université PARIS-SACLAY

NNT : 2018SACLS594

Thèse de doctorat

UNIVERSITÉ PARIS SUD
Comprendre le monde, construire l'avenir

Analyse du rapport bénéfices risques et des déterminants de la performance dans les sports collectifs

Thèse de doctorat de l'Université Paris-Saclay préparée à l'Institut de Recherche bioMédicale et d'Epidémiologie du Sport

École doctorale n°420 – Santé Publique – EDSP
Biostatistiques

Thèse présentée et soutenue à Paris, le 19 décembre 2018, par

Guillaume Saulière

Composition du Jury :

Mme. Katia Collomp PU, *Université d'Orléans*	Présidente
M. Franck Brocherie PhD, *INSEP*	Examinateur
M. Emmanuel Orhant Médecin, *Fédération Française de Football*	Examinateur
Mme. Adeline Leclercq-Samson PU, *Université de Grenoble*	Rapportice
M. Jean-Christophe Thalabard PUPH, *Université Paris Descartes*	Rapporteur
M. Jérôme Dedecker PU, *Université Paris Descartes*	Co-Directeur de thèse
M. Jean-François Toussaint PUPH, *Université Paris Descartes*	Directeur de thèse


UNIVERSITÉ
PARIS
SUD

université
PARIS-SACLAY

irmes
Institut de recherche biomédicale et d'épidémiologie du sport

INSEP

Remerciements

A la fin d'un doctorat, une seule personne est diplômée, alors qu'une thèse n'a rien d'un exercice individuel, mais résulte bien plus d'un travail collectif dont la réussite dépend de multiples facteurs. Je souhaite donc remercier l'ensemble des personnes impliquées dans cette aventure :

Les institutions indispensable à l'existence même de cette thèse: l'Ecole Doctorale de Santé Publique, l'Université Paris-Sud et l'Institut de Recherche bio-Médicale et d'Epidémiologie du Sport, laboratoire de recherche accueilli à l'Institut du Sport, de l'Expertise et de la Performance.

Messieurs les professeurs Jean-François Toussaint et Jérôme Dedecker. Votre codirection s'est révélée parfaitement complémentaire et vous m'avez offert à vous deux tout ce qu'il fallait pour mener à bien cette entreprise. Merci pour votre confiance jusque dans les ultimes instants.

Madame Adeline Leclercq-Samson et Monsieur Jean-Christophe Thalabard d'avoir accepté d'expertiser ce travail. Merci pour vos remarques et conseils. Madame Katia Collomp et Messieurs Franck Brocherie et Emmanuel Orhant d'avoir accepté d'être membres de mon jury de thèse en qualité de présidente et d'examinateurs. A vous cinq vous avez permis une discussion riche et extrêmement pertinente sur ce travail ainsi que de nombreuses pistes de réflexions très intéressantes pour l'avenir.

Un énorme merci à tous les *Jeans* de l'Irmes, qui ont le pouvoir de transformer de longues journées passées devant un écran en quelque chose de bien plus sympathique, et de travestir des congrès en remake palois, marseillais ou havrais de *Very Bad Trip* :

Geoffroy, nous travaillons ensemble depuis mon stage et j'espère simplement que ça va continuer ainsi !

Adrien, merci de partager ta passion du rugby, ton expérience de *sport scientist* m'a grandement aidé, merci pour la relecture du manuscrit et pour m'avoir ouvert les yeux sur ce qu'était réellement un petit déjeuner bien équilibré.

Juliana, merci de m'avoir accueilli dans ton bureau tropical au début puis à la fin de cette thèse.

Andy Zidaaaaaaaaane, merci d'avoir apporté un peu de sérieux et de calme à cet open space du bas, et pour ces histoires improbables que tu as pu nous raconter en lendemain de soirée.

Julien, un immense merci pour ton aide dans la dernière ligne droite. A se demander si sans ton escale de 2 ans au Vietnam je n'aurais pas terminer cette thèse en 6 mois, haha…

Adrien, chronologiquement je suis le doctorant irmessien qui t'a succédé, ton expérience, ta force et ta persévérance m'ont bien inspiré.

Laurie-Anne, merci pour ta contribution à la publication des articles qui composent cette thèse et de m'avoir fait découvrir l'importance de la nutrition et de la couleur du pipi dans le haut niveau.

Hélène *la boss*, merci pour la résolution de problèmes dont tu es la seule à avoir les solutions et les conseils que tu as pu m'apporter tout au long de cette thèse.

Philippe, merci pour l'intérêt que tu as porté sur ces travaux, et pour la jonquille : « on est en finaaaaaaaale !!! ».

Issa, merci pour les interludes musicaux, les découvertes de la scène rap francophone, l'instant wikipédia de 15h15, les pauses « brésilienne » dehors dans le froid,…

Thibault la moustache, merci pour la bonne humeur apportée sur le terrain de foot, les blagues douteuses qui arrivent comme une cerise et l'aide dans la résolution de casse-têtes propre à R.

Joana, merci d'avoir partagé ta passion pour le bon chocolat.

Arthur, merci pour les débats enflammés sur ce pauvre petit *p* qui n'a rien demandé, on pourrait régler ça sur le terrain de tennis.

Quentin de la Roche Carotte, merci sincèrement d'apporter un peu de naïveté, de candeur et de bienveillance dans cette équipe.

Nicolas, merci pour cette tranche de jambon.

A Grégory mon LCA, parti trop tôt.

Merci aux stagiaires que j'ai encadrés et qui m'ont bien aidé durant cette thèse : Camille, Imad, Naoufal, Rachelle, Roman.

Je souhaite vous remercier, vous qui sans forcément vous en rendre compte avez participé grandement à cette réussite. On se supporte depuis la maternelle, le collège, le lycée, la P1, le QG ou encore le master. Merci pour tous ces moments passés ensemble loin de la thèse, à boire des pintes à Paris, à Bruxelles, à la Rochelle, autour d'une table de poker ou d'un FIFA, le long d'un canal ou dans un parc, après un don du sang ou pendant un blind-test, à « danser », à m'écouter me plaindre des gens cités ci-dessus, à célébrer des anniversaires, des fiançailles, des mariages, des premiers boulots, des thèses, parfois à faire un peu de sport, à rénover un appartement,…, à vivre de bonnes choses. On a grandi ensemble, et c'est pas fini !

Merci à mes parents pour m'avoir permis d'arriver où j'en suis, de m'avoir toujours soutenu dans mes choix et laisser faire ce que je souhaitais sans trop poser de questions. Merci à mon frère et ma belle-sœur pour ma nièce, à ma sœur pour sa fraicheur. J'ai la chance d'avoir une famille assez géniale. Merci mes grands-parents pour votre soutien, votre confiance et le soleil. Merci à mon oncle à mon grand-père. Merci à toute ma famille, le plus gros moteur dans cette expérience. Merci aussi à ma belle-famille pour tous les moments de détente qu'on a pu partager ces dernières années.

Enfin, merci à toi Anne-Claire. Ta force, ta volonté et ton calme m'ont aidé tout long de ces dernières années à me dépasser et à franchir chaque étape.

« Le succès ne prend sens que lorsqu'on le partage avec les siens. »

Antoine Valentinelli, Cinq doigts

Liste des productions scientifiques issues de la thèse

Dedecker, J., & **Saulière, G**. (2017). 'The Mann–Whitney U-statistic for α-dependent sequences', *Mathematical Methods of Statistics*, 26/2: 111–33.

Saulière, G., Dedecker, J., Marquet, L.-A., Rochcongar, P., Toussaint, J.-F., & Berthelot, G. (2017). 'Z-scores-based methods and their application to biological monitoring: an example in professional soccer players', *Biostatistics (Oxford, England)*.

