



**HAL**  
open science

# Étude des facteurs de risque de blessures dans le Rugby Élite seniors

Jérémy Chéradame

► **To cite this version:**

Jérémy Chéradame. Étude des facteurs de risque de blessures dans le Rugby Élite seniors. Médecine humaine et pathologie. Université de Bordeaux, 2022. Français. NNT : 2022BORD0358 . tel-04064954

**HAL Id: tel-04064954**

**<https://theses.hal.science/tel-04064954v1>**

Submitted on 11 Apr 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE PRÉSENTÉE  
POUR OBTENIR LE GRADE DE  
**DOCTEUR DE**  
**L'UNIVERSITÉ DE BORDEAUX**

ÉCOLE DOCTORALE : Sociétés, Politique et Santé Publique

SPÉCIALITÉ : Santé publique Option Biostatistiques

Par **Jérémy CHÉRADAME**

**Étude des facteurs de risque de blessures dans le Rugby Élite seniors**

Sous la direction de : **Hélène JACQMIN-GADDA**  
co-directeur : **Philippe DECQ**

Soutenue le 6/12/2022 à Marcoussis

Membres du jury :

M. EDOUARD, Pascal – Université de Saint Etienne - Professeur, praticien hospitalier	Rapporteur
M. LEY, Christophe – Université du Luxembourg – Professeur	Rapporteur
M. SALAMON, Roger – Université de Bordeaux - Professeur émérite	Examineur
Mme. GELEIN, Brigitte – IRMAR – Maitresse de conférences	Examinatrice
Mme. JACQMIN-GADDA, Hélène – Université de Bordeaux - Directrice de recherche	Directrice de thèse
M. DECQ, Philippe – ENSAM – Professeur, praticien hospitalier	Co-directeur de thèse
M. HAGER, Jean Philippe – Centre orthopédique Santy - Médecin	Invité
M. PISCIONE, Julien – Fédération Française de Rugby - Docteur	Invité



# **Titre : Étude des facteurs de risque de blessures dans le Rugby Élite seniors**

---

## **Résumé :**

L'étude du risque de blessures est un enjeu important dans le sport, d'autant plus dans un sport comme le Rugby avec de fortes sollicitations de déplacement et de combat. Ces dernières années, la dangerosité du Rugby a été mise en avant et particulièrement vis-à-vis des commotions qui seraient en forte augmentation. Pour tenter de réduire ce risque de blessures il est important de bien gérer la charge de travail imposée aux joueurs. Cependant bien que ce soit un sujet fortement présent dans la littérature des sciences du sport, il n'existe à ce jour aucune mesure ne faisant consensus pour suivre le risque de blessure. Un Observatoire Médical du Rugby mené par la Ligue Nationale de Rugby et la Fédération Française de Rugby a été relancé en 2018 dans le but d'observer, de prévenir et surtout d'adapter le Rugby pour pouvoir limiter les accidents. C'est dans ce cadre que s'inscrit cette thèse.

Après un point sur la littérature existante et les limites de celles-ci, une première étude sera présentée sur le cas particuliers des commotions. Cette étude permet d'identifier des facteurs de risque de commotions en match dans le championnat français de Rugby (Top 14) durant 5 saisons. Pour cela, nous avons utilisé un modèle log-linéaire de Poisson à effets mixtes incluant le joueur en effet aléatoire et le logarithme du temps d'exposition en offset. Ainsi, à poste comparable, les plus légers sont plus à risque d'être commotionnés en plaquant et les plus grands de subir une commotion dans d'autres circonstances. De plus, les joueurs ayant accumulés un temps de jeu plus important depuis le début de la saison seraient moins à risque d'être commotionnés.

Dans un second temps, cette fois à l'échelle d'un club, nous avons analysé en quoi la planification hebdomadaire de la charge de travail influence le risque de blessure le jour de match. Les données manquantes dans la charge de travail hebdomadaire ont été traitées par imputation multiple par équations chaînées. L'impact de la charge de travail sur le risque en match a été étudié par un modèle logistique mixte pour chaque journée d'entraînement précédant le match. Nous avons donc pu constater que le volume, l'intensité et la programmation des séances d'entraînement sur le terrain et en musculation ont des effets distincts selon le moment où elles sont pratiquées. Il existe, par exemple, un effet protecteur de la haute intensité deux jours avant le match.

Enfin la dernière partie de cette thèse est consacrée aux opportunités que pourront offrir les données que devrait obtenir l'Observatoire Médicale du Rugby et aux méthodologies d'analyses qui pourront être employées. Pour sélectionner les mesures les plus en lien avec le risque de blessures, des méthodes comme la Sparse Group Lasso pourront par exemple être utilisées. De plus, pour identifier les périodes où la charge de travail a le plus d'influence sur le risque de blessure sans hypothèse a priori, la méthode Weighted Cumulative Index Exposure pourrait être utilisée.

Cette thèse a ainsi permis d'identifier certains facteurs de risque de blessure et notamment de commotion. Elle a également pu mettre en évidence l'influence de la charge de travail à l'entraînement sur le risque en match. Enfin les méthodes envisagées pour les données que pourrait collecter l'Observatoire Médical peuvent être de formidables opportunités pour tirer des conclusions robustes sur l'impact de la charge de travail et, de ce fait, mieux suivre le risque de blessures des joueurs de Rugby.

**Mots clés :** Biostatistiques, Blessures, Rugby, Facteurs de risque, Modèles mixtes

---

## **Title : Study of injury risk factors in senior elite rugby**

### **Abstract :**

The study of the risk of injury is an important consideration in sport, especially in a sport like Rugby with high demands on movement and combat. In recent years, the danger of rugby has been highlighted, particularly with regard to concussions, which are reported to be on the increase. To try and reduce the risk of injury it is important to manage the workload imposed on players. However, although this is a topic that is strongly present in the sports science literature, there is currently no agreed measure to monitor the risk of injury. A Rugby Medical Observatory led by the National Rugby League and the French Rugby Federation was relaunched in 2018 with the aim of observing, preventing and above all adapting Rugby to be able to limit accidents. It is within this framework that this thesis falls.

After a review of the existing literature and its limitations, a first study will be presented on the particular case of concussions. This study identifies risk factors for concussions during matches in the French rugby championship (Top 14) over 5 seasons. To do this, we used a log-linear Poisson mixed-effects model including the player as a random effect and the logarithm of the exposure time as an offset. Thus, at comparable positions, lighter players are more likely to be concussed when tackling and taller players to be concussed in other circumstances. In addition, players who have accumulated more playing time since the start of the season would be less likely to be concussed.

In a second step, this time at the club level, we analysed how the weekly workload planning influences the risk of injury on match day. Missing data in the weekly workload were treated by multiple imputation by chained equations. The impact of workload on match risk was investigated by a mixed logistic model for each training day prior to the match. We found that the volume, intensity and timing of field and strength training sessions had distinct effects depending on the time of day. For example, there is a protective effect of high intensity two days before the match.

Finally, the last part of this thesis concerns the opportunities that the data to be obtained from the Rugby Medical Observatory could offer and the analysis methodologies that could be used. To select the measures most related to injury risk, methods such as the Sparse Group

Lasso could be used. In addition, to identify the periods where workload has the greatest influence on injury risk without any a priori hypothesis, the Weighted Cumulative Index Exposure method could be used.

This thesis identified some risk factors for injury and concussion in particular. It also highlighted the influence of training workload on the risk in matches. Finally, the methods envisaged for the data that could be collected by the Medical Observatory could be a great opportunity to draw robust conclusions on the impact of workload and, therefore, to better monitor the risk of injury to rugby players.

**Keywords :** Biostatistics, Injuries, Rugby, Risk factors, Mixed models

# Remerciements

---

## **Mes sincères remerciements vont :**

A la directrice de cette thèse, Hélène Jacqmin-Gadda, pour son accompagnement tout au long de celle-ci. Toute son implication et ses recommandations ont été précieux pour la bonne réalisation de ces travaux.

Au co-directeur, le Professeur Philippe Decq, pour ses conseils aux différentes étapes de ce projet. Sa rigueur et ses connaissances ont été de véritables atouts.

Au responsable du département d'accompagnement de la Performance de la Fédération Française de Rugby, Julien Piscione qui a initié cette thèse. Son avis d'expert a été très utile pour cette thèse et les autres projets du département. J'ai beaucoup appris à ses côtés durant ces 4 années et je lui en suis grandement reconnaissant.

Aux membres du jury, d'avoir accepté ce rôle et permettre d'enrichir, par leurs expériences et leur regard critique, mes réflexions sur le sujet. En particulier à Roger Salamon d'avoir accepté de présider ce jury et à Christophe Ley ainsi que Pascal Edouard de prendre part à ce jury en qualité de rapporteur.

Aux personnes qui m'ont accompagné dans la réalisation de ce projet. Leur aide de par leur expertise s'est avérée importante pour conduire à l'aboutissement de cette thèse.

Aux doctorants de la Fédération Française de Rugby, Alexis Peeters, Bertrand Mathieu et Maxime Brick pour leur solidarité dans nos activités respectives. Plus généralement je tiens à remercier toutes les personnes de la direction sportive et les staffs des équipes de France pour la richesse de tous ces moments partagés depuis mon arrivée à la Fédération.

A ma famille, pour leur soutien depuis toutes ces années et sans qui rien n'aurait été possible.



# Valorisation Scientifique

---

## Article publié:

Chéradame, Jérémy, Julien Piscione, Christopher Carling, Jean-Pierre Guinoiseau, Bernard Dufour, Hélène Jacqmin-Gadda, et Philippe Decq. « Incidence and Risk Factors in Concussion Events: A 5-Season Study in the French Top 14 Rugby Union Championship ». *The American Journal of Sports Medicine* 49, no 7 (juin 2021): 1921-28.  
<https://doi.org/10.1177/03635465211006480>.

## Article soumis:

Chéradame Jérémy, Romain Loursac, Julien Piscione, Christopher Carling, Philippe Decq, Hélène Jacqmin-Gadda. «Impact of weekly training workload structure and content on the risk of injury in professional Rugby Union match-play »

## Communications orales:

- Facteurs de risque en Top 14. *FFR : Science et Rugby*, Juin 2018, Marcoussis, France
- Utilisation et exploitation des données dans une structure sportive. *Société Française de Statistiques : Statistiques et Sport*, Mai 2019, Paris, France
- Commotions et facteurs de risque en Top 14 : Etude prospective sur 5 saisons. *FFR : Science et Rugby*, Juin 2019, Marcoussis, France
- Ma thèse en 7 minutes : relation entre charge de travail et risque de blessure. *FFR : Science et Rugby*, Mai 2022, Marcoussis, France

## Formations dispensées:

- Epidémiologie des blessures en Rugby. *Diplôme Inter-Universitaire Pathologie du Rugby*, 2018 et 2022, FFR, Université de Bordeaux, Université de Versailles
- Avantages et limites des ratios aigue : chronique. *Certificat de Capacité de Préparateur Physique en Rugby*, 2021, Institut National Emploi et Formation



# Table des matières

---

Chapitre 1 : Introduction .....	17
Chapitre 2 : Etat de l'art .....	23
1) Revue de littérature des blessures dans le Rugby .....	24
1.1 Epidémiologie des blessures .....	24
1.1.1 Incidence des blessures .....	24
1.1.2 Sévérité des blessures .....	27
1.1.3 Types de blessures.....	28
1.1.4 Les commotions .....	29
1.2 Facteurs de risque de blessures .....	30
1.2.1 Circonstance de survenue.....	30
1.2.2 Les antécédents de blessures .....	32
1.2.3 Le profil des joueurs.....	33
1.2.4 Autres éléments contextuels.....	34
1.3 Impact de la charge de travail sur le risque.....	35
2) Méthodologie statistique .....	40
2.1 Traitement des données manquantes.....	40
2.1.1 Types de données manquantes .....	40
2.1.2 Technique de traitement de données manquantes .....	41
2.1.2.1 <i>Analyse des cas complets et des cas disponibles</i> .....	41
2.2 Modèles d'analyse de données corrélées .....	45
2.2.1 Approche conditionnelle .....	46
2.2.2 Approche marginale .....	48
2.3 Modèles de durée .....	49
2.3.1 Choix du temps de base.....	50
2.3.2 Modèle de Cox .....	51
2.3.3 Modèle de durée pour événements récurrents.....	52
2.4 Modèles de prédiction.....	55
Chapitre 3 : Etude des facteurs de risque de commotions.....	63
1) Résumé.....	64
2) Article.....	66

Chapitre 4 : Etude de l'impact de la charge de travail sur le risque de blessure .....	87
1) Résumé .....	88
2) Article soumis .....	90
Chapitre 5 : Réflexions sur l'analyse statistique des données de l'Observatoire Médical du Rugby .....	109
1) Données attendues pour l'Observatoire Médical du Rugby.....	110
2) Objectifs des analyses .....	112
3) Identifier les périodes d'exposition critiques .....	113
4) Sélection de variables.....	116
4.1 Méthode de régression pénalisé .....	116
4.2 Méthode de réduction de dimension .....	119
4.2.1 Régression par composantes principales.....	120
4.2.2 Partial Least Square.....	120
4.2.3 sPLS et gPLS.....	121
Chapitre 6 : Discussion .....	125
1) Les facteurs de risque en Rugby .....	126
2) Prévenir le risque de blessures .....	128
3) Perspectives .....	129
Chapitre 7 : Conclusion.....	131
Bibliographie.....	133

# Liste des figures

---

<i>Figure 1 : Taux d'incidence des blessures en match en Premiership avec moyenne <math>\pm 2</math> écart-type. Professional Rugby Injury Surveillance Project – RFU 2019/2020<sup>4</sup></i>	25
<i>Figure 2 : Indisponibilité en fonction de la localisation. Les lignes verte, orange et rouge représentent respectivement le 25<sup>ème</sup>, 50<sup>ème</sup> et 75<sup>ème</sup> percentiles.<sup>5</sup></i>	28
<i>Figure 3 : Taux d'incidence des commotions en match Premiership sur la période d'étude avec moyenne <math>\pm 2</math> écart-type. Professional Rugby Injury Surveillance Project – RFU 2019/2020<sup>4</sup></i>	29
<i>Figure 4 : Evolution des incidences de commotion dans les compétitions Elites Nationales en Rugby. Communication Symposium World Rugby 2019, Marcoussis</i>	30
<i>Figure 5 : Taux d'incidence du porteur de balle ou du plaqueur en Premiership sur la période d'étude avec moyenne. Professional Rugby Injury Surveillance Project – RFU 2019/2020<sup>4</sup></i>	31
<i>Figure 6 : Guide pour appliquer et interpréter les ratios aigue/chronique. La zone verte représente la zone où le risque de blessure est faible et la rouge celle où le risque est élevé. Gabbett 2015, redessiné à partir de Blanch et Gabbett 2015<sup>49</sup>.</i>	36
<i>Figure 7: Diagramme des variables testées dans la publication de West et al., 2020<sup>8</sup></i>	38
<i>Figure 8 : Schéma des étapes de l'imputation<sup>77</sup></i>	43
<i>Figure 9: Incidence of injury according to age, player position, number of weeks since last game and return from injury which led to an absence of more than 7 days. Values are presented as incidence (95% CI).</i>	99

# Liste des tables

---

*Table 1 : Risque de blessure dans rugby amateur masculin et féminin. Communication personnel Rémi Kaboré (ISPED)*\_\_\_\_\_29

*Table 2: Données anthropométrique des joueurs de la coupe du monde 2019. Fuller et al.*\_36

*Table 3: Machine learning methods in sport injury prediction and prevention: a systematic review Hans Van Eetvelde, Luciana D. Mendonça, Christophe Ley, Romain Seil and Thomas Tischer*\_\_\_\_\_63

*Table 4: Taille d'échantillon espérée sur une saison. Le nombre de joueurs est déterminé par rapport au nombre de joueurs qui ont participé à chacune des compétitions sur la saison 2021/2022.*\_\_\_\_\_114

# Liste des abbréviations

---

ACWR : Acute Chronic Workload Ratio  
ADASYN : Adaptive synthetic  
AG : Andersen et Gill  
AIC : Akaike Information Criterion  
AUC : Area under the ROC Curve  
BIC : Bayesian Information Criterion  
DMI : Dossier Médical Informatisé  
EWMA : Exponentially Weighted Moving Average  
FFR : Fédération Française de Rugby  
GEE : Generalized Estimating Equation  
GLMM : Generalized Linear Mixed Model  
gPLS : Group Partial Least Squares  
GPS : Global Positioning System  
HIA : Head Injury Assessment  
HSD : High Speed Distance  
IRR : Incidence Rate Ratio;  
LNR : Ligue Nationale de Rugby  
MAR : Missing At Random  
MCAR : Missing Completely At Random  
Mice : Multivariate Imputation by Chained Equations  
MNAR : Missing Not At Random  
OR : Odds Ratio  
PCR : Principal Component Regression  
PLS : Partial Least Squares  
PMM : Predictive Mean Matching  
PRISP : England Professional Rugby Injury Surveillance Project  
PWP : Prentice-Williams-Peterson

RFU : Rugby Football Union

ROC : Receiver Operating Characteristic

RPE : Rating Perceived Effort

SMOTE : Synthetic Minority Over-sampling Technique

sPLS : Sparse Partial Least Squares

sRPE : Session Rating Perceived Effort

SVM : Support Vector Machine

TD : Total distance

VIF : Variance Inflation Factor

VMA : Vitesse Maximale Aérobie

WCIE : Weighted Cumulative Index Exposure

WLW : Wei-Lin-Weissfeld

# **Chapitre 1 : Introduction**

---

L'étude du risque de blessures est un enjeu majeur pour toute pratique et notamment sportive. Quel que soit le niveau, il est important de pouvoir appréhender ce risque inhérent à n'importe quelle activité pour pouvoir juger si celui-ci est acceptable. Dans un sport complet comme le Rugby, que ce soit en termes de déplacement mais également de contact, il est d'autant plus important d'avoir des informations précises et pratiques à ce sujet. Ces dernières années, à la suite de plusieurs accidents graves et d'une médiatisation plus importante de certaines blessures comme les commotions, la défiance vis-à-vis de la dangerosité du Rugby s'est accrue. Dans ce sens, une enquête menée en 2019 par une société de marketing a mis en évidence que la France est un des seuls pays qui perçoit le Rugby comme plus dangereux qu'il y a 5 ans<sup>1</sup>. Les commotions peuvent en effet avoir des conséquences inquiétantes à court et long terme<sup>2,3</sup>. D'autant plus que leur nombre serait en forte augmentation depuis quelques années. Dans le championnat professionnel anglais il a quadruplé entre les saisons 2002-2012 et les saisons 2016-2020<sup>4</sup>. Par ailleurs, le nombre de blessures graves de manière générale (>84 jours d'indisponibilité) est également à la hausse<sup>5</sup>. En France, nous ne savons pas si ce constat est identique puisqu'aucune publication ne permet de savoir ce qu'il en est.

Aujourd'hui dans le Rugby de haut niveau français, les clubs ont un ou des médecins permettant de recueillir et suivre les blessures des joueurs à l'échelle du club. A un niveau plus large, deux systèmes de recueil des blessures existent mais ils ne sont que peu exploités et n'ont fait l'objet d'aucune valorisation scientifique. Le premier est le Dossier Médical Informatisé (DMI) développé par la Ligue Nationale de Rugby. Il est complété par les médecins des clubs pour chaque blessure entraînant une indisponibilité du joueur et contient de nombreuses informations sur les blessures. Cependant, celui-ci, comme son nom l'indique a pour but de contenir les informations médicales relatives aux blessures des joueurs mais n'a pas vocation à être un programme épidémiologique. Certaines variables qu'on peut considérer importantes pour une analyse des facteurs de risque, telles que le poste, ne sont que partiellement remplies (68%, 56%, 62%, 54% de données manquantes respectivement pour les saisons 2016 à 2019). La seconde source, plus récente puisque mise en place en 2012-2013 est un programme de surveillance des blessures en match. L'approche de ce programme est originale puisque ce sont des observateurs indépendants qui sont rémunérés pour identifier les événements des matchs étant potentiellement des blessures. Ceux-ci sont ensuite envoyés à une technicienne d'étude clinique qui récupère également le clip vidéo de l'évènement et les envoie au médecin du club du joueur concerné pour confirmation et précision. Le médecin du club peut également indiquer des blessures que les observateurs auraient pu oublier. Intéressant initialement toutes les

blessures, ce programme s'est limité au bout de 4 ans au périmètre des commotions. Certains pays comme l'Angleterre, via la fédération de rugby anglaise (Rugby Football Union, RFU) ont su se structurer depuis des années pour avoir des programmes de surveillance adéquats. Ainsi, depuis près de 20 ans (saison 2002-2003), ils ont mis en place le programme « England Professional Rugby Injury Surveillance Project » (PRISP) qui leur permet de suivre l'évolution des blessures sur leur championnat élite anglais et leur sélection nationale. Depuis maintenant quelques années, ils ont également inclus d'autres populations dans ce projet avec notamment le suivi des populations féminines (depuis 2017), des jeunes (depuis 2017) et des amateurs (depuis 2015). Outre le suivi épidémiologique<sup>4</sup>, le laboratoire de Bath exploite ces données pour en tirer une richesse supplémentaire via des analyses plus exploratoires comme l'étude de l'impact du nombre de matchs sur le risque de blessure<sup>6</sup>, l'influence des commotions sur les blessures suivantes<sup>7</sup> ou encore tester la validité de certains indicateurs du risque de blessures<sup>8</sup> et ceux sur un volume important de données. Autant de conclusions qui peuvent être importantes à considérer par les clubs puisqu'au-delà de cet aspect sécuritaire forcément essentiel, les absences de joueurs dues aux blessures ont généralement un impact en termes de performance collective<sup>9</sup>. Il est donc de l'intérêt de chaque club de tenter de limiter ces blessures.

Le Rugby est un sport relativement jeune puisqu'il est devenu professionnel seulement depuis 1995. Néanmoins, très rapidement après cette professionnalisation, les staffs se sont étoffés avec l'arrivée de kinésithérapeutes qui jouent un rôle primordial dans la prévention des blessures mais également dans la réathlétisation après une blessure. Par la suite, certains clubs ont commencé à recruter des personnes spécialistes de la préparation physique, domaine pourtant encore très peu répandu dans le sport français. L'objectif étant donc d'optimiser les qualités physiques des joueurs pour qu'ils soient le plus performant au moment souhaité et de limiter leur risque de blessures. Pour arriver à ce but, les préparateurs physiques doivent ainsi gérer la charge de travail de leurs athlètes. A ce jour, pour suivre le travail quotidien réalisé par les joueurs, de nombreux outils sont utilisés par les clubs. Un des principaux est le GPS (Global Positioning System). Placé entre les deux scapulas du joueur dans une brassière ou directement dans le maillot, et suivant des normes sécuritaires, ces GPS sont portés à chaque match et entraînement dans le but de quantifier l'activité de déplacement. Certains capteurs permettent également de collecter des données de cardiofréquence et offre même la possibilité d'avoir une quantification de plus en plus précise du nombre de contacts. Au total, pour certaines technologies, plus d'un millier de paramètres sont collectés pour chaque joueur à chaque séance. Au-delà de ces données de déplacement ou même plus généralement liée à

l'activité Rugby du joueur, la musculation prend une part essentielle dans la charge d'entraînement imposée au joueur. Il est donc également important de la quantifier. Pour cela des appareils peuvent être utilisés pour suivre les performances des joueurs sur de multiples exercices de leur séance. La perception de l'effort par l'athlète au travers de l'échelle RPE (Rating Perceived Effort) est fréquemment utilisée. Mise en place par Borg en 1982<sup>10</sup>, cette échelle permet via un questionnaire subjectif rempli par le joueur de noter sa perception de la difficulté de la séance allant de 0 pour une séance négligeable à 10 pour une séance de difficulté extrême. Celle-ci permet ainsi de quantifier sur une même échelle une séance de musculation et une séance de terrain. De plus, c'est une solution peu coûteuse qui peut donc être utilisée à n'importe quel niveau de pratique. D'autres questionnaires subjectifs comme le suivi d'état de forme (Wellbeing), basés sur divers critères comme le sommeil, les douleurs ou encore l'état émotionnel sont régulièrement recueillis pour évaluer le niveau de fatigue d'un joueur et ainsi pouvoir adapter la charge de travail qui lui sera proposée. En complément, des marqueurs biologiques comme les CPK (créatine phosphokinase) sont parfois mesurés pour évaluer le niveau de dommage musculaire ou encore des tests physiques permettant d'avoir une indication sur la forme d'un joueur à un instant t. Avec cette accumulation de données, la grande majorité des clubs professionnels ont intégrés dans leur effectif des « sport scientists » pour les analyser. Ces profils sont pour la quasi-totalité diplômés des sciences du sport avec donc de solides compétences scientifiques pour interpréter et exploiter les données qu'ils collectent. A ce jour, il n'existe cependant dans les clubs français que très peu d'expert des sciences des données ou des statistiques pour permettre d'exploiter des volumes importants de données et de construire des modèles statistiques adéquats. D'autant plus que pour le cas de la prévention des blessures la tâche est complexe et plusieurs limites méthodologiques sont importantes à considérer.

Le risque de blessures est multifactoriel<sup>11</sup>, il est donc nécessaire de prendre en compte dans les modèles les différents paramètres qui peuvent influencer sur ce risque. Néanmoins, de nombreuses variables peuvent être disponibles dans ce but, avec notamment des groupements de variables corrélées, des méthodes adaptées doivent donc être utilisées pour sélectionner celles d'intérêt. Les données étant bien souvent des mesures répétées par joueur ou par équipe, la structure des données est également importante à considérer. A ce jour, l'intervalle de temps impactant le risque de blessure, et donc sur laquelle calculer les différents indicateurs de suivi reste inconnu. Il est d'autant plus difficile à estimer que le risque à l'entraînement et en match est différent il peut donc subsister des interrogations sur le choix du temps de base d'analyse avec des temps d'expositions et d'évènements qui peuvent ne pas être identiques. Bien qu'il

existe de nombreuses publications sur l'éventuel lien entre la charge de travail et le risque de blessures, il n'y a aucune mesure qui fasse consensus pour pouvoir suivre ce risque au quotidien. De plus, en pratique, il existe bien souvent des données manquantes dans le suivi des joueurs ce qui peut représenter une limite supplémentaire. Au-delà de ce qu'il est réalisable de faire à l'échelle d'un club, avoir les données de plusieurs clubs sur un suivi régulier pourrait permettre d'accroître la connaissance sur la survenue de blessures. Cependant, les données peuvent faire transparaître des stratégies d'entraînement ou encore faire émerger les qualités physiques de certains joueurs, l'aspect concurrentiel inter-club oblige donc à passer par une structure indépendante pour réaliser ce type d'étude.

La Ligue Nationale de Rugby (LNR) et la Fédération Française de Rugby (FFR) ont conjointement décidé de relancer en 2018 un Observatoire Médical du Rugby dans le but d'observer, de prévenir et surtout d'adapter le Rugby pour pouvoir limiter les accidents. Ainsi à la suite de différents travaux réalisés par des médecins, techniciens, experts, arbitres, chercheurs, dont nous avons fait partie et sous l'égide de cet Observatoire, 45 mesures ont été prises. Celles-ci ont concerné différents thèmes que sont les écoles de rugby, le rugby amateur, la filière de formation, le rugby professionnel ainsi qu'un thème technique et transversal. Cet Observatoire a ainsi pour rôle de mettre en place des études scientifiques pour tenter de rendre le Rugby plus sûr.

C'est dans ce contexte que s'inscrit cette thèse. Les objectifs de ce travail sont d'identifier les facteurs de risque de blessures dans le rugby professionnel avec un intérêt particulier pour les facteurs liés à la charge de travail et d'autre part, de proposer une méthodologie pour analyser au mieux la survenue de blessure lorsque davantage de données seront disponibles dans le cadre de l'Observatoire.

Pour répondre à cela, une première étude a été réalisée sur les facteurs de risque de commotions dans l'ensemble du championnat professionnel français de Rugby. Dans un second temps, avec cette fois une approche plus micro à l'échelle d'un club, nous avons analysé la relation entre la planification hebdomadaire de la charge de travail et le risque de blessure le jour de match. Enfin, la dernière partie de cette thèse est consacrée aux opportunités que pourront offrir les données que pourrait recueillir l'Observatoire Médical du Rugby et aux méthodologies d'analyses qui pourront être employées dans ce cadre.



## **Chapitre 2 : Etat de l'art**

---

# 1) Revue de littérature des blessures dans le Rugby

## 1.1 Epidémiologie des blessures

### 1.1.1 Incidence des blessures

Dans la majorité des programmes épidémiologiques, la collecte et la définition des blessures est celle définie par le consensus World Rugby de 2007<sup>12</sup>. Ainsi, est défini comme blessure : « *toute atteinte empêchant le joueur de prendre part à toute activité du match ou de l'entraînement pendant plus d'une journée après la potentielle blessure* ». La fréquence est exprimée en termes d'incidence, c'est-à-dire le nombre de blessures pour 1000 heures-joueur. Cela permet de prendre en compte la durée d'exposition et avoir ainsi une mesure comparable.

Dans le rugby de haut niveau l'incidence de blessure en match est de 91 blessures pour 1000 heures de jeu d'après une méta-analyse de Williams et al. (2021)<sup>13</sup>. A titre d'exemple, au plus haut niveau, la coupe du monde 2015 a connu une incidence de 90.1 blessures pour 1000 heures-joueur<sup>14</sup> et celle de 2019 une incidence de 79.4<sup>15</sup>.

Le PRISP de la RFU, avec des années de reculs et sur toutes les équipes d'un championnat majeur, donne des conclusions plus robustes quant au réel risque en Rugby. Depuis 2002, l'incidence de blessures est du même ordre que les résultats précédents puisqu'elle est au total de 87 blessures pour 1000 heures-joueur<sup>4</sup>. Bien qu'elle oscille d'une année à l'autre, il n'y a aucune tendance à la hausse ou à la baisse. Sur la saison 2019-2020, elle est de 88 blessures pour 1000 heures-joueur, soit très proche de la moyenne globale (Figure 1).

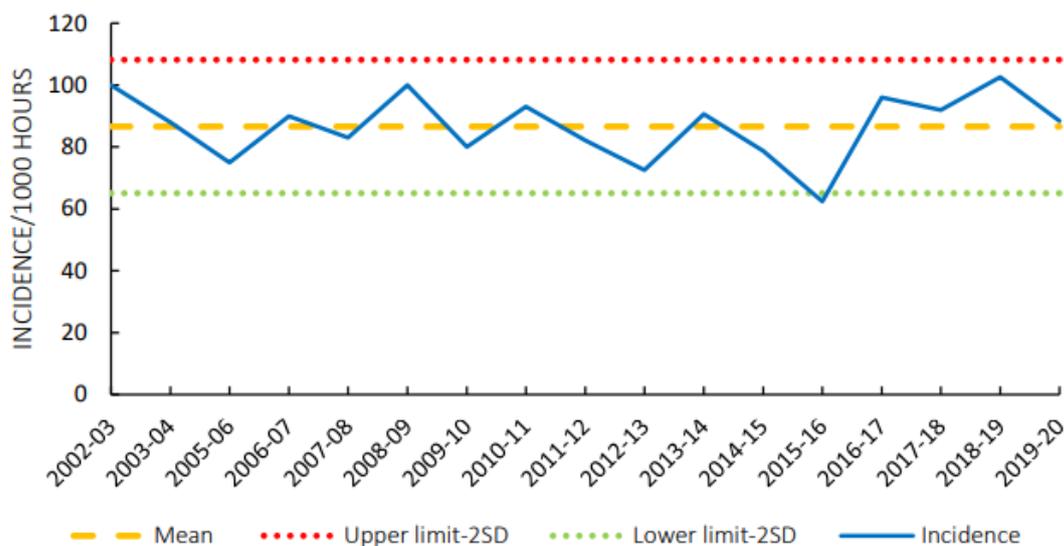


Figure 1 : Taux d'incidence des blessures en match en Premiership avec moyenne  $\pm 2$  écart-type. Professional Rugby Injury Surveillance Project – RFU 2019/2020<sup>4</sup>

A titre de comparaison dans d'autres sports comme le football par exemple, l'incidence en match a été estimée à 36.0 blessures pour 1000 heures-joueur<sup>16</sup>, tandis que le football Australien a une incidence de 41.7 blessures pour 1000 heures-joueur<sup>17</sup>.

A l'entraînement les incidences sont bien évidemment moins élevées puisqu'elles sont de l'ordre de 2.8 blessures pour 1000 heures-joueur<sup>13</sup>. Bien que cette incidence soit plus faible, le temps d'exposition étant beaucoup plus important, elle reste tout de même à considérer.

Concernant la pratique féminine, il existe moins d'études permettant de quantifier précisément le risque de blessures mais il semble moindre par rapport à leurs homologues masculins. Pour exemple, dans une étude menée sur le XV de France féminin entre 2006 et 2010<sup>18</sup> on retrouve une incidence de 41 blessures pour 1000 heures de jeu. Le programme épidémiologique de la RFU sur cette population indique des résultats comparables avec une incidence de 39 sur la saison 2019/2020<sup>4</sup>. A un niveau de compétition moindre, la fédération irlandaise a lancé depuis quelques années un programme épidémiologique (Irish Rugby Surveillance Project), sur son championnat national à différents niveaux et regroupant des équipes amateurs hommes et femmes. Une nouvelle fois l'incidence des femmes est légèrement plus basse avec 34.9 blessures pour 1000 heures-joueur contre 49.1 pour 1000 heures-joueur pour les hommes<sup>19</sup>. Une méta analyse de Yeomans et al. (2017)<sup>20</sup> sur le rugby amateur séniors suggère un risque plus faible qu'au haut niveau avec une incidence de 46.8 blessures pour 1000 heures de jeu.

Enfin, au niveau amateur, une étude nationale néozélandaise<sup>21</sup>, réalisée à partir des déclarations d'assurance de 2005 à 2017 a permis d'évaluer le risque sur une population très conséquente : 29% des joueurs ont eu à déclarer au moins une blessure lors de la saison. Encore une fois les femmes ont moins de blessures relativement à leur nombre. Cependant, ce rapport entre le risque des femmes et celui des hommes s'inverse avec l'âge. Les femmes de plus de 30 ans ont 49% de risque en plus par rapport aux hommes de la même tranche d'âge. De plus, on constate dans cette étude que, selon l'âge de pratique, le risque diffère. A l'âge adulte (>18 ans) le risque est presque 4 fois plus important que pour les plus jeunes. Chez les hommes, le risque croît jusqu'à atteindre un pic de blessure entre 20 et 25 ans suivi d'une réduction progressive. Au contraire pour les femmes le risque reste plutôt stable pour les plus âgées.

En France, la FFR et l'ISPED (Institut de Santé Publique d'Epidémiologie et de Développement) ont mené une étude observationnelle, dans le but de tester l'effet d'un programme de prévention par la suite et qui nous permet aujourd'hui d'avoir quelques chiffres sur le risque en rugby amateur (Table 1).

Catégorie	Nombre de blessures	Temps d'exposition (h-j)	Incidence pour 1000 h-j [IC95%]	OR [IC95%]	P-value
M-16	82	5325,83	15,40 [12,40 - 19,12]	0,61 [0,49 - 0,77]	<0,001
M-19	60	3540	16,95 [13,16 - 21,83]	0,68 [0,52 - 0,88]	0,0037
M+18	696	27800	25,04 [23,24 - 26,97]	1	
F-18	12	350	34,29 [19,47 - 60,37]	1,37 [0,77 - 2,42]	0,280
F+18	30	1100	27,27 [19,07 - 39,01]	1,09 [0,76 - 1,57]	0,646

Table 4 : Risque de blessure dans rugby amateur masculin et féminin. Communication personnel Rémi Kaboré (ISPED)

Sur cette première année d'étude l'incidence est de 23 blessures pour 1000 heures-joueurs. Les effectifs sont pour l'heure encore insuffisants pour distinguer des différences significatives entre le risque chez les hommes et les femmes mais le risque est cependant plus faible pour les jeunes par rapports aux plus de 18 ans.

Les résultats sont assez disparates selon les études sur les jeunes<sup>22</sup>, bien que dans l'ensemble les incidences semblent plus faibles que celles des professionnels ou que les amateurs plus âgés. Ces résultats sont confirmés par les chiffres du Youth Rugby Injury Surveillance Project de la RFU qui indique des incidences respectivement de 17.3, 22.0 et 41.9 pour les U13, U15 et U18 de 26 équipes de 16 écoles<sup>23</sup>.

### 1.1.2 Sévérité des blessures

En plus de la fréquence de blessures, il est important de considérer la sévérité de celles-ci, qui au-delà des conséquences sur la santé des joueurs induira une indisponibilité qui sera forcément préjudiciable pour l'équipe. Est défini par World Rugby comme sévérité, « *le nombre de jours écoulés depuis la date de la blessure jusqu'à la date du retour du joueur à la participation complète à l'entraînement de l'équipe et la disponibilité pour la sélection du match* ».

