available sources incorporating approximately 25 million articles spanning from 2005 to 2018. Ethel is used to extract news content which is relevant to a given set of search terms, in this case "rugby". Ethel tracks the sentiment of these articles over time allowing it to be measured against other metrics such as churn. Ethel also provides details of the geographic location of the community, in which the article mentions (for example: suburb, school, club or ground), allowing the geocoded junior player addresses in the administrative dataset to be linked.

Segmemento (DOT loves data, 2019) is a data driven, New Zealand specific segmentation tool which groups small areas of New Zealand into regions with similar demographic composition via a regionalisation clustering, size-constrained algorithm, largely implemented using the *scclust* package in R (Savje, 2017). Updated annually, the September 2019 release of Segmemento had 25 segments covering all of New Zealand. Attributes such as family composition, education, urbanality and deprivation are used to shape these segments. Junior players are tagged to a segment using the meshblock associated with their addresses to investigate the junior drop out behaviour against the combination of urbanality and deprivation.

## METHODOLOGY

The status of each player at the time of registering for a new season of rugby was categorised as returned (churn = 1) or departed (churn = 0). For the purposes of this study, players who changed their club but still played within the association were considered to be "returned" players. This approach was used to provide broad understanding of what factors influence a player to continue participating in the sport of rugby. As improving player retention across all clubs was a priority for the rugby union, this approach is consistent with the intended use of this administratively collected data. Players who left because they were too old to play in the junior competition were not included in the analysis as they were ineligible to register for a new season of play because of age-restrictions of the junior competition.

Adjusted odds ratio (OR) estimates for the predictive variables were used to determine how much more or less likely a player was to remain in Auckland Rugby Union's junior competition. Typically fit using maximum-likelihood estimation, logistic regression is a form of generalized linear model with a logit link function and a binomial random component. The dependent variables can take any real value. The logit function restricts the

probability of churn,  $\pi_i$ , to between 0 and 1. However, one of the assumptions of ordinary logistic regression is independence of subject observations which is not the case here as the data tracks players across three seasons. Using the player identification numbers to place repeated observations into clusters, repeated measures framework is implemented. This allows for the repeated independent variables of each player, which we would expect to be correlated over time. We implement the Generalized Estimating Equations (GEE) approach to extend the Generalized Linear Model allowing for the repeated measurements.

The initial analysis of the weight limit effect on churn was extended for the 2017-2019 data. The same model as outlined by Campbell, Bracewell, Blackie & Patel (2018) was tested on the recent data:

$$Y_{\text{Churn}} = \alpha + \beta_{\text{Age}} + \beta_{\text{Weight}} + (\beta_{\text{Ind}} * \beta_{\text{Age}}) \quad (1)$$

In the multivariable logistic regression model 2 building process, all variables (squad size, club proximity, rugby sentiment in the media, segment and deprivation) were considered for initial inclusion. There was no theoretical-basis underpinning the choice of these additional predictive variables for initial inclusion, rather these were variables that could be derived by geo-coding the player address data supplied by Auckland Rugby Union. However, after fitting the logistic model, some variables no longer remained statistically significant in terms of predicting the probability of remaining in rugby and so were removed from the final model to improve its fit. To select the best model for predicting player churn the 'glmulti' test in R-Studio was used to decide which variables should be kept in the model. The model selection is based on the AIC value.

Of the five variables initially entered into the logistic model 2, only three were retained in the final optimised multivariable model (the model with the lowest AIC among the candidate models):

$$Y_{\text{Churn}} = \alpha + \beta_{\text{Sentiment}} + \beta_{\text{Deprivation}} + \beta_{\text{Squad-Size}} \quad (2)$$

### 3. RESULTS

Probability of churning was found to be statistically significant only for 10 year old children who have been moved above their age grade due to their weight (64% more likely to leave the game). Whilst the initial study for 2011-2016 found churn due to weight limit effect significant for ages 7 to 11 inclusively, there is evidence (outlined in the introduction) to suggest that in recent years, the sustained level of churn is driven by latent drivers other than the weight band effect.