Saulière, G., Dedecker, J., Toussaint, J. F., & Sedeaud, A. (2018). 'Quantifying collective performance in rugby union'. *In writting*.

Sedeaud, A., **Saulière, G**., Marquet, L.-A., Del Vecchio, S., Bar-Hen, A., & Toussaint, J.-F. (2017). 'Collective effectiveness in the XV de France: selections and time matter', *European Journal of Sport Science*, 17/6: 656–64.

Demande de brevet européen

Method for detecting abnormal values of biomarker. **Guillaume Saulière**, Jérôme Dedecker, Geoffroy Berthelot. Réf : 17306583.0 – 1118. 2017

Liste des communications orales

Analyse de la performance collective au rugby, l'exemple du XV de France. **Guillaume Saulière**, Laurie-Anne Marquet, Julien Schipman, Andy Marc, Avner Bar-Hen, Jean-François Toussaint. Symposium de l'IRMES, 2015

Blessures dans le football européen. **Guillaume Saulière**, Andy Marc, Jean-François Toussaint. Symposium de l'IRMES, 2016

Mesure de l'expérience collective et évolution de son impact sur la performance dans le rugby à XV. **Guillaume Saulière**, Jérôme Dedecker, Jean-François Toussaint, Adrien Sedeaud. Symposium de l'IMES, 2017

Mesure de l'expérience collective et évolution de son impact sur la performance dans le rugby à XV. **Guillaume Saulière**, Jérôme Dedecker, Jean-François Toussaint, Adrien Sedeaud. Symposium de l'IMES, 2017

Etude sur les blessures en Ligue 1 de football 2011-2014. **Guillaume Saulière**, Naoufal Abdallah, Dylan Omoike, Andy Marc Adrien Sedeaud, Jean-François Toussaint. 12ème Journée du médical, INSEP 2015

Collective performance in rugby union : XV de France VS New-Zealand. **Guillaume Saulière**, Adrien Sedeaud, Laurie-Anne Marquet, Andy Marc, Avner Bar-Hen, Jérôme Dedecker, Jean-François Toussaint. 21st annual Congress of the European College of Sport Science, 2016.

Collective performance in rugby union. **Guillaume Saulière**, Adrien Sedeaud, Laurie-Anne Marquet, Avner Bar-Hen, Jérôme Dedecker, Jean-François Toussaint. Conference Sports Physique, Ecole Polytechinque, 2016.

Analyse de la performance collective au rugby, l'exemple du XV de France. **Guillaume Saulière**, Laurie-Anne Marquet, Julien Schipman, Andy Marc, Avner Bar-Hen, Jean-François Toussaint. 8ème édition du congrès commun de la SFMES et de la SFTS, 2015

Liste des autres publications scientifiques

Antero, J., **Saulière, G**., Marck, A., & Toussaint, J.-F. (2018). 'A Medal in the Olympics Runs in the Family: A Cohort Study of Performance Heritability in the Games History', *Frontiers in Physiology*, 9. DOI: 10.3389/fphys.2018.01313

Berthelot, G., Sedeaud, A., Marck, A., Antero-Jacquemin, J., Schipman, J., **Saulière, G**., Marc, A., et al. (2015). 'Has Athletic Performance Reached its Peak?', *Sports Medicine (Auckland, N.Z.)*, 45/9: 1263–71. DOI: 10.1007/s40279-015-0347-2

Hourcade, J. C., **Saulière, G**., Noirez, P., Toussaint, J. F., & Desgorces, F. D. (2017). 'Quatre mois de charge d'entraînement globale et par exercice chez le footballeur professionnel', *Science & Sports*, 32/4: 221–8. DOI: 10.1016/j.scispo.2017.03.004

Marc, A., Sedeaud, A., Schipman, J., Antero, J. J., **Saulière, G**., Kryger, K. O., & Toussaint, J. F. (2015). 'Geographic enrollment of the top 100 in athletics running events from 1996 to 2012', *The Journal of Sports Medicine and Physical Fitness.*

Marc, A., Sedeaud, A., Schipman, J., Antero-Jacquemin, J., **Saulière, G**., Kryger, K., & Toussaint, J.-F. (2015). 'Geographic enrollment of the top 100 in athletics running events from 1996 to 2012', *Journal of Sports Medicine and Physical Fitness*, 57/4: 418–425.

Marck, A., Antero, J., Berthelot, G., **Saulière, G**., Jancovici, J.-M., Masson-Delmotte, V., Boeuf, G., et al. (2017). 'Are We Reaching the Limits of Homo sapiens?', *Frontiers in Physiology*, 8. DOI: 10.3389/fphys.2017.00812

Moussa, I., **Saulière, G**., Sène, J. M., Toussaint, J. F., & Sedeaud, A. (n.d.). 'Training load, fatigue level and injury: a short report on French high-level judokas. (Under review at International Journal of Sports Physiology and Performance)'.

Schipman J, **Saulière G**, Sedeaud A, Deschamps T, Ovigneur H, Maillet H, Berthelot G, et al. (2015). *Indice de masse corporelle et condition physique chez 49 600 collégiens et lycéens de six régions françaises, 2007-2014. Numéro thématique. Activité physique ou sportive : des bénéfices pour la santé à tout âge*.

Sedeaud, A., Sène, J. M., Krantz, N., **Saulière, G**., Moussa, I., & Toussaint, J. F. (2018). 'L'importance de la quantification de la charge d'entraînement : exemple d'un modèle', *Science & Sports*, 33/1: 22–32. DOI: 10.1016/j.scispo.2017.06.011

« Vaincu par l'exemple, maugréant tout bas, Porthos étendit la main, et les quatre amis répétèrent d'une seule voix la formule dictée par d'Artagnan :

Tous pour un, un pour tous. »

Alexandre Dumas,

Les Trois Mousquetaires

Sommaire

I. Introduction Générale

Le monde du sport, est, comme la santé, l'économie, la finance, un terrain fertile à la production massive de données (1,2). Les Fédérations sportives et les clubs se voient de plus en plus confrontés à des problématiques initialement rencontrées par les entreprises : comment rester compétitif, fidéliser les licenciés, gérer une marque. Logiquement, les données sont utilisées pour aider à répondre à ces problématiques sportives, que nous aborderons exclusivement ici, et extra-sportives.

Démocratisées depuis quelques années, les données ont investi massivement le monde du sport de haut niveau, le nombre de données produites, leurs types, la fréquence à laquelle elles sont distribuées augmentent parallèlement aux évolutions technologiques. Les sources se multiplient, les performances physiques des athlètes sont de mieux en mieux retranscrites, de plus en plus finement numérisées, de même que leurs capacités physiologiques sont plus clairement expliquées (3–18). Il existe des dizaines d'entreprises proposant des outils de géolocalisation, d'accéléromètres tridimensionnels, de tracking des athlètes, d'analyses vidéo, de cardio-fréquencemètre…, quantifiant et catégorisant le déplacement des joueurs sur un terrain, leurs interactions, leurs statistiques. La professionnalisation du sport a structuré un environnement autour des athlètes, environnement composé d'entraineurs, de médecins, de kinésithérapeutes, de préparateurs physiques, de chercheurs. Chaque corps de métier apportant des informations potentiellement convertibles génère des données complémentaires. Chaque membre du staff devenant ainsi demandeur et consommateur de données.