Aujourd'hui la sévérité moyenne des blessures dans la majorité des programmes épidémiologiques sur le Rugby de haut niveau est autour des 30 jours d'absences<sup>13,4,5,24</sup>. Par exemple pour le championnat anglais, lors de la saison 2019/2020, cette sévérité moyenne était située à 38 jours<sup>4</sup> mais celle-ci est très influencée par les arrêts de longue durée car la sévérité médiane n'est que de 11 jours d'absence pour cette même saison sur ce championnat. Cependant, on constate une tendance à la hausse de la sévérité moyenne, particulièrement ces dernières années due à la hausse du nombre de blessures graves (>84 jours)<sup>5</sup>. C'est le cas dans le championnat anglais mais pas seulement. Autre exemple, lors des coupes du monde 2015 et 2019, la sévérité moyenne était de 29 et 30 jours, tandis qu'elle était de 24 pour celle de 2011 et seulement 15 pour celle de 2007<sup>14,15</sup>. De plus, lors d'un symposium World Rugby en 2019, chaque fédération a présenté son suivi des blessures et les résultats qui en découlent. Ainsi, on a retrouvé cette tendance à la hausse de la sévérité dans d'autres championnats nationaux comme en France, au Pays de Galle ou encore en Nouvelle-Zélande. Une des hypothèses susceptibles d'expliquer cette augmentation est une plus forte considération de la prise en charge des blessures avec une extension du délai de retour à la compétition. La commotion en est un bon exemple où World Rugby a mis en place un programme de reprise progressive au jeu (programme RPDJ) pour s'assurer que le joueur reprenne le Rugby en toute sécurité.

« L'injury-burden » est un autre indicateur qui est couramment calculé pour mesurer le niveau d'indisponibilité due aux blessures. Il résulte simplement du produit de l'incidence et de la sévérité moyenne et permet ainsi d'avoir une meilleure perception de l'indisponibilité totale des joueurs. Etant donné que le nombre de blessures reste relativement stable au fil des années et que le temps de retour à l'entraînement est de plus en plus long, il est donc logique de voir que cette indisponibilité tend à croître au fil des années. Il y a des répercussions de plus en plus importantes des blessures sur la gestion de l'effectif d'un club, il est donc nécessaire d'analyser plus finement les blessures rencontrées.

### 1.1.3 Types de blessures

Dans un sport complet comme le Rugby et où la part de combat est très marquée, il est logique de retrouver une hétérogénéité dans les blessures plus importante que dans beaucoup d'autres sports.

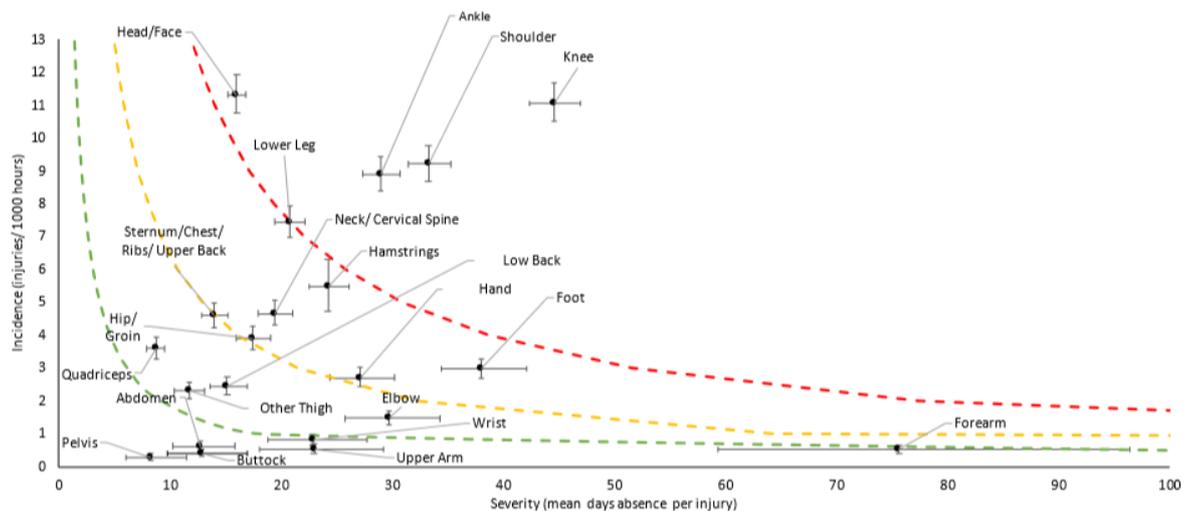


Figure 2 : Indisponibilité en fonction de la localisation. Les lignes verte, orange et rouge représentent respectivement le 25<sup>ème</sup>, 50<sup>ème</sup> et 75<sup>ème</sup> percentile.<sup>5</sup>

La zone majoritairement touchée est celle du membre inférieur. A elle seule, elle a une incidence de 45 blessures dans cette zone pour 1000 heures-joueurs<sup>5</sup>. Dans la Figure 2, on s'aperçoit que le genou est une articulation très sensible puisqu'elle fait partie des blessures les plus courantes (11.1 pour 1000 heures-joueurs) et avec la sévérité parmi les plus importantes (45 jours en moyenne). A moindre proportion, la cheville fait également partie des blessures les plus courantes. Dans cette zone on retrouve en grande partie des blessures musculaires aux ischio-jambiers, c'est même la première cause de blessure à l'entraînement (0.47 pour 1000 heures-joueur). Les autres blessures les plus communes dans cette zone sont ligamentaires et notamment du ligament collatéral médial (ou ligament latéral interne) ainsi que du ligament croisé antérieur<sup>4</sup>. Une étude de Montgomery et al. (2016)<sup>25</sup> a permis d'en savoir plus sur les mécanismes de survenues de ces blessures au ligament croisé antérieur. Dans cette étude, elles interviennent principalement quand le joueur est plaqué et lorsque le joueur réalise une course offensive. Dans ce second scénario c'est principalement lors d'une phase d'évitement.

Le membre supérieur est moins impacté avec une incidence de 15 blessures pour 1000 heures joueur. Dans cette zone, c'est l'épaule qui est particulièrement touchée. D'après Usman

et McIntosh (2012)<sup>26</sup>, le risque de blessures dans cette zone aurait tendance à décroître selon le niveau de jeu.

L'autre zone du corps qui est aussi vulnérable et qui mérite une attention particulière est la tête et notamment les commotions.

#### 1.1.4 Les commotions

La commotion cérébrale est une des préoccupations majeures dans le Rugby et accroît l'attention portée aux blessures. En effet, ces blessures peuvent avoir des effets délétères à long terme sur la santé des joueurs<sup>5</sup>. C'est aujourd'hui le type de blessures le plus fréquent dans le rugby. L'incidence estimée aujourd'hui dans le Rugby est de 12 commotions pour 1000 heures-joueurs<sup>13</sup>. Le plus inquiétant est l'augmentation de ce risque au fil des saisons. Pour exemple, lors des 9 premières saisons du PRISP, l'incidence était située entre 3 et 6 commotions pour 1000 heures-joueur<sup>4</sup>. Depuis la saison 2012-2013, cette incidence n'a cessé d'augmenter pour atteindre un pic en 2016-2017 de 21 commotions pour 1000 heures-joueur et reste aujourd'hui à un niveau équivalent de 19.8 (Figure 3). Cette augmentation intervient à la suite de l'instauration du protocole HIA (Head Injury Assessment) permettant une meilleure identification de la commotion via des évaluations en trois temps, pendant le match (HIA1), post-match (HIA2) et 36/48 heures après la blessure (HIA3). Ainsi on peut supposer que ce protocole soit une des explications de cette hausse de l'incidence observée. Cependant, outre cette augmentation de la vigilance sur ces traumatismes, une telle évolution reste source d'inquiétude surtout sur ce type de blessure.

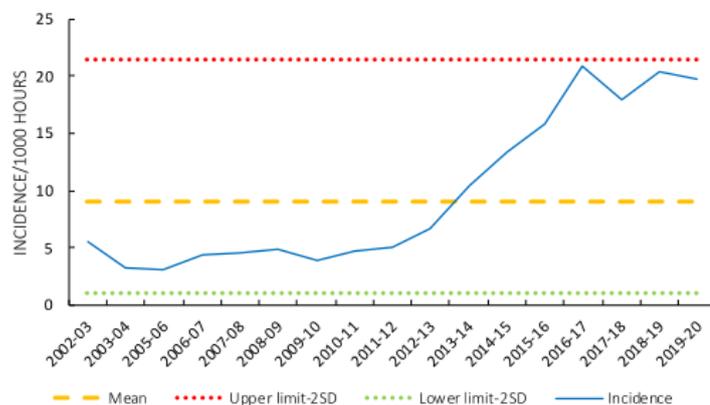


Figure 3 : Taux d'incidence des commotions en match Premiership sur la période d'étude avec moyenne  $\pm$  2 écart-type. Professional Rugby Injury Surveillance Project – RFU 2019/2020<sup>4</sup>

Le Royaume-Uni n'est pas un cas isolé puisque lors du symposium World Rugby de 2019, on a pu s'apercevoir qu'on avait un pic dans beaucoup de championnats nationaux (Argentine, Angleterre, Ecosse, France, Irlande, Nouvelle-Zélande) sur la même saison et une baisse par la suite (Figure 4).

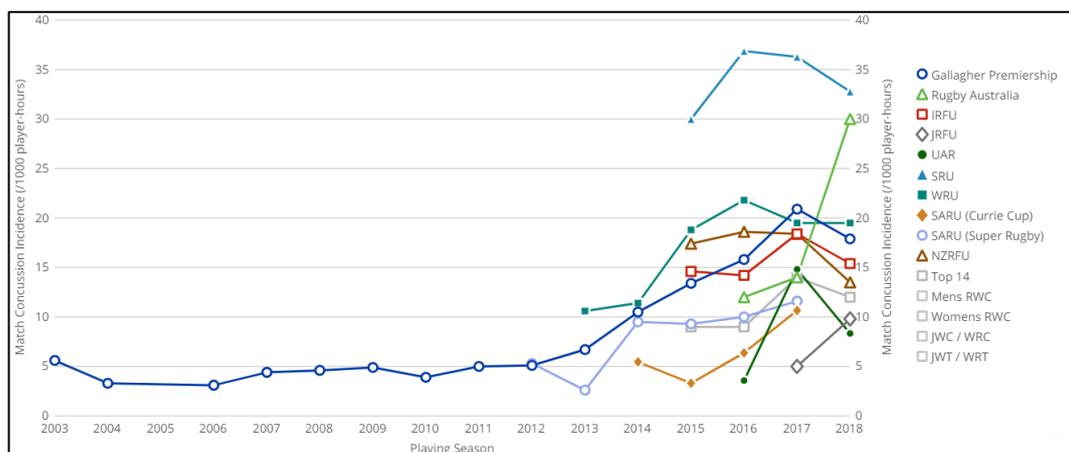


Figure 4 : Evolution des incidences de commotion dans les compétitions Elites Nationales en Rugby. Communication Symposium World Rugby 2019, Marcoussis

## 1.2 Facteurs de risque de blessures

### 1.2.1 Circonstance de survenue

Le combat physique est très présent au rugby, il est donc prévisible qu'il concentre une grande partie des blessures. En effet, la majorité des blessures en rugby ont lieu lors des phases de contacts et particulièrement sur le plaquage. Dans beaucoup d'études, le plaquage concerne près de 50% des atteintes totales et même 70% des commotions<sup>13,4</sup>.

En dehors du plaquage, et à un niveau globalement similaire, les collisions (sans le fait de ceinturer comme le plaquage), les rucks, puis les courses sont les autres causes de blessures les plus fréquentes en rugby.

Une alternative pour quantifier le risque de chaque action de jeu est de rapporter le nombre de blessures au nombre d'occurrences de ces actions. Cela permet ainsi de s'affranchir du fait que certaines actions surviennent plus fréquemment et d'ainsi d'identifier plus précisément la dangerosité de chacune. C'est la méthodologie utilisée par Roberts et al. (2014)<sup>27</sup>, sur 3 saisons et différents niveaux du rugby anglais. Le plaquage a donc ainsi un ratio de 2.3 blessures pour 1000 plaquages contre 0.5 blessures pour 1000 rucks. Certains types de

contact comme les plaquages illégaux impliquant une charge à l'épaule sont beaucoup plus à risque puisqu'on a 15 blessures pour 1000 plaquages.

Aujourd'hui, c'est donc le plaquage qui concentre toute l'attention et le risque est aussi bien encouru pour le plaqueur que le plaqué. La méta-analyse de Williams en 2013<sup>28</sup> montre que le plaqué est d'avantage touché avec une incidence de 29 blessures pour 1000 heures-joueur contre 19 pour le plaqueur<sup>28</sup>. Plus récemment, la nouvelle mise à jour de Williams et al.<sup>13</sup> démontre une proportion des blessures similaire entre le plaqueur (23.0%) et le plaqué (22.8%). De même, dans le programme épidémiologique mené sur le championnat anglais, le constat a été identique. Depuis quelques années l'incidence de blessures du plaqueur a fortement augmenté pour atteindre un niveau de risque comparable au porteur de balle. Pour les commotions, dans 76% des cas, c'est le plaqueur qui subit la commotion<sup>29</sup>.

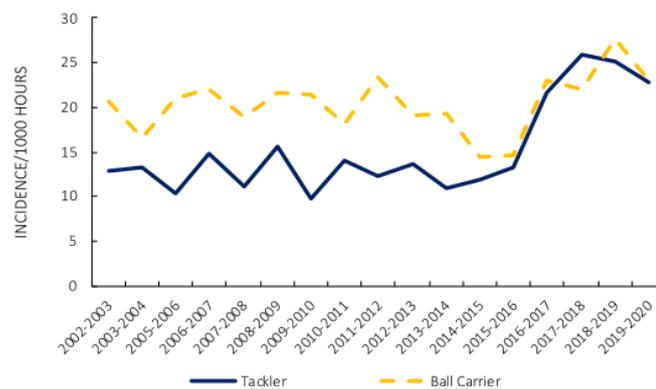


Figure 5 : Taux d'incidence du porteur de balle ou du plaqueur en Premiership sur la période d'étude avec moyenne. Professional Rugby Injury Surveillance Project – RFU 2019/2020<sup>4</sup>

Ces dernières années plusieurs analyses vidéo des plaquages ont été réalisées pour tenter de comprendre plus finement la survenue des commotions afin de tenter de les faire diminuer en réformant le plaquage.

Un des premiers facteurs qui est généralement cité est la technique de plaquage. Avoir les bonnes attitudes, le bon positionnement, notamment de la tête, lors du plaquage semble déterminant pour limiter le risque<sup>30,31</sup>. Outre la technique du plaqueur, celle du porteur de balle est également à considérer. Quand celle-ci est déficiente, il peut y avoir davantage de blessures pour lui ou même pour son vis-à-vis<sup>32,33,34</sup>.

En dehors de ces aspects techniques, la vitesse du plaquage serait aussi déterminante. Cross et al. (2017)<sup>35</sup> ont montré qu'il y avait un risque plus élevé quand le plaqueur accélérât

ou qu'il était déjà à une vitesse élevée au moment du plaquage. Également dans cette étude, et confirmé dans une seconde de Tucker et al. la même année<sup>29</sup>, ce sont surtout les plaquages tête contre tête qui occasionnent beaucoup de commotions comparé aux autres types de contacts (Odds ratio de 40).

C'est donc pour tenter de limiter ce type de contact qu'une expérimentation a été faite dans la Championship Cup 2018/2019 en interdisant les plaquages au-delà de la ligne située au niveau du sternum dans le but de baisser la hauteur du plaquage. Les résultats de cette expérimentation n'ont pas été concluants étant donné que l'incidence de commotions fut supérieure, bien que non significativement, pendant cet essai comparé aux autres périodes de la compétition (pré et post)<sup>36</sup>. Il semble donc que baisser la hauteur de plaquage ne soit pas la solution pour réduire le risque.

### 1.2.2 Les antécédents de blessures

Un des principaux facteurs de risque est la blessure elle-même. En effet, en dehors du fait que la blessure va occasionner une période où le joueur ne pourra ni jouer ni s'entraîner, au-delà de cette période, celle-ci entraînera probablement des conséquences dans la survenue d'une nouvelle atteinte. Une étude a été réalisée à partir des joueurs de 8 saisons de Premiership dans le but d'analyser le risque de rechute<sup>37</sup>. Les auteurs concluent que 70% des blessures sont de nouvelles blessures, 14% ont la même localisation que la précédente mais un type différent et 16% sont identiques à la précédente blessure. Parmi ces dernières, celles qui ont un risque plus élevé de rechute sont les tensions musculaires au cou, les entorses de la cheville et les névralgies cervicales. D'autres blessures comme celles aux ischio-jambiers sont aussi plus à risque pour un joueur en ayant déjà subi une récemment<sup>38</sup>. Dans cette même étude on a pu constater que le fait d'avoir une faible force excentrique du genou n'est pas forcément un facteur de risque mais le fait d'avoir un déséquilibre entre les deux jambes l'est davantage.

Certaines blessures comme les commotions, auraient, elles aussi, des répercussions dans la survenue d'autres traumatismes. D'après Cross et al. en 2015<sup>7</sup>, les joueurs ayant subi une commotion ont 60% de risque en plus de se blesser que les autres. Ils distinguent également une différence entre les commotionnés et les joueurs atteints par une blessure autre que la commotion. Les joueurs commotionnés se blessent de nouveau plus rapidement que les joueurs touchés par une autre blessure. En effet, l'intervalle de temps médian de survenue d'une

nouvelle blessure après avoir repris la compétition est de 53 jours pour les commotionnés contre 114 pour les autres joueurs.

### 1.2.3 Le profil des joueurs

Selon leurs postes, les profils physiques des joueurs sont dissemblables et les tâches qui leur sont demandés en match le sont tout autant. En exemple, la table 2 représente le profil des joueurs de la dernière coupe du monde. On s'aperçoit qu'en moyenne les secondes lignes mesurent 197.3 cm contre seulement 179 cm pour les joueurs de la charnière et en termes de poids les avants pèsent 111.8 kg en moyenne contre 91.2 pour les arrières<sup>15</sup>. Concernant leurs tâches sur le terrain, les avants subissent significativement plus d'impacts, sur les plaquages et les rucks que les arrières<sup>39</sup>. A contrario les arrières réaliseront plus de courses et notamment à haute intensité. De ce fait, on peut supposer que le risque de blessure et peut-être les facteurs de risque varient selon le poste qui est donc un facteur important à considérer. Par exemple, Thornton et al. (2016)<sup>40</sup> ont montré que l'effet de la charge de travail variait selon les postes des joueurs, d'où l'importance de l'adapter à chacun.

Playing position (n)	Age (years)	Stature (cm)	Body mass (kg)
<b>Backs (n=284)</b>	27.3 (3.5)	182.6 (6.0)	91.2 (9.1)
Halves (n=101)	27.7 (3.6)	179.0 (5.9)	85.1 (7.5)
Inside backs (n=76)	27.7 (3.2)	185.5 (4.5)	97.5 (7.0)
Outside backs (n=107)	26.7 (3.6)	183.9 (5.4)	92.6 (8.1)
<b>Forwards (n=362)</b>	27.3 (3.8)	188.3 (7.0)	111.8 (9.2)
Front row (n=169)	27.2 (3.9)	183.7 (4.5)	114.4 (9.1)
Second row (n=80)	28.0 (3.6)	197.3 (4.6)	114.8 (6.6)
Back row (n=113)	27.1 (3.7)	188.9 (5.0)	105.8 (7.9)
<b>All players (n=646)</b>	27.3 (3.7)	185.8 (7.2)	102.8 (13.7)
<b>p value*</b>	1.000	<0.001	<0.001

*Data are expressed as Mean (SD). \* indicates Backs vs Forwards  
RWC, Rugby World Cup; n, number of players*

Table 5: Données anthropométrique des joueurs de la coupe du monde 2019. Fuller et al. 2020<sup>15</sup>

Au-delà du poste, les risques encourus peuvent varier selon les profils des joueurs. Gabbett, Ullah et Finch en 2012<sup>41</sup> ont réalisé une étude prospective sur 66 joueurs de rugby pendant 3 années et ont observé que les plus lourds, les plus rapides, ceux avec un développement plus pauvre de course fractionné à haute intensité et qui ont une force du haut du corps plus importante ont un risque supérieur de subir une blessure sur contact.

Il est donc important d'individualiser dans la mesure du possible le suivi des joueurs étant donné les capacités et les potentielles fragilités de chacun.

#### 1.2.4 Autres éléments contextuels

En dehors de ces aspects individuels, d'autres facteurs liés au contexte de la rencontre peuvent faire varier le risque.

La période de la saison doit en effet être prise en compte. Dans les travaux de Gabbett en 2010<sup>42</sup>, on a pu observer des relations différentes entre la charge de travail et le risque de blessures musculaires selon les différentes phases de la saison. Les joueurs peuvent donc avoir une charge de travail plus importante lors de la phase de pré-saison sans être réellement à risque, pour autant à partir de 3000 AU (unité arbitraire) de sRPE, on observe une forte augmentation de la probabilité de blessures. Par ailleurs, lors de la dernière partie de saison, le risque est, quel que soit le niveau de charge de travail, plus sensible à une augmentation de celle-ci.

Enfin, au cours de la saison d'autres éléments contextuels peuvent amener à être plus prudent lors d'un match. Depuis leur apparition, les terrains artificiels font débat quant à leur impact sur la sécurité des joueurs. Ces terrains étant encore minoritaires, il est difficile de trouver des études avec suffisamment de données pour avoir des conclusions claires. Une publication de Cousins et al. en 2021<sup>43</sup>, sur deux saisons d'un club anglais alerte sur la dangerosité des terrains hybrides ou artificiels. Ils ont obtenu un OR de 2.58 [95% CI 1.65–4.03] pour les synthétiques et un OR de 2.16 [1.07–4.37] pour les hybrides par rapport aux terrains naturels. A une plus grande échelle, depuis 2016, il y a un retour annuel sur les blessures qu'ils occasionnent dans le PRISP de la RFU. Contrairement à l'étude de Cousins et al. sur un seul club, on n'observe pas plus de blessures sur ce type de terrain par rapport à un terrain naturel. Toutefois la sévérité des blessures s'avère plus longue<sup>4</sup>. C'est particulièrement le cas sur les blessures au niveau postérieur de la cuisse pour lesquels la durée d'absence est deux fois supérieure quand elles ont lieu sur terrain artificiel (46 jours contre 23 en moyenne)<sup>44</sup>. On retrouve également des blessures plus graves au niveau du pied puisqu'il y a 10 jours d'absence en moyenne sur terrain naturel contre 91 jours sur un terrain artificiel mais ces résultats restent à confirmer. Il semblerait donc que bien que pas plus nombreuses, les blessures rencontrées s'avèrent avoir des conséquences différentes. Il faut donc continuer à analyser l'effet de ces pelouses pour pouvoir s'y préparer au mieux. D'autant qu'il existe une hétérogénéité dans les différents types de terrains artificiels avec probablement des répercussions diverses.

### 1.3 Impact de la charge de travail sur le risque

Dans le but d'avoir des joueurs performants et à leur pleine capacité, la gestion de la charge de travail est un enjeu clé. Elle l'est aussi pour essayer de limiter leur risque de blessure. C'est un des seuls facteurs qui soit maîtrisable et c'est donc sur celui-ci que les staffs doivent agir pour améliorer la sécurité de leur joueur. Beaucoup d'études ont été menées pour essayer de comprendre l'influence de la charge de travail sur les blessures.

La charge peut être décomposée en charge externe (stimulus externe appliqué à l'athlète) et interne (réponse physiologique et psychologique individuelle à la charge externe)<sup>45</sup>. Le GPS est la source première de quantification de la charge externe, il va permettre de quantifier le déplacement du joueur à différentes intensités de courses : distance dans chaque zone de vitesse, nombre de fois qu'il y est entré, durée qu'il y a passé, etc. Dans ces boîtiers GPS est également intégré un accéléromètre qui permet de quantifier les accélérations et décélérations des joueurs, encore une fois selon différents seuils. A partir d'algorithmes basés sur les données de cet accéléromètre il est désormais possible d'avoir une quantification du nombre de contacts qui est un paramètre important pour caractériser l'activité en Rugby. Pour ce qui est de la charge interne, il est possible de l'obtenir via des données de cardiofréquencemètres ou de tests physiques. Plus simplement, la note des questionnaires RPE donne une bonne indication de la perception individuelle de la difficulté d'une séance. Bien souvent elle est couplée à la durée de la séance en les multipliant simplement ensemble pour pouvoir avoir une notion de durée et d'intensité. Bien entendu la note induit déjà cette notion de durée mais Foster<sup>46</sup> a montré la validité de cette méthodologie pour quantifier la charge ressentie par le joueur notamment en ayant démontré sa relation avec le rythme cardiaque sur différentes séances. Cette méthode nommée session-RPE ou sRPE est couramment utilisée au quotidien dans le suivi de la charge et dans les publications scientifiques quel que soit le sport.

Sur une semaine d'entraînement, Cross et al. (2016)<sup>47</sup>, ont trouvé un seuil de sRPE de 1250 unité arbitraire (AU) au-delà duquel le risque de blessure augmente. Sur de plus longues périodes le lien semble non linéaire. En effet, dans ce même article de Cross et al., portant sur une cohorte de 173 rugbymans du championnat anglais, les auteurs montrent qu'en accumulant la charge d'entraînement lors des 4 dernières semaines, on distingue une courbe en U. Il y aurait donc un risque moindre de se blesser quand la charge d'entraînement du dernier mois est comprise dans l'intervalle [5932, 8651] et plus important au-dessus et en dessous de celui-ci. On retrouve un résultat semblable sur une période encore plus longue dans une publication de

Williams et al. en 2017<sup>6</sup>. Dans cette étude conduite sur 7 saisons et concernant une cohorte de 1253 joueurs du championnat anglais, les auteurs montrent que les joueurs ayant joué moins de 15 rencontres lors de la dernière année et ceux qui au contraire en ont disputé plus de 35 ont un risque plus élevé de se blesser.

Au-delà d'avoir eu une charge de travail trop faible ou trop forte, avoir une variation trop importante serait également un facteur d'augmentation du risque. Ainsi, Blanch et Gabbett ont proposé en 2015<sup>48</sup>, le principe de ratio aigue/chronique, communément nommé « Acute Chronic Workrate Ratio » (ACWR). Il est défini comme étant le rapport entre la charge moyenne de la semaine précédente et celle du mois précédent. L'idée sous-jacente étant donc de ne pas faire varier de manière trop brutale la charge de travail des joueurs mais plutôt de manière progressive.

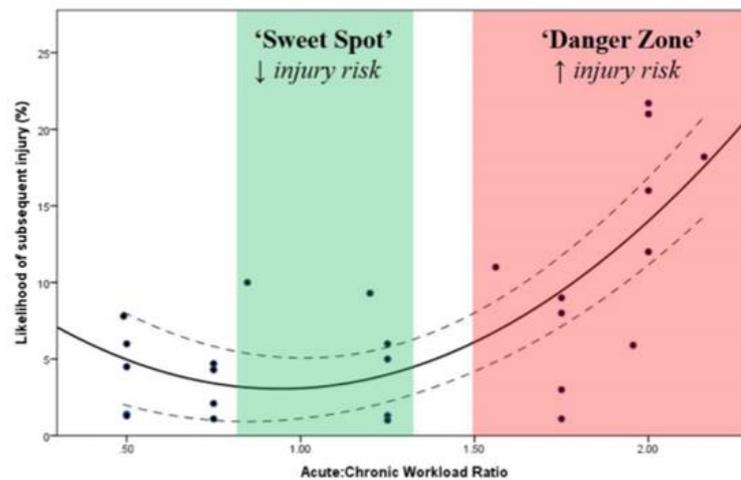


Figure 6 : Guide pour appliquer et interpréter les ratios aigue/chronique. La zone verte représente la zone où le risque de blessure est faible et la rouge celle où le risque est élevé. Gabbett 2015, redessiné à partir de Blanch et Gabbett 2015<sup>49</sup>.

Dans une publication suivante<sup>49</sup>, Gabbett définit le ratio idéal comme étant situé entre 0.8 et 1.3 reprenant les résultats de l'étude précédente. A partir d'un ratio dépassant 1.5 le joueur est considéré à risque (Figure 6). Ce ratio et ces seuils sont devenus viraux et ont fait partie de la norme du suivi de la charge de travail pour énormément de clubs du monde entier et de tous les sports.

Cependant des interrogations ont été soulevées quant à cet indicateur notamment dues aux limites de cette étude. Il a été construit à partir des données de 3 études sur du Rugby, du football australien et du cricket, des sports bien différents qui peuvent avoir une réponse diverse à la charge. De plus, par manque d'information, certaines variables de charge de travail ont dû être discrétisées, en prenant la valeur centrale de l'intervalle (0.75 pour une valeur située entre

0.5 et 1) ou la valeur limite pour ceux aux extrémités (2 pour une valeur >2). Elles ont ensuite été analysées comme une variable quantitative continue en appliquant une simple interpolation linéaire pour conduire aux conclusions de cette étude. De plus, il existe une hétérogénéité sur les variables utilisées pour cette publication puisque c'est à la fois à partir de charge externe comme la distance totale et de charge interne comme la sRPE que l'interpolation a été réalisée. Se pose ainsi la question du choix des variables à considérer sur lesquelles calculer ce ratio. Dans une étude de Cummins et al.<sup>50</sup> sur la charge externe en Rugby, seule la durée, les accélérations, les décélérations, le Player Load (un indicateur de charge mécanique) et la distance à très haute intensité ont été liés positivement avec le risque de blessure sans contact. Cependant d'après plusieurs revues systématiques<sup>51,52,53</sup> reprenant de nombreuses publications à ce sujet sur différents sports, les ratios aigue/chronique semblent pouvoir être utilisés sur différentes métriques de charge interne ou externe sans qu'aucune ne soit véritablement préférable.

Le calcul du ratio aigue/chronique a également été remis en cause. Le ratio aigue/chronique tel qu'il a été défini mesure non pas une variation mais davantage une proportion<sup>54</sup>. En effet la charge de la semaine précédente (numérateur) est incluse dans la moyenne de la charge des 4 dernières semaines (dénominateur). Cette définition provoque deux limites à ce ratio. La première est qu'il conduira forcément à un maximum théorique de 4<sup>54</sup>. Sur ce point, on peut noter l'aspect bénéfique qu'il évite les valeurs extrêmes du ratio qui surviendraient forcément pour un joueur de retour de blessure par exemple. La seconde limite est que par définition, il y a une corrélation entre le numérateur et le dénominateur qui crée une baisse de la variabilité inter-joueur<sup>55</sup>. Il y a un débat autour du fait d'utiliser un calcul du ratio couplé ou non couplé, c'est-à-dire avec ou sans la moyenne de la semaine précédente dans le dénominateur. Quel que soit le calcul il ne changerait pas les conclusions<sup>56</sup>, mais certains considèrent que le non couplé serait mathématiquement plus juste et donc recommandé<sup>57,55</sup>. De plus, Impellizzeri et al.<sup>58</sup> ont conclu qu'il n'existe pas de différence sur l'impact du ratio, que l'on utilise la charge chronique réelle, une charge chronique fixe ou simulée aléatoirement et remettent donc en cause l'intérêt de ce ratio. Par ailleurs Hulin et al (2016)<sup>59</sup> montrent que la charge chronique pourrait également apporter une information supplémentaire au ratio aigue/chronique. Dans cette publication, il semblerait que les fortes charges combinées à des ratios aigue/chronique raisonnables auraient plutôt un effet protecteur. Ainsi prendre uniquement la valeur du ratio serait insuffisante étant donné que pour une même valeur, la relation avec le risque de blessure diffère selon le volume de la charge de travail.

Les moyennes calculées au numérateur et au dénominateur ont également été remises en causes. Murray et al. (2016)<sup>60</sup> ont proposé de ne pas prendre de simples moyennes glissantes mais d'utiliser des moyennes pondérées exponentiellement afin de donner ainsi une importance plus grande aux observations les plus récentes. Cette méthode nommée Exponentially Weighted Moving Average (EWMA) serait en effet plus sensible pour la détection des blessures.

Que ce soit la méthodologie de calcul, la plage de temps sur laquelle effectuer ce calcul, les variables utilisées ou encore les seuils limites à ne pas dépasser, il y a donc beaucoup de façon de calculer ces ratios mais le concept sous-jacent reste identique.

Afin de questionner le lien entre ces ratios et les blessures, une étude qui sert de référentiel intéressant, notamment dans le Rugby, a été menée par West et al. en 2020<sup>8</sup> sur une population importante de 696 joueurs de 13 clubs professionnels anglais sur deux saisons. Comme décrit dans la figure 7, de nombreux indicateurs (64) ont été construits que ce soit couplé ou non, sur des moyennes glissantes ou pondérées exponentiellement et sur différentes périodes.

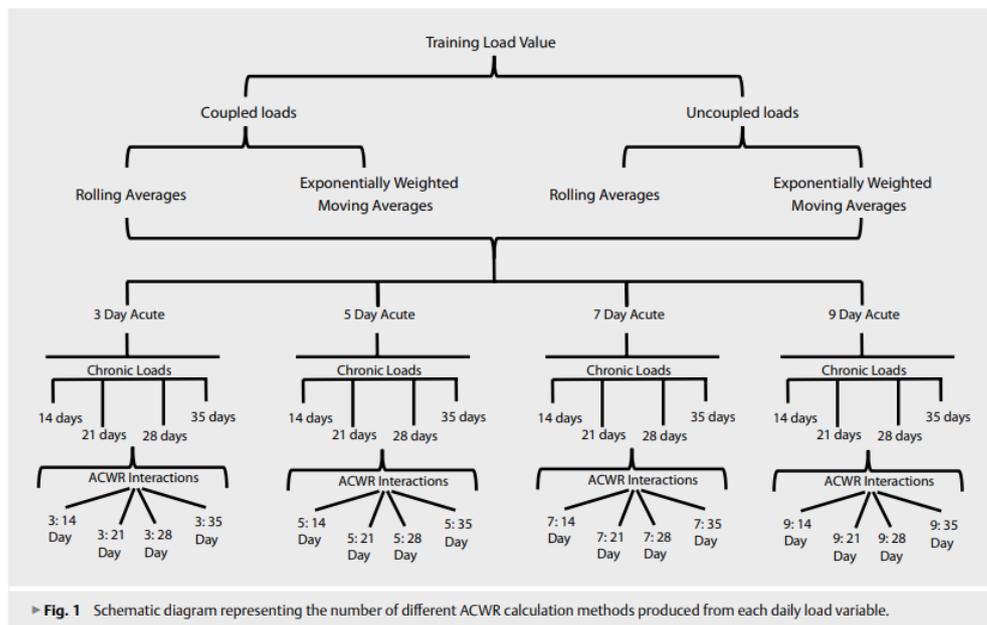


Figure 7: Diagramme des variables testées dans la publication de West et al., 2020<sup>8</sup>

Ajusté sur d'autres variables contextuelles comme les précédentes blessures et le temps de jeu cumulé, des comparaisons deux à deux ont été faites entre différents niveaux de ratio (faible, modérée et élevée) pour chacun des ratios de la Figure 7. Ces modèles ont été estimés pour expliquer soit l'ensemble des blessures soit uniquement les blessures musculaires, tendineuses ou ligamentaires sans contact. En conclusion de ce travail, seulement 40% des

comparaisons pour l'ensemble des blessures et 31% pour les blessures sans contacts se sont révélées significatives. Les meilleurs modèles selon l'aire sous la courbe ROC (AUC) (cf section 3.1) sont l'EWMA non couplé de 3 sur 14 jours (AUC=0.64) pour l'ensemble des blessures et l'EWMA couplé de 5 sur 14 jours sur les blessures sans contact (AUC=0.69). Ces résultats sont assez révélateurs du caractère aléatoire de ce lien avec les blessures où aucune période, pondération des moyennes ou éventuel couplage ne fait foi pour caractériser le lien avec les blessures. Le pouvoir prédictif reste dans tous les cas assez faible et la relation causale n'est pas prouvée<sup>57</sup>.

Une autre limitation évoquée par ces approches, que ce soit l'ACWR ou l'EWMA, est la sensibilité aux valeurs manquantes, courante en science du sport. Pour répondre à cette problématique, Moussa et al.<sup>61</sup> ont proposé un nouvel indicateur, le REDI (Robust Exponential Decreasing Index). Il est défini ainsi :

$$REDI_{today}^{\lambda} = \frac{1}{\sum_{i=0}^N \alpha_i^{\lambda}} \sum_{i=0}^N \alpha_i^{\lambda} \times WL_i$$

$$\text{et } \alpha_i^{\lambda} = \begin{cases} 0 & \text{si } WL_i \text{ est manquant} \\ e^{-\lambda i} & \text{sinon} \end{cases}$$

où  $WL_i$  est la charge de travail du  $i$ -ème jour avant le jour actuel.  $N$  est le nombre total de jours précédents dans le jeu de données ;  $\lambda$  est un paramètre qui peut être ajusté afin de modifier la pondération.

Le concept principal de cet indice est d'introduire un poids exponentiel explicite qui multiplie chaque charge de travail. Toutes les charges de travail pondérées sont additionnées puis divisées par les sommes des poids afin de normaliser l'indice. En ce sens, l'indice est défini comme une moyenne pondérée des charges de travail. Bien que les propriétés de cet indicateur semblent intéressantes, aujourd'hui il n'existe pas encore de publication ayant montré un lien entre celui-ci et la fatigue ou le risque de blessure.

Il existe donc aujourd'hui plusieurs méthodologies pour étudier la relation entre la charge de travail et le risque de blessure que ce soit en étudiant un cumul ou une variation mais aucune approche ne fait réellement consensus.

## **2) Méthodologie statistique**

Dans ce contexte, l'objectif est généralement d'identifier des facteurs de risque de blessure tout en considérant les spécificités de ces données. Nous sommes en présence de données répétées où plusieurs observations peuvent être liées à un même individu et/ou à une même équipe, d'où une corrélation intra-individuelle et/ou intra-équipe. De plus, les blessures sont des événements récurrents qui nécessitent donc des modèles appropriés à ce type d'évènement dans le cas où l'on souhaite analyser le délai de survenue de ces blessures. Il existe un grand nombre de prédicteurs possibles tandis que le nombre de blessures reste un évènement relativement rare ce qui peut nécessiter de recourir à des méthodes de sélection de variables ou de réduction de dimension. L'objectif de ces études peut être explicatif, pour identifier les facteurs de risque, mais également prédictifs afin de construire des modèles de prédictions du risque individuel. Autant de contraintes qu'il faudra considérer dans les analyses pour répondre à cette problématique.