Variable	Estimate	Standard Error	Z value	Pr(> z )	
(Intercept)	-0.4376	0.0737	35.29	<0.0001	***
Positive Rugby Sentiment in the media	-0.2557	0.1367	3.50	0.0613	*
Large Team	0.2517	0.0565	19.81	<0.0001	***
Deprivation Index 2	-0.1718	0.0096	3.23	0.0722	*
Deprivation Index 3	0.1082	0.0092	1.19	0.2754	
Deprivation Index 4	-0.0580	0.1182	0.24	0.6235	
Deprivation Index 5	-0.0681	0.1152	0.35	0.5545	
Deprivation Index 6	0.2254	0.1263	3.19	0.0743	*
Deprivation Index 7	0.2709	0.1172	5.34	0.0208	**
Deprivation Index 8	0.5014	0.1228	16.66	<0.0001	***
Deprivation Index 9	0.4437	0.1308	11.58	<0.0001	***
Deprivation Index 10	0.9669	0.1252	59.66	<0.0001	***

Table 1: Table of Coefficients for Repeated Measures Model 2 ( \*: p<0.05; \*\*: p<0.01; \*\*\*: p<0.001)

#### SQUAD SIZE

Juniors in large teams, where the team is larger than required team size by more than three, are 29% more likely to leave the game compared to a player not in a large team. That is, at U10 level, where the match is 10 aside, a squad of 14 or more is deemed to be large.

## RUGBY IN THE MEDIA

Juniors from communities where their community rugby is showcased positively in digital mainstream media are 23% less likely to leave the game compared to a player from a community where their community rugby is portrayed negatively.

## DEPRIVATION

Juniors from communities with the lowest socio-economic status (the most deprived: deprivation index = 10) are 2.6 times more likely to leave the game compared to a player from one of the most affluent communities (the least deprived: deprivation index = 1).

Variable	Odds Ratio
Positive Rugby Sentiment in the media	0.7744
Large Team	1.2862
Deprivation Index 2	0.8421
Deprivation Index 6	1.2528
Deprivation Index 7	1.3111
Deprivation Index 8	1.6510
Deprivation Index 9	1.5585
Deprivation Index 10	2.6300

Table 2: Odds ratios (OR) for the significant variables from model 2

## SEGMOMENTO SEGMENT

Whilst the geo-spatially tagged segment provided by DOT loves data Segmomento product was not included in the final model, most likely due to the strong relationship with deprivation, the Segmomento grid system (Figure 1) revealed the impact of deprivation on player retention, with additional insight into the influence of urbanality on churn. The grid system is driven by the two dimensions (urban to rural across the top, low deprivation to high deprivation down the side). Figure 1 reveals how juniors that churn are populated in the bottom left hand corner of the matrix (representing the poorer urban communities whilst the top right hand corner represents the better off communities). In 2019, 53% of players from segment '03' left the game compared to only 30% of players from segment '05' who left the game. Using data provided by the University of Otago, Sport New Zealand also found that the most deprived areas are typically urban.

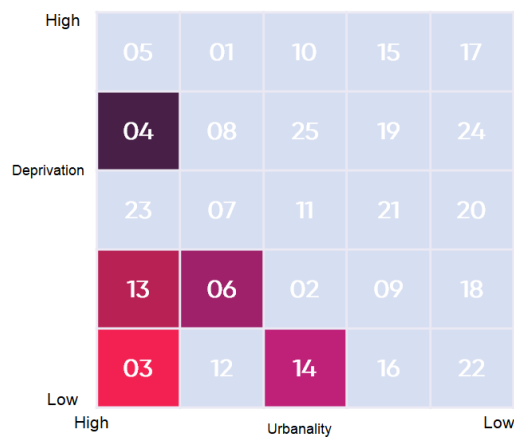


Figure 1: Segmomento Matrix for 2019 rate of players who churned in each segment

Overall, the repeated measures analysis clearly supports the hypothesis that social factors; large squad size, negative rugby sentiment in the media and high deprivation are negatively affecting player retention in Auckland junior rugby.