La numérisation du sport de haut niveau amène un avantage considérable mais présente aussi certaines limites. L'avantage indiscutable est l'apport d'informations permettant une meilleure compréhension, une contextualisation et une interprétation des performances observées à travers ces filtres technologiques innovants. Le risque est l'arrivée de masses de données aux mains d'équipes non outillées pour répondre de façon pertinente aux questions posées. Le fait de ne pas disposer de *sport* ou *data scientist* intégré au staff permettant une visualisation scientifique, rationnelle et utile de ces masses de données diminue l'adhésion des staffs à participer au temps chronophage de la collecte de toutes ces données (19).

Répondre aux problématiques de terrain, aider les chercheurs de la très haute performance ou les médecins dans l'identification des risques de blessure sont les objectifs affichés des producteurs de données. Or la réalité est plus complexe, la seule réception de données multiples ne permet pas de répondre directement aux questions posées. Trop souvent les

données sont livrées brutes aux staffs, sans cadre ni explication de comment les lire, les confronter à d'autres informations et les interpréter. La production de données doit être contextualisée, accompagnée et passée au filtre des descriptions et associations statistiques dans un premier temps. En amont une réflexion commune à l'entourage du sportif et au sportif lui-même, doit permettre l'émergence d'une question. Viendra ensuite la production et la collecte de données nécessaires pour y répondre. Production et collecte qui doivent être préparées en amont afin d'organiser le stockage des données dans une base dont l'architecture optimisera le traitement. Suit alors l'étape d'analyse des données qui fournira une information claire, ciblée, adaptée, aidant à la prise de décision. Pour arriver à la mise en place d'un outil utile à l'aide à la décision, de nombreuses étapes de validation sont nécessaires (20,21). Une fois question, protocole, choix du type de données, choix des outils de mesures fiables valides et reproductibles et analyses descriptives réalisés ; une démarche de profilage individuel des données doit être mis en place. En effet, pour des recueils identiques les associations entre les variables sont différenciées. Cette thèse illustrera différentes étapes de ce processus spécifique à la recherche dans le sport de haut niveau, principalement par l'apport de méthodes biostatistiques pour mieux comprendre les déterminants de la performance et minimiser l'exposition aux risques de blessure dans les sports collectifs.

Les données produites par les sports collectifs sont particulières et diffèrent de celles issues des sports individuels (22). Les performances ne sont pas uniquement métriques, plus difficilement compréhensibles, deux matches de football avec le même résultat peuvent avoir deux physionomies, deux histoires complètement différentes. De même, contrairement à de nombreux sports individuels, il existe une infinité de possibilités différentes pour remporter un match alors que pour gagner un 100m, il n'en existe qu'une : courir plus vite que les autres. La complexité de la technique de course d'Usain Bolt pour remporter l'Or en 9 secondes 58 centièmes n'est pas questionnée. La qualité, la force et la vitesse de chaque foulée résultent de milliers d'heures d'entrainement. Les aptitudes physiques et physiologiques à leur pic, optimisées par une minimisation des risques de blessures, ont permis à Bolt d'établir ce record (7,13,23,24). Mais la compréhension de cette performance et celles d'une équipe de sport collectif ne posent pas les mêmes questions. Celles, au pluriel, car une performance collective résulte des relations d'équilibre et d'interactions entre les performances de chacun des joueurs qui composent l'équipe, le tout en adaptation permanente aux joueurs adverses et leurs propres interrelations (25–27).

Nous nous intéresserons ici principalement au football et au rugby à 15, deux sports collectifs dit « d'invasion » dans lesquels deux équipes se disputent la possession d'une balle au sein d'une zone de jeu. Résultent de cette lutte de nombreux échanges entre les joueurs, leurs adversaires et la balle. Les données spatio-temporelles issues des matches, les analyses vidéo effectuées a posteriori, les statistiques comme le nombre de frappes, de tacles, de plaquages, de mêlées, de passes ont un intérêt évident dans l'explication de la performance de l'équipe à un instant *t* (28–43). Contextualiser et préparer a priori cette performance via l'intégration d'informations issues de réflexions communes aux différents membres du staff est encore plus important. C'est l'analyse de données nécessaires à la préparation des matches au sens large qui va nous intéresser. Un sport collectif est une activité qui demande à un groupe de sportifs individuels de travailler ensemble. La réussite du groupe demande alors deux étapes : la préparation spécifique de chaque joueur et la mise en place d'une cohésion collective. De ce constat résulte le fait que les données produites par les sports collectifs vont devoir être analysées collectivement ou individuellement, selon leur nature. Une individualisation de l'étude de certains paramètres sera indispensable à leur compréhension alors que d'autres devront être observés à l'échelle du groupe.

Ces deux facettes de la performance dans les sports collectifs seront étudiées dans cette thèse à travers le développement et l'application d'outils biostatistiques adaptés à leur nature.

II. Objectif général de la thèse

L'objectif général de la thèse est d'aider à la compréhension des déterminants de la performance et de l'état de santé des sportifs dans les sports collectifs. Cet objectif est réalisé à travers le développement et l'application d'outils biostatistiques adaptés à la dimension individuelle et collective des sports collectifs.

III. Partie 1: Développement de nouveaux outils d'analyses statistiques pour application dans le sport *: l'individualisation au service de la performance*

1) Introduction générale

Comme établi précédemment, la compréhension des performances d'un sport collectif nécessite l'observation des individus et du groupe qu'ils composent. L'individualisation du suivi des joueurs est le thème traité dans cette première partie, à travers deux études distinctes.

La première est consacrée à l'analyse de données issues du suivi longitudinal de marqueurs biologiques chez des footballeurs de haut niveau. Ce suivi est mis en place dans le but de surveiller l'état de santé des sportifs. La deuxième aborde la correction d'un outil statistique qui permettra, entre autre, de comparer les performances de deux athlètes suivis sur une période définie.

Ces deux études s'inscrivent dans le processus de compréhension de la performance d'un collectif à travers l'observation et l'analyse de données longitudinales personnelles contextualisant la santé et la performance des individus.

2) Etude 1 : Méthode de détection de valeur atypique au sein d'un suivi longitudinal de marqueurs biologiques.

a) Introduction

La recherche constante du très haut niveau et de l'amélioration des capacités physiques et techniques des athlètes est indissociable de l'enchainement de séances d'entrainement à haute intensité et de compétitions. Dans sa recherche de progrès, le sportif de haut niveau se situe constamment sur une « ligne de crête » poussant ses capacités physiologiques à leurs confins et en conséquence au plus proches du basculement physiopathologiques. Les institutions sportives, quant à elles, sont de plus en plus désireuses de préserver la santé des athlètes et recherchent des outils nouveaux permettant de détecter précocement des atypicités cliniques et biologiques, synonymes de risque imminent. Atypicité et anormalité réfèrent dans notre contexte à un écart d'une valeur au sein d'une série longitudinale. Dans notre cas, les variables étudiées sont des biomarqueurs comparés au reste de la série. Si dans ce suivi longitudinal une ou des valeurs sont jugées singulières alors une investigation clinique peut être déclenchée pour comprendre les causes et prévenir les potentielles conséquences de cet écart à la norme individuelle (44–46).

Les staffs médicaux s'appuient sur ce suivi systématique des marqueurs biologiques au cours du temps pour précéder les alertes cliniques. Cette méthode de surveillance produit des données longitudinales individuelles dont l'analyse nécessite le développement de nouveaux outils biostatistiques. Des outils simples et reproductibles permettront ainsi une analyse fine des biomarqueurs individuels.