### **2.1 Traitement des données manquantes**

Dans le cas du suivi de la charge de travail, de nombreuses raisons peuvent expliquer que certaines mesures puissent manquer. Par exemple la non-réponse d'un joueur à un questionnaire ou encore un souci technique sur un GPS. C'est un problème courant dans les études longitudinales. Ces données manquantes vont entraîner une perte d'information et donc une réduction de la puissance de l'étude. Elles peuvent également induire un biais voir même rendre les analyses impossibles. L'impact de ces données manquantes sera dépendant notamment de leur type. Rubin en 1976<sup>62</sup> a proposé une classification des différents types de données manquantes.

#### **2.1.1 Types de données manquantes**

Les données manquantes sont considérées comme complètement aléatoires (missing completely at random, MCAR) si la probabilité d'être absente est la même pour toutes les observations. Les causes des données manquantes ne sont ici pas dépendantes de données observées ou de paramètres d'intérêt non observés.

Dans le cas où la probabilité d'être manquante est liée à des données observées, on définit les données comme manquantes aléatoirement (missing at random, MAR). Cette hypothèse est donc plus générale et réaliste que les données MCAR.

Enfin si la probabilité d'observation dépend de la valeur de la variable en question, les données manquantes sont définies comme non aléatoires (missing not at random, MNAR). Ces données manquantes pourront provoquer un biais qui sera important à considérer.

## 2.1.2 Technique de traitement de données manquantes

### *2.1.2.1 Analyse des cas complets et des cas disponibles*

Une des méthodes les plus commune pour traiter des données manquantes est l'analyse des données complètes elle consiste simplement à supprimer les observations qui possèdent des données manquantes. Si les données ne sont pas MCAR, cette suppression peut entraîner un biais du fait que les individus conservés dans l'échantillon d'analyse ne sont pas représentatifs de l'échantillon initial. De plus, cette méthode induit une perte d'information qui peut être conséquente. Une possibilité pour limiter la perte d'information peut tout simplement être de supprimer la ou les variable(s) possédant un taux de valeurs manquantes trop important.

Pour limiter la perte d'information, la suppression par paire ou analyse de cas disponibles peut être employée. Cette méthode permet de calculer les moyennes de chacune des variables sur leurs données observées et les corrélations et covariances sur les paires de variables sans donnée manquante. Si les données ne sont pas MCAR ces estimations peuvent être biaisées<sup>63</sup>. Cette méthode entraîne des tailles d'échantillons différentes selon les variables étudiées, il faudra passer par des techniques d'optimisations avancées pour les calculer<sup>64</sup>.

### *2.1.2.2 Imputation simple de données manquantes*

Au lieu de supprimer des données, plusieurs méthodes permettent de donner une valeur de substitution à ces données manquantes. Ces différentes méthodes varient en termes

d'approches, d'avantages et d'inconvénients il est donc important d'en sélectionner une cohérente par rapport à notre contexte.

Une première méthode plutôt simpliste pourrait être d'imputer par la moyenne, la médiane ou le mode. Cette méthode est bien entendue limitée, elle sous-estimera la variance de la variable et biaisera également les corrélations avec les autres variables<sup>65</sup>. Cette méthode est donc déconseillée surtout dans le cadre d'analyse multivariée. Des méthodes intuitives peuvent être envisagées via des règles qui peuvent avoir un sens du point de vue du contexte. Par exemple, Griffin et al<sup>66</sup> ont testé différentes règles d'imputation basées sur les données de l'équipe, du joueur ou des deux simultanément pour imputer des données de charge de travail en football. C'est en imputant par la charge moyenne de l'équipe de la session d'intérêt qu'ils obtiennent les meilleurs résultats avec l'erreur quadratique moyenne normalisée la plus faible, et ce, quel que soit le pourcentage de données manquantes. Cette méthodologie est intéressante en pratique d'autant qu'elle est simple à mettre en place, cependant elle fait abstraction d'autres facteurs comme des caractéristiques individuels du joueur.

D'autres méthodes comme celles des k plus proches voisins peuvent également être employées<sup>67</sup>. Cette méthode utilise un algorithme pour déterminer les k voisins les plus proches, c'est-à-dire les observations ayant les caractéristiques les plus semblables, puis prend pour valeur la moyenne de la valeur des k voisins. Le nombre de voisins k à sélectionner peut-être déterminé par validation croisée. Il est nécessaire de choisir également la distance permettant de déterminer la similarité entre les observations. Le choix de la distance influence les imputations d'où son importance<sup>68,69</sup>, la plus classique étant la distance euclidienne. La méthode des k plus proches voisins est facile à implémenter cependant elle peut être trop coûteuse en temps de calcul avec trop de variables ou d'observations et peut donc être difficilement envisageable pour notre problématique.

Pour pouvoir considérer plusieurs facteurs dans l'imputation, des méthodes de régressions peuvent être employées. Dans ce cas pour chaque variable à imputer, un modèle de régression est estimé à partir d'autres variables sur les valeurs observées. Les prédictions sont ensuite utilisées pour imputer les données manquantes de la variable réponse. Les valeurs imputées sont ainsi situées sur une droite de régression ce qui implique que la corrélation entre les prédicteurs et la variable réponse est surestimée<sup>70,71</sup>. Une façon de contourner cette limite est d'ajouter un terme résiduel à la valeur imputée par la regression classique, cette méthode est

nommée imputation par régression stochastique<sup>72</sup>. Le résidu est tiré d'une distribution normale avec une moyenne nulle et un écart type égale à la variance résiduelle de la régression. Cette méthode permet donc d'ajouter du bruit aux prédictions et d'ainsi obtenir des estimations non biaisées sous l'hypothèse MAR<sup>65</sup>.

Dans le cas de l'imputation simple présenté jusqu'ici, quel que soit la méthode, la donnée manquante est remplacée par une valeur unique. Par conséquent l'imputation ne tiens pas compte de l'incertitude sur les valeurs imputées et sous-estime la variance des estimations réalisées sur l'échantillon imputé. Des méthodes comme l'imputation multiple pourront ainsi permettre de résoudre ce problème.

### 2.1.2.3) Imputation multiple

L'imputation multiple consiste à imputer plusieurs fois les données manquantes afin de tenir compte de cette incertitude<sup>73,74,75</sup>. Pour cela, Rubin<sup>76</sup> a développé une méthode en trois étapes (Figure 8).

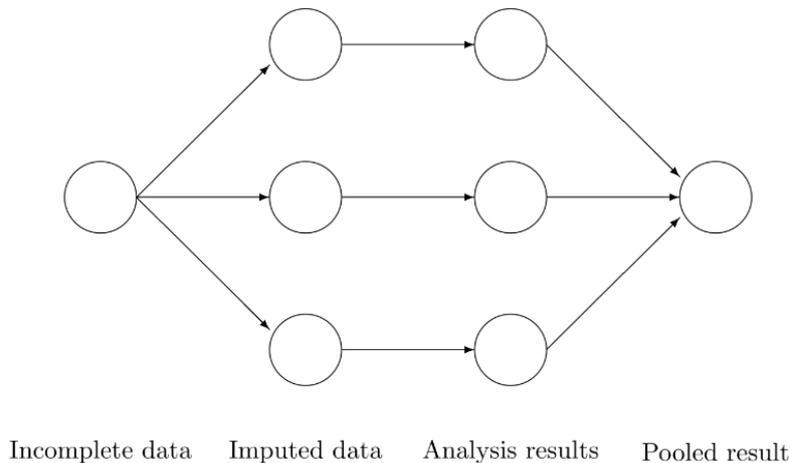


Figure 8 : Schéma des étapes de l'imputation<sup>77</sup>

La première consiste à réaliser  $m$  imputations par tirage aléatoire dans la distribution de la variable à imputer conditionnellement aux données observées en tenant compte de la variance des paramètres estimés du modèle d'imputation et de l'erreur de prédiction. Dans un second temps, les modèles statistiques envisagés pour l'analyses sont estimés sur chacun des  $m$  jeux de données complets. Enfin dans la dernière étape, les estimations sont combinées pour obtenir

les estimations globales, les variances et les intervalles de confiance selon la règle de Rubin<sup>76</sup>. Ainsi l'estimation combiné  $\bar{\theta}$  est la moyenne des estimations des données imputées :

$$\bar{\theta} = \frac{1}{m} \left( \sum_{i=1}^m \theta_i \right)$$

La variance de cet estimateur combiné est obtenue à partir de la variance intra-imputation et inter-imputation :

$$V(\bar{\theta}) = V(\bar{\theta})_{intra} + \left( 1 + \frac{1}{m} \right) V(\bar{\theta})_{inter}$$

où la variance intra-imputation est la moyenne des variances des imputations :

$$V(\bar{\theta})_{intra} = \frac{1}{m} \left( \sum_{i=1}^m V(\theta_i) \right)$$

et la variance inter-imputation est la somme de l'écart au carré entre l'estimation de chaque jeu de données imputés et le paramètre combiné pondéré par  $m - 1$  :

$$V(\bar{\theta})_{inter} = \frac{1}{m - 1} \left( \sum_{i=1}^m (\theta_i - \bar{\theta})^2 \right)$$

Utiliser 3 à 5 jeux de données d'imputation sont généralement suffisants<sup>78,79</sup> mais des méthodes plus avancées peuvent permettre de déterminer le nombre d'imputations à réaliser en fonction de la part de données manquantes<sup>80</sup>.

Généralement plusieurs variables possèdent des données manquantes selon une structure non emboîtée. Dans ce cas, l'estimation des modèles d'imputation sur l'échantillon des sujets avec données complètes conduit à une perte d'information importante et un risque de biais. L'imputation multiple par équation chaînée (mice)<sup>72</sup> permet de remédier à ce problème. Cet algorithme consiste à imputer chaque variable de manière itérative via un modèle d'imputation à partir de toutes les autres variables, y compris celles dont les valeurs sont manquantes, en utilisant leurs valeurs imputées à l'itération précédente. Chaque variable à imputer est modélisée selon une distribution appropriée ce qui offre une bonne flexibilité<sup>81</sup>. Pour éviter des biais dans le modèle d'analyse, toutes les variables qui seront incluses doivent être présentes dans le modèle d'imputation pour que leurs relations puissent rester identiques<sup>82</sup>. Pour les mêmes raisons, il est donc important d'inclure également la variable réponse dans le modèle d'imputation<sup>83,84</sup>.

Pour être plus robuste face à un éventuel écart par rapport à l'hypothèse gaussienne, des méthodes comme l'appariement moyen prédictif (predictive mean matching, PMM)<sup>85,86</sup> peuvent être utilisées. Cette procédure consiste à imputer une observation non pas par la prédiction mais par une valeur observée sur un autre individu choisi au hasard parmi l'ensemble des individus dont la valeur prédite par le modèle d'imputation est proche de la prédiction de la valeur manquante. Il a été montré que cette méthode est robuste quel que soit la distribution de la variable imputée<sup>87,88</sup>. Néanmoins cette méthode peut avoir pour risque de dupliquer trop souvent les mêmes valeurs dans le cas de petits échantillons ou quand le taux de valeur manquantes est trop conséquent.

Dans la littérature scientifique en science du sport, assez peu d'articles reportent des données manquantes. Dans une étude sur 137 articles de recherche en football, seulement 11% mentionnent avoir des données manquantes et parmi ceux-ci seulement 4 les ont imputées<sup>89</sup>. Il est difficile de vérifier si cette faible proportion est bien représentative de la réalité mais elle semble peu crédible. Les quelques articles qui ont imputé des données manquantes l'ont fait à partir de règles simples basés sur des moyennes<sup>90,91</sup>, méthode très courante en pratique. Bien que recommandée<sup>92</sup>, à ce jour il n'y a pas à notre connaissance aucune publication dans ce contexte qui a utilisé l'imputation multiple.

## **2.2 Modèles d'analyse de données corrélées**

L'approche la plus classique pour modéliser une variable binaire (blessé vs non blessé) en fonction de plusieurs facteurs explicatifs est la régression logistique, qui suppose que les observations sont indépendantes. Cependant, dans notre contexte, que ce soit des analyses longitudinales, des études transversales ou même encore des méta-analyses, les observations sont fréquemment groupées. Il est donc nécessaire de prendre en compte la corrélation intra-sujet, intra-équipe ou intra-étude des données répétées afin d'éviter des biais sur les estimations des variances des  $\hat{\beta}$  et donc sur les tests associés.

Dans certaines publications nous n'avons cependant aucune indication sur la prise en compte de cette corrélation. Ainsi il semblerait que de simples régressions logistiques ont été estimées sur des données longitudinales, notamment en football Australien dans les études de Rogalski et al.<sup>93</sup> et celle de Colby et al.<sup>90</sup> qui ont modélisé le risque de blessure en relation avec

des charges cumulées ou des variations de charges d'une semaine à l'autre. Gabbett<sup>42</sup> a également utilisé la même méthodologie pour étudier l'influence de la charge de travail sur le risque de blessure musculaire sans contact et pour en déduire par la suite un modèle de prédiction.

Dans la présentation des différentes approches envisageables nous nous plaçons dans le cadre de données longitudinales qui est le cas le plus courant dans les études sur le risque de blessures.

### 2.2.1 Approche conditionnelle

Une première possibilité de modélisation pour considérer les mesures répétées est l'approche conditionnelle utilisant les modèles mixtes. Cette méthode, dite « subject specific » permet de prendre en compte la corrélation entre les unités statistiques via des effets aléatoires. Il est donc composé d'effets fixes qui sont identiques pour tous les individus et représentent la tendance moyenne de la population, et d'effets aléatoires propre au groupement des données qui représenteront la déviation par rapport à cette tendance de la population. Pour une variable dépendante gaussienne, le modèle linéaire mixte peut s'écrire pour l'individu  $i$  au temps  $j$  par :

$$Y_{ij} = X_{ij}^T \beta + Z_{ij}^T \gamma_i + \varepsilon_{ij} \text{ avec } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ et } \gamma_i \sim N(0, B)$$

où  $X_{ij}$  est le vecteur des variables explicatives de dimension  $p$ ,  $\beta$  le vecteur des effets fixes,  $Z_{ij}$  un vecteur représentant la structure des effets aléatoires de dimension  $q$ ,  $\gamma_i$  le vecteur des effets aléatoires spécifiques au sujet et  $\varepsilon_{ij}$  les erreurs résiduelles supposées indépendantes et identiquement distribuées (iid). Conditionnellement à  $\gamma_i$ , les variables  $Y_{ij}$  sont indépendantes. La distribution multivariée de la variable réponse est entièrement spécifiée contrairement à l'approche marginale. Il a également été démontré sur la base de simulations, que l'estimation des effets fixes et de leurs variances était relativement robuste à une mauvaise spécification de la distribution de l'erreur ou des effets aléatoires<sup>94</sup>. De plus, les modèles mixtes sont également robustes aux données manquantes aléatoirement<sup>95</sup>, un argument intéressant étant donné qu'elles peuvent être courantes dans le suivi de la charge de travail.

Dans notre contexte, la variable dépendante est généralement binaire (indicatrice de blessure) ou quantitative discrète (nombre de blessures). On suppose donc que, conditionnellement aux effets aléatoires, la variable réponse suit un modèle linéaire généralisé à effets mixtes définit par :

$$g(\mathbb{E}[Y_{ij}|\gamma_i]) = X_{ij}^T\beta + Z_{ij}^T\gamma_i$$

Dans le cadre d'un modèle logistique avec un lien logit et une distribution binomiale, qui convient aux données binaires, la fonction de lien est définie par  $g(\mu) = \ln\left(\frac{\mu}{1-\mu}\right)$ . Pour un modèle de Poisson, adapté aux variables quantitatives discrètes, la fonction de lien est définie par  $g(\mu) = \ln(\mu)$ .

Dans la majorité des études sur le risque de blessure en sport qui utilisent des modèles mixtes, c'est une distribution de Poisson qui est choisie. La régression de Poisson est adaptée pour modéliser des taux représentant un nombre d'évènements sur un temps d'exposition. Il suffit pour cela d'insérer le logarithme du temps d'exposition dans le modèle dans un terme dit offset. En prenant ensuite l'exponentielle des coefficients  $\beta$  on obtient la valeur de l'incidence rates ratio (IRR). Ce modèle a par exemple été utilisé dans le basketball pour étudier l'influence de variables de déplacement, insérées en effets fixes, sur le risque de blessures en prenant le joueur en intercept aléatoire<sup>96</sup>. Pour des méta-analyses, ces régressions mixtes de Poisson avec offset sont également utilisées avec cette fois des effets aléatoires pour tenir compte de la corrélation intra-étude<sup>28,13,16</sup>. Williams et al. ont également utilisé un modèle mixte de Poisson avec le temps de jeu individuel en offset dans une stratégie à deux étapes<sup>97</sup>. Premièrement, une analyse en composante principale a été employée pour réduire le nombre de variables explicatives en permettant de regrouper celles corrélées en composantes. Ensuite, pour chaque composante, des modèles linéaires généralisé mixte ont été estimés à partir de chacune des variables participant à la construction de la composante pour déterminer celle la plus associée avec le risque de blessure en fonction du coefficient de détermination conditionnel. Chaque variable a ainsi été modélisée indépendamment en simple effet fixe et en ajoutant un terme au carré. Des modèles mixtes ont également été employés pour tenir compte de la corrélation intra-club dans une étude transversale qui avait pour but d'étudier la relation entre la charge d'entraînement, les performances et l'indisponibilité en rugby professionnel<sup>98</sup>. Concernant les relations entre charge d'entraînement et performance, des modèles linéaires mixtes ont été estimés tandis que des modèles linéaires généralisés avec une distribution gamma ont été employés pour étudier les facteurs associés à l'indisponibilité qui a une distribution étalée à gauche.

## 2.2.2 Approche marginale

Une seconde possibilité pour l'analyse de données corrélées est l'approche marginale utilisant les équations d'estimation généralisées (Generalized Estimating Equation, Liang & Zeger 1986, Prentice & Zhao 1991). Elle consiste à modéliser l'espérance marginale de la variable réponse en traitant la corrélation intra-sujet comme une nuisance en posant une hypothèse de travail sur la structure de la corrélation au sein des groupes. En utilisant l'estimateur sandwich ou robuste de la variance des  $\hat{\beta}$ , l'approche GEE permet d'obtenir des estimations des  $\beta$  et de leurs variances convergents même si la matrice de travail est mal spécifiée. Cette méthode est définie comme une méthode « population average » car les betas représentent les effets moyens des variables explicatives dans la population considérée. On pose  $\mu_{ij} = \mathbb{E}(Y_{ij})$  avec  $g(\mathbb{E}(Y_{ij})) = X_{ij}^T \beta$ . Les paramètres  $\beta_k$  sont estimés en résolvant :

$$\sum_{i=1}^N \frac{\partial \mu_i}{\partial \beta} V_i(\delta)^{-1} (Y_i - \mu_i(\beta)) = 0$$

où  $V_i(\delta)$  est une matrice de covariance de travail de  $Y$  dépendant d'un vecteur de paramètre  $\delta$ ;  $\hat{\beta}$  est asymptotiquement multivarié gaussien d'espérance  $\beta$  et de matrice de covariance robuste :

$$V_R(\hat{\beta}) = \left\{ \sum_{i=1}^K D_i^T V_i^{-1} D_i \right\}^{-1} \left\{ \sum_{i=1}^K D_i^T V_i^{-1} \text{cov}(Y_i) V_i^{-1} D_i \right\} \left\{ \sum_{i=1}^K D_i^T V_i^{-1} D_i \right\}^{-1}$$

où  $D_i = \frac{\partial \mu_i}{\partial \beta}$ .

Divers choix sont possibles pour la structure de corrélation de travail. Une possibilité est de ne pas définir de structure particulière ; dans ce cas tous les éléments de la matrice de corrélation seront estimés. Une autre option peut être de faire une hypothèse de travail d'indépendance. Si on considère que les observations d'un même groupe n'ont pas d'ordre particulier on peut choisir une matrice de travail échangeable qui suppose que toutes les paires d'observation d'un groupe ou d'un sujet ont la même corrélation. Enfin si on suppose qu'il peut exister une dépendance temporelle entre les observations d'un même sujet une éventualité est de définir une matrice autorégressive d'ordre 1 en supposant donc que les observations les plus proches dans le temps auront une corrélation plus importante. Toutefois, comme indiqué précédemment, quelle que soit la structure choisie, les estimateurs GEE sont convergents. Lorsque la matrice de covariance est mal spécifiée, les estimateurs GEE sont asymptotiquement

non biaisés. Cependant, contrairement aux modèles mixtes, la convergence des estimateurs GEE n'est assurée que si les données manquantes sont complètement aléatoires.

Cette approche a été employée dans plusieurs études sur le risque de blessures dans divers sports<sup>40,99,47,50,100</sup>. Thornthorn et al.<sup>40</sup>, en Rugby à treize, ont utilisé l'approche GEE pour estimer les paramètres de charge de travail les plus en lien avec le risque de blessure pour différents groupes de postes. Ils ont défini une matrice de corrélation intra-joueur non structurée. Colby et al.<sup>100</sup> ont également utilisé une approche GEE pour identifier en quoi la charge de travail réalisée sur différentes période de la présaison pouvait avoir un impact sur le risque de blessure sans contact dans la saison. L'étude de Cummins et al.<sup>50</sup> ainsi que celle de Cross et al.<sup>47</sup>, se sont intéressées à l'impact d'une semaine d'entraînement sur le risque de blessures de la semaine suivante. Les deux ont utilisé une structure autorégressive d'ordre 1 pour considérer la temporalité des répétitions. Cross et al., qui ont réalisé l'étude sur plusieurs clubs, ont considéré à la fois la corrélation intra-joueur et intra-club. Ils ont également ajouté le temps d'exposition en match en terme offset. En football, des méthodes d'estimations GEE ont également été utilisées pour modéliser l'impact de différentes variables contextuelles du match (résultat, compétition, lieu) sur le nombre de blessures par équipe<sup>99</sup>. Cette étude longitudinale a été menée sur 9 saisons et 26 clubs avec des données groupées au niveau équipe-match. Deux modèles ont été estimés avec comme variable réponse 0 vs 1 blessure et 0 vs 2 blessures ou plus. Pour considérer les mesures répétées sur l'équipe, une structure indépendante est spécifiée pour la matrice de travail.

## **2.3 Modèles de durée**

Une autre approche est de modéliser le temps avant la survenue de la blessure. Dans ce contexte, plusieurs défis peuvent survenir par rapport aux cas d'utilisation plus classique de ce type de méthodes. Tout d'abord concernant le choix du temps de base, un joueur n'est pas à risque tout le temps mais uniquement quand il pratique du Rugby, d'où la question de l'unité de temps à sélectionner. De plus, un joueur peut subir plusieurs blessures au cours de la saison, il est donc important de considérer ces évènements récurrents.

### 2.3.1 Choix du temps de base

Trois formulations sont possibles pour déterminer l'intervalle de risque, c'est-à-dire quand le joueur est à risque de se blesser sur une échelle de temps donnée. En raisonnant en temps total, l'échelle de temps est la même pour tous les évènements, elle est toujours référencée par rapport au premier temps d'observation. Le temps calendaire utilise la même échelle de temps qu'en total time mais cette fois le sujet est à risque du  $k^{\text{ème}}$  évènement qu'à partir du temps du  $k - 1^{\text{ème}}$ . Dans ce cas, le début de la période à risque correspond à la date de l'évènement précédent. Enfin, l'écart de temps (« gap time ») représente le temps depuis l'évènement précédent, il est donc remis à zéro à chaque nouvelle occurrence. Les modèles en gap time et en temps calendaire sont conditionnels puisqu'un joueur est à risque de subir une nouvelle blessure qu'à condition d'en avoir subi une antérieure. A contrario, le modèle en temps total est défini comme marginal puisque l'individu est à risque dès le début de la période et pour toutes les blessures indépendamment de toute blessure antérieure. Cette définition est contre intuitive pour le cas des évènements récurrents et ne semble pas adaptée<sup>101</sup>. L'intervalle jusqu'à la première blessure est identique quel que soit la formulation choisie.

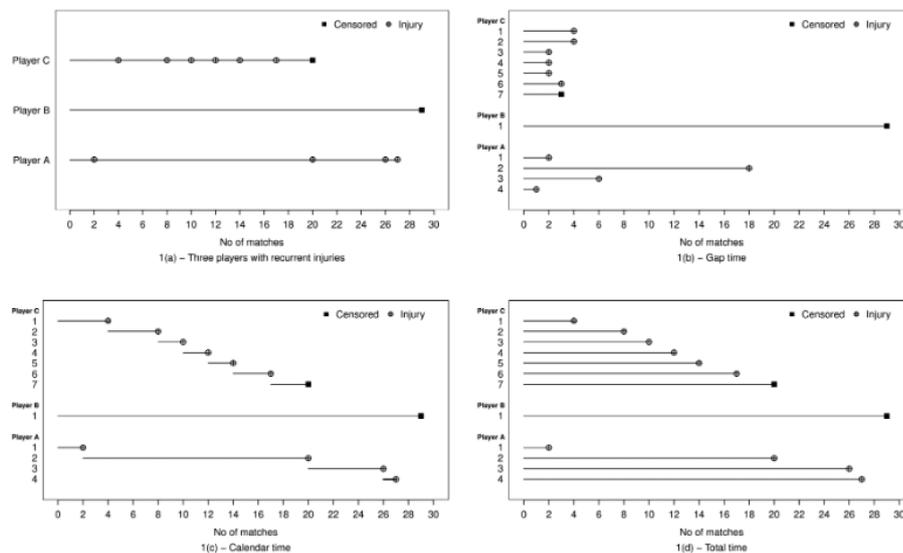


Figure 9 : Illustrations des formulations de l'intervalle de risque : (A) trois joueurs présentant des blessures récurrentes ; (B) gap time ; (C) temps calendaire ; (D) temps total. Un cercle indique un évènement de blessure et un carré plein indique une censure. Chaque temps jusqu'à un évènement ou une censure est un intervalle de risque distinct<sup>102</sup>.

Dans la littérature, différentes approches ont été appréhendées. Rizi et al.<sup>103</sup> ont utilisés un modèle de Cox en modélisant le temps d'exposition jusqu'à la première blessure à partir du temps total de pratique. Ils ont donc pris en compte à la fois le temps d'exposition à l'entraînement et en match. Williams et al.<sup>6</sup> ont défini la même unité de temps à la différence

près que c'était le temps de pratique de l'équipe à l'entraînement qui était considéré, probablement dû à une difficulté de récupérer la durée d'entraînement individuel. Cela permet ainsi de considérer toutes les périodes où le joueur est à risque mais pas de tenir compte du fait que le risque est bien plus important en match qu'à l'entraînement. D'autres ont fait le choix de considérer uniquement les matchs en temps de base. Ainsi Gabett et al.<sup>41</sup>, sur une cohorte de 66 joueurs de rugby à treize pendant 3 saisons ont utilisé le nombre de matchs par saison auxquels les joueurs ont participé pour définir leur temps d'exposition. Chaque saison a été traitée comme une période d'observation indépendante. Enfin Ullah et al.<sup>102</sup> ont pris le numéro du match comme temps de base sur lequel ils ont ensuite appliqués différents modèles de durée. Dans ce cas, le temps de base est commun à tous les joueurs malgré qu'ils ne soient pas tous à risque sur chacun des matchs s'ils n'y ont pas participé. Dans cet article qui avait pour but de comparer plusieurs méthodes, ils ont raisonné en temps calendaire pour les méthodes Anderson-Gill et le modèle de fragilité, en temps total sur le modèle Wei-Lin-Weissfeld et en gap time pour l'approche Prentice-Williams-Petersen. Toutes ces méthodes vont être présentées dans la section suivante.

### 2.3.2 Modèle de Cox

Une des méthodologies les plus classique en modèle de durée est le modèle à risques proportionnels de Cox. Il permet d'exprimer la fonction de risque instantanée de blessure en fonction de différentes covariables et du temps :

$$\lambda_i(t) = \lambda_0(t)e^{\beta^T X_i}$$

avec  $\lambda_0(t)$  la fonction de risque de base au temps  $t$ . Une hypothèse importante de ce modèle est que l'association entre chaque covariable et l'évènement est constante au cours du temps (hypothèse des risques proportionnels).

Cette technique a été utilisée pour identifier d'éventuels prédicteurs de blessures graves (>28 jours) en Rugby à 7<sup>103</sup>. Cette étude a été réalisée sur une cohorte de 104 joueurs de Rugby de 3 universités pendant une saison ; 31 blessures ont eu lieu durant cette période dont 9 blessures graves. Les auteurs ont ainsi pu montrer que les joueurs plus lents, moins agiles et ceux présentant une capacité d'extension des fléchisseurs de hanche faible, étaient des sujets plus enclins à subir une blessure sévère.

Une limite importante à ce type d'analyse est qu'elle ne considère que le temps jusqu'au premier évènement. Dans un contexte comme celui des blessures qui sont des évènements récurrents, d'autres approches plus adaptées peuvent être employées<sup>104,105</sup>.

### 2.3.3 Modèle de durée pour évènements récurrents

Andersen et Gill<sup>106</sup> ont proposé une méthode qui généralise le modèle à risques proportionnels de Cox en utilisant un processus de comptage pour permettre l'analyse de données récurrentes. Elle est basée sur l'hypothèse que le risque instantané de subir un évènement au moment  $t$  depuis l'entrée dans l'étude reste identique, que des évènements antérieurs se soient produits ou non. Ce modèle ne tient donc pas compte de l'ordre des évènements, ce qui fait que chaque sujet est à risque pour tout évènement tant qu'il est sous observation au moment de l'évènement. La méthode AG utilise donc une fonction de risque de base commune pour tous les évènements et estime un paramètre global pour les facteurs d'intérêt. La fonction de risque est définie par :

$$\lambda_i(t) = \lambda_0(t)e^{\beta_k^T X_{ik}}$$

où  $k$  est le  $k^{\text{ème}}$  évènement. Cependant une des hypothèses de ce modèle est que les évènements sont tous indépendants, et donc qu'après ajustement sur les covariables il n'y a pas d'hétérogénéité du risque de blessures entre les sujets. Ce modèle peut être approprié lorsque les corrélations entre les évènements de chaque individu sont induites par des covariables mesurées, comme par exemple le nombre de blessures précédentes<sup>107,105,108</sup>.

Une autre possibilité pour analyser les évènements récurrents est le modèle Prentice-Williams-Peterson (PWP)<sup>109</sup>. Les modèles PWP sont une version stratifiée du modèle AG, la stratification est faite selon le nombre d'évènements précédents. Ces modèles peuvent être spécifiés en temps calendaire ou en gap time. Dans le premier cas le risque de base varie d'un évènement à l'autre. La fonction de risque pour l'individu  $i$  et le  $k^{\text{ème}}$  évènement est :

$$\lambda_{ik}(t) = \lambda_{0k}(t)e^{\beta_k^T X_{ik}}$$

En gap time elle est défini par :

$$\lambda_{ik}(t - t_{k-1}) = \lambda_{0k}(t - t_{k-1})e^{\beta_k^T X_{ik}}$$

où  $\lambda_{0k}$  représente la fonction de risque de base spécifique à l'évènement  $k$  au temps  $t$ . Dans les deux cas ce sont des modèles conditionnels car un individu n'est pas considéré à risque pour le  $k^{\text{ème}}$  évènement tant qu'il n'a pas subi le  $k - 1^{\text{ème}}$ . Contrairement au modèle AG, l'effet des covariables peut varier d'un évènement à l'autre dans les modèles PWP. Par conséquent, les modèles PWP pourraient être préférables au modèle AG lorsque le risque de base et les effets des covariables sont différents en fonction du nombre de récurrences. Conditionnellement aux variables explicatives, ce modèle suppose qu'il n'existe pas d'hétérogénéité du risque entre les sujets.

Wei-Lin-Weissfeld (WLW)<sup>110</sup> ont proposé un modèle de survie marginal estimé par GEE adapté au modèle de Cox. Cette méthode consiste à traiter la dépendance des temps de survie comme une nuisance en spécifiant la fonction de risque marginale des temps de survie sans modéliser la structure de dépendance entre les temps de survie. L'idée principale du WLW est de restreindre l'analyse des événements récurrents à  $K$  événements, et d'appliquer le modèle des risques proportionnels de Cox à  $K$  événements distincts. La fonction de risque est identique au modèle PWP en calendar time :

$$\lambda_{ik}(t) = \lambda_{0k}(t)e^{\beta_k^T X_{ik}}$$

Un individu est à risque pour chaque évènement tant qu'il est sous observation et donc il est à risque pour un évènement ultérieur même si aucun évènement précédent ne s'est produit. Si aucune covariable dépendante du temps n'est incluse dans le modèle AG pour tenir compte du nombre d'évènements passés, les estimations du modèle AG et WLW seront identiques<sup>105</sup>. Néanmoins, le modèle WLW tient compte de la corrélation intra-sujet en ajustant la variance via un estimateur « sandwich ». Cette méthode propose ainsi une estimation robuste de la matrice de covariance<sup>110</sup>. Les estimations sont également non biaisées même si l'hypothèse de travail d'indépendance n'est pas valide<sup>111</sup>. Un des inconvénients est également de ne pas considérer tous les évènements étant donné que si  $K$  est pris comme le nombre maximum d'évènements rencontrés par un individu, tous les autres seront à risque pour les  $K$  évènements même s'ils en ont connu peu ce qui entrainera des estimations moins précises<sup>112,113</sup>.

Enfin une dernière possibilité est d'opter pour un modèle de fragilité (frailty model)<sup>114</sup>. Le modèle de fragilité est une extension du modèle de Cox dans lequel on ajoute une variable de fragilité ou effet aléatoire, de manière comparable aux modèles mixtes. Ce terme de fragilité permet de prendre en compte qu'au-delà des covariables considérées, il peut subsister une hétérogénéité du risque selon le joueur de par des variables non mesurées. Dans le cas

d'évènements récurrents le modèle à fragilités partagées est pertinent du fait que chaque observation d'un même individu partagent la même fragilité<sup>115</sup>. Les modèles de fragilités peuvent être définis en calendar time ou en gap time<sup>101</sup>. La fonction de risque en temps calendaire est définie par :

$$\lambda_i(t) = \lambda_0(t)U_i e^{\beta_k^T X_{ik}}$$

avec  $U_i$  le terme de fragilité de l'individu  $i$ . La fonction de risque pour un modèle de fragilité en gap time est définie par :

$$\lambda_i(t - t_{ik-1}) = \lambda_0(t - t_{ik-1})U_i e^{\beta_k^T X_{ik}}$$

Le choix le plus classique de la distribution des effets aléatoires est la distribution gamma qui permet une simplification de l'écriture de la vraisemblance. D'autres distributions peuvent être utilisées pour l'estimation de la fragilité comme la distribution normale, la distribution log-normale et la distribution uniforme. Plusieurs publications ont montré que si la fragilité est présente dans les données mais ignorée, les effets des covariables sont sous estimés<sup>116,117</sup>.

Les modèles de fragilités partagées sont cependant limités qu'à un seul niveau de regroupement. De ce fait, des modèles de fragilités emboîtés ou nested frailty models ont été développés pour permettre de modéliser des données de survies hiérarchiques à plusieurs niveaux<sup>118</sup>. Pour un modèle hiérarchique à deux niveaux de regroupement, en considérant la  $k^{\text{ème}}$  observation du sous-groupe  $i$  et du groupe  $j$ , la fonction de risque peut ainsi s'écrire :

$$\lambda_{ijk}(t) = \lambda_0(t)u_i w_{ij} e^{\beta_k^T X_{ijk}}$$

avec  $u_i$  et  $w_{ij}$  les deux effets aléatoires emboîtés. Ce modèle peut être intéressant dans le cas où l'on souhaite modéliser les données de plusieurs clubs, dans ce cas nous pouvons ainsi inclure en plus de l'effet sur le joueur, un effet équipe.

Une publication de Ullah, Gabbett et Finch en 2012<sup>102</sup> a eu pour objectif de comparer l'utilisation de ces différents modèles de durée sur des évènements récurrents. Pendant une saison de Rugby à treize (29 matchs) et spécifiquement sur les blessures avec contact, ils ont testé un modèle de Cox à risques proportionnels, un modèle d'Andersen-Gill, un modèle de fragilité, un modèle WLW sur temps total et un modèle PWP en gap-time. Le modèle de Cox a démontré un meilleur ajustement pour les premiers matchs. Néanmoins il entraîne une perte d'information étant donné qu'il ne considère que la première blessure, c'est une limite

importante dans ce cadre d'évènement récurrents. Les modèle AG et le modèle de fragilité ont montré un meilleur ajustement sur ces données. Néanmoins, comme le modèle de fragilité permet d'inclure des variables explicatives spécifiques à chaque période à risque contrairement au modèle AG, les auteurs recommandent de le privilégier. Kelly et Lim préconisent également d'utiliser un modèle de fragilité par rapport à la méthode PWP dans le cas où la corrélation intra-joueur est significative<sup>112</sup>.

Un modèle de fragilité a été appliquée dans une nouvelle publication de Gabbett, Ullah et Finch en 2012<sup>41</sup> également sur 3 saison d'un club de Rugby à treize. L'objectif de cette étude est d'identifier des facteurs de risque de blessures parmi des données de tests physiques réalisés au début de chacune des saisons. Seules les blessures sur contact en matchs étaient conservées et l'exposition a été définie par le nombre de matchs. Les blessures récurrentes n'ont été prises en compte qu'au cours d'une saison par individu et chaque saison a été traitée comme une période d'observation indépendante dans le modèle de fragilité.