#### **4. DISCUSSION**

In recent years, New Zealand media have addressed how children's sport can be an "unequal playing field" where some simply miss out due to unaffordability (Keogh, 2018). This study provides evidence of its reality in the junior rugby competition in Auckland, supporting that to retain junior players from struggling communities, Rugby Union must seriously consider how to increase accessibility of New Zealand's national sport for all New Zealanders. This means putting in place financial support systems for poorer families to afford keeping their children playing in the sport. Furthermore, our study provides motivation to increase government initiatives such as the recent agreement between New Zealand's Ministry for Vulnerable Children and Sport New Zealand to help children depending on government care to engage in sport, by reducing the preventive barriers (Stuff NZ, 2019).

"A process by which people are inspired by elite sport, sports people or sports events to participate themselves" is known as the trickledown effect (Weed, 2009). Wicker & Sotiriadou (2013) describes this concept as a "result of athlete performances, sports stars as personalities, and major sporting events" and establishes the positive relationship between hosting major sports events on sport participation using the 2006 Melbourne Commonwealth Games. Regression results showed that younger people, less educated people, and Indigenous people are more likely to spend more time participating in sport as a result. Our study has established how sensitive the trickledown effect is, even at a community level; the impact of local rugby media on junior rugby engagement. England's Football Association encourages that in the social media context, clubs could provide hash tags for players, to widely promote and positively engage clubs with the wider community actively and in real time (Kirkham, 2015). With greater club rugby experiences voiced online this would drive greater positive news articles around the sport, motivating players to stay in the game.

Our study had some limitations which in turn represent directions for future research. Trickledown effect is analysed only through the consumption of local rugby news media whilst further insight can be gained by incorporating the magnitude of how young kiwis are inspired by elite sports and elite sports figures. We anticipate this can be quantified by investigating viewing patterns with subscriptions to paid television that gives access to international-level sports matches. An interesting experience Marlborough Cricket Association general manager Ed Gilhooly shared was whilst many were articulating the declining of children playing team sports in New Zealand, signups for Cricket in the region increased which he attributed to the 'Cricket World Cup effect' and the continued high performance of the Black Caps (Lewis, 2016). Establishing the magnitude of the influence of elite sports, such as world-class rugby, on junior players would encourage coaches and clubs to showcase national level sport to young New Zealanders. As an example, could sponsor live international matches in public spaces, encouraged by the evidence found here of its ability to attract and retain more young kiwis in the sport.

Sports for children must ensure that children registered to play be given an opportunity to in fact participate. Kirkham, Participation Manager of the West Australian Football Commission (2015) describes how children "want equality in a sports experience, they do not want to be the player that sits on the bench with little game time and do not want their friends sitting on the bench either." Cases such as these increase when large teams are formed than the standard squad size. Kirkham argues youth sports is no place for coaches that fail to provide a fair and equitable environment for all juniors and recommends an equal player rotation policy for all coaches. Coaches, often with a 'win at all costs mentality', that fail to include certain participants must be held to account by parents and the club. Visek's (2016) work on the 'Not Fun Maps' reveal that "coaches that favour some kids" discourages children from sport and often leads to youth sport participants dropping out of sport. Junior rugby coaches must actively ensure that team sizes are appropriate otherwise ensure regardless of talent level, that players get equal time on field as our study has found significant evidence that juniors in large teams are more likely to churn.

#### **5. CONCLUSIONS**

The negative impact of high deprivation, large team size and negative rugby media on junior rugby retention is a new knowledge area, where once existed a gap in the literature. The contribution from this study for this area positions junior rugby coaches and clubs to more informed evidence-based decisions in improving player retention alongside age-weight grading systems. Rugby Union must seriously consider how to increase accessibility of New Zealand's national sport for all New Zealanders.

## Acknowledgements

This research is part of a larger body of work, “Data informed decision bias”, a PhD Thesis supported by Callaghan Innovation and DOT loves data, supervised by Associate Professor Ivy Liu of Victoria University of Wellington and Dr Paul Bracewell.