Ainsi, la détection de valeurs anormales dans le sport de haut niveau est nécessaire pour préserver l'intégrité physique des athlètes et l'équité des compétitions, permettant aussi de révéler certaines pratiques dopantes. Avec ce double objectif, l'Agence Mondiale Antidopage (AMA) a adopté l'*Athletic Biological Passport* (APB), principal outil à ce jour utilisé par les différentes instances dirigeantes sportives pour l'analyse du suivi longitudinal de leurs athlètes (44,47–49). L'APB est une représentation statistique du suivi longitudinal de ces marqueurs biologiques. L'idée est d'individualiser ce suivi sous la forme d'un passeport pour chaque athlète permettant l'élaboration de valeurs de références personnalisées. La méthode utilisée se base sur une approche bayésienne. Le calcul des seuils de références se fait à partir de distribution a priori du marqueur biologique, obtenue dans une population de référence (50,51). Avec l'ajout des valeurs mesurées chez un athlète donné, le calcul des seuils intègre les paramètres individuels de dispersion aux paramètres de la distribution populationnelle, personnalisant ainsi les bornes de détection. Cependant cette approche bayésienne possède quelques limites (52). La première est l'implication de paramètres issus de populations de références dans le calcul des seuils individuels. L'intégration rapide des paramètres individuels dans le calcul des seuils balaye le potentiel problème inhérent à la vérité biologique qui est que la variabilité interindividuelle est supérieure à la variabilité intra-individuelle (53–56). Cependant ce recours indispensable à des valeurs issues de population de références pose un problème logistique non négligeable : il faudrait constituer de larges populations de références dans les conditions standardisées et contrôlées la cohérence spécifique à chaque sport, discipline, poste,…

C'est dans ce contexte et en partenariat avec la Fédération Française de Football que nous avons développé des méthodes permettant de détecter des valeurs anormales au sein d'une série de mesures longitudinales individuelles, cherchant à mettre en évidence les variations significatives des principaux paramètres sériques.

b) Z-scores-based methods and their application to biological monitoring: an example in professional soccer players (47–49,53,54,58–86)

Biostatistics (2019) **20**, 1, *pp.* 48–64
doi:10.1093/biostatistics/kxx044
Advance Access publication on November 15, 2017

Z-scores-based methods and their application to biological monitoring: an example in professional soccer players

GUILLAUME SAULIÈRE*

Institut National du Sport de l'Expertise et de la Performance, Institut de Recherche bioMédicale et d'Épidémilogie du Sport (IRMES), France, Université Paris-Sud Orsay, France and Université Paris Descartes, EA 7329 Paris, France

guillaume.sauliere@insep.fr

JÉRÔME DEDECKER

Université Paris Descartes, Sorbonne Paris Cité, Laboratoire MAP5 (UMR 8145), Paris, France

LAURIE-ANNE MARQUET

Institut National du Sport de l'Expertise et de la Performance, Institut de Recherche bioMédicale et d'Épidémilogie du Sport (IRMES), France and Université Paris Descartes, EA 7329 Paris, France

PIERRE ROCHCONGAR†

Fédération Française de Football, Ligue de Football Professionnel (LFP) Paris, France

JEAN-FRANCOIS TOUSSAINT

Institut National du Sport de l'Expertise et de la Performance, Institut de Recherche bioMédicale et d'Épidémiologie du Sport (IRMES), France, Université Paris Descartes, EA 7329 Paris, France and Assistance Publique—Hôpitaux de Paris, Centre d'Investigation en Médecine du Sport, Hôtel-Dieu (CIMS), Paris, France

GEOFFROY BERTHELOT

Institut National du Sport de l'Expertise et de la Performance, Institut de Recherche bioMédicale et d'Épidémiologie du Sport (IRMES), France, Université Paris Descartes, EA 7329 Paris, France and REsearch LAboratory for Interdisciplinary Studies (RELAI), Paris, France

SUMMARY

The clinical and biological follow-up of individuals, such as the biological passport for athletes, is typically based on the individual and longitudinal monitoring of hematological or urine markers. These follow-ups aim to identify abnormal behavior by comparing the individual's biological samples to an established baseline. These comparisons may be done via different ways, but each of them requires an appropriate

*To whom correspondence should be addressed.
†Deceased 16 December 2016.

extra population to compute the significance levels, which is a non-trivial issue. Moreover, it is not necessarily relevant to compare the measures of a biomarker of a professional athlete to that of a reference population (even restricted to other athletes), and a reasonable alternative is to detect the abnormal values by considering only the other measurements of the same athlete. Here we propose a simple adaptive statistic based on maxima of Z-scores that does not rely on the use of an extra population. We show that, in the Gaussian framework, it is a practical and relevant method for detecting abnormal values in a series of observations from the same individual. The distribution of this statistic does not depend on the individual parameters under the null hypothesis, and its quantiles can be computed using Monte Carlo simulations. The proposed method is tested on the 3-year follow-up of ferritin, serum iron, erythrocytes, hemoglobin, and hematocrit markers in 2577 elite male soccer players. For instance, if we consider the abnormal values for the hematocrit at a 5% level, we found that 5.57% of the selected cohort had at least one abnormal value (which is not significantly different from the expected false-discovery rate). The approach is a starting point for more elaborate models that would produce a refined individual baseline. The method can be extended to the Gaussian linear model, in order to include additional variables such as the age or exposure to altitude. The method could also be applied to other domains, such as the clinical patient follow-up in monitoring abnormal values of biological markers.

Keywords: Biological monitoring; Biological passport; Clinical follow-up; Longitudinal monitoring; Z-score.

1. Introduction

The systematic monitoring of biological variables over time allows for the early detection of physiological or biological changes. Such longitudinal studies are currently used for the identification of atypical biological observations in elite athletes (see in particular Saugy *and others*, 2014; Silva *and others*, 2013; Sottas *and others*, 2007, 2011; Zorzoli *and others*, 2014; Zorzoli and Rossi, 2010; Damsgaard *and others*, 2006; Parisotto *and others*, 2000; Lobigs *and others*, 2017b). It is known that the biological markers are impacted by both the frequency and intensity of training, the environment or health issues (Malcovati *and others*, 2003; Nikolaidis *and others*, 2014; Thirup, 2003; Hecksteden *and others*, 2016b; Lombardi *and others*, 2011; Sharpe *and others*, 2002). This biological monitoring allows for determining the baseline value and its associated variability, provided a sufficient follow-up, and a large cohort. Several sport organizations and researchers already developed such approaches in order to establish the biological profile of elite athletes for doping or medical concerns (Sallet *and others*, 2008; Malm *and others*, 2016). A well-known implementation is the athlete biological passport (APB) that is designed to indirectly reveal the effects of doping (Zorzoli and Rossi, 2010; Robinson *and others*, 2007; Pottgiesser *and others*, 2011). It was officially introduced in 2009 and consists of a normalized follow-up of the athlete's hematological parameters. The data generated by the APB can be analyzed using a Bayes procedure (Sottas *and others*, 2007; Pottgiesser *and others*, 2011), which uses both an *a priori* distribution on the individual parameters and the progressive inclusion of individual's values to learn the characteristics of the individual. This Bayes procedure is now investigated in other biomarkers, including the monitoring of muscle recovery (Hecksteden *and others*, 2016a). However, Malcovati *and others* (2003) and Egger *and others* (2016) underlined that the inter-individual variability is higher than the intra-individual one, suggesting that the use and design of population data for estimating intra-individual variability is an issue. Other concerns were raised by Banfi *and others* (2011) who pointed out that the APB methodology does not take any environmental seasonality into account. We shall say more on this question, and discuss other possible applications of a Bayes procedure to our context, at the end of Section 5. The aim of this work is to provide a tractable, yet easy to implement, method for detecting abnormal values in a series of measures from the same individual, complementary to biological analyses. We introduce three new methods based on appropriate Z-scores and their multivariate versions to take care of several markers simultaneously

(as suggested by Parisotto *and others*, 2000; Julian *and others*, 2017; Hecksteden *and others*, 2016b; Malm *and others*, 2016). The simplicity of these methods allows for a clear understanding of the inner workings. These methods are evaluated on the 3-year follow-up of five biological markers for elite male soccer players practicing in the French soccer league. In the section below, we describe the methodological context and explain our three methods.