Dans l'étude de Williams et al. en 2017<sup>6</sup>, c'est un modèle de fragilité niché qui a été utilisé. Le modèle de fragilité niché permet donc d'inclure deux effets aléatoires pour décrire la structure hiérarchique des données<sup>118</sup>. Pour prendre en compte les corrélations intra-groupes, un facteur aléatoire sur l'équipe et un deuxième sur le joueur ont été inclus dans le modèle. Le modèle a été ajusté sur le poste, l'âge, le poids, la taille et l'historique de blessures. Les variables d'intérêt ajoutés dans le modèle sont la charge cumulée depuis le précédent mois et la précédente année ainsi qu'une interaction des deux avec la charge de la dernière année discrétisée en quartiles. Les résultats ont été citée dans une section précédente (cf. section 1.3).

## **2.4 Modèles de prédiction**

Au-delà d'identifier les facteurs pouvant prédisposer un joueur à une blessure, la prédiction du risque individuel est un challenge important dans le quotidien d'un club. Une nouvelle fois plusieurs difficultés peuvent survenir dans cet objectif. Plusieurs points pourront en effet être limitant comme le grand nombre de prédicteurs potentiels, possiblement corrélés entre eux, des évènements rares notamment à l'entraînement, ou encore l'aspect plus aléatoire des blessures dues au contact.

Les arbres de décisions sont parmi les méthodes les plus courantes en data mining et sont largement utilisés dans de multiples disciplines<sup>119</sup>. Ils consistent en une série d'étapes

séquentielles où l'objectif est de produire des groupes d'individus les plus homogènes possibles de la variable à prédire. A chaque itération, l'algorithme sépare les individus en k groupes selon la variable qui permettra la meilleure séparation. Si nous sommes dans le cas d'un arbre de classification, l'algorithme cherchera à minimiser une mesure d'impureté (généralement entropie de Shannon ou indice de Gini)<sup>120,121</sup> et si nous sommes dans le cas d'un arbre de régression on cherchera à maximiser la variance inter-classe. Il n'est cependant pas recommandé d'avoir un arbre trop détaillé qui aura des feuilles permettant une séparation parfaite de la variable réponse entraînant un risque de sur-apprentissage<sup>122</sup>. Pour cela l'arbre sera élagué, selon un principe de parcimonie ; il est nécessaire de réaliser un arbitrage entre performance et complexité de l'arbre. Cette méthode simple, utilisée seule peut cependant donner des résultats biaisés et avec une variance importante. C'est pourquoi on utilise généralement des méthodes dites d'ensemble pour combiner plusieurs arbres. Le principe du modèle d'ensemble est qu'un groupe d'apprenants faibles se rassemblent pour former un apprenant fort.

Une des méthodes les plus communes est le bagging. L'agrégation bagging (ou bootstrap aggregating) permet de réduire la variance et donc de rendre les prédictions plus robustes. Le bagging consiste à sous-échantillonner les données, entraîner le modèle sur chaque sous-échantillon pour ensuite combiner les résultats. C'est le principe des forêts aléatoires (Random Forest), proposée par Breiman et Cutler en 2001<sup>123</sup>, qui ajoute en plus de sous-échantillonner les données, une sélection aléatoire des variables à chaque nœud de chaque arbre

Le boosting est une autre technique d'ensemble pour créer une collection de prédicteurs<sup>124</sup>. Cette technique est cette fois séquentielle, tous les arbres construits ne sont pas indépendants étant donné que chaque modèle va chercher à corriger les erreurs du précédent. Pour cela, un poids plus important est attribué aux individus pour lequel la valeur a été mal prédite par le modèle précédent pour mieux les prédire par la suite. Une extension du boosting est le gradient boosting<sup>125</sup>. Cet algorithme utilise le gradient de la fonction de perte pour le calcul des poids des individus lors de la construction de chaque nouveau modèle. Cependant le gradient boosting a tendance à provoquer du sur-apprentissage. Plusieurs hyperparamètres comme le nombre d'arbres, la profondeur des arbres, le nombre d'observations minimum dans un nœud terminal ou encore le minimum d'amélioration apporté par les séparations, peuvent être ajustés. Cet algorithme et notamment l'extrême gradient boosting (XGBoost), permettant

une utilisation optimisée du gradient boosting, sont des solutions très populaires pour la prédiction<sup>126</sup>.

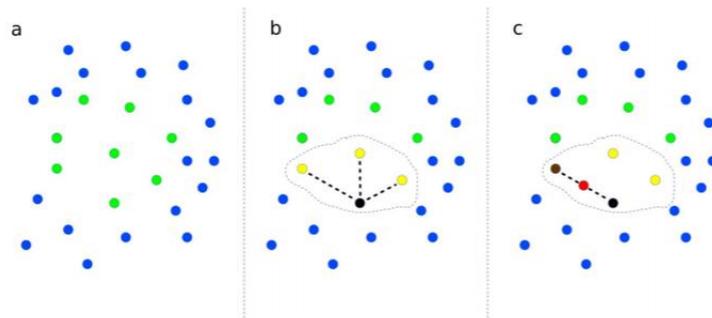
Bien que la majorité des méthodes de bagging ou boosting s'effectuent sur des arbres de décisions elles peuvent également être mise en place sur d'autres classifieurs comme la méthode SVM (Support Vecteur Machine)<sup>127</sup>. C'est également une méthode d'apprentissage automatique supervisée qui a pour objectif de séparer au mieux les données par un hyperplan pour obtenir la marge maximale entre les différents groupes de données. L'idée étant de transformer l'espace de représentation des données en un espace de plus grande dimension, pour pouvoir traiter des données difficiles à séparer linéairement.

Enfin une des dernières méthodes de machine learning parmi les plus classiques est l'utilisation de réseaux de neurones artificiels<sup>128</sup>. Le principe se veut semblable à celui du cerveau humain. Il y a trois types de couches, une couche d'entrée, plusieurs couches cachées et une couche de sortie. Chaque neurone va avoir une valeur particulière qui lui sera transmise par des neurones de la couche précédente et qui transformera cette valeur via une fonction d'activation pour ensuite l'assigner à un ou plusieurs autres neurones. Chaque neurone sera ainsi connecté par des liens pondérés avec d'autres neurones. Il existe plusieurs variations aux réseaux de neurones, définissant leur structure ou encore leur profondeur.

Dans le cas d'étude des blessures en sport, la proportion d'observations correspondant à une blessure est souvent beaucoup plus faible que celle des non blessés, on dit dans ce cas qu'on est dans le cas d'un événement rare et donc que les données sont déséquilibrées. Cela peut être une limite pour une bonne prédiction des blessures. Des méthodes d'échantillonnages peuvent permettre de résoudre ce problème.

Le sous-échantillonnage (undersampling) consiste à diminuer le nombre d'observations de la classe majoritaire pour rééquilibrer le jeu de données. Il peut être aléatoire ou en fonction d'algorithmes (Tomek links, cluster centroid). A l'inverse, le sur-échantillonnage (oversampling) va générer de nouvelles observations de la classe minoritaire. Une nouvelle fois, il peut être réalisé aléatoirement ou via des méthodes plus avancées comme SMOTE ou ADASYN. L'algorithme SMOTE pour Synthetic Minority Over-sampling Technique, a été développé par Chawla et al. en 2002<sup>129</sup>. Il consiste à identifier pour chaque point de la classe minoritaire ses k plus proches voisins parmi les autres points de la classe minoritaire. Il choisit ensuite aléatoirement un des points de ces k plus proches voisins, et génère un nouveau point sur le segment qui relie le point à celui choisi aléatoirement. ADASYN (Adaptive synthetic)

fonctionne sous le même principe que SMOTE mais avec un nombre d'observations qui est inversement proportionnel à la densité des observations de classe minoritaire. Il génère plus d'observations dans les régions où la classe minoritaire est faible.



**Figure 2.** Graphical representation of the SMOTE algorithm. (a) SMOTE starts from a set of positive (green points) and negative (blue points) examples; (b) It then selects a positive example (black) and its  $k$  nearest neighbors among the positives (yellow points, with  $k=3$ ). (c) Finally one of the  $k$  nearest neighbours is randomly selected (brown point) and a new synthetic positive example is added, by randomly generating an example (red point) along the straight line that connects the black and brown points. The procedure depicted in (b,c) is repeated for all the positives, by adding each time a new synthetic example similar (in an Euclidean sense) to the other positive examples.

Figure 9: Représentation graphique de l'algorithme SMOTE. Schubach et al., 2017<sup>130</sup>

Lorsque le nombre de prédicteurs est trop important il peut être préférable de les réduire pour améliorer l'interprétabilité et le pouvoir prédictif du modèle. Ainsi en prédiction, avoir des modèles parcimonieux, donc avec moins de variables explicatives permettra de réduire la variance. Des algorithmes de sélection pas à pas comme les méthodes backward, forward ou stepwise peuvent être utilisés pour minimiser des critères comme l'AIC, le BIC ou le Cp de Mallows. Une alternative pour des données à grande dimension est d'utiliser des méthodes de pénalisation qui permettront de créer un modèle de régression en incluant un terme de pénalisation. Une des méthodes de régression pénalisée est la méthode Lasso (Least Absolute Shrinkage and Selection Operator)<sup>131</sup>. Cette méthode utilise un critère de pénalisation par la norme  $L^1$  qui est la somme des valeurs absolues des coefficients. Le Lasso permet de forcer l'estimation de certains coefficients à 0 et peut ainsi également être considéré comme une méthode de sélection.

Quelle que soit la méthode utilisée pour construire un outil de prédiction, celui-ci doit ensuite être évalué, notamment en termes de sensibilité et spécificité. La sensibilité est le taux de blessés correctement prédits blessés et la spécificité le taux de non blessés correctement prédits non blessés. Ces deux indicateurs se calculent donc comme ceci :

$$\text{Sensibilité} = \frac{\text{Vrais positifs}}{\text{vrais positifs} + \text{faux négatifs}} \text{ et } \text{Spécificité} = \frac{\text{Vrais négatifs}}{\text{vrais négatifs} + \text{faux positifs}}$$

Cette classification est dépendante du seuil fixé pour déterminer à partir de quelle prédiction on considère qu'une observation a l'évènement. Ainsi la courbe ROC (Receiver Operating

Characteristic) permet de représenter l'évolution simultanée de la sensibilité et d'un moins la spécificité (taux de faux positifs) quand le seuil varie. L'aire sous la courbe ROC (AUC) est calculée afin de donner une mesure quantitative de la qualité de prédiction du modèle. La valeur de l'AUC est située entre 0.5 et 1 avec 1 un pouvoir prédictif parfait. Elle peut être interprétée comme la probabilité que, pour deux individus sélectionnés aléatoirement dans chacune des classes, le classifieur ait une valeur plus importante pour celui ayant réellement l'évènement par rapport à l'autre. L'AUC permet ainsi d'évaluer et de comparer différents modèles pour prédire des réponses binaires comme les blessures.

Van Eetvelde et al.<sup>132</sup> ont réalisé une revue de littérature de l'utilisation des méthodes de machine learning appliquées à la prédiction de blessure dans le sport. Une dizaine d'études sont ressorties de leur recherche dont une seule concerne le Rugby mais à treize<sup>40</sup>. Sur les 11 papiers conservés dans la revue, 9 utilisent des méthodes d'arbre de décisions, 4 les machines à vecteur support et 2 des réseaux de neurones artificiels. Concernant ceux sur les arbres de décisions, tous sauf un utilisent une stratégie de bagging avec pour la majorité des forêts aléatoires. Sur plusieurs études il y a une sélection préalable de variables avant d'appliquer l'algorithme d'apprentissage. Par exemple, Carey et al.<sup>133</sup> utilisent une analyse en composante principale pour réduire le nombre de variables. Rodas et al.<sup>134</sup> ont utilisé une méthode Lasso pour sélectionner des variables significatives. Enfin pour rééquilibrer les proportions de blessés/non blessés, 6 articles ont utilisé des méthodes de sous-échantillonnage aléatoire et/ou de sur-échantillonnage dont les méthodes SMOTE et ADASYN. La qualité méthodologique de ces études a été évaluée à partir de l'échelle NOS<sup>135</sup>. Toutes les études ont été considérées comme étant de qualité très faible à modérée. Les limites soulevées concernent notamment l'indépendance entre les jeux d'apprentissages et de tests, tout d'abord le fait que les joueurs soient présents dans ces deux échantillons ce qui remet en cause l'hypothèse d'indépendance. De plus, dans certaines publications il y a une phase de discrétisation/standardisation en amont du partitionnement ce qui fait donc rentrer les données de l'échantillon test dans le traitement des données d'apprentissage.

**Table 5** Study results characteristics

Authors, Year	Performance Measures (+ Values for Best ML Model)	Predictive Performance of ML Methods	Measures of Feature Importance	Most Important Injury Predictors
Ayala et al., 2019 [2]	AUC (0.873), Sensitivity (77.8%), Specificity (83.8%)	An alternating decision tree, combined with synthetic minority oversampling and boosting gave the best results	The frequency with which each of the features appears across the tree classifiers	Sleep Quality
Carey et al., 2018 [10]	(Median) AUC (all below 0.65), Sensitivity, Specificity, Precision, False Discovery Rate, Likelihood Ratios	The proposed ML models perform only marginally better than would be expected by random chance	NR	NR
López-Valenciano et al., 2018 [25]	AUC (0.747), Sensitivity (65.5%), Specificity (79.1%)	An alternating decision tree, combined with synthetic minority oversampling and boosting gave the best results	The frequency with which each of the features appears across the tree classifiers	sport devaluation, history of muscle injury in last season
McCullagh et al., 2013 [27]	Accuracy (82.9%), Sensitivity (94.5%), Specificity (81.1%)	Indication that Artificial Neural Networks are able to derive meaningful information from the vast amount of data available to assist in the injury prediction process	NR	NR
Oliver et al., 2020 [32]	AUC (0.663), Sensitivity (55.6%), Specificity (74.2%)	The machine learning model provided improved sensitivity to predict injury	The frequency with which each of the features appears across the tree classifiers	interactions of asymmetry, knee valgus angle and body size
Rodas et al., 2019 [34]	Accuracy (52%), Sensitivity (75%), Specificity (23%)	There is low prediction potential for presence or absence of tendinopathy	The number of times that a feature (genetic predictor) received a non-zero coefficient in the LASSO analysis	rs10477683 in the fibrillin 2 gene was the most robust SNP (single-nucleotide polymorphism)
Rommers et al., 2020 [35]	F1-score (85%), Sensitivity (85%), Precision (85%)	It is possible to predict injury with high accuracy	SHAP (SHapley Additive exPlanations) summary plot	Higher predicted age at PHV (peak height velocity), longer legs, higher body height, lower body fat percentage
Rossi et al., 2018 [36]	(Mean) AUC (0.76), F1-score (64%), Sensitivity (80%), Specificity (87%), Precision (50%), Negative Predicted Value (96%)	The single Decision tree performs best in terms of precision	Mean decrease in Gini coefficient	Previous injury (exponential weighted moving average), total distance (monotony of workload feature) and high-speed running distance (exponential weighted moving average)
Ruddy et al., 2018 [38]	(Median) AUC (0.58, 0.57 and 0.52)	Eccentric hamstring strength, age, and previous hamstring strain injury (HSI) data cannot be used to identify athletes at an increased risk of HSI with any consistency	NR	NR
Thornton et al., 2017 [41]	AUC (0.74, 0.65, 0.64 and 0.64)	Machine learning techniques can appropriately monitor injury risk amongst professional team sport athletes	Number of times that each feature appears in the ensemble of decision trees	The relative importance of each training load variable varied for each player
Whiteside et al., 2016 [46]	Accuracy (75%), Sensitivity (74%), Specificity (75%), Precision (75%), False Omission Rate (26%)	Machine learning models can predict future ulnar collateral ligament surgeries with high accuracy	The frequency with which each feature appeared in the optimized models in the fivefold cross-validation	Mean days between consecutive games, pitches in repertoire, mean pitch speed, horizontal release location

*Table 6: Machine learning methods in sport injury prediction and prevention: a systematic review Hans Van Eetvelde, Luciana D. Mendonça, Christophe Ley, Romain Seil and Thomas Tischer*

Le tableau 3 représente le résumé des études et leur qualité prédictive. Sur l'ensemble des articles, les AUC vont d'au minimum 0.52 jusqu'à 0.87. Les meilleures prédictions sont sur des blessures aux ischio-jambiers en football. Il est impossible à la vue de cette revue de littérature de distinguer si une méthode est plus performante que les autres, d'autant que les jeux de données, les sports et les blessures sont dissemblables. Cependant, Carey et al. ont réalisé une étude comparative de différents modèles de prédiction<sup>133</sup>. L'étude porte sur une cohorte de 75 joueurs de football australien pendant 3 saisons pendant lesquelles ont été récoltées différentes variables de charge de travail à partir des GPS et des questionnaires RPE. Les variables à modéliser étant les blessures aux ischio-jambiers, plus généralement les blessures sans contact et également le temps d'indisponibilité due à une blessure sans contact. Le jeu de données a été séparé en deux parties, les deux premières saisons servant à l'apprentissage et la troisième étant le test. Ils ont ainsi évalué les qualités de prédiction d'une régression logistique régularisée par la méthode elastic-net, d'une estimation par GEE, d'une forêt aléatoire et d'un modèle SVM. Ils ont testé de plus, différentes étapes de pré-traitement des données avec une analyse en composante principale, du sous-échantillonnage, l'algorithme SMOTE ou simultanément une ACP et une méthode SMOTE. Les capacités prédictives des modèles sont assez pauvres, les AUC sont situées entre 0.50 et 0.65. Seule la régression

logistique régularisée pour les blessures aux ischio-jambiers obtient un AUC à 0.76 après qu'une ACP ait été appliquée. Un autre travail de recherche a également eu pour but de déterminer la meilleure méthode de prédiction de blessures en football. Ils ont obtenu le meilleur résultat avec un modèle SVM agrégé par bagging à la suite d'une méthode de sur-échantillonnage.

Concernant l'étude sur le Rugby à treize de Thornthorn et al.<sup>40</sup>, les forêts aléatoires ont été appliquées à chaque joueur individuellement, l'objectif étant d'identifier les variables les plus prédictives de blessure pour chaque joueur. Des similitudes de variables ont pu être observées selon les groupes de postes bien qu'il y ait à chaque fois des individus faisant office d'exception. Pour chaque groupe de postes les AUC vont de 0.64 à 0.75 mais le fait d'avoir réalisé des modèles spécifiques à chaque individu entraîne un phénomène important de sur-apprentissage qui limite ces résultats.

Outre cet article, une seule étude, qui était par ailleurs une des premières sur la prédiction de blessure dans le sport, a été réalisée sur du Rugby mais une nouvelle fois à treize<sup>42</sup>. Dans cette publication de Gabbett en 2010, un modèle prédictif a été construit par une simple régression logistique basé sur la charge prévue et réalisée par semaine pour identifier des blessures musculaires sans contact. Les résultats de ce modèle semblent excellents en termes de prédiction étant donné que la spécificité et la sensibilité sont de 98.8% et 87.1% respectivement. Cependant plusieurs éléments interrogent. Tout d'abord il semblerait qu'il puisse y avoir un effet de causalité inversé qui puisse expliquer ces très bons résultats. En effet, une blessure impute la charge réalisée et c'est donc cette blessure qui va expliquer une faible charge par rapport à celle prévue et non l'inverse. De plus, aucune précision n'est donnée pour savoir exactement comment les données répétées ont été prises en compte. Enfin, la matrice de confusion présentée n'est pas cohérente avec les résultats énoncés.

Le challenge reste donc encore totalement ouvert pour tenter de déterminer un modèle de prédiction robuste en Rugby.



## **Chapitre 3 : Etude des facteurs de risque de commotions**

---

## 1) Résumé

Les commotions sont une source de préoccupation majeure dans le Rugby de par leurs conséquences<sup>2,3</sup>. D'autant qu'elles seraient en forte augmentation depuis une dizaine d'années<sup>4</sup>. Cependant, en France, l'évolution de ce risque reste inconnue. Bien qu'il existe des études sur les risques de commotions sur des phases spécifiques comme les plaquages<sup>29,35</sup>, il n'existe pas d'études multivariées des facteurs de risque de commotions et notamment en lien avec la charge de travail.

Cette étude devait permettre de déterminer l'incidence de commotion en France au plus haut niveau et son évolution sur ces dernières saisons. Elle avait surtout pour objectif de déterminer des facteurs de risque qui peuvent provoquer que certains joueurs sont plus enclins à subir une commotion en match. Pour cela nous nous sommes appuyés sur les données du programme de surveillance des blessures en que nous avons joint avec des données de fournisseur de données sportives et des données anthropométriques. Toutes ces données ont été collectées sur le championnat de France de Rugby (Top 14) pendant 5 saisons de 2013 à 2018. Les facteurs de risque considérés ont donc été la charge de travail représentée par le temps de jeu en match, le poids, la taille, l'âge ainsi que le poste du joueur.

Dans ce but, pour tenir compte des données répétées nous avons utilisé un modèle linéaire généralisé mixte avec le joueur en effet aléatoire. Nous avons privilégié les modèles mixtes à des modèles marginaux estimés par GEE par rapport à leur interprétation plus adaptée dans une optique explicative et parce que le nombre d'observations par joueur était variable. La distribution utilisée est celle de Poisson et le logarithme du temps d'exposition a été ajouté en offset. Ce choix de modélisation est adapté pour modéliser un taux d'évènement sur une durée d'exposition variable. Le modèle de Poisson à effets mixtes pour un individu  $i$  au match  $j$  est défini par :

$$P(Y_{ij} = y) = \frac{\lambda_{ij}^y e^{-\lambda_{ij}}}{y!}$$

et

$$\log(\lambda_{ij}) = X_{ij}^T \beta + \gamma_i + \log(\delta_{ij}) + \varepsilon_{ij} \text{ avec } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ et } \gamma_i \sim N(0, B)$$

où  $\gamma_i$  est l'effet aléatoire joueur et  $\delta_{ij}$  le temps de jeu. Un deuxième effet aléatoire sur l'équipe a été testé mais la variance de celui-ci n'était pas significative et son ajout induisait des problèmes de convergence donc nous ne l'avons pas conservé. Deux modèles supplémentaires ont été réalisés en fonction de la circonstance de survenue de la commotion (en plaquant ou sur un autre évènement de match). Dans chaque modèle, les observations où le joueur a subi une commotion sous une autre circonstance que celle étudiée ont été retirés de l'analyse.

Le modèle de poisson avec offset permet ainsi de tenir compte du temps de participation au match contrairement à un modèle logistique mixte. Un modèle de survie pour évènements récurrents aurait pu être envisagé comme Williams et al. Néanmoins le choix du temps de base en nombre de matchs ne permet pas de considérer le temps de jeu du joueur et, à l'inverse, en minutes cumulées ne permet pas de distinguer sa répartition en nombre de matchs.

L'incidence de commotion sur le championnat français pendant la période d'étude est de 10.4 [9.3,11.4] commotions pour 1000 heures d'exposition. La période de play-off en fin de saison où l'enjeu est plus important est plus à risque de commotions (OR = 3.96 [2.10,7.35] par rapport au premier tiers de saison). Certains profils ont également pu être mis en avant avec un risque plus important de commotions. Ainsi les joueurs de la charnière (demi de mêlée et demi d'ouverture) ont un risque plus élevé que les autres postes. De plus, toutes choses égales par ailleurs, les joueurs plus légers ont un risque plus important d'être commotionnés en plaquant et les plus grands dans les autres circonstances. Enfin il existe un effet protecteur du temps de jeu depuis le début de la saison. Sur ce dernier point une explication plausible de ce résultat est que les joueurs habitués à jouer au haut niveau depuis le début de la saison ont probablement une meilleure technique pour diminuer le risque.

## 2) Article

### **Incidence and risk factors in concussion events: A 5-season study in the French Top 14 Rugby Union Championship**

Jérémy Chéradame<sup>1</sup>, Julien Piscione<sup>1</sup>, Christopher Carling<sup>2</sup>, Jean-Pierre Guinoiseau<sup>2</sup>, Bernard Dufour<sup>3</sup>, Hélène Jacqmin-Gadda<sup>4\*</sup>, Philippe Decq<sup>5\*</sup>

#### **Author affiliation:**

<sup>1</sup> Research Department, French Rugby Union Federation, Marcoussis, France

<sup>2</sup> Research Department, French Football Federation, Clairefontaine, France<sup>3</sup> Medical Department, French Rugby Union Federation, Marcoussis, France

<sup>4</sup> Medical Department, French National Rugby League, Paris, France

<sup>5</sup> Univ. Bordeaux, Inserm, Bordeaux Population Health U1219, ISPED, Bordeaux, France

<sup>6</sup> Institut de Biomécanique humaine Georges Charpak, Arts et Métiers ParisTech et Université de Paris, assistance Publique hôpitaux de paris, hôpital Beaujon, neurochirurgie, Clichy, France

\* Equally contributed

Correspondence to: Jérémy Chéradame, Research Department, French Rugby Union Federation, Marcoussis, France; Jeremy.cheradame@ffr.fr

#### **ABSTRACT**

**Background:** Concussions are a source of major concern in Rugby and a limited number of studies have attempted to identify risk factors for these injuries.

**Purpose:** This study investigated the incidence of match concussion and associated risk factors including individual workload, anthropometric variables, playing position, and season phase in elite rugby union players.

**Study design:** Cohort study; Level of evidence 2b

**Methods:** All concussions and explanatory variables were collected for every match over 5 consecutive seasons (2014-2018) in 1334 professional players participating in the French Top 14 championship. Concussion risk was estimated using mixed effects Poisson regression.

**Results:** Mean match concussion incidence equated to 10.4 with 95% confidence interval [9.3;11.5] concussions for 1000 hours of game exposure. A peak was reached in the 2016/2017 season (13.7 [11.0;16.5]). A greater risk was observed in the playoffs compared to the first phase of the season (IRR= 3.96 [2.10,7.35]). In comparison to other positions, half-backs incurred the highest rate of concussion events (incidence: 16.1 [11.8;20.3]). Irrespective of playing position, those with greater height and lower body mass reported a higher risk of concussions (p=0.02), especially during tackling actions for lighter players (p=0.01) and during other match events for taller players (p=0.03). When adjusted for season phase, players who had accumulated a higher amount of playing time since the beginning of the season demonstrated a lower risk of concussion (p=0.005).

**Conclusions:** Inter and intra-seasonal variations in concussion rates were observed. Within positional groups lighter and taller players were more at risk with the highest incidence generally observed in half-backs. Workload as measured by the number of matches played prior to a concussion event appeared to have a protective rather than a deleterious effect on concussion risk.

**Keywords:** concussion, rugby, epidemiology, longitudinal study, incidence, multivariate analysis

## **INTRODUCTION**

The risk of concussion is of particular concern in full-contact collision team sports such as rugby union owing to its high incidence and deleterious consequences on player health and performance.[16,3] The risk and burden of concussion injuries observed at elite standards in

rugby union is also high. Indeed, occurrence rates have reportedly increased over the past decade especially in match-play conditions. In English Premiership players for example, an incidence of 3.0 to 6.0 concussion events per 1000 player hours was observed between seasons 2002/2003 and 2011/12. Rates started to rise from 2012/2013 onwards notably reaching a peak of 20.9 concussions per 1000 player hours during the 2016/2017 season.[15] Accordingly, research to determine the general risk of concussion in other equivalent standard elite rugby union populations and whether rates have also evolved over time is warranted.

In addition to the potential severe short- and long-term health consequences following concussion, an increased risk of sustaining another injury has also been observed[4,6,18] while players frequently report a recurrence of symptoms during return to play protocols.[6] It is therefore clearly of interest to identify factors, particularly those that are potentially controllable, that could influence the risk of sustaining concussion among elite rugby players. To the best of our knowledge, no study has focused on the possible association between workload and concussion occurrence in elite rugby union. Several studies have analyzed the relationship between workload and general injury risk with discrepancies observed across findings. A higher risk has been reported in players deemed to have played too much, not enough, or who experienced a sudden sharp change in workload.[8,13,21] Similar research exploring the association between concussion rates and match-play participation (e.g., total playing time, number of matches accumulated, in substitutes compared to starting players) as well as the effects of season phase is arguably warranted.

In rugby union, studies have been conducted in an attempt to increase understanding of the circumstances of concussion incidents during match-play. For example risk factors identified include the zone of the body impacted during collisions, the height of tackling actions, and tackling technique generally.[7,9,19,20] Also, while research has demonstrated substantial differences in injury risk generally across individual playing positions[2] there is limited

information on the risk of concussion associated to player position at elite standards. Similarly, investigations examining the potential effects of player age, height and weight on concussion rates are lacking in elite rugby populations.

Studies investigating the risk of injury and concussion in rugby union tend to utilize univariate analyses to individually examine different risk factors. Potentially confounding effects that might bias results are therefore neglected. Accounting for the associations among potential risk factors in a multivariate analysis is essential to reduce such biases. Yet, even an incidence of 20.9 concussions per 1000 hours of play [15] only equates to less than a 3% risk of concussion per player-match, thus it is difficult to collect sufficiently large datasets to explore and aid understanding of the different factors that might increase the risk of concussion in rugby players. Accordingly, longitudinal research programs including multiple clubs are necessary.

In 2013, a prospective longitudinal injury surveillance program in France's Top 14 elite rugby union Championship (the highest professional league in France) was implemented. As a key part of this program, the aim of the present study was to determine the rate of concussion in this Championship, assess its evolution over time (5 seasons), and identify potential risk factors associated with concussion incidents using multivariate analyses.

## **METHODS**

### **Participants**

Data were collected during a prospective longitudinal study including all teams competing in the elite French rugby union championship (Top 14) over five seasons between 2013/14 and 2017/18. Owing to a promotion and relegation system in this championship, data were gathered

in 14 teams per season with a total of 18 different participating teams over the whole period. A total of 1334 professional rugby union players and 935 official Top14 matches were included in the analysis.

### **Data collection**

Injury data were prospectively collected via an injury surveillance program. This program involved observation by trained independent observers who viewed Top14 match recordings and identified potential concussion incidents. Thereafter, a clinical technician created video clips of these incidents which were subsequently sent to each team's physician who confirmed whether or not a concussion episode had occurred. Team physicians could also report concussion episodes not identified by the match observers.

The playing events leading to these injuries were retrospectively added using digital video match footage. Additional data were collected via a commercial sport data provider (Opta®) to obtain information on playing position and player participation time from all 1734 official matches (Top14, European championships, international matches) over the entire duration of the study. While information on injuries incurring in European club competitions and international matches was not accessible, data on player participation in these was used to calculate the number of matches for each player as a measure of workload.

Anthropometric data (height and weight) for each player was provided by the National Rugby League. The data obtained for the first year of each player's participation in the study were used for the entire duration of the study. All data were anonymized.

The players provided their consent to participate in the study. Permission for the study was obtained from the institutional review board (CEERB PARIS NORD) while authorization was obtained from the French national data protection authorities (CCTIRS and CNIL).

## Statistical analysis

Player-season characteristics were compared according to playing position using analyses of variance with a random player effect. The incidence of concussion and associated confidence intervals were calculated for 1000 hours of game play. Incidence Rate Ratios (IRR) and 95% confidence intervals were calculated using a Poisson mixed model with a random player effect. The response variable was the number of observed concussions and the duration of exposure was taken into account by adding the logarithm of game time as an offset. Explanatory variables included time variables such as season and phase of the championship (divided into 3 roughly equivalent periods and analyzed independently from the final phase of the championship), player characteristics (age, weight and height, playing position), and workload. As the playing position of the substitutes was unknown to the present authors, the playing position variable was split into 4 categories based on shirt number: forwards (shirt numbers 1 to 8), half-backs (numbers 9 and 10), backs (numbers 11 to 15) and substitutes (numbers 16 to 23). To observe the effects of participating in a substitute role, the same model was created with only two categories: substitute and starting players. For workload, different variables, using game time only, with different temporality were built and tested to evaluate their association with the risk of concussion. All these variables were calculated for the workload preceding matches irrespective or not of whether a concussion event occurred. We tested the cumulated number of matches, and cumulated game time, over the previous one, two or three months or from the beginning of the season prior to the event. Using the method proposed by Williams *et al*[22], we computed the exponential weighted moving average (EWMA) playing time for a match over a period of 4, 8 and 12 weeks. This allowed more recent playing time to be given greater importance. All variables per minute were divided by 80 (match duration) to avoid scale issues and make parameter interpretation easier (one unit is equivalent to the duration of one match). Bonferroni correction was applied due to the multiplicity of these tests.

Two supplementary models were performed to distinguish risk factors associated with concussions incurred during tackling events and those in other circumstances. In each of these analyses, the observation (player-match) corresponding to concussions sustained in a match event other than that analyzed were removed from the data. These models were adjusted for the average number of tackles performed per 80 match minutes over the season for each player. Potential multicollinearity issues were verified using the variance inflation factor (VIF).[1] The VIF for all variables in the multivariable model was less than 2.7 demonstrating that there was no multicollinearity issue. The statistical program R was used for all these analyses and the package *lme4 v1.1-21* for the Poisson mixed model with 9 quadratures points and the Nelder\_Mead algorithm for the optimizer.

## **RESULTS**

### **Population description**

The characteristics of the cohort are described in Table 1. Over the 5 seasons, 1334 players participated, representing a sample of 3199 player-seasons. On average, players were  $27.4 \pm 4.5$  years of age, weighed  $102.4 \pm 14.2$  kg and measured  $186.1 \pm 7.7$  cm during their first season in the French Top 14 Championship. Values for age, height and particularly weight, significantly differed according to player position ( $p < 0.001$ ). Forwards were generally older, taller and heavier ( $28.1 \pm 4.5$  years,  $111.9 \pm 9.1$  kg and  $189.3 \pm 7.4$  cm) than backs ( $27.2 \pm 4.2$  years,  $93.0 \pm 8.4$  kg and  $183.8 \pm 5.3$  cm) and half-backs ( $26.8 \pm 4.4$  years,  $83.7 \pm 6.8$  kg and  $178.3 \pm 5.2$  cm).

TABLE 1  
Study Population by Playing Position at Player-Season Level<sup>a</sup>

	All	Forwards	Half-backs	Backs	Unknown <sup>b</sup>	P Value <sup>c</sup>
Players	1334	675	193	355	111	
Player-seasons	3199	1739	455	881	124	
Player information						
Age, y <sup>d</sup>	27.4 (4.5)	28.1 (4.5)	26.8 (4.4)	27.2 (4.2)	22.1 (3.2)	<.001
Weight, kg <sup>e</sup>	102.4 (14.2)	111.9 (9.1)	83.7 (6.8)	93.0 (8.4)	104.4 (14.1)	<.001
Height, cm <sup>e</sup>	186.1 (7.7)	189.3 (7.4)	178.3 (5.2)	183.8 (5.3)	185.4 (6.8)	<.001
Workload information						
No. of games	13.4 (7.3)	14.1 (7.0)	14.4 (7.6)	13.3 (7.0)	2.1 (2.2)	.02
Playing time	697.2 (486.6)	680.6 (449.0)	675.7 (472.7)	835.1 (514.2)	29.2 (34.2)	<.001

<sup>a</sup>Values are presented as No. or mean (SD).

<sup>b</sup>Playing position unknown.

<sup>c</sup>P value for the mixed effects analysis of variance for the interpositional comparison.

<sup>d</sup>At the beginning of the season.

<sup>e</sup>At the beginning of the study.

Table 7 Description of study population by playing position at player season level. Values are mean (SD).

The mean number of matches played in any one season per player was 13.4 (range 1 to 38 games) for a total of 697.2 minutes of play. Matches played ( $p=0.02$ ) and total playing time ( $p<0.001$ ) varied across positions with backs participating in fewer games overall than forwards and half-backs but reporting the highest total playing time.

## Concussion rates

A total of 388 concussion events was reported over the five-season study period. The overall incidence of concussion observed during this period was 10.4 concussions for 1000 hours of match exposure with 95% confidence interval [9.3;11.5]. Overall, 22% of all players sustained at least one match concussion.

### Concussion rates: inter-season evolution

Figure 1a reports an increase in rates during the final two seasons versus the first season with respective incidences of 13.7 [11.0,16.5] and 12.1 [9.5,14.6] concussions per 1000 hours of exposure in 2016/17 and 2017/18 compared to 7.9 [5.9,10.0] in 2013/14.

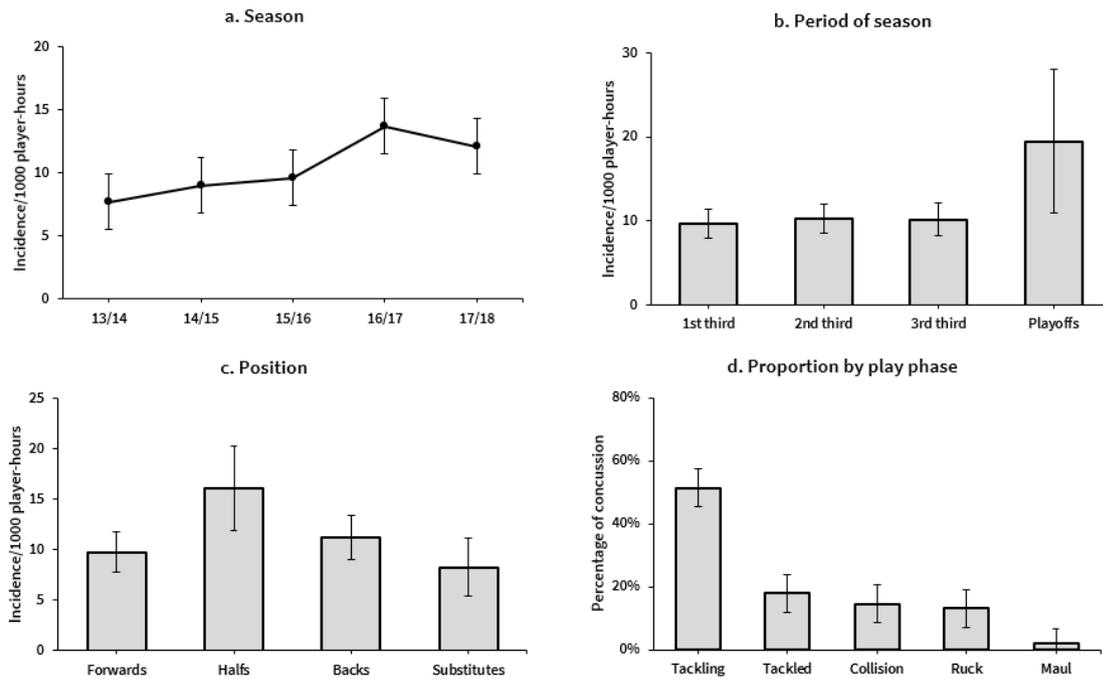


Figure 1 Incidence of concussion sustained according to season, seasonal period, player position and match event. Values are presented as incidence (95% CI).