## References

- Bracewell, P.J., McNamara, T.S., & Moore, W.E. (2016). How Rugby Moved the Mood of New Zealand. *Journal of Sport and Human Performance*. 4(4). pp. 1-9.
- Campbell, E. C., Bracewell, P. J., Blackie, E. & Patel, A. K. (2018). The Impact of Auckland Junior Rugby Weight Limits on Player Retention. *Journal of Sport and Health Research*. In press (accepted January 3, 2018).
- Campbell, E. C., Patel, A. K. & Bracewell, P. J. (2018). Optimising Junior Rugby Weight Limits in New Zealand. Retrieved from: <http://www.mathsportinternational.com/anziam/Mathsport%2014.pdf>
- Hart, M. & Macandrew, R.(2019). National sporting body warns of drop in pupils playing sport. Retrieved from:<https://www.stuff.co.nz/sport/113355381/national-sporting-body-warns-of-drop-in-pupils-playing-sport>
- Keogh, B. (2018). Kids missing out on playing sport due to cost, families re-mortgaging homes to pay fees. Retrieved from: <https://www.stuff.co.nz/sport/108818305/kids-missing-out-on-playing-sport-due-to-cost-families-re-mortgaging-homes-to-pay-fees>
- Kirkham, T. (2015). To investigate innovative and inclusive retention strategies for youth participants in community sport. The Winston Churchill Memorial Trust of Australia. Retrieved from: [https://www.churchilltrust.com.au/media/fellows/Kirkham\\_T\\_2015\\_Retention\\_strategies\\_for\\_youth\\_participants\\_in\\_community\\_sport.pdf](https://www.churchilltrust.com.au/media/fellows/Kirkham_T_2015_Retention_strategies_for_youth_participants_in_community_sport.pdf)
- Lewis, O. (2016). Participation rates down as people move away from team sports. Retrieved from: <https://www.stuff.co.nz/marlborough-express/sport/78031272/participation-rates-down-as-people-move-away-from-team-sports>
- Savje, F., (2017). sccust. RDocumentation. <https://www.rdocumentation.org/packages/sccust/versions/0.2.2> Accessed: 7 November 2019.
- Sport New Zealand. (2019). Active NZ Spotlight on the impact of deprivation on participation. Retrieved from: <https://sportnz.org.nz/assets/Uploads/Active-NZ-Spotlight-on-Deprivation-August-2019.pdf>
- Sport New Zealand. (2011). 2011 Young People’s Survey Series. Retrieved from: <https://www.srknowledge.org.nz/researchseries/sport-new-zealands-young-peoples-survey-series/>
- Stuff NZ. (2019). Oranga Tamariki and Sport New Zealand to increase sport participation for kids. Retrieved from: <https://www.stuff.co.nz/sport/112866621/oranga-tamariki-and-sport-new-zealand-to-increase-sport-participation-for-kids>
- Thomas, J. (2019). Sir Graham Henry's silver bullet to fix New Zealand Rugby's participation issues. Retrieved from: <https://www.stuff.co.nz/sport/rugby/114329935/sir-graham-henrys-silver-bullet-to-fix-new-zealand-rugbys-participation-issues>
- Visek, A. J., Achrafi, S. M., Manning, H., McDonnell, K., Harris, B. S., & DiPietro, L. (2014). The Fun Integration Theory: Towards sustaining children and adolescents sport participation. *Journal of Physical Activity & Health*. 12, 424.
- Visek, et al. (2016). Not Fun Maps. Preliminary analyses of unpublished data. George Washington University.
- Ward, A., Trowland, H., & Bracewell. (2019). The Dynamic Deprivation Index: measuring relative socio-economic deprivation in NZ on a monthly basis. *Kōtuitui: New Zealand Journal of Social Sciences Online*. <https://doi.org/10.1080/1177083X.2019.1578807>
- Weed, M. (2009). The potential of the demonstration effect to grow and sustain participation in sport. Review Paper for Sport England. Canterbury Christ Church University: Centre for Sport, Physical Education & Activity Research (SPEAR).
- Wicker, P. and Sotiriadou, P. (2013), "The trickle-down effect: What population groups benefit from hosting major sport events?", *International Journal of Event Management Research*, 8, 2, 25-41.
- Widmeyer, W. Neil; Brawley, Lawrence R.; Carron, Albert V. (1990). The Effects of Group Size in Sport. *Journal of Sport & Exercise Psychology*. Jun1990, Vol. 12 Issue 2, p177-190. 14p.
- World Rugby. (2013). Relative Age Effect. Retrieved from World Rugby: <https://playerwelfare.worldrugby.org/?documentid=110>. Accessed: 7 November 2019.