2. METHODS

Let $X_1, \dots, X_n$ be n independent Gaussian random variables, with $X_i \sim \mathcal{N}(\mu_i, \sigma^2)$. We shall first consider a preliminary method (Method 0), to see if the "new" observation x_n is abnormal, given $x_1, \dots, x_{n-1}$. Method 1 generalizes this approach, by testing if one of the observations in the sample is abnormal given the other ones. Method 2 is designed to detect if a series of consecutive observations in the sample are abnormal, given the rest of the sample. For these three methods (Method 0, Method 1, and Method 2), we propose a simple statistic T_n whose distribution is known under the appropriate null hypothesis, and a decision rule associated with this statistic.

Method 3 is somewhat different, because we want to account for possible seasonality (the behavior of the biomarkers may differ according to the time of year (summer or winter for the data set presented in Section 4.1)). In that case, we split the sample into two subsamples $X_1, \dots X_{n_1}$ and $Y_1, \dots, Y_{n_2}$ with $n = n_1 + n_2$, and we denote by μ_{X_i} and μ_{Y_j} the expectations of X_i and Y_j. Here again, we propose a simple statistic T_n to detect if the observation x_n is abnormal given the previous observations in the same subsample.

In the following subsections, we describe these methods in detail. For methods 0, 1, and 3, we shall also present a multivariate extension in the case where $X_1, \dots, X_n$ are n independent Gaussian random vectors. These extensions allow us to deal with the situation where several biomarkers are systematically measured on the same individual, in order to avoid multi-test issues.

2.1. *Method 0*

To detect if the observation x_n is abnormal given the mean of the previous observations, the usual way is to consider the statistic

$$T_n = \frac{X_n - \bar{X}_{n-1}}{\hat{\sigma}_{n-1}\sqrt{1 + \frac{1}{n-1}}},$$

where $\bar{X}_{n-1}$ and $\hat{\sigma}^2_{n-1}$ are the empirical mean and variance of the sample $X_1, \dots, X_{n-1}$, that is

$$\bar{X}_{n-1} = \frac{1}{n-1}\sum_{k=1}^{n-1} X_k \quad \text{and} \quad \hat{\sigma}^2_{n-1} = \frac{1}{n-2}\sum_{k=1}^{n-1}(X_k - \bar{X}_{n-1})^2.$$

Sottas *and others* (2007) quoted that, for small n, one cannot use this statistic to detect an abnormal value of X_n by comparing the observed value $|t_n|$ of $|T_n|$ to the quantile $1 - \alpha/2$ of the $\mathcal{N}(0, 1)$ distribution. This remark is correct, because for small n, the distribution of T_n is far from the $\mathcal{N}(0, 1)$ distribution. However, this issue can easily be overcome by taking the quantile of the exact distribution of T_n under the null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_n = \mu$. Indeed, under H_0, it follows from Cochran's Theorem that $\hat{\sigma}_{n-1}$ is independent of the couple $(X_n, \bar{X}_{n-1})$. From this simple fact, we easily deduce that, under H_0, the statistic T_n has the Student$(n - 2)$ distribution.

The decision rule is then the following: at level α, we consider that the value x_n is abnormal if $|t_n| > c_{\alpha/2,n-2}$, where $c_{\alpha/2,n-2}$ is the quantile of order $1-\alpha/2$ of the Student$(n-2)$ distribution. This method can be applied as soon as $n \geq 3$.

Remark. If there are N different samples (corresponding to N different individuals) Sharpe *et al.* propose to use the statistic

$$Z_n = \frac{X_n - \bar{X}_{n-1}}{\hat{\sigma}_{\text{uni}}\sqrt{1+\frac{1}{n-1}}},$$

for each sample $X_1, \ldots, X_n$ (the size n may not be the same for all samples). Here, the standard error σ_{uni} is assumed to be the same for all the variables of all the samples, and $\hat{\sigma}_{\text{uni}}$ is the estimator of σ_{uni} built from all the variables. In that case, if N is large, the distribution of Z_n is close to the $\mathcal{N}(0,1)$ distribution, and, at level α, the quantile $1-\alpha/2$ of the $\mathcal{N}(0,1)$ can be used to detect abnormal values. However, as quoted in Sottas *and others* (2007), the assumption that σ_{uni} is the same for all the variables is unrealistic for many biomarkers (in general, the variance depends on the individuals).

Multivariate extension. In this paragraph, we briefly present the multivariate extension of Method 0. Here $X_1, \ldots, X_n$ are n independent Gaussian $\mathbb{R}^d$-valued random vectors, where $X_i \sim \mathcal{N}(\mu_i, C)$ and C is assumed to be invertible. Again, we want to detect if the observation x_n is abnormal given the previous observations. This corresponds to the case where the same d markers are observed at each medical visit. One could of course apply the previous procedure to each marker separately, but a single decision rule is preferable in order to avoid multi-test issues. In that case, one can use the statistic

$$T_n = \frac{(n-1)}{nd}(X_n - \bar{X}_{n-1})' C_{n-1}^{-1}(X_n - \bar{X}_{n-1}),$$

where

$$\bar{X}_{n-1} = \frac{1}{n-1}\sum_{k=1}^{n-1} X_k \quad \text{and} \quad C_{n-1} = \frac{1}{n-1-d}\sum_{k=1}^{n-1}(X_k - \bar{X}_{n-1})(X_k - \bar{X}_{n-1})'.$$

Under $H_0 : \mu_1 = \mu_2 = \cdots = \mu_n = \mu$, the statistic T_n has the Fisher$(d, n-1-d)$ distribution (see Proposition 8.14 in Eaton, 1983).

2.2. *Method 1*

Given the observations $x_1, \ldots, x_n$ from the sample $X_1, \ldots, X_n$ there seems to be no good reason to consider only the case where the last observation could be abnormal given the previous ones. A much more natural question is: how to detect an abnormal observation among all the observations? Even if one focuses only on the last observation, this question is of interest, because the presence of one (or more) abnormal observation in the past could lead to a wrong decision about the new observation x_n.

The first way is to iterate Method 0 as follows: one checks if x_3 is abnormal given x_1, x_2, then one checks if x_4 is abnormal given x_1, x_2, x_3, and so on up to the last step where one checks if x_n is abnormal given $x_1, \ldots, x_{n-1}$. This approach is not satisfactory for at least two reasons: First, the procedure is not invariant by time inversion (the forward and backward procedures can lead to two distinct results). Next, there is a multi-test problem, because $n-2$ tests are performed. If each test is performed at level α, then the level of the series of tests will be greater than α. Moreover the level correction seems difficult, since these tests are not independent, and the Bonferroni correction is known to be too conservative.

Hence, to answer this question, a global test is preferable. We propose the statistic

$$T_n = \max_{i\in\{1,\dots,n\}} \left| \frac{X_i - \bar{X}_{n,-i}}{\hat{\sigma}_{n,-i}\sqrt{1+\frac{1}{n-1}}} \right|,$$

where

$$\bar{X}_{n,-i} = \frac{1}{n-1}\sum_{k=1,k\neq i}^{n} X_k \quad \text{and} \quad \hat{\sigma}^2_{n,-i} = \frac{1}{n-2}\sum_{k=1,k\neq i}^{n} (X_k - \bar{X}_{n,-i})^2.$$

This method, based on the maximum of n (non-independent) variables with Student$(n-2)$ distribution, can be seen as a straightforward extension of Method 0.