Estimates of the Poisson mixed effects regression for the risk of concussion are presented in Table 2. After adjusting for season, number of matches played, player age, height, weight and position), the incidence risk ratios for the 2016/2017 (IRR=1.67, 95% confidence interval [1.20,2.31]) and 2017/2018 (IRR=1.46 [1.04,2.05]) seasons demonstrated a significantly greater risk of concussion compared to the first season (2013/14).

TABLE 2  
Adjusted IRR for the Risk of Concussion<sup>a</sup>

	Coefficient	SE	IRR (95% CI)	P Value
Intercept	-11.063	1.66		<.001
Season				
2013-2014	Reference			
2014-2015	0.191	0.18	1.13 (0.79-1.60)	.51
2015-2016	0.116	0.18	1.12 (0.79-1.60)	.52
2016-2017	0.511	0.17	1.67 (1.20-2.31)	.002
2017-2018	0.379	0.17	1.46 (1.04-2.05)	.03
Season part				
First part	Reference			
Second part	0.379	0.15	1.46 (1.09-1.94)	.01
Third part	0.534	0.19	1.71 (1.18-2.46)	.004
Playoffs	1.377	0.32	3.96 (2.10-7.35)	<.001
Position				
Half-backs	Reference			
Forwards	-0.361	0.23	0.70 (0.45-1.09)	.11
Backs	-0.407	0.18	0.67 (0.47-0.94)	.02
Substitutes	-0.641	0.24	0.53 (0.33-0.84)	.007
Age	-0.003	0.01	1.00 (0.97-1.02)	.81
Weight	-0.015	0.01	0.98 (0.97-1.00)	.02
Height	0.023	0.01	1.02 (1.00-1.04)	.02
Total playing time	-0.038	0.01	0.96 (0.93-0.99)	.005

<sup>a</sup>Multivariate analysis using mixed effects Poisson (n = 388 concussions). IRR, incidence rate ratio.

Table 8 Adjusted incidence risk ratio for the risk of concussion (multivariate analysis using Mixed effect poisson (n=388 concussions))

### Concussion rates: intra-season evolution

The incidence of concussions was highest during the playoffs of the Top 14 championship reaching 19.5 [10.9,28.1] per 1000h (Figure 1b). At multivariate level, concussion event occurrence in the second (IRR=1.46 [1.09,1.94]), the third (IRR= 1.79 [1.19,2.67]) and especially the playoffs of each season (IRR= 3.96 [2.10,7.35]) was significantly greater than in the first phase of the season (Table 2).

### Effect of accumulation of playing time

Table 3 presents the regression estimates for all the workload variables considered adjusting for the season phase and all the covariates presented in Table 2. The IRR computed for one standard deviation of each workload variable and the associated p-value showed that none of the three workload measures (number of games, total playing time or EWMA on playing time) computed

over the preceding 1-month period was associated with concussion risk in the current game. In contrast, the associations generally increased and became significant when the period of time considered increased; over the preceding two or three months. Following the Bonferoni correction, only the total playing time accumulated since the beginning of the season was significantly associated with the concussion risk in the current game. Irrespective of season phase, the total playing time since the beginning of the season was negatively related to the risk of sustaining concussion in the current game ( $p=0.005$ ).

TABLE 3  
Estimates of the Regression Coefficient for the Workload Variables in the Mixed Effects Poisson Regression<sup>a</sup>

	Coefficient	SE	IRR	P Value
Played in the past month				
No. of games	-0.058	0.05	0.94 (0.86-1.03)	.19
Game time	-0.083	0.05	0.92 (0.82-1.02)	.12
EWMA (n = 4)	-0.404	0.22	0.67 (0.43-1.03)	.07
Played in the past 2 mo				
No. of games	-0.051	0.03	0.95 (0.90-1.00)	.07
Game time	-0.078	0.03	0.92 (0.86-0.99)	.02
EWMA (n = 8)	-0.617	0.28	0.54 (0.31-0.94)	.03
Played in the past 3 mo				
No. of games	-0.044	0.02	0.96 (0.93-0.99)	.03
Game time	-0.063	0.03	0.94 (0.89-0.99)	.01
EWMA (n = 12)	-0.783	0.31	0.625 (0.27-0.97)	.009
Played since the beginning of the season				
No. of games	-0.030	0.011	0.96 (0.92-0.99)	.007
Game time	-0.038	0.013	0.97 (0.95-0.99)	.005

<sup>a</sup>Adjusted for all the covariates in Table 2. All variables were calculated per minute of match time (divided by 80). Significance threshold with Bonferroni correction:  $P < .005$ . EWMA, exponentially weighted moving average; IRR, incidence rate ratio.

Table 9 Estimates of the regression coefficient for the workload variables in the mixed effect poisson regression adjusting for all the covariates in table 2. All variables were calculated per minute match time (divided by 80).. Significance threshold with Bonferoni correction:  $p<0.005$ .

\* Exponentially weighted moving average

## Effect of player position

The half-backs reported the highest risk of concussion with an incidence of 16.1 [11.8,20.3] concussions for 1000 hours of play (Figure 1c). The risk of concussion was significantly lower in other back-line players compared to half-backs (IRR= 0.67 [0.47,0.94]). There was no significant difference between half-backs and forwards ( $p=0.11$ ). Overall, the IRR of concussion in starting players versus substitutes was 1.38 [0.99,1.91] ( $p=0.06$ ). Concussion risk

was significantly lower in substitutes compared to half-backs (IRR=0.53 [0.33,0.84],  $p=0.007$ ) but no differences were observed compared to forwards ( $p=0.13$ ) and backs ( $p=0.26$ ).

### **Effect of anthropometric characteristics**

Player weight affected the risk of concussion ( $p=0.02$ ) (Table 2). Heavier players were less affected; an increase of 10 kilograms decreased the risk by 14%. For players in the same position and with comparable weight, height was also a risk factor ( $p=0.02$ ). Taller players incurred more concussions and for an increase in height of 10 centimeters, the risk of concussion increased by 24%.

### **Effect of game events**

Across all players, the majority of concussions occurred during tackling phases (69.2%) with 51.3% of events incurred by the tackler, 17.9% by the player tackled, while 14.6% followed a ruck, 13.1% a collision between players and 2.1% a maul (Figure 1d).

For sub analyses dependent upon the two models used to explore rates according to the circumstances of the concussion (concussions sustained during tackling and those incurred in other circumstances), no differences between the two models analyzed were reported between seasons or in relation to player position and age (Table 4). The risk of concussion was greater in the playoffs in both models but the risk reported in the second and third phases of the season was significantly different to the first phase only for concussions sustained in circumstances other than tackling. Total playing time was a risk factor only in the second model ( $p=0.02$  versus  $p=0.15$ ). Lighter players reported a higher risk of concussions only in tackling events ( $p=0.01$

versus  $p=0.11$  in other circumstances). Conversely, taller players were more at risk of incurring concussion in other circumstances compared to in tackling events ( $p=0.03$  versus  $p=0.10$ ).

TABLE 4  
Estimates From the Mixed Effects Poisson Regression for the Risk of Concussions Sustained in Tackling Events and Other Circumstances<sup>a</sup>

	Concussion Sustained in Tackling Events (n = 169)				Concussion Sustained in Other Circumstances (n = 160)			
	Coefficient	SE	IRR (95% CI)	P Value	Coefficient	SE	IRR (95% CI)	P Value
Intercept	-11.993	2.43		<.001	-14.790	2.58		<.001
Mean No. of tackles	0.119	0.03	1.12 (1.06-1.19)	<.001	0.014	0.04	1.01 (0.94-1.09)	.70
Season								
2013-2014	Reference				Reference			
2014-2015	-0.002	0.28	0.98 (0.57-1.73)	.99	0.035	0.28	1.04 (0.60-1.78)	.89
2015-2016	0.183	0.27	1.20 (0.70-2.05)	.50	-0.033	0.28	0.97 (0.56-1.68)	.91
2016-2017	0.411	0.26	1.51 (0.91-2.51)	.11	0.462	0.25	1.59 (0.96-2.61)	.07
2017-2018	0.145	0.27	1.16 (0.68-1.98)	.60	0.207	0.28	1.23 (0.71-2.12)	.46
Season								
First part	Reference				Reference			
Second part	0.295	0.23	1.34 (0.87-2.07)	.19	0.575	0.23	1.78 (1.13-2.79)	.01
Third part	0.443	0.28	1.56 (0.90-2.70)	.12	0.707	0.29	2.03 (1.14-3.60)	.02
Playoffs	1.291	0.47	3.64 (1.44-9.21)	.006	1.679	0.48	5.36 (2.10-13.66)	<.001
Position								
Half-backs	Reference				Reference			
Forwards	-0.513	0.33	0.60 (0.31-1.15)	.12	-0.569	0.36	0.57 (0.28-1.15)	.11
Backs	-0.334	0.25	0.72 (0.44-1.18)	.19	-0.429	0.27	0.65 (0.39-1.10)	.11
Substitutes	-0.879	0.36	0.42 (0.20-0.84)	.02	-0.981	0.39	0.37 (0.17-0.80)	.01
Age	-0.012	0.02	0.99 (0.95-1.03)	.54	0.021	0.02	1.02 (0.98-1.06)	.32
Weight	-0.025	0.01	0.97 (0.96-0.99)	.01	-0.017	0.01	0.98 (0.96-1.00)	.11
Height	0.024	0.01	1.02 (0.99-1.05)	.10	0.034	0.02	1.03 (1.00-1.07)	.03
Total playing time	-0.030	0.02	0.97 (0.93-1.00)	.15	-0.047	0.02	0.95 (0.91-0.99)	.02

<sup>a</sup>IRR, incidence rate ratio.

Table 10 Estimates from the mixed effects poisson regression for the risk of concussions sustained in tackling events and other circumstances.

## DISCUSSION

In this study, concussion rates in French Top 14 rugby union players and risk factors potentially linked to these were investigated over a 5-season period. A mean concussion incidence of 10.4 [9.3,11.5] was observed overall but this increased substantially during the final two seasons (13.7 [11.0,16.5] in 2016/2017 and 12.1 [9.5,14.6] in 2017/2018). Over the course of the season, players were more at risk during the playoffs with a reported incidence of 19.5 [10.9,28.1] incidents per 1000h. However, the number of match minutes accumulated during the season seemingly had a protective effect against the risk of sustaining concussion. The risk and

circumstances of concussions were related to players' anthropometric characteristics, with greater incidences notably reported in lighter and taller players.

### **Concussion evolution**

In the present cohort of Top 14 players, a mean concussion rate of 10.4 [9.3,11.5] events per 1000h match-play was observed over a 5-season period from 2013/14 to 2017/18. This rate is lower than that reported (15.7) in the equivalent top-flight rugby Championship in England over the same period.[14] It is also lower in comparison to rates observed in international standard players participating in the last two Rugby World Cup tournaments (respectively 12.5 and 12.2 for the RWC 2015 and 2019) [11,12]. It is difficult to suggest reasonable explanations for discrepancies in rates reported across studies. Despite the present lower overall rate, is noteworthy that in this study concussion incidences increased from season to season and were highest during the final two Top 14 seasons investigated reaching a peak of 13.7 concussions per 1 000 player-hours in the penultimate season (2016/2017). This trend could be considered cause for concern yet the reported peak incidence is still substantially lower than the highest value observed (20.9) over a 5-season period in English Rugby Premiership and which also occurred during the 2016/2017 season.[15] Similarly, a single season study again conducted during the same season but in professional Irish players reported a concussion rate of 18.4 events per 1000 player hours.[5] Although improvements made in the assessment, identification and reporting of concussion events might partly explain the present season-to-season trend in elite French Rugby, it nevertheless reflects the constant duty of medical and coaching practitioners and game stakeholders to implement concussion risk reduction strategies to ensure contemporary player welfare.

## **Match participation**

Over the four phases of the championship season, a progressive increase in concussion rates was observed and particularly during the final knockout playoff phase where the highest incidence (19.5 [10.9;28.1]) was observed. A reasonable explanation could be the physical and mental fatigue accumulated over the course of the season. Research has shown that physical fatigue following an acute bout of physically fatiguing exercise negatively affected tackling technique in amateur rugby union players.[10] Similar work in elite players comparing tackling technique and the circumstances of concussion events across different phases of the season whilst accounting for individual match participation time would be pertinent to investigate the potential effects of fatigue accumulated over the season. Another explanation for this finding might be the increased stakes of matches later in the season and which are particularly high during the final knockout playoff phase. Indeed, the menace of relegation, elimination or alternatively qualification and the opportunity to be Championship winners could give rise to players adopting more risky behavior.

No association was observed between either the cumulated playing time or number of matches played over the previous month and concussion rates. It is also noteworthy that the risk progressively diminished as the length of time of the analysis preceding the current match increased and that players reporting higher cumulative playing time since the season start had a significantly lower risk. These findings suggest that players with greater accumulated playing time since the beginning of the season were not more likely to suffer a concussion. Although the present methodological design is different, these conclusions tend to disagree with those of Rafferty *et al*,[18] who reported that elite rugby players were at a greater risk of sustaining concussion than not following prolonged periods of exposure to match-play (e.g., 25 matches). Indeed, the current results infer that increased match-play participation might offer a protective effect against the risk of concussion injury. Related research has shown that professional players

involved in fewer than 15 matches and over 35 matches over the preceding 12-month period had an increased risk generally of injury whereas this was not the case for involvement in a 15-35 match participation window period.[21] Collectively these results suggest the importance of appropriate preparation and sufficient participation in match-play to develop physical fitness and robustness for the demands of elite competition. However, it is important to note that the aforementioned effect of the increased risk during the different phases of the season and the lesser risk related to playing time accumulated since the start of the season is only significant if these factors are adjusted to each other in the statistical model. These two opposite effects on the risk of concussion observed during over the course of the season must therefore be considered simultaneously otherwise the possible effect of one factor will be hidden by the effect of the other.

### **Individual risk factors**

Cross-sectional data in elite rugby populations regarding the effects of player age on concussion injury are scarce. Here, no effect of age on concussion incidence was reported which is in agreement with findings reported on general injury risk in players of a similar standard.[21] In contrast, position-specific differences in concussion rates were reported with a statistically higher risk notably observed in half-backs compared to other players in the back line while the rate was also higher than in forwards (albeit not significantly). In previous studies, authors tended to compare back versus forward positional groups only with the former demonstrating a higher, albeit non-significantly greater incidence of concussion than forwards.[14] As such, we recommend that in future epidemiological studies, concussion rates in players in the half-back positions are quantified separately and not as part of the back line as a whole. A reasonable explanation for this higher rate of concussion in half-backs could be their body mass values which were generally lower in comparison to the other playing positions although the risk in these players nevertheless increased even when adjusted for this factor. Indeed, across the

present cohort generally, lighter players, irrespective of playing position, were more at risk of sustaining a concussion injury and especially when tackling opponents. Additional research is required to scrutinize more closely the circumstances of these tackling actions (e.g., video analysis) as tackle technique is commonly recognized as a key risk factor in concussion occurrence [7,9,19,20]. Finally, it is also noteworthy that taller players reported significantly more concussions although these episodes tended to occur in game events other than during tackling actions. Again, we suggest a need for more detailed information on the circumstances of these incidents.

### **Strengths and limitations of the study**

This study quantified the risk of concussion in an elite senior rugby championship (French Top 14) in which no information is available to date. In addition, the longitudinal nature of this investigation examined all matches played over a 5-season period; corresponding to 43 010 player-matches, thereby providing a sufficiently large dataset for analyzing potential risk factors linked to concussion incidents. Indeed, methodological issues regarding sample size and study power are frequent in the current literature and this study partly fulfills these shortfalls (Prien et al., 2018).[17] This investigation also shows the importance of employing an adjusted risk analysis taking into account different factors in the same statistical model and not simply performing univariate analyses that may be subject to confounding biases.

A limitation of this study is the lack of information regarding concussions occurring in the present players in match-play out of Top 14 competition (European cups, international matches). Similarly, the incidence and potential impact of concussions sustained in training settings were not covered here although their occurrence is generally lower than in competition.[14]

Finally, in order to improve knowledge and understanding of potential risk factors and the mechanisms that can lead to concussions, information on the effects of physical loading would be pertinent. Here, information relating to accumulated playing time was used to determine the impact of workload. Data on internal (e.g., heart-rate or ratings of perceived exertion) and external (running activity) workloads in training and match settings should be included in future investigations.

## **CONCLUSION**

This longitudinal study has quantified the risk of sustaining concussion and identified associated risk factors in French Elite rugby players over five consecutive seasons. An overall incidence of concussion of 10.4 concussions per 1000 hours was observed with highest values reported in the last two seasons. Rates were highest during the matches played at the end of the championship, possibly due to the increased stakes of this knockout phase although a higher workload measured by accumulated playing time from the beginning of the season tended to reduce the risk. The majority of concussions occurred during a tackle event and particularly in the tackling player. Half-backs sustained more concussions than other positions in the backs. Irrespective of playing position, lighter and taller players reported a higher risk of concussions with these sustained in different match circumstances (tackling versus non tackling events). These results enable a greater understanding of the risk of concussions and associated risk factors in top-flight French rugby and in elite game generally and potentially could be used as a foundation to help develop and inform strategies to help reduce their occurrence. It would also be pertinent to reproduce this analysis in other championships to determine whether the same risk factors also exist whilst accounting for potential regional specificities in tactical and physical characteristics of match-play.

## REFERENCES

- 1 Akinwande M, Dikko H and Samson A. Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis. *Open J Stat.* 2015;5:754-7672
- Brooks J and Kemp S. Injury-prevention priorities according to playing position in professional rugby union players. *Br J Sports Med.* 2011;45:765–775. PMID: 20484316
- 3 Burger N, Abrahams S, Brown J *et al.* Long-term effects of concussion in rugby union versus other contact sports. [www.BokSmart.com](http://www.BokSmart.com).
- 4 Burman E, Lysholm J, Shahim P *et al.* Concussed athletes are more prone to injury both before and after their index concussion: a data base analysis of 699 concussed contact sports athletes. *BMJ Open Sport Exerc Med.* 2016;2:e000092. PMID: 27900166
- 5 Cosgrave M and Williams S. The epidemiology of concussion in professional rugby union in Ireland. *Phys Ther Sport.* 2018. 35. PMID: 30513491
- 6 Cross M, Kemp S, Smith A *et al.* Professional Rugby Union players have a 60% greater risk of time loss injury after concussion: a 2-season prospective study of clinical outcomes. *Br J Sports Med.* 2016;50:926–31. PMID: 26626266
- 7 Cross, MJ, Tucker R, Raftery M *et al.* Tackling concussion in professional rugby union: a case–control study of tackle-based risk factors and recommendations for primary prevention. *Br J Sports Med.* 2017;53:1021–1025. PMID: 29021244
- 8 Cross, MJ, Williams S, Trewartha G *et al.* The Influence of In-Season Training Loads on Injury Risk in Professional Rugby Union. *Int J Sports Physiol Perform.* 2016;11:350–355. PMID: 26309331

- 9 Davidow D, Quarrie K, Viljoen W *et al.* Tackle technique of rugby union players during head impact tackles compared to injury free tackles. *J Sci Med Sport.* 2018;21:1025–1031. PMID: 29803736
- 10 Davidow D, Redman M, Lambert M. *et al.* The Effect of Physical Fatigue on Tackling Technique in Rugby Union. *J Sci Med Sport.* PMID: 32359940
- 11 Fuller C, Taylor A, Kemp S *et al.* Rugby World Cup 2015: World Rugby injury surveillance study. *Br J Sports Med.* 2017;51:51–57. PMID: 27461882
- 12 Fuller C, Taylor A, Douglas M *et al.* Rugby World Cup 2019 injury surveillance study. *South African Journal of Sports Medicine.* 2020;32:1-6.
- 13 Gabbett T. The training-injury prevention paradox: should athletes be training smarter and harder? *Br J Sports Med.* 2016;50:273–280. PMID: 26758673
- 14 Gardner A, Iverson G, Williams W *et al.* A Systematic Review and Meta-Analysis of Concussion in Rugby Union. *Sports Med.* 2014;44:1717-1731. PMID: 25138311
- 15 Kemp S, West S, Brooks J, *et al.* England Professional Rugby Injury Surveillance Project 2017-2018 Season Report. [www.englandrugby.com/participation/playing/player-welfare-rugby-safe/rugbysafe-research](http://www.englandrugby.com/participation/playing/player-welfare-rugby-safe/rugbysafe-research) 2019.
- 16 Leddy J, Baker J, and Willer B. Active Rehabilitation of Concussion and Post-concussion Syndrome. *Phys Med Rehabil Clin N Am.* 2016;27:437-454. PMID: 27154855
- 17 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;48:953–969. PMID: 29349651

18 Rafferty J, Ranson G, Oatley G *et al.* On average, a professional rugby union player is more likely than not to sustain a concussion after 25 matches. *Br J Sports Med.* 2019;53:969–973. PMID: 29530941

19 Tierney GJ, Denvir K, Farrell G, *et al.* The Effect of Tackler Technique on Head Injury Assessment Risk in Elite Rugby Union. *Med Sci Sports Exerc.* 2018;50:603–608. PMID: 29049096

20 Tucker R, Raftery M, Kemp S *et al.* Risk factors for head injury events in professional rugby union: a video analysis of 464 head injury events to inform proposed injury prevention strategies. *Br J Sports Med.* 2017;51:1152–1157. PMID: 28642222

21 Williams S, Trewartha G, Kemp S *et al.* How Much Rugby is Too Much? A Seven-Season Prospective Cohort Study of Match Exposure and Injury Risk in Professional Rugby Union Players. *Sports Med.* 2017;47:2395–2402. PMID: 28361327

22 Williams S, West S, Cross MJ *et al.* Better way to determine the acute:chronic workload ratio? *Br J Sports Med.* 2017;51:209–210. PMID: 27650255

## **Chapitre 4 : Etude de l'impact de la charge de travail sur le risque de blessure**

---

## 1) Résumé

La gestion de la charge de travail est un paramètre important dans le sport pour avoir des joueurs performants en match et avec un risque de blessures réduit. De nombreux travaux ont exploré l'impact de cette charge de travail sur le risque de blessures. Néanmoins assez peu de publications, et avec des conclusions hétérogènes, ont étudié l'impact à court terme de la charge de travail sur le risque de blessures. Au-delà du volume, il n'y a pas d'information sur les conséquences du type de séance planifier au cours de la semaine d'entraînement en Rugby.

L'objectif de cette étude était de comprendre en quoi l'activité des joueurs à l'entraînement au cours des différents jours précédant le match impacte le risque de blessures le jour du match. Cette étude a été réalisée sur les données de 72 joueurs d'un club professionnel du championnat de France de Rugby lors des saisons 2017/2018 jusqu'à 2019/2020. Les données utilisées pour définir l'activité des joueurs sont la distance totale parcourue et la distance à haute intensité (supérieur à la VMA) mesurées à partir des GPS, ainsi que des données de questionnaires RPE pour suivre la difficulté des séances de terrain et de musculation.

Due à la présence de données manquantes, des méthodes d'imputations ont dues être employées. Nous avons utilisé l'approche d'imputation multiple par équation chaînée<sup>72</sup> qui est une méthode flexible et qui fournit des imputations non biaisées sous l'hypothèse de données manquantes aléatoires (MAR)<sup>136,137,138</sup>. L'approche Predictive Mean Matching (PMM) a été utilisée pour avoir une imputation robuste quelle que soit la distribution de la variable imputée<sup>87,88</sup>. Le modèle d'imputation devait tenir compte des mesures répétées par joueur mais également de la charge journalière qui a montré être un des meilleurs indicateurs d'imputation<sup>66</sup>. Les modèles avec deux effets aléatoires sur le joueur et la journée d'entraînement ne convergeaient pas étant donné qu'il y avait un nombre de jours d'entraînements trop élevé. Nous avons donc utilisé un modèle d'imputation à deux niveaux incluant un intercept aléatoire spécifique au joueur et la moyenne journalière par poste de la variable à imputer :

$$Y_{ij} = \beta_0 + \beta_1 \bar{Y}_j + \beta^T X_{ij} + \gamma_i + e_{ij} \text{ avec } e_{ij} \sim N(0, \sigma^2) \text{ et } \gamma_i \sim N(0, B)$$

Les facteurs de risque de blessure en match ont été analysés sur chaque échantillon de données imputées avec un modèle logistique à intercept aléatoire spécifique au sujet pour tenir compte de la corrélation intra-joueur. Afin de déterminer les facteurs d'ajustement, la relation

entre le risque de blessure et différents facteurs individuels (âge, poids, taille et poste) et contextuels (période de la saison, match à domicile ou à l'extérieur, retour de blessure) a d'abord été testé. Le nombre de semaines depuis le précédent match a également été testé mais il provoquait une difficulté d'interprétation par sa relation directe avec le retour de blessure. Bien qu'il soit significatif également, l'ajustement du modèle étant moins bon avec cette variable, nous avons fait le choix de ne pas l'inclure dans l'analyse.

Dans l'objectif de modéliser l'impact de la charge de travail de chaque jour précédent le match sur le risque de blessures le jour de match, un modèle logistique mixte a été estimé pour chacun des 5 jours précédant le match avec le risque de blessures le jour de match en variable réponse. Les modèles pour un individu  $i$  au match  $j$  étaient définis par :

$$P(Y_{ij} = 1) = \frac{e^{X_{ij}^T \beta + \gamma_i}}{1 + e^{X_{ij}^T \beta + \gamma_i}}$$

et  $\gamma_i \sim N(0, B)$

Des modèles mixtes de Poisson ont été testés au préalable mais une sur-dispersion a été observée. Que ce soit avec la distribution de Poisson ou binomiale, les conclusions des modèles étaient pour autant identiques. De plus, il n'y a pas de différence de charge d'entraînement au cours de la semaine entre les titulaires et les remplaçants du prochain match. Cela aurait pu conduire à une prédiction davantage de ce statut et donc du temps de jeu davantage que de la blessure. C'est pourquoi nous avons privilégié la régression logistique.

Les résultats montrent qu'une charge importante en musculation en début de semaine augmente le risque de subir une blessure le jour de match (J-5:  $p < 0.001$  et J-4:  $p = 0.03$ ). On constate également un effet protecteur de la haute intensité deux jours avant le match. Concernant la veille de match, une RPE plus élevée est associée à un risque plus important. Sur ce dernier point, l'entraînement à J-1 étant standardisé, cet effet de la RPE semble plus évocateur d'une fatigue plus importante par rapport aux autres semaines, plus qu'à une difficulté plus élevée de l'entraînement. De plus, les joueurs qui sont de retour d'une blessure supérieur à 7 jours d'indisponibilité ont un OR entre 1.52 et 1.61 sur les 5 journées d'analyse.

Cette étude a permis de mettre en avant que le volume, l'intensité et la programmation des séances de musculation ont des effets distincts selon le moment où ils sont pratiqués au cours de la semaine d'entraînement.

## 2) Article soumis

### **Impact of weekly training workload structure and content on the risk of injury in professional Rugby Union match-play**

Brief title: Impact of weekly training workload on the risk of injury in Rugby Union match-play

Laboratory: Univ. of Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR 1219, ISPED, F-33000 Bordeaux, France

Authors:

Jérémy Chéradame, Research Department, French Rugby Union Federation, Marcoussis, France.

Romain Loursac, Centre Orthopédique Santy, FIFA Medical Center of Excellence, Groupe Ramsay-Generale de Santé, Lyon, France.

Julien Piscione, Research Department, French Rugby Union Federation, Marcoussis, France.

Christopher Carling, FFF Research Centre, French Football Federation, Clairefontaine National Football Centre, Clairefontaine-en-Yvelines, France. Laboratory Sport, Expertise and Performance (EA 7370), French Institute of Sport (INSEP), Paris, France.

Philippe Decq, Institut de Biomécanique Humaine Georges Charpak, Arts et Métiers ParisTech et Université de Paris, Assistance Publique–Hôpitaux de Paris, Hôpital Beaujon, Neurochirurgie, Clichy, France

Hélène Jacqmin-Gadda, Univ. of Bordeaux, Inserm, Bordeaux Population Health U1219, ISPED, Bordeaux, France

## **ABSTRACT**

The aim of this study was to investigate the impact of different components of daily training workload during the week preceding the match on the risk of match injury in professional rugby union players.

A cohort of 72 players from a single professional French club participated in the study. GPS-derived data including total distance and high-speed distance in addition to Ratings of Perceived Effort (RPE) for both on- and off-pitch (gym-based strength conditioning work) training were collected for each training session over 3-seasons (2017-2020). The association between the daily measures of workload over the week preceding the day of the match (MD) and the risk of injury in subsequent match-play was estimated using a mixed effects logistic model adjusted for contextual and individual factors.

A total of 184 injuries were sustained in 128 matches (incidence: 81.2 injuries per 1000 participation hours). A higher RPE for the strength conditioning session on MD-5 ( $p < 0.001$ ) and MD-4 ( $p = 0.03$ ) and for the on-pitch session on MD-1 ( $p = 0.01$ ) was associated with an increased risk of injury in matches. On MD-2, a higher total distance and that run at high-speed were respectively associated with a higher ( $p = 0.01$ ) and lower risk ( $p = 0.04$ ) of injury in matches played.

This study shows that different components of external and internal load had varying influences on injury risk and particularly in relation to the day on which these were performed in the week leading up to the next match. At MD-2, training load favouring intensity rather than volume could be beneficial in reducing the risk of match-play injury.

**Keywords:** injury prevention, biostatistics, workload monitoring, training planning, multiple imputation

## INTRODUCTION

Rugby Union is a team sport in which the risk of sustaining injury is one of the highest(5). The incidence of injury at professional standards has been estimated at 91 injuries per 1000 player hours in match-play and 3 injuries per 1000 player hours in training(39). As such, avoiding injury is important, above all for the health of the players, and also for the impact that subsequent absence has upon team performance (41). An increased understanding of the risk factors associated with injury is therefore essential to try to limit their effect. Albeit controllable, the workload performed in matches and training is shown to be associated with injury risk(13,15). The quantification of workload, whether internal or external, typically involves the use of Global Positioning Systems (GPS), heart rate monitors and subjective questionnaires. Several studies have used these methods, either singly or in combination, to enhance understanding of the relationship between workload and injury. Workload is often analysed longitudinally, whether cumulatively or as a variation with indicators such as chronic acute ratio (ACWR(2)), weighted ratios (EWMA(43)) or Robust Exponential Decreasing Index (REDI(30)) considered over different time scales.

In summary, studies converge towards the notion that a large variation in the workloads experienced by players over the previous weeks can be a risk factor for injury(7,13,31). In addition, players considered not to have played enough or conversely, too much, over the previous year, are at greater risk of injury(42). In the shorter term, Cross et al.(8) demonstrated a significant nonlinear effect with injury risk for cumulative workload over a 4-week period in Rugby. The authors determined a cumulative RPE session threshold of 1245 arbitrary units at which the risk of injury became greater in the subsequent week(8). However, Brooks (2008) reported that a large volume of training time over the week did not lead to a greater number of injuries sustained in training or in the following match in rugby(4). Nevertheless, planification of the training week has been the subject of much debate (1,10,35). Indeed, the different types of training sessions commonly utilised place varying demands on players. In a large cohort covering 11 seasons of the English national championship, West et al. (2017) demonstrated marked differences in injury rates according to the type of session undertaken. Notably, the authors reported an incidence of 9.6 injuries for 1000 hours during full contact rugby skills training

whereas only 0.7 injuries per 1000 player hours were observed for gym-based conditioning sessions(38). However, to our knowledge, the potential impact on match-play injury risk of training session type and the day of the week upon which it is performed has not been explored in elite rugby. Further knowledge of the training components specifically associated with injury risk could aid optimisation of workload management and potentially help to reduce the risk(32). In addition, analysis of the potential contribution of individual factors such as age, anthropometric data, previous injury, and playing position(16,29,36,42) and contextual variables such as the time of the season or the location of the match(9,14,22) is necessary to understand the injury risk. Arguably these should be investigated individually and then included in the statistical model used to explore the workload/injury risk association.

The main objective of this study over three consecutive seasons in a single professional rugby club was to identify whether different components of the daily training load over the week preceding the match were associated with the risk of injury in the following match whilst taking into account various contextual factors and the players' individual characteristics.

## **METHODS**

### **Experimental Approach to the problem**

Athletic workload, self-reported RPE questionnaires and injuries collated over 3 consecutive seasons from a cohort of an elite professional rugby union club were analysed retrospectively. To examine the effect upon injury risk of the workload from field- or gym-based training sessions performed during the week and the effects of the individual training day prior to official match-play, multivariate models were used.

### **Subjects**

Data was collected in 72 professional Rugby players (mean  $\pm$  SD: age : 26,8  $\pm$  5,2 years, weight : 103,2  $\pm$  13,5 kg, and height : 186,3  $\pm$  7,6 cm) who played at least 5 matches for a single Top 14 club in France over the seasons 2017/2018 to 2019/2020. The players participated in a total of 128 matches (48 in 2017/2018, 43 in 2018/2019, 36 in 2019/2020 while 23 participated in all 3 seasons).

Participation data was aggregated: 21 752 player training days and 2 461 player match days. Injury data was collected via the Electronic Medical Record declared to the French national data protection authority (n°1461311) for which the players gave their written consent. The study is in accordance with the principles outlined in the Declaration of Helsinki. The players were informed about the study and were allowed to withdraw from it at any moment.

## **Procedure**

### *Quantifying training load*

Individual training load was measured using several indicators for all the training sessions over the three seasons. Internal training load measures were collected by the club's physical conditioning staff using the Rating of Perceived Exertion (RPE)(3) method. Within 30 minutes of each training session, participants rated the perceived difficulty of their session, from 1 to 10, using the Borg scale. RPE is demonstrated to be a valid method for estimating exercise intensity(20) and has returned positive correlations of 0.89 and 0.86 with training heart rate and training blood lactate concentrations, respectively, during typical rugby league training activities(14) and particularly when scores are multiplied by the duration of the session (sRPE). Data were collected individually following both on-pitch and gym-based strength training sessions.

External training load measures were collected during each training session using 10 Hz GPS devices with an in-built 100 Hz tri-axial accelerometer, gyroscope and magnetometer (Optimeye S5, Catapult Innovations, Melbourne, Victoria). The manufacturer's software was used to download and store all sessions. Previous research has demonstrated the reliability and validity of these devices to determine external workload (37). Devices were worn in a fitted vest under the playing jersey and located between the scapulae. The total distance (TD) and high-speed distance (HSD) were used with values for the latter adjusted according to the individual's maximal aerobic speed threshold determined during a laboratory test of aerobic fitness. The duration of each field session was retrieved via the GPS or from the time duration entered in the RPE survey for the strength sessions.

Information on injury was collected prospectively by the club's medical staff. In line with the injury consensus statement defined by the International Rugby Board in 2007(12), an injury was defined as

any physical complaint that resulted in the individual being unable to take full part in any field- or gym-based training session or match. Additional information on the nature of the injury was recorded including location, type, and severity.

The same workload and injury variables were also collected for the training sessions and matches for 8 players who were selected to represent their national teams (senior and youth) over the same period.

### *Missing Data*

Of the total number of player training participation days, there were 13.7% of days where GPS data was unavailable even though a training session had taken place. Regarding RPE, data for 5.5% of the total number of days was missing. Multiple imputation was used to work around this missing data issue. Multiple imputation relies on 3 steps: (i) imputation of the missing data to build  $m$  complete data sets (here  $m = 5$ ) from models specific to each missing variable, (ii) estimation of model parameters on each imputed data set, (iii) computation of the mean of the  $m$  estimated parameters and estimation of their variance accounting for both inter and intra imputation variance(33).

As data were missing for several variables without any monotonic structure, data were imputed by Multiple Imputation by Chained Equations (MICE(6)). This algorithm involves imputing each variable iteratively via an imputation model from all the other variables including those with missing values using their imputed values at the previous iteration. To account for intra-player correlation and characteristics of the session, the imputation models included a player-specific random intercept and the average of the day derived from the other players of the same positional group (forward, half-back or back) for the variable to predict ( $\bar{X}_{kj.position}$ ), in addition to all other player and session specific variables (age, weight, height, GPS data, RPE for field and training, day in reference to the match and injury). For instance, for imputing  $X_1$  for subject  $i$  at day  $j$ , the model is:

$$X_{1ij} = \beta_0 + \beta_1 \bar{X}_{1j.position} + \sum_{k=2}^p \beta_k X_{kij} + u_{1i} + e_{1ij} \text{ with } u_{1i} \sim N(0, \sigma_1)$$

In order to be robust under a possible departure from the Gaussian assumption, we used the Predictive mean matching (PMM) imputation procedure proposed by Rubin (1986)(34) and Little (1988)(25). This involves imputing a randomly selected observation among the set of observed values whose

values predicted by a linear imputation model are close to the one to be imputed. Imputations were performed separately for the five days before the match, and for days solely including a gym session versus other training days. The imputation was conducted using R software version 4.0.5 and the mice package v3.13.0.