Under the null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_n = \mu$, one can easily check that the distribution P_n of T_n does not depend on the unknown parameters (μ, σ^2), and can therefore be tabulated.

The decision rule is then the following: at level α, we decide that there is (at least) one abnormal observation if $t_n > c_{\alpha,n}$, where $c_{\alpha,n}$ is such that $P_n([c_{\alpha,n}, \infty[) = \alpha$. This method can be applied as soon as $n \geq 3$.

Multivariate extension. Let $X_1, \dots, X_n$ be n independent Gaussian $\mathbb{R}^d$-valued random vectors, where $X_i \sim \mathcal{N}(\mu_i, C)$ and C is assumed to be invertible. We want to detect if vector x_ℓ is abnormal given the others. We propose the statistic:

$$T_n = \frac{(n-1)}{nd} \max_{i\in\{1,\dots,n\}} (X_i - \bar{X}_{n,-i})' C^{-1}_{n,-i} (X_i - \bar{X}_{n,-i}),$$

where

$$\bar{X}_{n,-i} = \frac{1}{n-1}\sum_{k=1,k\neq i}^{n} X_k \quad \text{and} \quad C_{n,-i} = \frac{1}{n-1-d}\sum_{k=1,k\neq i}^{n} (X_k - \bar{X}_{n,-i})(X_k - \bar{X}_{n,-i})'.$$

Under the null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_n = \mu$, one can easily check that the distribution P_n of T_n does not depend on μ nor C, and can therefore be tabulated.

The decision rule is then the following: at level α, we decide that there is (at least) one abnormal observation if $t_n > c_{\alpha,n}$, where $c_{\alpha,n}$ is such that $P_n([c_{\alpha,n}, \infty[) = \alpha$. This method can be applied as soon as $n \geq d+2$. Note that, if $d = 1$, the statistic T_n is exactly the square of the statistic described before.

2.3. *Method 2*

Recall that, in our setting, the index i of the observation x_i represents the order of observations (i.e. the observation x_i is the i-th measurement of a biomarker on the same individual). In this context, another interesting question is: how to detect a series of consecutive observations that are abnormal given the rest of the series?

More precisely, we want to know if there is a set $I = \{i, \dots, j\}$ of consecutive integers, for which the expectation of the $X_k, k \in I$ is different from the expectation of the other variables, that is: $\mu_k = \mu_\ell = \mu_A$ if k, ℓ do not belong to I and $\mu_k = \mu_\ell = \mu_B$ if k, ℓ belong to I.

To answer this question, we propose the statistic

$$T_n = \max_{I \in \mathcal{I}} \left| \frac{\bar{X}_I - \bar{X}_{\bar{I}}}{\hat{\sigma}_{n,I} \sqrt{\frac{1}{|I|} + \frac{1}{n-|I|}}} \right|,$$

where $\mathcal{I}$ is the collection of all possible intervals I included in $\{1, \dots, n\}$ with length $1 \leq |I| < n$,

$$\bar{X}_I = \frac{1}{|I|} \sum_{k \in I} X_k, \quad \bar{X}_{\bar{I}} = \frac{1}{n - |I|} \sum_{k \notin I} X_k,$$

and

$$\hat{\sigma}^2_{n,I} = \frac{1}{n-2} \left(\sum_{k \in I} (X_k - \bar{X}_I)^2 + \sum_{k \notin I}^{n} (X_k - \bar{X}_{\bar{I}})^2 \right).$$

Under the null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_n = \mu$, one can easily check that the distribution P_n of T_n does not depend on the unknown parameters (μ, σ^2), and can therefore be tabulated.

The decision rule is then the following: at level α, we decide that there is (at least) one consecutive series of abnormal observations if $t_n > c_{\alpha,n}$, where $c_{\alpha,n}$ is such that $P_n([c_{\alpha,n}, \infty[) = \alpha$. This method can be applied as soon as $n \geq 3$.

2.4. *Method 3*

Let us recall the context: we have two series of observations $x_1, \dots, x_{n_1}$ and $y_1, \dots, y_{n_2}$ and we want to know if one of these observations is abnormal given the other ones in the same subsample.

We propose the statistic

$$T_n = \max \left\{ \max_{i \in \{1,\dots,n_1\}} \left| \frac{X_i - \bar{X}_{n_1,-i}}{\hat{\sigma}_{X,-i,Y} \sqrt{1 + \frac{1}{n_1 - 1}}} \right|, \max_{j \in \{1,\dots,n_2\}} \left| \frac{Y_j - \bar{Y}_{n_2,-j}}{\hat{\sigma}_{Y,-j,X} \sqrt{1 + \frac{1}{n_2 - 1}}} \right| \right\},$$

where

$$\bar{X}_{n_1,-i} = \frac{1}{n_1 - 1} \sum_{k=1, k \neq i}^{n_1} X_k, \quad \bar{Y}_{n_2,-j} = \frac{1}{n_2 - 1} \sum_{k=1, k \neq j}^{n_2} Y_k,$$

$$\hat{\sigma}^2_{X,-i,Y} = \frac{1}{n-3} \left(\sum_{k=1, k \neq i}^{n_1} (X_k - \bar{X}_{n_1,-i})^2 + \sum_{k=1}^{n_2} (Y_k - \bar{Y}_{n_2})^2 \right),$$

and

$$\hat{\sigma}^2_{Y,-j,X} = \frac{1}{n-3} \left(\sum_{k=1}^{n_1} (X_k - \bar{X}_{n_1})^2 + \sum_{k=1, k \neq j}^{n_2} (Y_k - \bar{Y}_{n_2,-j})^2 \right).$$

Under the null hypothesis

$$H_0 : \mu_{X_1} = \mu_{X_2} = \cdots = \mu_{X_{n_1}} = \mu_X \text{ and } \mu_{Y_1} = \mu_{Y_2} = \cdots = \mu_{Y_{n_2}} = \mu_Y,$$

one can easily check that the distribution P_n of T_n does not depend on the unknown parameters (μ_X, μ_Y, σ^2), and can therefore be tabulated.

The decision rule is then the following: at level α, we decide that there is (at least) one abnormal observation among the two samples if $t_n > c_{\alpha,n}$, where $c_{\alpha,n}$ is such that $P_n([c_{\alpha,n}, \infty[) = \alpha$. This method can be applied as soon as $n_1 \geq 2$ and $n_2 \geq 2$.

Multivariate extension. Here $X_1, \ldots, X_{n_1}$ and $Y_2, \ldots, Y_{n_2}$ are two independent series of independent Gaussian $\mathbb{R}^d$-valued random vectors, with $X_i \sim \mathcal{N}(\mu_{X_i}, C)$, $Y_j \sim \mathcal{N}(\mu_{Y_j}, C)$ and C is assumed to be invertible. We want to detect if the vector x_k or y_ℓ is abnormal given the other ones in the same subsample. We propose the statistic

$$T_n = \max \left\{ \frac{(n_1 - 1)}{n_1 d} \max_{i \in \{1,\ldots,n_1\}} (X_i - \bar{X}_{n_1,-i})' C^{-1}_{X,-i,Y} (X_i - \bar{X}_{n_1,-i}), \right.$$
$$\left. \frac{(n_2 - 1)}{n_2 d} \max_{j \in \{1,\ldots,n_2\}} (Y_j - \bar{Y}_{n_2,-j})' C^{-1}_{Y,-j,X} (Y_j - \bar{Y}_{n_2,-j}) \right\},$$

where

$$\bar{X}_{n_1,-i} = \frac{1}{n_1 - 1} \sum_{k=1, k\neq i}^{n_1} X_k, \quad \bar{Y}_{n_2,-j} = \frac{1}{n_2 - 1} \sum_{k=1, k\neq j}^{n_2} Y_k,$$

$$C_{X,-i,Y} = \frac{1}{n - d - 2} \left(\sum_{k=1, k\neq i}^{n_1} (X_k - \bar{X}_{n_1,-i})(X_k - \bar{X}_{n_1,-i})' + \sum_{k=1}^{n_2} (Y_k - \bar{Y}_{n_2})(Y_k - \bar{Y}_{n_2})' \right),$$

and

$$C_{Y,-j,X} = \frac{1}{n - d - 2} \left(\sum_{k=1}^{n_1} (X_k - \bar{X}_{n_1})(X_k - \bar{X}_{n_1})' + \sum_{k=1, k\neq j}^{n_2} (Y_k - \bar{Y}_{n_2,-j})(Y_k - \bar{Y}_{n_2,-j})' \right).$$

Under the null hypothesis

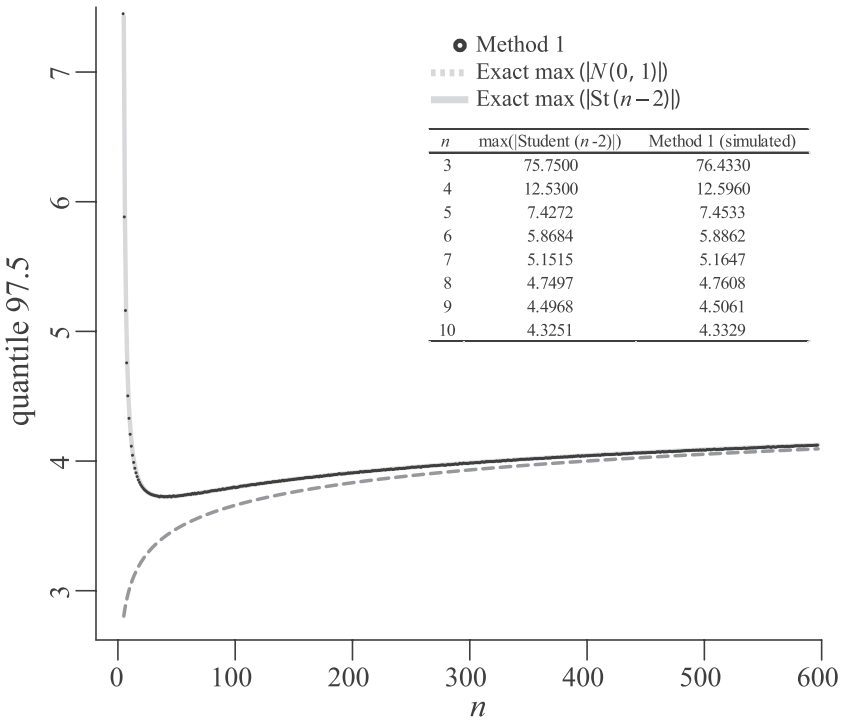
$$H_0 : \mu_{X_1} = \mu_{X_2} = \cdots = \mu_{X_{n_1}} = \mu_X \text{ and } \mu_{Y_1} = \mu_{Y_2} = \cdots = \mu_{Y_{n_2}} = \mu_Y,$$

one can easily check that the distribution P_n of T_n does not depend on the unknown parameters (μ_X, μ_Y, C) and can therefore be tabulated.

The decision rule is then the following: at level α, we decide that there is (at least) one abnormal observation among the two samples if $t_n > c_{\alpha,n}$, where $c_{\alpha,n}$ is such that $P_n([c_{\alpha,n}, \infty[) = \alpha$. This method can be applied as soon as $n_1 \geq 2$, $n_2 \geq 2$, and $n \geq d + 3$.

3. QUANTILES, POWER, AND ASYMPTOTIC DISTRIBUTION

In the following sections, we use MATLAB R2011b to compute the quantiles and assess the power of the methods. The tests and graphs were done using R v3.3.2.



n	max(|Student (*n*-2)|)	Method 1 (simulated)
3	75.7500	76.4330
4	12.5300	12.5960
5	7.4272	7.4533
6	5.8684	5.8862
7	5.1515	5.1647
8	4.7497	4.7608
9	4.4968	4.5061
10	4.3251	4.3329

Fig. 1. The levels estimated for Method 1 ($\alpha = 2.5\%$), along with the exact levels for max ($|\mathcal{N}(0,1)|$) and max ($|\text{Student}(n-2)|$) for $n = 3$ to 597 observations (10^6 individuals). The levels are detailed in the inset table for the first 10 observations.

3.1. *Computation of the quantiles*

In this section, we explain how to compute the quantiles of the distributions of the statistics T_n under the null hypothesis, for methods 0, 1, 2, and 3. For Method 0, the quantile of order α of distribution of the statistic $|T_n|$ is simply the quantile of order $1-\alpha/2$ of the Student$(n-2)$ distribution. The statistic T_n of Method 1 is the maximum of the absolute value of non-independent variables with Student$(n-2)$ distribution (respectively Fisher $(d, n-1-d)$ distribution in the multivariate case). To compute the quantiles of the distribution of this statistic, we use a standard Monte-Carlo procedure with 20×10^6 independent experiments, for each level α. For each n from 3 to 20, the quantiles are computed from 80% to 99.99% (see Supplementary material available at *Biostatistics* online). The same method is used for computing the quantiles of the distributions of the statistics of Methods 2 and 3 (please refer to Supplementary material available at *Biostatistics* online). For Method 1, we also compare the quantiles obtained by simulation to the quantiles of the distribution of the maximum of the absolute values of n independent variables with Student$(n-2)$ distribution. For the quantile 97.5%, we see that the simulated quantiles and the exact quantiles of this approximating distribution are almost identical as soon as $n > 5$. Hence, at this level, the exact quantile of the approximating distribution can be used as soon as $n > 5$. In Figure 1, we plot the quantile 97.5% obtained by simulation, the exact quantile of the approximating distribution, and also the exact quantiles of the distribution of the absolute values of n independent random variables with distribution $\mathcal{N}(0,1)$. We see that the convergence to the quantiles based on the Gaussian distribution is quite slow.

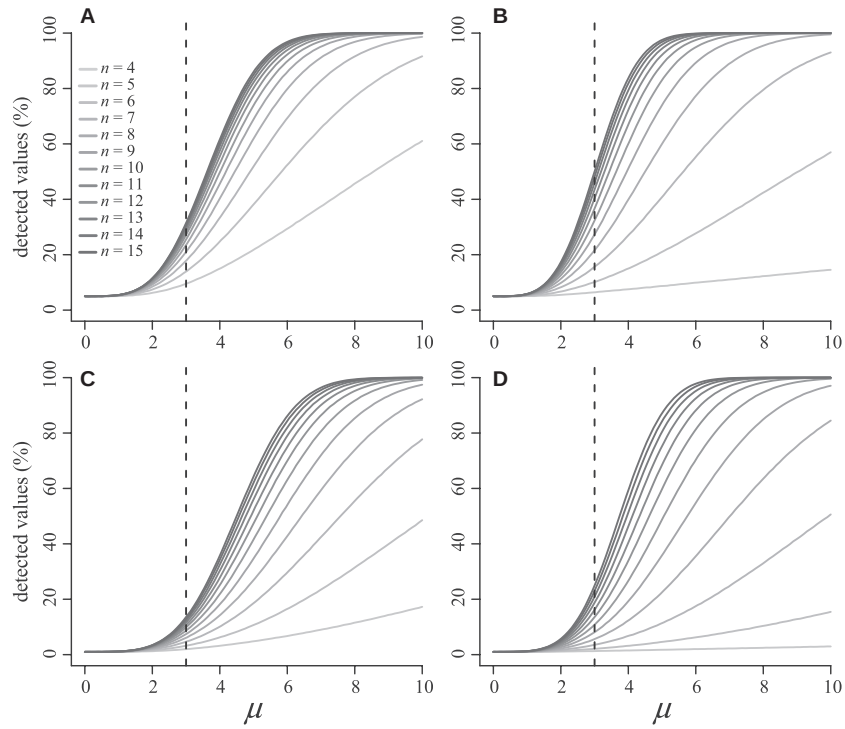


Fig. 2. Frequency of abnormal values detected by Method 1 (A, C) and its multivariate extension (B, D) as a function of μ and μ_d, for the levels $\alpha = 5\%$ (A, B) and $\alpha = 1\%$ (C, D). The dashed line at $\mu = 3$ is a visual indicator.