### **Statistical Analyses**

The incidence of injury and associated 95% confidence intervals were calculated for 1000 hours of match-play. In a first step, relationships were tested between injuries and different contextual match- or player-specific variables: age, weight, height, season, phase of season (first versus second phase), position (forwards, half-backs and backs), location (home versus away match) and whether or not the player had returned from a layoff period of over 7 days due to an injury. These factors were added as fixed effects in a mixed logistic model and the player as a random effect. Among them, variables significantly associated with the injury risk were included as adjustment variables when studying the risk of injury in relation to workload. Player position and phase of the season that may have had an influence on workload were forced as adjustment variables. Adjusted models for the risk of injury during a match according to the workload for each day of the previous week, were estimated separately for each day. Daily workload variables included the total distance covered (TD) during the daily training sessions, high speed distance (HSD), two measures of RPE for the field and gym sessions and the total training duration. TD and HSD were converted into kilometres and the duration of the training session into hours to obtain equivalent scales. Finally, a model including all the significant variables at the threshold of 0.10 from the 5 day-specific models was estimated. Players with less than 5 games were removed from the study to avoid convergence issues. The model fit was validated using the Hosmer Lemeshow test(19). Predictive capabilities of the model were examined by 5-fold cross-validation with 80% training and 20% test samples. Area under the ROC curve (AUC) and the Brier score were calculated by merging the 5 validation samples. For these calculations the individual predictions were averaged from the five imputation datasets.

## RESULTS

Altogether, 184 match injuries were sustained corresponding to an incidence of 81.2 injuries per 1000 player hours (Table 1). In matches and training combined, a total of 284 injuries (corresponding to an incidence of 5.3 in training and 14.6 match and training combined) were reported of which 244 (84.7%) required a layoff of over seven days.

Position	Player (n)	Player-game (n)	All injuries (n)	Game injuries (n)	Height (cm)	Weight (kg)	Age (y)	Injuries >7 days (n)
<b>Forwards</b>	42	1422	155	97	188,4 ± 7,8	111,8 ± 9,1	27,1 ± 5,2	136
<b>Half-Backs</b>	13	360	48	36	180,6 ± 7,1	86,9 ± 7,3	26,9 ± 6,1	36
<b>Backs</b>	17	680	81	51	185,7 ± 4,9	94,4 ± 8,2	26,2 ± 4,9	72
<b>Total</b>	72	2462	284	184	186,3 ± 7,6	103,2 ± 13,5	26,8 ± 5,2	244

Table 1: Population description.

### Workload analysis

The characteristics of the weekly training load are presented in Table 2. The lowest values for volume ( $1630 \pm 740$ m) and RPE ( $3.3 \pm 1.3$ ) and fewer strength training sessions were reported for MD-1. Two thirds of all the MD-3 were rest days with a null workload reported both on the field and in the gym. When a training session occurred on MD-3 the total and high-speed distance covered were the highest across the week. On 87% and 77% of MD-4, the players participated in a field training session and a gym session respectively. The sRPE for field ( $602 \pm 293$  AU) and gym sessions ( $281 \pm 113$  AU) were higher on MD-4 compared to the other training days.

	TD	HSD	RPE	Field duration	SRPE Field	Gym Duration	SRPE Gym
<b>Day-1</b>	1630 ± 740	186 ± 171	3,3 ± 1,3	31,8 ± 12,7	111 ± 70	48,5 ± 14,5	170 ± 91
	9%	9%	9%	9%	10%	94%	94%
<b>Day-2</b>	3275 ± 946	470 ± 224	4,6 ± 1,3	56,3 ± 17,0	412 ± 193	44,9 ± 13,2	197 ± 88
	26%	26%	24%	26%	27%	32%	31%

<b>Day-3</b>	4192 ± 1777	526 ± 367	4,9 ± 1,4	83,0 ± 28,9	585 ± 246	56,0 ± 18,7	278 ± 140
	66%	66%	65%	66%	68%	70%	70%
<b>Day-4</b>	3871 ± 1575	447 ± 303	4,6 ± 1,4	76,2 ± 26,4	602 ± 293	60,0 ± 19,7	281 ± 113
	13%	14%	12%	13%	13%	23%	21%
<b>Day-5</b>	2058 ± 1256	267 ± 259	3,9 ± 1,5	48,5 ± 21,8	210 ± 144	50,4 ± 14,0	238 ± 105
	44%	46%	38%	44%	46%	48%	46%

Table 1 : Description of daily workload measures for the five days preceding the match: Mean ± standard deviation excluding the null values and percentage of null value for Total Distance (TD), High Speed Distance (HSD), Rating Perceived Effort (RPE), session RPE (sRPE) for field and gym sessions and duration of field and gym sessions.

#### **Effect of age, playing position, season phase and previous injury on injury incidence**

Players aged between 28 and 33 years reported the highest rate of injury with incidences of 93.6 (95% confidence interval: [82.1,105.1]) for 28-30 and 105.4 [91.3,119.5] for 31-33 years old players (Figure 1). Players in the half-back positions were the most affected reporting an incidence of 115.9 [107.3, 124.5]. No significant season-to-season changes in injury risk was observed over the 3 seasons. The incidence of injury in the first phase of the season was 89.6 [85.9, 93.3] compared to 69.2 [65.3, 73.1] for the second phase. Among players returning from more than 7 days off following an injury the incidence was 136.7 [114.5,159.0] injuries versus 74.5 [69.3,79.7] for other players. No significant difference in injury rates in home versus away matches was observed.

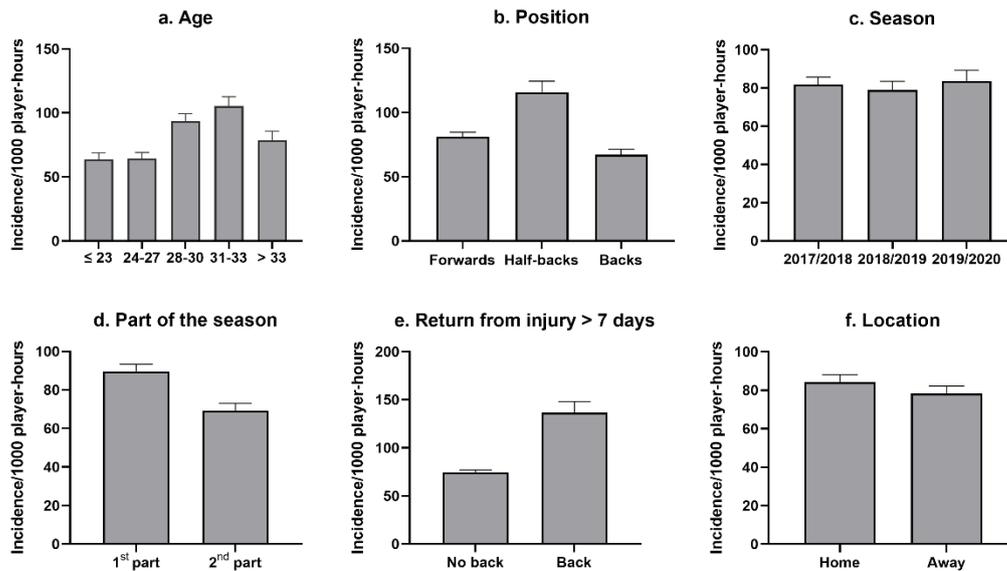


Figure 9: Incidence of injury according to age, player position, number of weeks since last game and return from injury which led to an absence of more than 7 days. Values are presented as incidence (95% CI).

In the adjusted analysis using a logistic mixed model including all match and player specific variables excluding workload, only age ( $p=0.02$ ) and return from an injury greater than 7 days ( $p=0.03$ ) were significantly associated with the risk of injury (Table 3). The reduced model included only significant variables and potential confounders for the relation between workload and injury that were forced even if not significant (position and season phase).

	Full model			Reduced model		
	p	OR	IC 95%	p	OR	IC 95%
<b>Age</b>	0,02	1,05	1,01 - 1,09	0,03	1,04	1,00 - 1,08
<b>Weight</b>	0,18	0,82	0,61 - 1,10	-	-	-
<b>Height</b>	0,92	0,99	0,80 - 1,22	-	-	-
<b>Position - Half-backs</b>	Ref.	-	-	Ref.	-	-
<b>Position - Forward</b>	0,95	0,98	0,49 - 1,96	0,34	0,66	0,41 - 1,07
<b>Position - Backs</b>	0,74	0,87	0,50 - 1,51	0,09	0,77	0,45 - 1,32
<b>Saison - 2017/18</b>	Ref.	-	-	-	-	-
<b>Saison - 2018/19</b>	0,21	0,79	0,55 - 1,14	-	-	-
<b>Saison - 2019/20</b>	0,70	0,93	0,63 - 1,37	-	-	-

<b>Part of season - First</b>	Ref.			Ref.		
<b>Part of season - Second</b>	0,21	0,82	0,59 - 1,12	0,16	0,79	0,58 - 1,10
<b>Location - Home</b>	Ref.			-		
<b>Location - Away</b>	0,74	0,95	0,70 - 1,29	-		
<b>Back inj. &gt;7 days</b>	0,03	0,82	0,61 - 1,10	0,03	1,61	1,04 - 2,51

Table 1: Estimates of mixed effects logistic models for the risk of injury in game according to match and players' characteristics.

### Impact of weekly workload on the risk of match-day injury

Estimates of the mixed effects logistic regressions for the risk of injury according to daily workload are presented in Table 4. Older players and players returning from injuries with an absence of more than 7 days were still at higher risk of incurring injury in a match. On MD-5 and MD-4 a high gym RPE significantly increased the risk of injury during the following match (OR= 1.20 [1.09;1.32] and 1.18 [1.02;1.37] for a 1-point increase in RPE with  $p = <0.001$  and 0.03 respectively). On MD-2, the total distance covered in training was associated with an increased injury risk (OR= 1.56 [1.08;2.24],  $p = 0.04$ ) while the distance run at high speed was associated with a decreased risk (OR = 0.31 [0.10;0.99],  $p = 0.01$ ). Pre-match day, a high RPE for field training was associated with an increased risk of injury during the match (OR= 1.17 [1.04;1.31],  $p = 0.01$ ). According to the AUC ROC, the model considering the training load at MD-5 was the most predictive of injury risk with AUC=0.60 while the Brier scores of the 5 models were very similar.

	Day-5			Day-4			Day-3			Day-2			Day-1		
	p	OR	IC 95%	p	OR	IC 95%	p	OR	IC 95%	p	OR	IC 95%	p	OR	IC 95%
<b>Age</b>	0,04	1,04	1,00 - 1,08	0,05	1,04	1,00 - 1,08	0,04	1,04	1,00 - 1,08	0,02	1,04	1,00 - 1,09	0,09	1,03	0,99 - 1,08
<b>Position - Half-back</b>	Ref.			Ref.			Ref.			Ref.			Ref.		
<b>Position - Forward</b>	0,14	0,69	0,42 - 1,13	0,28	0,75	0,45 - 1,27	0,08	0,64	0,39 - 1,05	0,36	0,79	0,48 - 1,30	0,14	0,69	0,43 - 1,13
<b>Position - Back</b>	0,44	0,80	0,46 - 1,40	0,36	0,77	0,44 - 1,34	0,35	0,77	0,45 - 1,33	0,65	0,88	0,50 - 1,53	0,46	0,81	0,47 - 1,41
<b>Part of season - First</b>	Ref.			Ref.			Ref.			Ref.			Ref.		
<b>Part of season - Second</b>	0,21	0,81	0,59 - 1,12	0,29	0,84	0,61 - 1,16	0,14	0,78	0,57 - 1,08	0,19	0,81	0,58 - 1,11	0,08	0,75	0,54 - 1,04
<b>Back inj. &gt;7 days</b>	0,07	1,52	0,97 - 2,38	0,05	1,58	1,01 - 2,47	0,04	1,59	1,02 - 2,48	0,04	1,61	1,03 - 2,51	0,04	1,61	1,03 - 2,52
<b>TD</b>	0,27	1,20	0,86 - 1,68	0,47	1,10	0,85 - 1,42	0,41	0,87	0,62 - 1,21	0,02	1,56	1,08 - 2,24	0,80	1,06	0,67 - 1,67
<b>HSD</b>	0,23	0,49	0,15 - 1,59	0,95	0,97	0,40 - 2,35	0,49	1,48	0,49 - 4,51	0,05	0,31	0,10 - 0,99	0,41	0,49	0,08 - 2,78
<b>RPE-field</b>	0,98	1,00	0,85 - 1,18	0,84	0,99	0,86 - 1,13	0,60	0,95	0,77 - 1,17	0,94	0,99	0,85 - 1,16	0,01	1,17	1,04 - 1,31
<b>RPE-gym</b>	<0,001	1,20	1,09 - 1,32	0,02	1,18	1,02 - 1,37	0,67	1,04	0,86 - 1,26	0,33	1,08	0,92 - 1,27	0,83	1,04	0,75 - 1,42

<b>Duration</b>	0,55	0,77	0,33 - 1,81	0,57	0,82	0,41 - 1,64	0,76	1,16	0,45 - 2,97	0,07	0,42	0,16 - 1,09	0,24	0,52	0,17 - 1,57
Hosmer-Lemeshow p		0,67		0,06			0,08			0,08				0,14	
Brier Score		0,0692		0,0692			0,0691			0,0691				0,0692	
AUC		0,597		0,569			0,564			0,564				0,575	

Table 4: Mixed logistic models for the risk of injury in match according to the training workload of each of the 5 days preceding the match adjusted for players' and match characteristics.

The significant variables at the threshold of 10% from the models presented in Table 4 were simultaneously included in a mixed effects logistic model for the risk of injury (Table 5). Only MD-4 RPE for gym sessions was no longer significantly adjusted on the other variables. The AUC for this final model was only slightly improved (value 0.61).

	p	OR	IC 95%
<b>Age</b>	0,01	1,03	0,99 - 1,08
<b>Position - Half-backs</b>	Ref.		
<b>Position - Forward</b>	0,25	0,80	0,48 - 1,32
<b>Position - Back</b>	0,72	0,91	0,51 - 1,59
<b>Part of season - First</b>	Ref.		
<b>Part of season - Second</b>	0,03	0,83	0,60 - 1,15
<b>Back inj. &gt;7 days</b>	0,05	1,56	0,99 - 2,43
<b>D-5 RPE-gym</b>	0,001	1,19	1,07 - 1,32
<b>D-4 RPE-gym</b>	0,16	1,11	0,96 - 1,29
<b>D-2 DT</b>	0,03	1,49	1,05 - 2,11
<b>D-2 HSD</b>	0,02	0,24	0,08 - 0,77
<b>D-2 Duration</b>	0,15	0,5	0,19 - 1,27
<b>D-1 RPE-field</b>	0,04	1,11	1,01 - 1,24
Hosmer-Lemeshow p		0,40	
Brier Score		0,690	
AUC		0,614	

Table 5: Mixed logistic models of the risk of injury in match on the significant variables at p<0.10 of each training day (Table 4).

## DISCUSSION

The main findings in this study were that the risk of sustaining injury in match-play was dependent upon the workload performed on different training days during the week prior to the match. Higher RPE values for the gym-based strength session performed at the beginning of the week or for the training session the day before a match were linked to a higher risk of injury in the next match. In contrast, the distance covered in high-speed running performed on MD-2 demonstrated a protective effect against injury. In addition, individual factors such as player age and return to competition after a time-loss injury increased the risk of injury.

Here, analysis of the distribution of the workload over the week generally follows the tactical periodisation principles previously proposed by Tee et al.(35) for Rugby. Although the authors provided no information on training load volume or intensity, the weekly structure observed in the present team is the same. A rest day was usually positioned on MD-3 while the highest training load, represented by sRPE, was reported on MD-4. Regarding, MD-2 a focus on 'speed' was observed with the highest proportion of high-speed running reported on this day (14.3% of the total).

Prior to examining the effects of training day and content on injury risk, relationships were tested between injuries and different contextual match- or player-specific variables and the effects of previous time-loss injury (whether or not the player had returned from a layoff period of over 7 days owing to injury). Although the descriptive analysis suggested a different incidence between half-backs and other positions, this result was not significant in the adjusted analysis considering other contextual and individual factors. However, depending on the position, the physical demands are quite different in Rugby(24), hence the importance of considering these in the statistical models. Findings on the effect of age on injury risk vary across the literature(17,21,28,42). In the current study, older players appeared at greater risk of incurring injury. The question of returning to competition after injury is also a point of interest. In the present players, those who played their first game following an injury which resulted in a layoff >7 days exhibited about a 50% higher risk of injury. We can suggest that these new injuries may be, in part, a reoccurrence of the previous injury although information to

corroborate this was unavailable (40). This result nevertheless suggests that the return to competition phase could be too early, potentially owing to a too greater workload for a player who has suffered a time-loss injury(2). Once all these potential confounders were identified, it was important to insert them into the statistical model to identify the effect of workload on injury risk.

Regarding the association between training load and match injury risk, RPE values were associated with an increased risk of injury in three out of the five training days prior to the following match. This finding demonstrates their usefulness as a measure for investigating the potential training workload/injury association in elite rugby. A higher RPE for the gym-based strength session performed on MD-4 and MD-5 notably were linked with a greater risk of injury. However, it is difficult to fully interpret this increased risk of injury during matches linked to perceived difficulty of the gym-based strength sessions, because we did not have access to an objective measure of the workload of gym session (number of repetitions, total weight lifted etc.). Indeed, study by Gabbett and Jenkins (15) reported that high strength and power training loads may contribute to field training injuries. In a future study, information relating to the program contents for each gym session would be useful to examine how for example, differences in the type and frequency of exercises performed might influence the risk of match injury.

Regarding the external load performed on MD-4, none of the field workload variables emerged significantly from the statistical model. Thus, it would seem that practitioners can modulate more freely the running in field on this day without any real consequence on the risk of injury. For MD-2, the sessions typically contained less volume (total distance) and were based more upon 'speed' represented by greater values for high-speed running distance. This programming of high-speed session on MD-2 appeared to be protective against the risk of match injury. Research has demonstrated that exposing players to large and rapid increases in high-speed running distances above their 2-year session average augments the risk of hamstring strain injuries(11). It is therefore important to ensure players perform a regular volume of high-speed running. In addition, weeks with a moderate volume of running at high-speed were previously found to have a protective effect on lower limb

injuries compared to weeks with less volume(26). The present work suggests that the training session two days before the game is the most conducive one to for performing high-speed running.

Following adjustment for all other factors including external load, higher RPE values reported in field-based training field the day before the match generally had a higher risk of incurring injury in the next day's match. Yet, the training session on MD-1 is generally quite standardized so the variation in the load observed between weeks is lower for this day (lower SD values) and the external load is also generally lower than on the other days. Therefore, the RPE for this day might arguably be considered more a measure of the players' wellbeing rather than representing the difficulty of the session. This is therefore an important criterion to consider when evaluating the injury risk.

In sum, it seems that globally the workload performed in field training in the final two days before the match had a greater impact on the risk of injury in match-play than that performed at the beginning of the week. In contrast, the reverse was observed for gym-based strength training.

### **Limits**

The rate of missing data is a limitation of the current study. However, to handle this issue, we used the multiple imputation method accounting for both the intra-subject correlation and training session characteristics. Griffin et al (2021)(18) have previously shown that daily average imputation was the most efficient and convenient method. Given the heterogeneity of the workload per position, considering these averages per position improves the imputation model. The Predictive Mean Matching has been proven to be a robust imputation procedure whatever the distribution of the imputed variable(23,27).

In addition, this investigation was conducted in a single club and thus on a limited number of players. However, the staff remained unchanged during this period which avoids some possible confounding effects, but the results may nevertheless not be generalizable to all clubs. Moreover, the comparison of the associations with the workload between the 5 days may be blurred by the large difference in the volume, intensity and the type of training session each day. To more clearly distinguish the effect of the time interval between the daily training sessions and the match from the effect of the

characteristics of the training session it would be necessary to have identical programs of training session each day.

Finally, while this study has enabled identification of risk factors and possible factors of the risk for the injuries incurred, the predictive abilities of the models, with a maximum AUC value of 0.61, is not sufficient to use this model as an individual prediction model. Predictive abilities could potentially be improved in future work by adding other measures of workload such as higher running speed thresholds (e.g., sprinting), accelerations and the number of player-to-player contacts. We were not able to include these variables either because the algorithms to calculate these variables changed during this period or these were not available when the players were on international duty. Finally, the number of injuries in our data set was not sufficient to study associations between workload and injury type.

## **PRACTICAL APPLICATIONS**

This study has arguably provided novel information on the impact of workload distribution and content during the week on the risk of injury in following match-play. An increase in loading from gym-based conditioning measured by RPE at the beginning of the week was associated with a higher risk of match injury. It also seems that a field training session performed on MD-2 during which players perform a lower volume of running activity but at a higher speed, might be beneficial in reducing the risk of injury. Finally, subjective perceptions of the difficulty of the session using RPE on the day before the match was associated with injury risk. While, this study would need to be reproduced on a larger scale (multiple club settings notably), we can conclude that objective and subjective measures of workload provided useful information that can be useful to consider in the planning and evaluation of the training week and aid in trying to limit the risk of injury in professional rugby match-play.

## BIBLIOGRAPHY

1. Afonso, J, Bessa, C, Nikolaidis, PT, Teoldo, I, and Clemente, F. A systematic review of research on Tactical Periodization: absence of empirical data, burden of proof, and benefit of doubt. *Hum Mov* 21: 37–43, 2020.
2. Blanch, P and Gabbett, TJ. Has the athlete trained enough to return to play safely? The acute:chronic workload ratio permits clinicians to quantify a player's risk of subsequent injury. *Br J Sports Med* 50: 471–475, 2016.
3. Borg, GA. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc* 14: 377–381, 1982.
4. Brooks, JHM, Fuller, CW, Kemp, SPT, and Reddin, DB. An assessment of training volume in professional rugby union and its impact on the incidence, severity, and nature of match and training injuries. *J Sports Sci* 26: 863–873, 2008.
5. Brooks, JHM and Kemp, SPT. Recent Trends in Rugby Union Injuries. *Clin Sports Med* 27: 51–73, 2008.
6. Buuren, S van and Groothuis-Oudshoorn, K. **mice** : Multivariate Imputation by Chained Equations in R. *J Stat Softw* 45, 2011. Available from: <http://www.jstatsoft.org/v45/i03/>
7. Cousins, BEW, Morris, JG, Sunderland, C, Bennett, AM, Shahtahmassebi, G, and Cooper, SB. Match and Training Load Exposure and Time-Loss Incidence in Elite Rugby Union Players. *Front Physiol* 10: 1413, 2019.
8. Cross, MJ, Williams, S, Trewartha, G, Kemp, SPT, and Stokes, KA. The Influence of In-Season Training Loads on Injury Risk in Professional Rugby Union. *Int J Sports Physiol Perform* 11: 350–355, 2016.
9. Dalton-Barron, NE, McLaren, SJ, Black, CJ, Gray, M, Jones, B, and Roe, G. Identifying Contextual Influences on Training Load: An Example in Professional Rugby Union. *J Strength Cond Res* 35: 503–511, 2021.
10. Delgado-Bordonau, JL and Mendez-Villanueva, A. Tactical Periodization: Mourinho's best-kept secret? 7
11. Duhig, S, Shield, AJ, Opar, D, Gabbett, TJ, Ferguson, C, and Williams, M. Effect of high-speed running on hamstring strain injury risk. *Br J Sports Med* 50: 1536–1540, 2016.
12. Fuller, CW, Molloy, MG, Bagate, C, Bahr, R, Brooks, JHM, Donson, H, et al. Consensus statement on injury definitions and data collection procedures for studies of injuries in rugby union. *Br J Sports Med* 41: 328–331, 2007.
13. Gabbett, TJ. The training—injury prevention paradox: should athletes be training smarter and harder? *Br J Sports Med* 50: 273–280, 2016.
14. Gabbett, TJ and Domrow, N. Relationships between training load, injury, and fitness in sub-elite collision sport athletes. *J Sports Sci* 25: 1507–1519, 2007.
15. Gabbett, TJ and Jenkins, DG. Relationship between training load and injury in professional rugby league players. *J Sci Med Sport* 14: 204–209, 2011.

16. Gabbett, TJ, Ullah, S, and Finch, CF. Identifying risk factors for contact injury in professional rugby league players – Application of a frailty model for recurrent injury. *J Sci Med Sport* 15: 496–504, 2012.
17. Ganse, B, Degens, H, Drey, M, Korhonen, MT, McPhee, J, Müller, K, et al. Impact of age, performance and athletic event on injury rates in master athletics – First results from an ongoing prospective study. 7
18. Griffin, A, Kenny, IC, Comyns, TM, Purtill, H, Tiernan, C, O’Shaughnessy, E, et al. Training load monitoring in team sports: a practical approach to addressing missing data. *J Sports Sci* 1–11, 2021.
19. Hosmer, DW, Lemeshow, S, and Sturdivant, RX. Applied logistic regression. Third edition. Hoboken, New Jersey: Wiley, 2013.
20. Impellizzeri, FM, Rampinini, E, Coutts, AJ, Sassi, A, and Marcora, SM. Use of RPE-Based Training Load in Soccer: *Med Sci Sports Exerc* 36: 1042–1047, 2004.
21. Kallinen, M and Markku, A. Aging, Physical Activity and Sports Injuries: An Overview of Common Sports Injuries in the Elderly. *Sports Med* 20: 41–52, 1995.
22. King, D, Hume, P, Clarke, T, and Gissane, C. Does Playing Away From Home Influence the Number and Severity of Impacts in Amateur Rugby Union Players: Analyses by Home/Away, Won/Lost and First/Second Season Halves. 7, 2020.
23. Kleinke, K. Multiple Imputation Under Violated Distributional Assumptions: A Systematic Evaluation of the Assumed Robustness of Predictive Mean Matching. *J Educ Behav Stat* 42: 371–404, 2017.
24. Lindsay, A, Draper, N, Lewis, J, Giesege, SP, and Gill, N. Positional demands of professional rugby. *Eur J Sport Sci* 15: 480–487, 2015.
25. Little, RJA. Missing-Data Adjustments in Large Surveys. 11, 2021.
26. Malone, S, Owen, A, Mendes, B, Hughes, B, Collins, K, and Gabbett, TJ. High-speed running and sprinting as an injury risk factor in soccer: Can well-developed physical qualities reduce the risk? *J Sci Med Sport* 21: 257–262, 2018.
27. Marshall, A, Altman, DG, and Holder, RL. Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. *BMC Med Res Methodol* 10: 112, 2010.
28. McIntosh, AS. Rugby Injuries. In: *Medicine and Sport Science*. Maffulli, N and Caine, DJ, eds. . Basel: KARGER, 2005 [cited 2022 Aug 18]. pp. 120–139 Available from: <https://www.karger.com/Article/FullText/85394>
29. Meeuwisse, WH, Tyreman, H, Hagel, B, and Emery, C. A Dynamic Model of Etiology in Sport Injury: The Recursive Nature of Risk and Causation. *Clin J Sport Med* 17: 215–219, 2007.
30. Moussa, I, Leroy, A, Sauliere, G, Schipman, J, Toussaint, J-F, and Sedeaud, A. Robust Exponential Decreasing Index (REDI): adaptive and robust method for computing cumulated workload. *BMJ Open Sport Exerc Med* 5: e000573, 2019.
31. Murray, NB, Gabbett, TJ, Townshend, AD, and Blanch, P. Calculating acute:chronic workload ratios using exponentially weighted moving averages provides a more sensitive indicator of injury likelihood than rolling averages. *Br J Sports Med* 51: 749–754, 2017.

32. Quarrie, KL, Raftery, M, Blackie, J, Cook, CJ, Fuller, CW, Gabbett, TJ, et al. Managing player load in professional rugby union: a review of current knowledge and practices. *Br J Sports Med* 51: 421–427, 2017.
33. Rubin, DB (ed.). Frontmatter. In: Wiley Series in Probability and Statistics. Hoboken, NJ, USA: John Wiley & Sons, Inc., 1987 [cited 2021 Oct 5]. pp. i–xxix Available from: <https://onlinelibrary.wiley.com/doi/10.1002/9780470316696.fmatter>
34. Rubin, DB. Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations. 9, 2021.
35. Tee, JC, Ashford, M, and Piggott, D. A Tactical Periodization Approach for Rugby Union. *Strength Cond J* 40: 1–13, 2018.
36. Thornton, HR, Delaney, JA, Duthie, GM, and Dascombe, BJ. Importance of Various Training-Load Measures in Injury Incidence of Professional Rugby League Athletes. *Int J Sports Physiol Perform* 12: 819–824, 2017.
37. Thornton, HR, Nelson, AR, Delaney, JA, Serpiello, FR, and Duthie, GM. Interunit Reliability and Effect of Data-Processing Methods of Global Positioning Systems. *Int J Sports Physiol Perform* 14: 432–438, 2019.
38. West, SW, Williams, S, Kemp, SPT, Cross, MJ, McKay, C, Fuller, CW, et al. Patterns of training volume and injury risk in elite rugby union: An analysis of 1.5 million hours of training exposure over eleven seasons. *J Sports Sci* 38: 238–247, 2020.
39. Williams, S, Robertson, C, Starling, L, McKay, C, West, S, Brown, J, et al. Injuries in Elite Men’s Rugby Union: An Updated (2012–2020) Meta-Analysis of 11,620 Match and Training Injuries. *Sports Med*, 2021. Available from: <https://link.springer.com/10.1007/s40279-021-01603-w>
40. Williams, S, Trewartha, G, Kemp, S, Cross, M, Brooks, J, Fuller, C, et al. Subsequent Injuries and Early Recurrent Diagnoses in elite Rugby Union Players. *Int J Sports Med* 38: 791–798, 2017.
41. Williams, S, Trewartha, G, Kemp, SPT, Brooks, JHM, Fuller, CW, Taylor, AE, et al. Time loss injuries compromise team success in Elite Rugby Union: a 7-year prospective study. *Br J Sports Med* 50: 651–656, 2016.
42. Williams, S, Trewartha, G, Kemp, SPT, Brooks, JHM, Fuller, CW, Taylor, AE, et al. How Much Rugby is Too Much? A Seven-Season Prospective Cohort Study of Match Exposure and Injury Risk in Professional Rugby Union Players. *Sports Med* 47: 2395–2402, 2017.
43. Williams, S, West, S, Cross, MJ, and Stokes, KA. Better way to determine the acute:chronic workload ratio? *Br J Sports Med* 51: 209–210, 2017.

## **Chapitre 5 : Réflexions sur l'analyse statistique des données de l'Observatoire Médical du Rugby**

---

## 1) Données attendues pour l'Observatoire Médical du Rugby

Bien qu'il n'ait à ce jour pas une activité quotidienne, l'objectif de l'Observatoire lorsqu'il a été créé était d'améliorer la prévention, la prise en charge et le suivi des blessures. Dans ce but, 45 préconisations ont été formulées le 29 mars 2018. Parmi celles-ci, il était préconisé d'améliorer la saisie des données des programmes épidémiologiques existants et notamment celle du dossier médical informatisé. De plus, il a été formulé le besoin de mettre en place un outil commun d'évaluation de la charge de travail du joueur. Une consultation a donc été réalisée pour déterminer le choix de cet outil commun. Ainsi nous avons pu participer au test de validité et reproductibilité de différentes solutions GPS. Elles ont été testées sur différents exercices (sprints et parcours avec changements de directions et de vitesses<sup>139</sup>) et dans différentes conditions (4 stades différents). Fin 2020, la Ligue Nationale de Rugby a donc nommé Catapult comme fournisseur de solution GPS pour tous les clubs professionnels et l'équipe nationale pour les 4 prochaines années. Tous les clubs ont donc dû utiliser cette solution à l'issue des éventuels contrats en cours qu'ils pouvaient avoir avec d'autres sociétés. Depuis la saison 2021/2022 seuls quelques clubs n'avaient pas encore les GPS Catapult et à partir de la saison 2022/2023, les 30 clubs professionnels français collectent quotidiennement l'activité de leurs joueurs à chaque séance d'entraînement ou de match avec la même technologie. L'intérêt principal de cet outil commun est le partage simplifié de données brutes qui permet à chaque utilisateur de pouvoir configurer toutes les variables comme il le souhaite et ainsi analyser les données selon leur propre modèle. Le partage de données s'effectue aujourd'hui uniquement entre les clubs et la sélection nationale pour une liste de joueurs ciblés. Pour répondre à l'objectif initial, la suite logique de cette démarche serait donc que l'ensemble des données puissent remonter à l'Observatoire dans le but de réaliser des études pour mieux comprendre la relation entre l'activité et le risque de blessure.

Le volume de données disponible ainsi que l'hétérogénéité des différents clubs pourraient être une vraie richesse pour mener des études pertinentes et précises avec une puissance statistique suffisante. Ainsi les données disponibles seraient donc :

- Informations relatives aux joueurs (date de naissance, poids, taille)
- Programme de suivi des blessures :
  - o Circonstance
  - o Localisation

- Type
- Charge externe via les GPS Catapult (en match et à l'entraînement) :
  - Distance totale et selon différents seuils de vitesse
  - Accélération / Décélération
  - Contacts
  - Charge mécanique
- Statistiques de match :
  - Temps de jeu
  - Activité en match

Le Tableau 3 indique des ordres de grandeurs de la population et de l'exposition de blessures en match qui devraient être disponibles sur chaque saison. Sur une saison de championnat, plus de 1400 joueurs des 30 équipes (47 joueurs par équipe en moyenne) pourront constituer la population. Ils disputent au cours d'une saison au total 432 matchs auxquels viendront s'ajouter des matchs de coupe d'Europe et des matchs internationaux. Cependant pour les internationaux étrangers les données de charge de travail seront probablement difficilement disponibles, ce qui sera à considérer dans les analyses futures.

	Equipes (n)	Matchs (n)	Joueurs (n)	Exposition (h)
Top 14	14	187	644	7480
Pro D2	16	245	765	9840
Total	30	432	1409	17320

*Table 3: Taille d'échantillon espérée sur une saison. Le nombre de joueurs est déterminé par rapport au nombre de joueurs qui ont participé à chacune des compétitions sur la saison 2021/2022.*

Sur l'ensemble des matchs que représente une saison de Rugby à XV dans les championnats professionnels français, si la tendance reste similaire à la saison 2019/2020 on s'attend à avoir plus de 2300 blessures (programme EAM).

## 2) Objectifs des analyses

A ce jour et à la suite des travaux de cette thèse, il n'existe encore aucune métrique qui fait consensus pour mesurer au mieux la relation entre la charge de travail et le risque de blessures. Plusieurs questions restent ainsi ouvertes.

Le premier objectif visera à identifier les variables de charge de travail les plus associées au risque de blessures parmi l'ensemble des mesures collectées : distance totale, à haute intensité selon différents seuils absolus ou relatifs au joueur, nombre d'accélération et décélérations, nombre de contacts, etc. Dans nos travaux nous n'avons malheureusement pas pu utiliser toutes les variables souhaitées, que ce soit dû à des changements d'algorithmes ou à des seuils différents entre ceux définis par le club et les équipes nationales. Ainsi, des données d'accélération, de décélérations ou encore la quantification du nombre de contacts peuvent être des variables intéressantes à considérer dans ce type d'analyse. Questionner également plusieurs seuils de vitesses qu'ils soient en absolu ou en relatif peut être une possibilité pour déterminer lesquels sont les plus adéquats pour répondre à cette problématique.

Au-delà de cette sélection de variables, il est nécessaire d'identifier de quelle manière cette charge de travail influence le risque de blessure. Il faut ainsi comprendre si les indicateurs de charge de travail doivent être considérés en termes de charge cumulée, de charge moyenne éventuellement pondérée ou encore en variation de charge de travail. Enfin, la période depuis laquelle cette charge a un effet sur le risque de blessure et donc sur laquelle calculer ces indicateurs sera également à déterminer.

Dans la littérature scientifique sur le lien entre charge de travail et blessure, les variables de charge de travail ont toujours été construites a priori. En général, plusieurs périodes sont testées pour identifier la plus prédictive du risque mais les conclusions de ces travaux ne convergent pas vers des indicateurs communs. De plus, la pondération exponentielle a été proposée pour attribuer une importance plus grande à la charge de travail récente mais une nouvelle fois aucune méthode n'a été publiée pour estimer la meilleure pondération à appliquer à chaque temps. Dans nos travaux nous avons tenté de répondre à cette problématique comme décrit dans la section suivante. Cependant, en raison de la faible taille de l'échantillon qui concernait un seul club, nous n'avons pas réussi avec les données à disposition à estimer la forme de la relation entre la charge de travail et le risque de blessures en fonction du temps. D'autant qu'il est probable que la forme de cette relation soit spécifique à chaque type de

blessures. La taille de l'échantillon des données de l'Observatoire Médical du Rugby devrait permettre de réaliser de telles analyses.

### 3) Identifier les périodes d'exposition critiques

Pour estimer la relation entre le risque de blessures et une charge cumulée d'exposition des méthodes comme les indices cumulés d'exposition (CIE) pourraient être utilisées<sup>140,141</sup>. Le principe est d'inclure dans le modèle une variable représentant l'exposition cumulée sur une période. Cependant cette approche suppose que l'effet de l'exposition est constant sur la période. Dans le but de modéliser les effets cumulés d'expositions variant dans le temps, la méthode Weighed Cumulative Index Exposure (WCIE) a été développée<sup>142,143,144,145</sup>. Le WCIE incorpore dans sa mesure une fonction de pondération qui permet d'attribuer des poids différents aux intensités d'exposition survenue à différents temps. Les premiers concepts d'expositions cumulées pondérées dans le temps ont été introduits par Breslow et al.<sup>142</sup> et Thomas<sup>143</sup> pour modéliser simultanément l'effet de la durée d'exposition, de l'intensité et de la période à laquelle elle a lieu. Ce concept a ensuite été développé en utilisant une modélisation paramétrique de la fonction de poids définie *a priori*<sup>144,145</sup>. Les résultats peuvent être non valides si la fonction de poids est mal spécifiée ; de ce fait l'absence de connaissances fortes pour la spécifier peut être limitant<sup>144</sup>. Hauptmann<sup>146</sup> a proposé une méthode plus flexible en utilisant des splines de régression cubiques pour modéliser la forme de la fonction de poids sans *a priori*. L'approche a été étendue par la suite pour des données de cohorte par Sylvestre et Abrahamowicz<sup>147</sup>. Il semble donc possible d'estimer à partir des données, la pondération de l'exposition passée sur le risque à un temps donné. La méthode WCIE pourrait ainsi être utilisée pour identifier la forme de la courbe représentant la relation entre charge de travail et blessure en fonction du temps. Nous avons testé cette approche sur les données de la seconde étude qui sont collectées quotidiennement dans les clubs. Ces données provenant de GPS et de questionnaires subjectifs permettent ainsi d'avoir un suivi longitudinal de la charge de travail des joueurs.

Le modèle WCIE définit l'effet cumulé de l'exposition dépendant du temps comme une somme pondérée des expositions passées :

$$WCIE(u) = \sum_{t=a}^u w(u-t)X(t)$$

où  $u$  est le temps courant,  $a$  la période maximum précédant le temps courant sur laquelle l'exposition est considérée,  $X(t)$  la charge de travail au temps  $t$  et  $w(u-t)$  la pondération attribuée à la charge de travail en  $t$  selon le délai entre le temps courant  $u$  et le temps  $t$ . Il est recommandé de choisir  $a$  suffisamment grand pour limiter les hypothèses a priori. Si les expositions lointaines n'ont pas d'impact, la pondération estimée pour ces délais devrait être nulle.

La forme de la fonction de pondération est déterminée à partir des données à l'aide de fonctions paramétriques. L'utilisation de splines cubiques naturels permet une estimation lisse de la fonction de poids  $w(u-t)$  sans faire d'hypothèses paramétriques fortes. Le choix du nombre de nœuds et la nature de la fonction spline utilisée déterminent le nombre de paramètres à estimer et donc la flexibilité de la fonction. L'estimation de splines cubiques naturels avec  $m$  nœuds intérieurs consiste en la création de  $m+1$  courbes dont la combinaison linéaire fournit l'estimation de la fonction de spline. Un nombre trop grand de nœuds entraînera un risque de sur-ajustement<sup>148</sup> et une courbe très irrégulière tandis qu'un nombre trop faible une estimation trop lissée<sup>149</sup>; 5 nœuds maximum seraient considérés comme suffisants<sup>150</sup>. Le choix peut être réalisé en se basant sur le critère d'information d'Akaike (AIC)<sup>151</sup>. Sans connaissance a priori sur la forme de la relation, il est recommandé de placer les nœuds intérieurs de manière équidistante dans l'intervalle  $[u-a, u]$ <sup>147</sup>.

L'estimation de la fonction de poids par des B-splines cubique<sup>152</sup> se fait en créant  $m$  covariables artificielles dépendantes du temps:

$$D_{ik}(u_{ij}) = \sum_{t=u_{ij}-a}^{u_{ij}-1} B_k(u_{ij}-t) \times X_i(t)$$

où  $B_k$ ,  $k = 1, \dots, m$  représente les  $m$  fonctions de la base de spline. Le temps courant  $u_{ij}$  est défini pour l'individu  $i$  et la date du match  $j$ . En effet, de par l'amplification du risque de blessures en match, et avec des blessures qui ont potentiellement des causes diverses par rapport à l'entraînement, nous avons fait le choix de ne considérer comme évènement d'intérêt uniquement les blessures en match. Les covariables  $D_{ik}(u_{ij})$  sont définies dans notre cas sur la somme jusqu'à  $u_{ij}-1$ . Cela permet ainsi d'éviter de considérer l'activité du jour d'intérêt qui

pourrait être influencée par la survenue d'une blessure. Le risque de blessure pour un match est alors estimé uniquement en fonction de la charge de travail passée.

Les modèles WCIE ont généralement été appliqués pour modéliser des évènements de survie<sup>147,153</sup>. Dans notre cas il est complexe d'utiliser une méthode à évènements récurrents. La difficulté vient du fait que le temps d'exposition et le temps d'évènement ne sont pas identiques. En effet, comme explicité précédemment en raison de l'hétérogénéité du risque entre le match et l'entraînement nous avons fait le choix de tenir compte uniquement des blessures en match. Cependant la charge induite par l'entraînement est, elle, importante à considérer d'où cette distinction entre le temps d'exposition et le temps d'évènement. Nous avons par conséquent fait le choix d'utiliser un modèle logistique mixte pour la probabilité de blessure en match au temps  $u_{ij}$  :

$$\text{logit} \left( P(Y_{ij} = 1) \right) = \beta_0 + \alpha_{0i} + \sum_{l=1}^p \beta_l Z_{l,ij} + \sum_{k=1}^m \gamma_k D_{ik}(u_{ij}) + \varepsilon_{ij}$$

où  $\alpha_{0i}$  est l'effet aléatoire sur les joueurs et les  $\beta_l$  représentent les effets fixes des covariables  $Z_l$  dont on ne cherche pas à modéliser l'effet cumulé.

Pour quantifier la force de l'effet de l'exposition sur le risque de blessure on peut ensuite calculer :

$$\hat{\beta}_{WCE} = \sum_{t=u}^a \sum_{k=1}^m \hat{\gamma}_k B_k(u-t)$$

où  $\hat{\gamma}_k$  représente l'estimation des paramètres associés aux covariables dépendantes du temps  $D_{ik}(u_{ij})$  obtenus à partir du modèle logistique mixte.  $\hat{\beta}_{WCE}$  représente donc l'effet global de la charge cumulée depuis  $a$  jours avant la date du match et  $e^{\hat{\beta}_{WCE}}$  est l'OR associé à l'augmentation d'une unité de l'indice cumulé d'exposition pondérée. Enfin on obtient l'estimation de la fonction de poids en chaque temps par :

$$\hat{w}(u-t) = \sum_{k=1}^m \theta_k B_k(u-t)$$

avec  $\theta_k = \frac{\hat{\gamma}_k}{\hat{\beta}_{WCE}}$ . La quantité  $\hat{w}(u - t)$  représente donc l'importance relative assignée à la charge de travail du  $t^{\text{ème}}$  jour avant le match du jour  $u$ . Les intervalles de confiance de la fonction de poids peuvent ensuite être obtenus par échantillonnage bootstrap<sup>154</sup>.

La faible puissance statistique de notre étude portant sur un seul club durant un peu moins de trois saisons n'a pas permis d'obtenir des résultats intéressants avec cette méthode. La forme de la relation, quelle que soit la variable de charge de travail considérée, était très influencée par le nombre de nœuds sélectionnés et le critère AIC conduisait à retenir le modèle avec 0 nœud, soit un simple modèle linéaire. De plus, les intervalles de confiance autour des courbes  $w(u - t)$  selon  $u - t$  étaient très larges, incluant la valeur 0 et suggérant donc une absence d'effet de la charge de travail. Nous avons donc fait le choix de construire plusieurs variables de quantification de charge sur différentes périodes et de tester leur relation avec le risque de blessures dans notre jeu de données.

Cependant la richesse des données auxquelles pourrait avoir accès l'Observatoire pourrait permettre d'utiliser la méthode WCIE pour évaluer sans hypothèse a priori l'impact de la charge de travail des semaines précédant le match.

## 4) Sélection de variables

La charge d'entraînement peut être quantifiée par de nombreuses métriques notamment via les GPS qui fournissent de très nombreuses variables (> 1000 pour la solution Catapult). Il est donc nécessaire de sélectionner les variables à considérer dans les modèles. D'autant qu'en grande dimension des problèmes de multicollinéarité peuvent être rencontrés et la variance de l'estimateur des moindres carrés  $\hat{\beta}^{MCO}$  devient trop élevée. De plus, un modèle trop complexe aura une flexibilité qui entrainera un sur-ajustement à l'échantillon de données et aura de moins bonnes performances sur d'autres échantillons. Pour cela plusieurs approches peuvent être envisagées.

### 4.1 Méthode de régression pénalisée

Le principe des méthodes de régression pénalisées est d'ajouter un terme de pénalisation pour obtenir un estimateur avec une variance plus faible mais légèrement biaisée.

La régression Ridge utilise une pénalisation par la norme L2 correspondant à la distance euclidienne<sup>155</sup>. Elle permet de diminuer la valeur des coefficients de régression (shrinkage) mais pas jusqu'à 0. Cette méthode permet donc de répondre aux problèmes de multicollinéarité et de réduire la variance des prédictions mais pas de réaliser de la sélection de variables. De part cette limite nous n'avons pas utilisé cette méthode mais utilisé une alternative déjà évoquée dans la section 2.4 de l'état de l'art, la méthode Lasso<sup>131</sup>. Cette méthode applique cette fois une pénalité de type L1 qui permettra de forcer des coefficients à 0 et donc de réaliser de la sélection de variables. Dans la méthode Lasso, le vecteur des coefficients  $\hat{\beta}^L$  est également obtenu par minimisation de la somme des carrés des résidus mais avec une pénalité supplémentaire. Le Lasso consiste donc à résoudre :

$$\hat{\beta}^L = \arg \min_{\beta} (\|y - X\beta\|_2^2 + \lambda \|\beta\|_1)$$

où  $\lambda$  est un paramètre de régularisation. Si  $\lambda = 0$  on retrouve l'estimateur des moindres carrés. Plus  $\lambda$  augmente plus le nombre de coefficients nuls sera important.

Pour le cas des données de charge de travail, on peut généralement supposer qu'il existe des groupes de variables. Différents indicateurs calculés sur le volume de déplacement, l'intensité ou la fatigue par exemple peuvent être considérés comme appartenant à des groupes communs. De ce fait, la méthode Group Lasso est une approche permettant de sélectionner ou rejeter l'ensemble d'un groupe<sup>156</sup>. L'estimateur Group Lasso est obtenu en résolvant :

$$\hat{\beta}^{GL} = \arg \min_{\beta} (\|y - \sum_{l=1}^L X_l \beta_l\|_2^2 + \lambda \sum_{l=1}^L \sqrt{p_l} \|\beta_l\|_2)$$

où  $l = 1, \dots, L$  représente les différents groupes de variables et  $p_l$  le nombre de variables dans le groupe  $l$ . Le terme  $\sqrt{p_l}$  permet de tenir compte de la taille des groupes et d'éviter que les groupes plus importants soient privilégiés. Néanmoins l'objectif étant d'identifier les caractéristiques de la charge de travail les plus associées au risque de blessures, une sélection au sein des groupes est nécessaire. Pour cela l'estimateur Sparse Group Lasso<sup>157</sup> semble le plus approprié à la sélection des facteurs de risque de blessures incluant les nombreuses variables de charge de travail. Il est obtenu en résolvant le problème d'optimisation :

$$\hat{\beta}^{SGL} = \arg \min_{\beta} (\|y - \sum_{l=1}^L X_l \beta_l\|_2^2 + \lambda(\alpha \|\beta\|_1 + (1 - \alpha) \sum_{l=1}^L \sqrt{p_l} \|\beta_l\|_2))$$

où  $\alpha \in [0,1]$  est une combinaison convexe des pénalités du Lasso et du Group Lasso. Pour  $\alpha = 0$ ,  $\hat{\beta}^{SGL} = \hat{\beta}^{GL}$  et pour  $\alpha = 1$ ,  $\hat{\beta}^{SGL} = \hat{\beta}^L$ . Cette méthode permet ainsi de retenir les variables les plus prédictives au sein des groupes les plus prédictifs.

Cependant dans le cadre d'analyse de données longitudinales la méthode Lasso a besoin d'être adaptée. Elle a donc été développée pour des modèles linéaires généralisés<sup>158</sup> puis étendue par la suite aux modèles à effets mixtes<sup>159,160</sup>. Groll et Tutz<sup>160</sup> ont ainsi étendu l'algorithme d'optimisation du Lasso proposé par Goeman<sup>161</sup>, qui utilise un algorithme de descente du gradient, aux modèles linéaires généralisés mixtes. Contrairement aux algorithmes de Lasso proposés précédemment<sup>162,163</sup>, l'algorithme de descente du gradient permet de maximiser la log-vraisemblance pénalisée en produisant des modèles moins complexes. Quand l'algorithme se rapproche de la valeur optimale il passe sur une procédure de scoring de Fisher pour améliorer la rapidité de convergence. A l'étape finale, un modèle qui inclut toutes les variables non-nulles est estimé pour déterminer la valeur finale des coefficients. Cet algorithme a notamment pour avantage de pouvoir être utilisable sur des données avec plus de variables explicatives que les autres procédures<sup>160</sup>. Celui-ci a été implémenté dans le package R *glmmLasso*. Il offre la possibilité d'utiliser la méthode Lasso à partir de modèles linéaires généralisés mixtes donc avec une distribution logistique ou de Poisson par exemple. Néanmoins rien n'est précisé quant à la possibilité de pouvoir spécifier un terme en offset. De plus, la solution développée par Groll et Tutz permet de considérer une structure de groupe de variables mais pas de réaliser de la sélection au sein de ces groupes. Un package semblable (*glmmixedlasso*<sup>159</sup>) basé sur une approximation de Laplace de la vraisemblance pénalisée est également disponible mais la sélection au sein des groupes ne semble pas possible également. A contrario, la méthode Sparse Group Lasso peut être employée avec le package *MSGGLasso* mais il ne permet pas de spécifier des modèles avec effets aléatoires. Il n'existe à ce jour pas d'implémentation sous R de la méthode Sparse Group Lasso pour des modèles linéaires généralisés mixtes.

Nous avons donc fait le choix d'utiliser le package *glmmLasso* pour sélectionner les variables explicatives associées au risque de blessures sur les données collectées dans un club professionnel de la deuxième étude. La procédure Lasso a été réalisée sur un modèle logistique mixte avec le joueur en effet aléatoire. Le paramètre de pénalité  $\lambda$  a été déterminé de manière itérative en débutant par une valeur haute où tous les coefficients sont nuls, puis réduit pas à pas en utilisant les estimations de l'ajustement précédent comme valeur de départ pour la suivante. La valeur minimisant le BIC a ensuite été sélectionnée comme  $\lambda$  optimal. Une soixantaine de variables construites à priori ont été insérées dans le modèle parmi lesquelles des facteurs contextuels et individuels, des variables de charges journalières sur les jours précédant le match, des variables de charges cumulées et pondérées sur différentes périodes ou

encore des variables représentant une variation de charge. La procédure a été testée dans un premier temps en Group Lasso en définissant les groupes en fonction du type de représentation de la charge (cumul, variation, etc.) et dans un second temps en fonction de la variable de charge considérée (distance totale, haute intensité, RPE, etc.). Concernant la première méthodologie, aucune variable ne fut conservée par le modèle minimisant le BIC. Pour ce qui est de la seconde, seul l'âge fut conservé par la procédure. En sélectionnant un paramètre de pénalité moins restrictif les groupes de variables non nulles comprenaient majoritairement des variables non significatives. Ces résultats n'étant pas satisfaisants, nous avons dans la suite de nos analyses retiré cette structure de groupe. Néanmoins, une nouvelle fois peu de variables étaient conservées par le modèle. Une des explications possibles vient du fait qu'il y ait une corrélation importante entre toutes les variables de charge de travail insérées dans le modèle et il a été démontré que le Lasso a tendance à ne sélectionner qu'une variable aléatoirement parmi un ensemble de variables corrélées<sup>164</sup>. La fonction *glmlasso* peut aussi avoir tendance à obtenir trop de coefficients nuls<sup>165</sup>. Dans ce sens également, bien que cette méthode puisse permettre d'inclure des variables catégorielles dans les variables explicatives<sup>160,156</sup>, la variable du poste n'était pas sélectionnée tandis qu'elle était pourtant significative quand sa sélection était forcée. Malgré cela, les principales variables sélectionnées par ces approches concernaient des charges journalières. Nous avons donc fait le choix de nous concentrer sur les indicateurs de la charge de travail quotidienne des 5 jours précédant le match (Chapitre 4).

Pour résumer, la méthode Lasso nous a amené à ne conserver que peu de variables indiquant que la relation entre charge de travail et blessures est difficilement identifiable au travers de notre jeu de données. De ce fait, la structure par groupe n'a pas permis d'obtenir des résultats exploitables dans notre cas mais considérer cette structure semble recommandable par rapport à ce type de données bien qu'à ce jour il n'existe pas encore de solution pour réaliser une méthode SGLasso avec effets aléatoires.

## **4.2 Méthode de réduction de dimension**

Au-delà de la sélection, une autre approche envisageable peut être de réduire le nombre de dimensions. Ces approches sont particulièrement utiles dans le contexte où il y a de nombreuses variables explicatives. L'objectif de ces méthodes est généralement plus prédictif

mais il pourrait également permettre d'obtenir de nouveaux indicateurs permettant de synthétiser la charge de travail<sup>166</sup>.

#### 4.2.1 Régression par composantes principales

La régression sur composantes principales (PCR) consiste à résumer l'ensemble des variables explicatives  $X_1, \dots, X_p$  par un nombre réduit de combinaisons linéaires de ces variables. Le principe sera ensuite de sélectionner les  $k$  composantes principales qui vont expliquer le plus d'information via une méthode de validation croisée. Enfin, un modèle linéaire généralisé sera utilisé pour prédire le risque de blessures en fonction des composantes sélectionnées. Cette méthode pourra permettre d'améliorer les performances prédictives du modèle en réduisant les problèmes de multicolinéarité des données. Une limite néanmoins importante est que le modèle sera construit sur les composantes les plus représentatives des données de charge de travail mais pas forcément les plus en lien avec le risque de blessures.

#### 4.2.2 Partial Least Square

De même que pour l'approche PCR, la régression Partial Least Square (PLS) a pour objectif de réduire le nombre de prédicteurs en un nombre plus petit de composantes non corrélées et de réaliser une régression des moindres carrés sur ces composantes<sup>167,168</sup>. Outre maximiser la variance expliquée par les variables de charge de travail, elle va réaliser un compromis en cherchant également à maximiser la corrélation entre cette combinaison linéaire des variables explicatives et le risque de blessures. Contrairement à la méthode précédente, la régression PLS est donc une méthode supervisée puisqu'elle tient compte de l'information présente dans la variable à expliquer.

Le problème d'optimisation de la régression PLS consiste à résoudre<sup>169</sup> :

$$\max_{\|w_h\|=1} \text{cov}(X_{h-1}w_h, y)$$

où les  $w_h$  sont les poids (loadings) associés aux  $H$  composantes latentes,  $X_{h-1}$  est la matrice dite « déflatée » de  $X$ , qui consiste à retirer la variabilité déjà expliquée par  $X_h$ . La régression PLS peut s'écrire :

$$y = Tc^T + \varepsilon$$

où  $T = (t_1, \dots, t_H)$  sont les composantes latentes,  $c^T$  le vecteur des coefficients des composantes et  $\varepsilon$  le vecteur des résidus. Les composantes latentes de  $X$  sont calculées comme des combinaisons linéaires des variables d'origines tel que  $T = XW^*$  où  $W^* = (w_1, \dots, w_H)$  est la matrice des coefficients des variables  $x_j$  dans chaque composante  $t_h$ . On peut donc noter  $t_h = \sum_{j=1}^p w_{hj}^* x_j$ . Pour un modèle qui retient donc  $H$  composantes, on obtient :

$$y = \sum_{h=1}^H c_h \left( \sum_{j=1}^p w_{hj}^* x_j \right)$$

Dans le cadre d'un modèle linéaire généralisé la régression PLS peut s'écrire <sup>170</sup> :

$$g(\theta) = \sum_{h=1}^H c_h \left( \sum_{j=1}^p w_{hj}^* x_j \right)$$

Avec  $g(\cdot)$  la fonction de lien et  $\theta$  le vecteur de probabilités. L'algorithme proposé par Bastien et al.<sup>170</sup> pour un modèle PLS-GLM réside en 4 étapes :

- 1) Calcul des  $H$  composantes de la PLS  $t_h$
- 2) Régression linéaire généralisée de  $y$  sur les composantes PLS retenues
- 3) Expression de la PLS-GLM en termes de variables explicatives originales
- 4) Validation Bootstrap des coefficients dans le modèle final

Comme le cas de la régression Ridge, cette méthode va réduire les effets de certaines variables explicatives mais sans procéder à de la sélection. Ainsi, quand le nombre de variables est très grand, les effets des variables explicatives importantes pourront être masquées dans les composantes par celles plus nombreuses de potentielles variables moins influentes<sup>171</sup>. De ce fait, certaines méthodes permettent de combiner sélection et réduction de dimension.

### 4.2.3 sPLS et gPLS

Pour répondre à cet objectif de sélection de variables, une pénalisation par la norme L1 est ajoutée au problème d'optimisation de manière similaire à la régression Lasso. Cette méthode nommée Sparse Partial Least Square permettra ainsi de conserver uniquement les variables les plus prédictives dans la construction des variables latentes. Cette approche a été proposée par Lê Cao et al.<sup>172</sup> et repose sur l'utilisation d'un algorithme de décomposition en valeurs singulières (SVD)<sup>173</sup>. Elle passe donc par la décomposition d'une matrice  $M = X^T Y$

permettant l'approximation de matrices de rang inférieur. De manière itérative la régression sPLS cherchera à résoudre le problème d'optimisation :

$$\min \|M_{h-1} - w_h\|_F^2 + P_{\lambda_1, h}(w_h)$$

où  $P_{\lambda_1, h}(w_h)$  est le terme de pénalisation par la norme L1 des loadings de  $X$ .

De même que la méthode Lasso, une structure de groupe peut être définie au sein des variables. Dans ce cas le problème d'optimisation de la Group PLS est :

$$\min \sum_{l=1}^L \|M_l - w_l\|_F^2 + \lambda_1 \sum_{l=1}^L \sqrt{p_l} \|w_l\|_2$$

Enfin, au-delà de la sélection des groupes les plus prédictifs, une sélection des variables au sein de ces groupes est également possible avec la méthode Sparse Group PLS. Dans ce cas une pénalisation par la norme L2 est ajoutée en plus à la fonction objectif à minimiser :

$$\sum_{l=1}^L \|M_l - w_l\|_F^2 + \lambda_1 ((1 - \alpha_1) \sum_{l=1}^L \sqrt{p_l} \|w_l\|_2 + \alpha_1 \|w_l\|_1)$$

où l'hyper-paramètre  $\alpha_1 \in [0,1]$  permet de réaliser un compromis entre les pénalités sPLS et gPLS.

Une limite néanmoins à considérer est que la méthode PLS n'a pas encore été développée dans le cadre des modèles linéaires généralisés à effets mixtes. En effet, il n'existe pas encore de package disponible à ce jour sous R pour inclure des effets aléatoires à une régression PLS.

Ces méthodes permettent donc de réduire la quantité de variables pour notamment un meilleur pouvoir prédictif des modèles. Cependant l'interprétation des coefficients associés aux composantes pourra être plus complexe. Elles pourront être ainsi envisagées dans une approche davantage prédictive qu'explicative. Le choix de la méthode dépend donc du but de l'étude, pour une étude à visée explicative, qui est probablement l'objectif principal dans un premier temps, la sélection de variables sera recommandée. Pour ce cas, la méthode Sparse Group Lasso pour modèle linéaire généralisé mixte semble la plus adaptée à ce contexte. Néanmoins, de par le fait qu'elle ne soit à ce jour pas encore implémentée pour le logiciel R, le Lasso simple peut être une alternative intéressante.

En conclusion, plusieurs méthodologies ont été évoquées dans ce chapitre dans le but de permettre une meilleure identification des paramètres à suivre pour limiter le risque de blessures. Ces méthodes statistiques ont été testées sur les données d'un club, seulement à cette échelle, nous n'avons pas pu en extraire des conclusions robustes. Ces résultats renforcent donc le besoin d'avoir une structure indépendante rassemblant un échantillon plus important de données et pouvant éventuellement permettre de cibler certains types de blessures en particulier.



## **Chapitre 6 : Discussion**

---

## 1) Les facteurs de risque en Rugby

En amont de la construction d'un modèle d'analyse pour prédire le risque de blessure il est important d'identifier les variables à considérer dans celui-ci. Nos travaux ont ainsi permis de mettre en avant l'influence de différents facteurs de risques sur les blessures. Nous avons également pu étudier le cas spécifique des commotions et notamment remarquer une augmentation de leur incidence dans le championnat français lors des saisons 2016/2017 et 2017/2018.

Concernant le sujet des commotions nous avons constaté un effet protecteur du temps de jeu en match depuis le début de la saison. Cette conclusion va à l'inverse de celle de Rafferty et al. qui préconisent de ne pas dépasser le seuil de 25 matchs pour lequel un joueur aurait plus de risque d'avoir subi une commotion que de ne pas en avoir eu. Leur méthodologie consistait à calculer le nombre de matchs pour lequel la probabilité cumulée d'avoir une commotion dépassait le seuil arbitraire de 0.5. De ce fait, quel que soit le résultat, la conclusion aurait donc été dans le sens de limiter le nombre de matchs. D'autant plus que dans leur méthodologie, ils ont considéré une augmentation constante du risque au fil des matchs tandis que nous avons justement montré que celui-ci décroît avec l'accumulation du temps de jeu. Bien entendu, cet effet de l'accumulation du temps de jeu n'est probablement que la traduction d'une résultante qui a pour origine d'autres facteurs. On peut notamment supposer, par l'importance de la technique lors des plaquages déjà montrée dans la littérature<sup>35,29</sup>, que celle-ci indique des joueurs qui ont de meilleures habilités sur ces situations. Il est donc important de bien préparer à l'entraînement les joueurs au contact pour qu'ils puissent par la suite réduire leur risque de commotion en match.

Nos travaux ont également permis de constater que même à très court terme la programmation de la semaine pouvait avoir un impact sur le risque de blessures en match. Nous avons ainsi pu identifier au cours de la semaine les facteurs à cibler en fonction des journées d'entraînements. Les semaines d'entraînements sont composées de différentes journées avec des charges et des objectifs divers où le jeu et la préparation physique sont intelligemment associés. C'est d'autant plus le cas depuis l'essor du concept de périodisation tactique dont l'un des principes fondamentaux est que l'entraînement ne doit jamais séparer les éléments physiques, tactiques, techniques et mentaux de la préparation<sup>174</sup>. Tout d'abord appliqué au football, ce concept a été ensuite repris dans le Rugby avec notamment une adaptation proposée

par Tee et al.<sup>175</sup>. Nous avons eu la chance de travailler avec un club de haut niveau dont la programmation de la semaine semblait suivre ce modèle. Nos travaux ont ainsi permis de confirmer notamment que deux jours avant le match, où il est préconisé d'avoir une séance basée sur du travail de vitesse<sup>176,175</sup>, avoir une proportion de distance à haute intensité (supérieur à la VMA) plus élevé a un effet protecteur sur le risque de blessures en match. Le questionnaire RPE d'évaluation individuelle de la difficulté de la séance s'est révélé être très utile comme indicateur de suivi du risque. Il a en effet mené à constater qu'une charge importante en séance de musculation en début de semaine pouvait avoir un effet négatif sur le risque de blessure. Un intérêt tout particulier doit être porté à la note RPE complétée à la suite de l'entraînement de la veille de match. Sur cette journée d'entraînement qui est très standardisée, et ajusté cette fois-ci sur la charge réalisée, une note plus élevée indique un risque plus important de se blesser le lendemain. Ce questionnaire est donc un outil de mesure intéressant pour suivre le risque, que ce soit comme un substitut pour quantifier la charge ou même lorsque c'est possible, comme ressentie de cette charge perçue par le joueur en considérant celle réellement réalisée.

Dans ce sens, chaque joueur perçoit différemment la charge qui lui est imposée. C'est encore plus vrai dans le Rugby où selon leur poste, les joueurs ont des tâches bien distinctes. De plus, au sein même de ces postes il peut exister une certaine hétérogénéité entre différents profils. Il est donc important de considérer ces dissemblances dans les modèles. Nos travaux ont également permis de mettre en avant cet aspect. Nous avons pu constater des différences selon les postes avec notamment des joueurs de la charnière plus sujets aux commotions. Des distinctions ont aussi été observées selon les profils même à poste comparable. Ainsi, les plus légers étaient davantage à risque d'être commotionné en plaquant et les plus grand dans les autres circonstances de jeu et donc principalement en étant plaqué. De plus, la variance des effets aléatoires étant significatives dans les différents modèles mixtes que nous avons pu utiliser, même en ajustant sur tous ces facteurs, il subsiste toujours une variation inter-individuelle. C'est pourquoi il est important d'essayer d'individualiser la charge de travail.

Le retour à la pratique et particulièrement au match à la suite d'une blessure est également un point d'intérêt<sup>48,37</sup>. Dans la seconde étude nous avons pu constater un risque plus élevé de se blesser à nouveau pour les joueurs qui sont de retour d'une blessure ayant occasionné au moins sept jours d'indisponibilités. Bien que ce processus de retour en jeu soit contrôlé dans les clubs, il semble encore ne pas être suffisant. Pour certains cas comme les commotions, où les conséquences d'une nouvelle récurrence peuvent être importantes, des protocoles sont définis. Tout récemment, en juin 2022, World Rugby a une nouvelle fois

modifié ce protocole en obligeant notamment à un repos forcé de 7 jours minimum à la suite d'une commotion. A ce jour, de plus en plus de technologies permettent d'évaluer précisément l'état des capacités cognitives du joueur et peuvent donc être un bon moyen d'évaluer le moment propice de retour à la pratique. De manière similaire pour d'autres types de blessures, de nombreux tests physiques sont utilisés dans le but d'évaluer l'état de forme des joueurs et d'ainsi ne pas traiter la blessure comme une donnée uniquement binaire. Nous n'avons pas eu l'occasion d'avoir accès à des données de tests mais la littérature est déjà conséquente à ce sujet<sup>177,178,179,180,181</sup>.

Il existe donc de multiples facteurs qui influent sur le risque de blessures, la difficulté va être de rendre toutes ces informations exploitables en pratique en considérant notamment cet aspect multifactoriel.

## **2) Suivi du risque de blessures**

Au quotidien, les staffs doivent décider de la participation d'un joueur à un match ou un entraînement. Pour cela, ils ont donc besoin de s'appuyer sur des indicateurs pour prendre leurs décisions. Certains indicateurs couramment utilisés comme les ACWR ont été remis en cause et bien que d'autres comme les EWMA ont montré de meilleurs résultats, ils restent limités pour prévenir les blessures. Au cours de cette thèse, nous n'avons pas pu trouver un indicateur simple à implémenter qui pourrait permettre de réaliser ce suivi du risque de blessure. Bien qu'il existe à ce jour des limites quant à l'implémentation de certaines méthodes comme la Sparse Group Lasso avec mesures répétées pour déterminer les variables d'intérêts, celles que nous avons employées ne nous ont pas permis d'aboutir à des conclusions fiables. En effet, à l'échelle d'un club il semble difficile d'avoir suffisamment de puissance statistique pour tirer des conclusions robustes. Néanmoins nous avons mis en évidence les contraintes méthodologiques qu'il est nécessaire de considérer dans les analyses.

Une des premières difficultés souvent rencontrées en pratique pour réaliser un suivi longitudinal est le traitement des données manquantes. Une approche proposée par Griffin et al.<sup>66</sup> est d'imputer ces données manquantes par la moyenne de l'équipe ou d'un échantillon de joueurs. Bien que cette méthode soit simple à mettre en œuvre, dans cette thèse nous avons utilisé une méthode plus avancée qui permet notamment de considérer plusieurs variables

simultanément. L'imputation multiple par équations chaînées permet ainsi d'avoir une imputation non biaisée des valeurs manquantes. Nos travaux nous ont amené à penser qu'inclure au modèle d'imputation le joueur en effet aléatoire et la moyenne au poste de la variable à imputer semble la méthode la plus adaptée pour obtenir les meilleures estimations. Utiliser également l'appariement prédictif moyen permet d'obtenir des imputations robustes quelle que soit la distribution de la variable à imputer.

Les modèles que nous avons construits avaient avant tout un but explicatif. C'est une première étape importante pour tenter de comprendre les mécanismes qui conduisent à une blessure et pouvoir adapter certains paramètres comme la charge de travail pour réduire le risque. Outre cet aspect, l'objectif pour les staffs des équipes étant d'essayer d'anticiper la survenue de blessure, il est nécessaire d'avoir un modèle avec un bon pouvoir prédictif. Bien que ce ne soit pas l'objectif premier de notre étude, nous avons évalué les capacités prédictives de nos modèles. Nous avons obtenu un AUC de 0.61 seulement ce qui est insuffisant pour en faire un modèle de prédiction crédible. Bien que les blessures soient des événements difficilement prévisibles dans un sport de contacts comme le rugby, d'autres méthodes plus orientées sur cet objectif, comme celles explicitées dans la section 2.4, pourraient permettre d'obtenir de meilleures prédictions. La méthode Sparse Group PLS est aussi une option à envisager dans cet optique pour réduire les dimensions de l'ensemble des variables de la charge de travail. Par ailleurs, nous n'avons pas eu la possibilité d'exploiter au cours de cette thèse certaines variables comme le nombre de sprints, d'accélération, de décélération ou encore le nombre de contacts qui sont des mesures qui seraient liées avec le risque de blessure<sup>50,90,182</sup>.

Il est tout de même important qu'il y ait un suivi des blessures et des différents indicateurs de charge de travail au sein des clubs pour avoir une approche raisonnée de la préparation physique des joueurs. Néanmoins cette thèse a mis en évidence le besoin de centraliser ces données pour en tirer des conclusions robustes bénéfiques à chaque club.

### **3) Perspectives**

La blessure en Rugby est toujours un sujet d'actualité important. Pour preuve, en juillet 2022, 180 anciens joueurs victimes de commotions ont décidé d'attaquer en justice World

Rugby et les fédérations anglaises et galloises pour leur manque de protection vis-à-vis des risques de commotions. Bien que nos travaux aient permis d'identifier plusieurs pistes d'intérêt, ils ont soulevé les limites que peuvent avoir l'utilisation des données à l'échelle d'un club. La transmission des données de plusieurs clubs auprès d'un organisme indépendant est donc essentielle pour mener des analyses statistiques conduisant à des résultats fiables et généralisables. La démarche ainsi menée il y a quelques années par la Ligue Nationale de Rugby et la Fédération Française de Rugby en créant cet Observatoire Médicale du Rugby va dans le bon sens. La mise en place d'un système commun de suivi de l'activité de déplacement des joueurs dans tout le Rugby français est une réussite. Il reste néanmoins à pouvoir exploiter toutes ces données qui, associée à d'autres données comme le dossier médical informatisé, peuvent être une source d'information précieuse. Dans la dernière partie de ce manuscrit nous avons donné des pistes de méthodes statistiques adaptées pour ces données. Ainsi, la méthode d'indice cumulé d'exposition pondérée (WCIE) pourrait permettre d'estimer la forme de la relation entre la charge de travail aux différentes périodes et le risque de blessures. Ce type de méthodes et les résultats qui en découlent, notamment appliqués à différents types de blessures, peuvent être des indications intéressantes pour adapter au mieux le programme d'entraînements des joueurs. L'étude des différentes sources de données qu'elles soient GPS, RPE, d'état de forme ou encore de tests physiques doivent déjà permettre de réduire les blessures à un niveau club. Néanmoins, l'exploitation des données que pourrait recueillir l'Observatoire Médical du Rugby représente une source d'espoir important pour faire avancer significativement la recherche à ce sujet.

## **Chapitre 7 : Conclusion**

---

Cette thèse a permis de mettre en évidence l'influence de certains facteurs de risque de blessures et en particulier des commotions. Nous avons également pu identifier dans quelle mesure la charge de travail effectuée au cours de la semaine d'entraînement joue un rôle sur la survenue de blessures le jour de match. Au cours de ces travaux, nous avons constaté la difficulté pour un club d'identifier la relation entre blessure et charge de travail. De ce fait, les données que pourraient collecter l'Observatoire Médicale du Rugby est une perspective importante. Les méthodes que nous avons proposées dans ce cadre pourraient permettre d'obtenir des résultats robustes pour connaître de quelle manière la charge de travail influence le risque de blessures et ainsi que les préparateurs physiques puissent l'adapter pour réduire le nombre de blessures en Rugby.

# Bibliographie

---

1. *Global Reach of Rugby 2019*. Nielsen
2. Casson IR, Viano DC. Long-Term Neurological Consequences Related to Boxing and American Football: A Review of the Literature. *JAD*. 2019;69(4):935-952. doi:10.3233/JAD-190115
3. Tator CH. Concussions and their consequences: current diagnosis, management and prevention. *CMAJ*. 2013;185(11):975-979. doi:10.1503/cmaj.120039
4. Kemp S, West S, Brooks J, Starling L. England professional rugby injury surveillance project 2020-2021 season report. Published 2022. <https://www.englandrugby.com/participation/playing/player-welfare-rugby-safe/rugbysafe-research>
5. West SW, Starling L, Kemp S, et al. Trends in match injury risk in professional male rugby union: a 16-season review of 10 851 match injuries in the English Premiership (2002–2019): the Professional Rugby Injury Surveillance Project. *Br J Sports Med*. 2021;55(12):676-682. doi:10.1136/bjsports-2020-102529
6. Williams S, Trewartha G, Kemp SPT, et al. How Much Rugby is Too Much? A Seven-Season Prospective Cohort Study of Match Exposure and Injury Risk in Professional Rugby Union Players. *Sports Med*. 2017;47(11):2395-2402. doi:10.1007/s40279-017-0721-3
7. Cross M, Kemp S, Smith A, Trewartha G, Stokes K. Professional Rugby Union players have a 60% greater risk of time loss injury after concussion: a 2-season prospective study of clinical outcomes. *Br J Sports Med*. 2016;50(15):926-931. doi:10.1136/bjsports-2015-094982
8. West SW, Williams S, Cazzola D, Kemp S, Cross MJ, Stokes KA. Training Load and Injury Risk in Elite Rugby Union: The Largest Investigation to Date. *Int J Sports Med*. 2021;42(08):731-739. doi:10.1055/a-1300-2703
9. Williams S, Trewartha G, Kemp SPT, et al. Time loss injuries compromise team success in Elite Rugby Union: a 7-year prospective study. *Br J Sports Med*. 2016;50(11):651-656. doi:10.1136/bjsports-2015-094798
10. Borg GA. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc*. 1982;14(5):377-381.
11. Meeuwisse WH, Tyreman H, Hagel B, Emery C. A Dynamic Model of Etiology in Sport Injury: The Recursive Nature of Risk and Causation. *Clinical Journal of Sport Medicine*. 2007;17(3):215-219. doi:10.1097/JSM.0b013e3180592a48
12. Fuller CW, Molloy MG, Bagate C, et al. Consensus statement on injury definitions and data collection procedures for studies of injuries in rugby union. *British Journal of Sports Medicine*. 2007;41(5):328-331. doi:10.1136/bjism.2006.033282

13. 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*. Published online December 2, 2021. doi:10.1007/s40279-021-01603-w
14. Fuller CW, Taylor A, Kemp SPT, Raftery M. Rugby World Cup 2015: World Rugby injury surveillance study. *Br J Sports Med*. 2017;51(1):51-57. doi:10.1136/bjsports-2016-096275
15. Fuller C, Taylor A, Douglas M, Raftery M. Rugby World Cup 2019 injury surveillance study. *S Afr j sports med*. 2020;32(1):1-6. doi:10.17159/2078-516X/2020/v32i1a8062
16. López-Valenciano A, Ruiz-Pérez I, Garcia-Gómez A, et al. Epidemiology of injuries in professional football: a systematic review and meta-analysis. *Br J Sports Med*. 2020;54(12):711-718. doi:10.1136/bjsports-2018-099577
17. Saw R, Finch CF, Samra D, et al. Injuries in Australian Rules Football: An Overview of Injury Rates, Patterns, and Mechanisms Across All Levels of Play. *Sports Health*. 2018;10(3):208-216. doi:10.1177/1941738117726070
18. Gutierrez D, Roque P, Bagate C, et al. Étude épidémiologique prospective des pathologies dans le XV de France féminin de rugby de juillet 2006 à septembre 2010. *Journal de Traumatologie du Sport*. 2011;28(4):208-214. doi:10.1016/j.jts.2011.10.008
19. McLoughlin R, Liston M, Commins T, Yeomans C. The Irish Rugby Injury Surveillance Project. 2019 - 2020 Season Report. <https://www.irishrugby.ie/iris-schools-season-report-1920/>
20. Yeomans C, Kenny IC, Cahalan R, et al. The Incidence of Injury in Amateur Male Rugby Union: A Systematic Review and Meta-Analysis. *Sports Med*. 2018;48(4):837-848. doi:10.1007/s40279-017-0838-4
21. Quarrie K, Gianotti S, Murphy I. Injury Risk in New Zealand Rugby Union: A Nationwide Study of Injury Insurance Claims from 2005 to 2017. *Sports Med*. 2020;50(2):415-428. doi:10.1007/s40279-019-01176-9
22. Leahy TM, Kenny IC, Campbell MJ, et al. Injury surveillance in school Rugby: A systematic review of injury epidemiology & surveillance practices. *Physical Therapy in Sport*. 2019;38:170-178. doi:10.1016/j.ptsp.2019.05.005
23. Barden C, Hancock M, Stokes K, McKay C, Kemp SPT. Youth Rugby Injury Surveillance and Prevention Project. Season Report 2019-2020.
24. Bitchell CL, Mathema P, Moore IS. Four-year match injury surveillance in male Welsh professional Rugby Union teams. *Physical Therapy in Sport*. 2020;42:26-32. doi:10.1016/j.ptsp.2019.12.001
25. Montgomery C, Blackburn J, Withers D, Tierney G, Moran C, Simms C. Mechanisms of ACL injury in professional rugby union: a systematic video analysis of 36 cases. *Br J Sports Med*. 2018;52(15):994-1001. doi:10.1136/bjsports-2016-096425
26. Usman J, McIntosh AS. Upper limb injury in rugby union football: results of a cohort study. *Br J Sports Med*. 2013;47(6):374-379. doi:10.1136/bjsports-2012-091224

27. Roberts SP, Trewartha G, England M, Stokes KA. Collapsed scrums and collision tackles: what is the injury risk? *Br J Sports Med.* 2015;49(8):536-540. doi:10.1136/bjsports-2013-092988
28. Williams S, Trewartha G, Kemp S, Stokes K. A Meta-Analysis of Injuries in Senior Men's Professional Rugby Union. *Sports Med.* 2013;43(10):1043-1055. doi:10.1007/s40279-013-0078-1
29. Tucker R, Raftery M, Fuller GW, Hester B, Kemp S, Cross MJ. A video analysis of head injuries satisfying the criteria for a head injury assessment in professional Rugby Union: a prospective cohort study. *Br J Sports Med.* 2017;51(15):1147-1151. doi:10.1136/bjsports-2017-097883
30. Tierney GJ, Denvir K, Farrell G, Simms CK. The Effect of Tackler Technique on Head Injury Assessment Risk in Elite Rugby Union. *Medicine & Science in Sports & Exercise.* 2018;50(3):603-608. doi:10.1249/MSS.0000000000001461
31. Sobue S, Kawasaki T, Hasegawa Y, et al. Tackler's head position relative to the ball carrier is highly correlated with head and neck injuries in rugby. *Br J Sports Med.* 2018;52(6):353-358. doi:10.1136/bjsports-2017-098135
32. Burger N, Lambert MI, Viljoen W, et al. Mechanisms and Factors Associated With Tackle-Related Injuries in South African Youth Rugby Union Players. *Am J Sports Med.* 2017;45(2):278-285. doi:10.1177/0363546516677548
33. Davidow D, Quarrie K, Viljoen W, et al. Tackle technique of rugby union players during head impact tackles compared to injury free tackles. *Journal of Science and Medicine in Sport.* 2018;21(10):1025-1031. doi:10.1016/j.jsams.2018.04.003
34. Suzuki K, Nagai S, Iwai K, Furukawa T, Takemura M. How does the situation before a tackle influence a tackler's head placement in rugby union?: application of the decision tree analysis. *BMJ Open Sport Exerc Med.* 2021;7(1):e000949. doi:10.1136/bmjsem-2020-000949
35. Cross MJ, Tucker R, Raftery M, et al. Tackling concussion in professional rugby union: a case-control study of tackle-based risk factors and recommendations for primary prevention. *Br J Sports Med.* 2019;53(16):1021-1025. doi:10.1136/bjsports-2017-097912
36. Stokes KA, Locke D, Roberts S, et al. Does reducing the height of the tackle through law change in elite men's rugby union (The Championship, England) reduce the incidence of concussion? A controlled study in 126 games. *Br J Sports Med.* 2021;55(4):220-225. doi:10.1136/bjsports-2019-101557
37. Williams S, Trewartha G, Kemp S, et al. Subsequent Injuries and Early Recurrent Diagnoses in elite Rugby Union Players. *Int J Sports Med.* 2017;38(10):791-798. doi:10.1055/s-0043-114862
38. Bourne MN, Opar DA, Williams MD, Shield AJ. Eccentric Knee Flexor Strength and Risk of Hamstring Injuries in Rugby Union: A Prospective Study. *Am J Sports Med.* 2015;43(11):2663-2670. doi:10.1177/0363546515599633

39. Lindsay A, Draper N, Lewis J, Giese SP, Gill N. Positional demands of professional rugby. *European Journal of Sport Science*. 2015;15(6):480-487. doi:10.1080/17461391.2015.1025858
40. Thornton HR, Delaney JA, Duthie GM, Dascombe BJ. Importance of Various Training-Load Measures in Injury Incidence of Professional Rugby League Athletes. *International Journal of Sports Physiology and Performance*. 2017;12(6):819-824. doi:10.1123/ijsp.2016-0326
41. Gabbett TJ, Ullah S, Finch CF. Identifying risk factors for contact injury in professional rugby league players – Application of a frailty model for recurrent injury. *Journal of Science and Medicine in Sport*. 2012;15(6):496-504. doi:10.1016/j.jsams.2012.03.017
42. Gabbett TJ. The Development and Application of an Injury Prediction Model for Noncontact, Soft-Tissue Injuries in Elite Collision Sport Athletes. *Journal of Strength and Conditioning Research*. 2010;24(10):2593-2603. doi:10.1519/JSC.0b013e3181f19da4
43. Cousins BEW, Morris JG, Sunderland C, Bennett AM, Shahtahmassebi G, Cooper SB. Match and Training Load Exposure and Time-Loss Incidence in Elite Rugby Union Players. *Front Physiol*. 2019;10:1413. doi:10.3389/fphys.2019.01413
44. Kemp S, West S, Brooks J, Cross M. England professional rugby injury surveillance project 2017-2018 season report. Published 2019. <https://www.englandrugby.com/participation/playing/player-welfare-rugby-safe/rugbysafe-research>
45. Soligard T, Schwelunus M, Alonso JM, et al. How much is too much? (Part 1) International Olympic Committee consensus statement on load in sport and risk of injury. *Br J Sports Med*. 2016;50(17):1030-1041. doi:10.1136/bjsports-2016-096581
46. Foster C, Florhaug JA, Franklin J, et al. A New Approach to Monitoring Exercise Training: *Journal of Strength and Conditioning Research*. 2001;15(1):109-115. doi:10.1519/00124278-200102000-00019
47. Cross MJ, Williams S, Trewartha G, Kemp SPT, Stokes KA. The Influence of In-Season Training Loads on Injury Risk in Professional Rugby Union. *International Journal of Sports Physiology and Performance*. 2016;11(3):350-355. doi:10.1123/ijsp.2015-0187
48. Blanch P, Gabbett TJ. Has the athlete trained enough to return to play safely? The acute:chronic workload ratio permits clinicians to quantify a player's risk of subsequent injury. *Br J Sports Med*. 2016;50(8):471-475. doi:10.1136/bjsports-2015-095445
49. Gabbett TJ. The training—injury prevention paradox: should athletes be training smarter and harder? *Br J Sports Med*. 2016;50(5):273-280. doi:10.1136/bjsports-2015-095788
50. Cummins C, Welch M, Inkster B, et al. Modelling the relationships between volume, intensity and injury-risk in professional rugby league players. *Journal of Science and Medicine in Sport*. 2019;22(6):653-660. doi:10.1016/j.jsams.2018.11.028
51. Griffin A, Kenny IC, Comyns TM, Lyons M. The Association Between the Acute:Chronic Workload Ratio and Injury and its Application in Team Sports: A Systematic Review. *Sports Med*. 2020;50(3):561-580. doi:10.1007/s40279-019-01218-2

52. Andrade R, Wik EH, Rebelo-Marques A, et al. Is the Acute: Chronic Workload Ratio (ACWR) Associated with Risk of Time-Loss Injury in Professional Team Sports? A Systematic Review of Methodology, Variables and Injury Risk in Practical Situations. *Sports Med.* 2020;50(9):1613-1635. doi:10.1007/s40279-020-01308-6
53. Maupin D, Schram B, Canetti E, Orr R. The Relationship Between Acute: Chronic Workload Ratios and Injury Risk in Sports: A Systematic Review. *OAJSM.* 2020;Volume 11:51-75. doi:10.2147/OAJSM.S231405
54. Wang C, Vargas JT, Stokes T, Steele R, Shrier I. Analyzing Activity and Injury: Lessons Learned from the Acute:Chronic Workload Ratio. *Sports Med.* 2020;50(7):1243-1254. doi:10.1007/s40279-020-01280-1
55. Lolli L, Batterham AM, Hawkins R, et al. The acute-to-chronic workload ratio: an inaccurate scaling index for an unnecessary normalisation process? *Br J Sports Med.* 2019;53(24):1510-1512. doi:10.1136/bjsports-2017-098884
56. Gabbett TJ, Hulin B, Blanch P, Chapman P, Bailey D. To Couple or not to Couple? For Acute:Chronic Workload Ratios and Injury Risk, Does it Really Matter? *Int J Sports Med.* 2019;40(09):597-600. doi:10.1055/a-0955-5589
57. Impellizzeri FM, Tenan MS, Kempton T, Novak A, Coutts AJ. Acute:Chronic Workload Ratio: Conceptual Issues and Fundamental Pitfalls. *International Journal of Sports Physiology and Performance.* 2020;15(6):907-913. doi:10.1123/ijsp.2019-0864
58. Impellizzeri FM, Woodcock S, Coutts AJ, Fanchini M, McCall A, Vigotsky AD. What Role Do Chronic Workloads Play in the Acute to Chronic Workload Ratio? Time to Dismiss ACWR and Its Underlying Theory. *Sports Med.* 2021;51(3):581-592. doi:10.1007/s40279-020-01378-6
59. Hulin BT, Gabbett TJ, Lawson DW, Caputi P, Sampson JA. The acute:chronic workload ratio predicts injury: high chronic workload may decrease injury risk in elite rugby league players. *Br J Sports Med.* 2016;50(4):231-236. doi:10.1136/bjsports-2015-094817
60. Murray NB, Gabbett TJ, Townshend AD, Blanch P. Calculating acute:chronic workload ratios using exponentially weighted moving averages provides a more sensitive indicator of injury likelihood than rolling averages. *Br J Sports Med.* 2017;51(9):749-754. doi:10.1136/bjsports-2016-097152
61. Moussa I, Leroy A, Sauliere G, Schipman J, Toussaint JF, Sedeaud A. Robust Exponential Decreasing Index (REDI): adaptive and robust method for computing cumulated workload. *BMJ Open Sport Exerc Med.* 2019;5(1):e000573. doi:10.1136/bmjsem-2019-000573
62. Rubin DB. Inference and missing data. *Biometrika.* 1976;63(3):581-592. doi:10.1093/biomet/63.3.581
63. Baraldi AN, Enders CK. An introduction to modern missing data analyses. *Journal of School Psychology.* 2010;48(1):5-37. doi:10.1016/j.jsp.2009.10.001

64. Van Praag BMS, Dijkstra TK, Van Velzen J. Least-squares theory based on general distributional assumptions with an application to the incomplete observations problem. *Psychometrika*. 1985;50(1):25-36. doi:10.1007/BF02294145
65. Enders CK. *Applied Missing Data Analysis*. Guilford Press; 2010.
66. Griffin A, Kenny IC, Comyns TM, et al. Training load monitoring in team sports: a practical approach to addressing missing data. *Journal of Sports Sciences*. Published online May 10, 2021:1-11. doi:10.1080/02640414.2021.1923205
67. Batista GEAPA, Monard MC. An analysis of four missing data treatment methods for supervised learning. *Applied Artificial Intelligence*. 2003;17(5-6):519-533. doi:10.1080/713827181
68. Hu LY, Huang MW, Ke SW, Tsai CF. The distance function effect on k-nearest neighbor classification for medical datasets. *SpringerPlus*. 2016;5(1):1304. doi:10.1186/s40064-016-2941-7
69. Santos MS, Abreu PH, Wilk S, Santos J. How distance metrics influence missing data imputation with k-nearest neighbours. *Pattern Recognition Letters*. 2020;136:111-119. doi:10.1016/j.patrec.2020.05.032
70. Beale EML, Little RJA. Missing Values in Multivariate Analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1975;37(1):129-145. doi:10.1111/j.2517-6161.1975.tb01037.x
71. Olinsky A, Chen S, Harlow L. The comparative efficacy of imputation methods for missing data in structural equation modeling. *European Journal of Operational Research*. 2003;151(1):53-79. doi:10.1016/S0377-2217(02)00578-7
72. Buuren S van, Groothuis-Oudshoorn K. **mice** : Multivariate Imputation by Chained Equations in R. *J Stat Soft*. 2011;45(3). doi:10.18637/jss.v045.i03
73. Mislavy RJ, Little RJA, Rubin DB. Statistical Analysis with Missing Data. *Journal of Educational Statistics*. 1991;16(2):150. doi:10.2307/1165119
74. Rubin DB, Schenker N. Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Nonresponse. *Journal of the American Statistical Association*. 1986;81(394):366-374. doi:10.1080/01621459.1986.10478280
75. Rubin DB. Multiple Imputation after 18+ Years. *Journal of the American Statistical Association*. 1996;91(434):473-489. doi:10.1080/01621459.1996.10476908
76. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley; 1987.
77. Asif M. Imputation Methods for Multiple Regression with Missing Heteroscedastic Data. *Thailand statistician*. 2022;20(01):15.
78. Rubin DB, ed. Frontmatter. In: *Wiley Series in Probability and Statistics*. John Wiley & Sons, Inc.; 1987:i-xxix. doi:10.1002/9780470316696.fmatter

79. O’Kelly M. Multiple Imputation and Its Application. James Carpenter and Michael Kenward (2013). Chichester: John Wiley & Sons. 345 pages, ISBN: 9780470740521.: Book review. *Biom J.* 2014;56(2):352-353. doi:10.1002/bimj.201300188
80. von Hippel PT. How Many Imputations Do You Need? A Two-stage Calculation Using a Quadratic Rule. *Sociological Methods & Research.* 2020;49(3):699-718. doi:10.1177/0049124117747303
81. Bartlett JW, Seaman SR, White IR, Carpenter JR. Multiple imputation of covariates by fully conditional specification: accommodating the substantive model. *Stat Methods Med Res.* 2015;24(4):462-487. doi:10.1177/0962280214521348
82. Schafer JL. *Analysis of Incomplete Multivariate Data.*; 1997.
83. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Statist Med.* 2011;30(4):377-399. doi:10.1002/sim.4067
84. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009;338(jun29 1):b2393-b2393. doi:10.1136/bmj.b2393
85. Little RJA. Missing-Data Adjustments in Large Surveys. Published online 2021:11.
86. Rubin DB. Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations. Published online 2021:9.
87. Marshall A, Altman DG, Holder RL. Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. *BMC Med Res Methodol.* 2010;10(1):112. doi:10.1186/1471-2288-10-112
88. Kleinke K. Multiple Imputation Under Violated Distributional Assumptions: A Systematic Evaluation of the Assumed Robustness of Predictive Mean Matching. *Journal of Educational and Behavioral Statistics.* 2017;42(4):371-404. doi:10.3102/1076998616687084
89. Borg DN, Nguyen R, Tierney NJ. *Missing Data: Current Practice in Football Research and Recommendations for Improvement.* SportRxiv; 2020. doi:10.31236/osf.io/fhwcu
90. Colby MJ, Dawson B, Heasman J, Rogalski B, Gabbett TJ. Accelerometer and GPS-Derived Running Loads and Injury Risk in Elite Australian Footballers. *Journal of Strength and Conditioning Research.* 2014;28(8):2244-2252. doi:10.1519/JSC.0000000000000362
91. Bowen L, Gross AS, Gimpel M, Bruce-Low S, Li FX. Spikes in acute:chronic workload ratio (ACWR) associated with a 5–7 times greater injury rate in English Premier League football players: a comprehensive 3-year study. *Br J Sports Med.* 2020;54(12):731-738. doi:10.1136/bjsports-2018-099422
92. Benson LC, Stilling C, Owoeye OBA, Emery CA. Evaluating Methods for Imputing Missing Data from Longitudinal Monitoring of Athlete Workload. *jsportscimed.* Published online March 5, 2021:188-196. doi:10.52082/jssm.2021.188

93. Rogalski B, Dawson B, Heasman J, Gabbett TJ. Training and game loads and injury risk in elite Australian footballers. *Journal of Science and Medicine in Sport*. 2013;16(6):499-503. doi:10.1016/j.jsams.2012.12.004
94. Jacqmin-Gadda H, Sibillot S, Proust C, Molina JM, Thiébaud R. Robustness of the linear mixed model to misspecified error distribution. *Computational Statistics & Data Analysis*. 2007;51(10):5142-5154. doi:10.1016/j.csda.2006.05.021
95. Jacqmin-Gadda H, Commenges D, Dartigues J. [Analysis of longitudinal Gaussian data with missing data on the response variable]. *Rev Epidemiol Sante Publique*. 1999;47(6):525-534.
96. Caparrós T, Casals M, Solana Á, Peña J. Low External Workloads Are Related to Higher Injury Risk in Professional Male Basketball Games. :9.
97. Williams S, Trewartha G, Cross MJ, Kemp SPT, Stokes KA. Monitoring What Matters: A Systematic Process for Selecting Training-Load Measures. *International Journal of Sports Physiology and Performance*. 2017;12(s2):S2-101-S2-106. doi:10.1123/ijsp.2016-0337
98. West SW, Williams S, Kemp SPT, Eager R, Cross MJ, Stokes KA. Training Load, Injury Burden, and Team Success in Professional Rugby Union: Risk Versus Reward. *Journal of Athletic Training*. 2020;55(9):960-966. doi:10.4085/1062-6050-0387.19
99. Bengtsson H, Ekstrand J, Waldén M, Hägglund M. Match Injury Rates in Professional Soccer Vary With Match Result, Match Venue, and Type of Competition. *Am J Sports Med*. 2013;41(7):1505-1510. doi:10.1177/0363546513486769
100. Colby MJ, Dawson B, Heasman J, et al. Preseason Workload Volume and High-Risk Periods for Noncontact Injury Across Multiple Australian Football League Seasons. *Journal of Strength and Conditioning Research*. 2017;31(7):1821-1829. doi:10.1519/JSC.0000000000001669
101. Duchateau L, Janssen P, Kezic I, Fortpied C. Evolution of recurrent asthma event rate over time in frailty models. *J Royal Statistical Soc C*. 2003;52(3):355-363. doi:10.1111/1467-9876.00409
102. Ullah S, Gabbett TJ, Finch CF. Statistical modelling for recurrent events: an application to sports injuries. *Br J Sports Med*. 2014;48(17):1287-1293. doi:10.1136/bjsports-2011-090803
103. Mirsafaei Rizi R, Yeung SS, Stewart NJ, Yeung EW. Risk factors that predict severe injuries in university rugby sevens players. *Journal of Science and Medicine in Sport*. 2017;20(7):648-652. doi:10.1016/j.jsams.2016.11.022
104. Nielsen RO, Bertelsen ML, Ramskov D, et al. Time-to-event analysis for sports injury research part 2: time-varying outcomes. *Br J Sports Med*. 2019;53(1):70-78. doi:10.1136/bjsports-2018-100000
105. Amorim LD, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. *International Journal of Epidemiology*. 2015;44(1):324-333. doi:10.1093/ije/dyu222

106. Andersen PK, Gill RD. Cox's Regression Model for Counting Processes: A Large Sample Study. *Ann Statist.* 1982;10(4). doi:10.1214/aos/1176345976
107. Moulton L. Multivariate time-to-event models for studies of recurrent childhood diseases. *International Journal of Epidemiology.* 1997;26(6):1334-1339. doi:10.1093/ije/26.6.1334
108. Turkson AJ, Simpson T, Addor JA. An Introspective Overview of the Dynamics of Recurrent Events Data Analysis. *AJPAS.* Published online December 14, 2021:144-162. doi:10.9734/ajpas/2021/v15i430371
109. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika.* 1981;68(2):373-379. doi:10.1093/biomet/68.2.373
110. Wei LJ, Lin DY, Weissfeld L. Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions. *Journal of the American Statistical Association.* 1989;84(408):1065-1073. doi:10.1080/01621459.1989.10478873
111. Metcalfe C, Thompson SG. Wei, Lin and Weissfeld's marginal analysis of multivariate failure time data: should it be applied to a recurrent events outcome? *Stat Methods Med Res.* 2007;16(2):103-122. doi:10.1177/0962280206071926
112. Kelly PJ, Lim LLY. Survival analysis for recurrent event data: an application to childhood infectious diseases. *Statist Med.* 2000;19(1):13-33. doi:10.1002/(SICI)1097-0258(20000115)19:1<13::AID-SIM279>3.0.CO;2-5
113. Lintu MK, Kamath A. Performance of recurrent event models on defect proneness data. *Ann Oper Res.* Published online December 7, 2020. doi:10.1007/s10479-020-03884-2
114. Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography.* 1979;16(3):439-454. doi:10.2307/2061224
115. Clayton DG. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. :11.
116. Finkelstein DM, Schoenfeld DA, Stamenovic E. ANALYSIS OF MULTIPLE FAILURE TIME DATA FROM AN AIDS CLINICAL TRIAL. *Statist Med.* 1997;16(8):951-961. doi:10.1002/(SICI)1097-0258(19970430)16:8<951::AID-SIM547>3.0.CO;2-5
117. Pickles A, Crouchley R. A comparison of frailty models for multivariate survival data. *Statist Med.* 1995;14(13):1447-1461. doi:10.1002/sim.4780141305
118. Rondeau V, Filleul L, Joly P. Nested frailty models using maximum penalized likelihood estimation. *Statist Med.* 2006;25(23):4036-4052. doi:10.1002/sim.2510
119. Tibshirani S, Friedman H. Valerie and Patrick Hastie. :764.
120. Patel N, Upadhyay S. Study of Various Decision Tree Pruning Methods with their Empirical Comparison in WEKA. *IJCA.* 2012;60(12):20-25. doi:10.5120/9744-4304

121. Maszczyk T, Duch W. Comparison of Shannon, Renyi and Tsallis Entropy Used in Decision Trees. In: Rutkowski L, Tadeusiewicz R, Zadeh LA, Zurada JM, eds. *Artificial Intelligence and Soft Computing – ICAISC 2008*. Vol 5097. Lecture Notes in Computer Science. Springer Berlin Heidelberg; 2008:643-651. doi:10.1007/978-3-540-69731-2\_62
122. Han J, Kamber M. *Data Mining: Concepts and Techniques*. 3rd ed. Elsevier; 2012.
123. Breiman L. Random forests. *Machine Learning*. 2001;45(1):5-32. doi:10.1023/A:1010933404324
124. Freund Y, Schapire RE. Experiments with a New Boosting Algorithm. :9.
125. Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann Statist*. 2001;29(5). doi:10.1214/aos/1013203451
126. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Published online August 13, 2016:785-794. doi:10.1145/2939672.2939785
127. Vapnik VN. *The Nature of Statistical Learning Theory*. Qing hua da xue chu ban she; 2000. Accessed August 17, 2022. <https://doi.org/10.1007/978-1-4757-3264-1>
128. Jones SP, Jansen R, Fusaro RL. Preliminary Investigation of Neural Network Techniques to Predict Tribological Properties. *Tribology Transactions*. 1997;40(2):312-320. doi:10.1080/10402009708983660
129. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. *jair*. 2002;16:321-357. doi:10.1613/jair.953
130. Schubach M, Re M, Robinson PN, Valentini G. Imbalance-Aware Machine Learning for Predicting Rare and Common Disease-Associated Non-Coding Variants. *Sci Rep*. 2017;7(1):2959. doi:10.1038/s41598-017-03011-5
131. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1996;58(1):267-288. doi:10.1111/j.2517-6161.1996.tb02080.x
132. Van Eetvelde H, Mendonça LD, Ley C, Seil R, Tischer T. Machine learning methods in sport injury prediction and prevention: a systematic review. *J EXP ORTOP*. 2021;8(1):27. doi:10.1186/s40634-021-00346-x
133. Carey DL, Ong K, Whiteley R, Crossley KM, Crow J, Morris ME. Predictive Modelling of Training Loads and Injury in Australian Football. *International Journal of Computer Science in Sport*. 2018;17(1):49-66. doi:10.2478/ijcss-2018-0002
134. Rodas G, Osaba L, Arteta D, Pruna R, Fernández D, Lucia A. Genomic Prediction of Tendinopathy Risk in Elite Team Sports. *International Journal of Sports Physiology and Performance*. 2020;15(4):489-495. doi:10.1123/ijsp.2019-0431
135. Wells G, Shea B, O'Connell D, Peterson J. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. [https://www.ohri.ca/programs/clinical\\_epidemiology/oxford.asp](https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp)

136. Brand JPL. *Development, Implementation and Evaluation of Multiple Imputation Strategies for the Statistical Analysis of Incomplete Data Sets*. Erasmus Universiteit Rotterdam; 1999.
137. Raghunathan TE, Lepkowski JM, Hoewyk JV, Solenberger P. A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models. (12):12.
138. Van Buuren S, Brand JPL, Groothuis-Oudshoorn CGM, Rubin DB. Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*. 2006;76(12):1049-1064. doi:10.1080/10629360600810434
139. Beato M, Devereux G, Stiff A. Validity and Reliability of Global Positioning System Units (STATSports Viper) for Measuring Distance and Peak Speed in Sports. *Journal of Strength and Conditioning Research*. 2018;32(10):2831-2837. doi:10.1519/JSC.0000000000002778
140. Checkoway H, Rice CH. Time-weighted averages, peaks, and other indices of exposure in occupational epidemiology. *Am J Ind Med*. 1992;21(1):25-33. doi:10.1002/ajim.4700210106
141. Stranges S. Lifetime Cumulative Exposure to Secondhand Smoke and Risk of Myocardial Infarction in Never Smokers: Results From the Western New York Health Study, 1995-2001. *Arch Intern Med*. 2006;166(18):1961. doi:10.1001/archinte.166.18.1961
142. Breslow NE, Lubin JH, Marek P, Langholz B. Multiplicative Models and Cohort Analysis. *Journal of the American Statistical Association*. 1983;78(381):1-12. doi:10.1080/01621459.1983.10477915
143. Thomas DC. Models for Exposure-Time-Response Relationships with Applications to Cancer Epidemiology. :32.
144. Vacek PM. ASSESSING THE EFFECT OF INTENSITY WHEN EXPOSURE VARIES OVER TIME. *Statist Med*. 1997;16(5):505-513. doi:10.1002/(SICI)1097-0258(19970315)16:5<505::AID-SIM424>3.0.CO;2-Z
145. Abrahamowicz M, Bartlett G, Tamblyn R, du Berger R. Modeling cumulative dose and exposure duration provided insights regarding the associations between benzodiazepines and injuries. *Journal of Clinical Epidemiology*. 2006;59(4):393-403. doi:10.1016/j.jclinepi.2005.01.021
146. Hauptmann M, Wellmann J, Lubin JH, Rosenberg PS, Kreienbrock L. Analysis of Exposure-Time-Response Relationships Using a Spline Weight Function. *Biometrics*. 2000;56(4):1105-1108. doi:10.1111/j.0006-341X.2000.01105.x
147. Sylvestre MP, Abrahamowicz M. Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Statist Med*. 2009;28(27):3437-3453. doi:10.1002/sim.3701

148. Perperoglou A, Sauerbrei W, Abrahamowicz M, Schmid M. A review of spline function procedures in R. *BMC Med Res Methodol.* 2019;19(1):46. doi:10.1186/s12874-019-0666-3
149. Gomez-Rubio V. Generalized Additive Models: An Introduction with R (2nd Edition). *J Stat Soft.* 2018;86(Book Review 1). doi:10.18637/jss.v086.b01
150. Stone CJ. [Generalized Additive Models]: Comment. *Statist Sci.* 1986;1(3). doi:10.1214/ss/1177013607
151. Akaike H. A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika.* 1979;66(2):237-242. doi:10.1093/biomet/66.2.237
152. De Boor C. *A Practical Guide to Splines: With 32 Figures.* Rev. ed. Springer; 2001.
153. Sylvestre MP, Abrahamowicz M, Čapek R, Tamblyn R. Assessing the cumulative effects of exposure to selected benzodiazepines on the risk of fall-related injuries in the elderly. *Int Psychogeriatr.* 2012;24(4):577-586. doi:10.1017/S1041610211002031
154. Efron B, Tibshirani R. *An Introduction to the Bootstrap.* Nachdr. Chapman & Hall; 1998.
155. Hoerl AE, Kennard RW. Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics.* 1970;12(1):69-82. doi:10.1080/00401706.1970.10488635
156. Yuan M, Lin Y. Model Selection and Estimation in Regression with Grouped Variables. *Journal of the Royal Statistical Society Series B (Statistical Methodology).* 2006;68(1):49-67.
157. Simon N, Friedman J, Hastie T, Tibshirani R. A Sparse-Group Lasso. *Journal of Computational and Graphical Statistics.* 2013;22(2):231-245. doi:10.1080/10618600.2012.681250
158. Meier L, van de Geer S, Bühlmann P. The group lasso for logistic regression. :19.
159. Schelldorfer J, Meier L, Bühlmann P. GLMMLasso: An Algorithm for High-Dimensional Generalized Linear Mixed Models Using  $\ell_1$ -Penalization. *Journal of Computational and Graphical Statistics.* 2014;23(2):460-477. doi:10.1080/10618600.2013.773239
160. Groll A, Tutz G. Variable selection for generalized linear mixed models by L1-penalized estimation. *Stat Comput.* 2014;24(2):137-154. doi:10.1007/s11222-012-9359-z
161. Goeman JJ. L<sub>1</sub> Penalized Estimation in the Cox Proportional Hazards Model. *Biom J.* Published online November 24, 2009:NA-NA. doi:10.1002/bimj.200900028
162. Shevade SK, Keerthi SS. A simple and efficient algorithm for gene selection using sparse logistic regression. *Bioinformatics.* 2003;19(17):2246-2253. doi:10.1093/bioinformatics/btg308
163. Kim Y, Kim J. Gradient LASSO for feature selection. :8.

164. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J Royal Statistical Soc B*. 2005;67(2):301-320. doi:10.1111/j.1467-9868.2005.00503.x
165. Xie Y, Xu L, Li J, et al. Spatial Variable Selection and An Application to Virginia Lyme Disease Emergence. *Journal of the American Statistical Association*. 2019;114(528):1466-1480. doi:10.1080/01621459.2018.1564670
166. Weaving D, Beggs C, Dalton-Barron N, Jones B, Abt G. Visualizing the Complexity of the Athlete-Monitoring Cycle Through Principal-Component Analysis. *International Journal of Sports Physiology and Performance*. 2019;14(9):1304-1310. doi:10.1123/ijsp.2019-0045
167. Wold H. Estimation of principal components and related models by iterative least squares. *Journal of Multivariate Analysis*. 1966:391-420.
168. Tenenhaus M. L'approche PLS. *Revue de statistique appliquée*. 1999;tome 47(2):5-40.
169. Burnham AJ, Viveros R, MacGregor JF. Frameworks for latent variable multivariate regression. *J Chemometrics*. 1996;10(1):31-45. doi:10.1002/(SICI)1099-128X(199601)10:1<31::AID-CEM398>3.0.CO;2-1
170. Bastien P, Vinzi VE, Tenenhaus M. PLS generalised linear regression. *Computational Statistics & Data Analysis*. 2005;48(1):17-46. doi:10.1016/j.csda.2004.02.005
171. Chun H, KeleÅŸ S. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2010;72(1):3-25. doi:10.1111/j.1467-9868.2009.00723.x
172. Lê Cao KA, Rossouw D, Robert-Granié C, Besse P. A Sparse PLS for Variable Selection when Integrating Omics Data. *Statistical Applications in Genetics and Molecular Biology*. 2008;7(1). doi:10.2202/1544-6115.1390
173. Shen H, Huang JZ. Sparse principal component analysis via regularized low rank matrix approximation. *Journal of Multivariate Analysis*. 2008;99(6):1015-1034. doi:10.1016/j.jmva.2007.06.007
174. Delgado-Bordonau JL, Mendez-Villanueva A. Tactical Periodization: Mourinho's best-kept secret? :7.
175. Tee JC, Ashford M, Piggott D. A Tactical Periodization Approach for Rugby Union. *Strength & Conditioning Journal*. 2018;40(5):1-13. doi:10.1519/SSC.0000000000000390
176. Tsoukos A, Veligekas P, Brown LE, Terzis G, Bogdanis GC. Delayed Effects of a Low-Volume, Power-Type Resistance Exercise Session on Explosive Performance. *Journal of Strength and Conditioning Research*. 2018;32(3):643-650. doi:10.1519/JSC.0000000000001812
177. McKeag DB. Preseason Physical Examination for the Prevention of Sports Injuries: *Sports Medicine*. 1985;2(6):413-431. doi:10.2165/00007256-198502060-00003

178. Maulder P, Cronin J. Horizontal and vertical jump assessment: reliability, symmetry, discriminative and predictive ability. *Physical Therapy in Sport*. 2005;6(2):74-82. doi:10.1016/j.ptsp.2005.01.001
179. Tee JC, Klingbiel JFG, Collins R, Lambert MI, Coopoo Y. Preseason Functional Movement Screen Component Tests Predict Severe Contact Injuries in Professional Rugby Union Players. *Journal of Strength and Conditioning Research*. 2016;30(11):3194-3203. doi:10.1519/JSC.0000000000001422
180. Kenney RA. Reviewing the Use of Injury Screening Assessments and Identifying Risk of Injury. :49.
181. Hewett TE, Myer GD, Ford KR, et al. Biomechanical Measures of Neuromuscular Control and Valgus Loading of the Knee Predict Anterior Cruciate Ligament Injury Risk in Female Athletes: A Prospective Study. *Am J Sports Med*. 2005;33(4):492-501. doi:10.1177/0363546504269591
182. Fletcher BD, Twist C, Haigh JD, Brewer C, Morton JP, Close GL. Season-long increases in perceived muscle soreness in professional rugby league players: role of player position, match characteristics and playing surface. *Journal of Sports Sciences*. 2016;34(11):1067-1072. doi:10.1080/02640414.2015.1088166