3.2. *Power of the Method 1*

In this subsection, we show how Method 1 will detect an abnormal value when all the others are identically distributed. We proceed as follows: we simulate 1×10^6 sequences consisting of n independent Gaussian random variables. The third random variable has distribution $\mathcal{N}(\mu, 1)$ with $\mu \geq 0$, whereas all the other random variables have distribution $\mathcal{N}(0, 1)$. The simulations are performed for a number of observations n ranging from 4 to 15 and μ ranging from 0 to 10 with a step of 0.1. We use the same procedure for the multivariate analysis, with $d = 2$. The third random vector has distribution $\mathcal{N}(\tilde{\mu}, C)$, whereas all the other random vectors have distribution $\mathcal{N}(0, Id)$. Here $\tilde{\mu} = (\mu, \mu)$ with the same possible values of μ as in the univariate case, and the covariance matrix C is given by

$$C = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

The results are given in Figure 2 below.

As expected, the power curve is increasing with n and μ. For the univariate version of Method 1, percentages of abnormal sequences detected for $n = 9$ and μ, respectively, equals to 2, 4, and 6 are 10.53%, 51.06%, and 90.57%. Similar graphs can be obtained for Methods 2 and 3.

3.3. *Asymptotic result for Methods 1 and 3*

In this subsection, we give the asymptotic distribution (as n goes to infinity) of the (centered and normalized) statistic T_n of Method 1 under H_0. At the end, we shall also deal with the statistic T_n of Method 3. This asymptotic distribution is the Gumbel distribution. As n goes to infinity, the distribution of T_n is close to the distribution of the maximum of the absolute value of n *iid* random variables with $\mathcal{N}(0,1)$ distribution (see Figure 1). Recall the expression of the terms of standardization introduced by Hall (1979):

$$\beta_n = \sqrt{2\log n} - \frac{\log(4\pi\log n)}{2\sqrt{(2\log n)}}.$$

One can prove the following result: under the null hypothesis H_0, the distribution of the random variable $\beta_{2n}(T_n - \beta_{2n})$ converges to a Gumbel distribution, that is a distribution whose cumulative distribution function is given by $\Lambda(x) = \exp(-\exp(-x))$.

The proof of this proposition is not difficult. The main point is to replace the empirical means and variances in the definition of T_n by the true values of the parameters μ and σ^2. Then, apply a known result for the distribution of the maximum absolute value of n *iid* random variables with $\mathcal{N}(0,1)$ distribution (see for instance Gasull *and others*, 2015a). As for the Gaussian case, the convergence of the maximum to the Gumbel distribution is very slow (rate of order $\log n$). So the quantiles of the Gumbel distribution can be used for very large samples only.

Let us consider now the statistic T_n proposed at the end of Section 2.2 to deal with the multivariate case. Arguing as in the 1D case, we easily see that the asymptotic distribution of T_n under H_0 is the same as that of the maximum of n *iid* random variables $\xi_1, \ldots, \xi_n$ such that $d\xi_i \sim \chi^2(d)$. The standard norming sequences are given for instance in the book by Embrechts *and others* (2013) (see p. 155). It follows that, under H_0 and for $d \geq 2$, $d(T_n - b_n)/2$ converges in distribution as $n \to \infty$ to the Gumbel distribution, where

$$b_n = \frac{2}{d}\left(\log n + \frac{(d-2)}{2}\log\log n - \log\Gamma\left(\frac{d}{2}\right)\right).$$

We refer to the article by Gasull *and others* (2015b), Section 5, for more accurate norming sequences.

The same proofs can be done for the statistic T_n of Method 3, with straightforward modifications. In the 1D case, we find that, under H_0, the distribution of the random variable $\beta_{2n}(T_n - \beta_{2n})$ converges to a Gumbel distribution, provided that $\min(n_1, n_2) > C(\log(n))^a$ for some $C > 0$ and some $a > 1$. In the multivariate case (see the end of Section 2.4), the distribution of $d(T_n - b_n)/2$ converges under H_0 to a Gumbel distribution under the same condition on $\min(n_1, n_2)$. Note that this last restriction on n_1, n_2 is very mild and should be realized in all practical situations.

4. Application to soccer players data

4.1. *Description of the data base*

After obtaining the approval of the Institutional Ethics Committee, our three methods were applied to a database of elite soccer players. It consists of five typical biological markers from 2577 male soccer players from the French elite leagues 1 and 2. Markers include concentrations of ferritin (μmol/L), serum iron (μmol/L), hemoglobin (g/L), erythrocytes (T/L), and hematocrit levels (%). The biomarkers were collected every 6 months in July/August and in January/March from 2006 to 2012 for a total of 12 collections. The large interval between two measures (around 6 months) allowed for independent sampling (Sharpe *and others*, 2006). We kept the players that had at least 5 measurements over the 12 possible

Table 1. *Percentage of individual series significantly non-normal with the Shapiro test and corresponding mean of P-values. And results of the sign tests. (L) indicates a log-transformation of the data*

Marker	% of non-normal series	Mean *P*-value of Shapiro tests	*P*-value of the sign test
Ferritin (L)	7.05	0.47	0.804
Serum iron (L)	7.67	0.49	0.703
Hemoglobin	6.91	0.46	3×10^{-8}
Erythrocytes	6.53	0.48	3×10^{-8}
Ferritin	5.25	0.48	1×10^{-10}

measurements, totalizing: 757 players for ferritin and serum iron, 799 (erythrocytes and hematocrit), and 807 for hemoglobin because of the high number of transfers between clubs, injuries, and the progressive inclusion of new clubs in the elite league. Moreover, a technical issue with the sampling instrument resulted in the loss of the data in the markers of the 2009 July/August collection. According to Custer *and others* (1995) and Tufts *and others* (1985) the measure of the ferritin and serum iron have a log-normal distribution, so we take the logarithm of the observations for these two biomarkers.

4.2. *Preliminary analysis*

The individual series of biomarkers must comply with the conditions described in Section 2. Regarding the assumption of normality: for each individual and each biomarker, we use the Shapiro test to confirm that it is not unrealistic to assume that the observations are drawn from a normal sample.

We note that, for each biomarker, the percentage of non-normal sequences detected by the test is not significantly different from the 5% level (that is the false-discovery rate of the test under the null hypothesis) (see Table 1). Regarding the assumption of equidistribution for Methods 1 and 2, it is known that the training intensity and frequency, the environment, and the period are among the factors inducing a possible variability in biological markers (Malcovati *and others*, 2003; Thirup, 2003). The impact of seasonal changes on hematologic markers has already been investigated in both the general and athletic population. However, some uncertainty remains, especially regarding hematocrit as reported in Malcovati *and others* (2003), Meyer and Meister (2011), and Thirup (2003), opposing Heisterberg *and others* (2013). Hence, to see if the period of the year has a significant influence on the biomarkers, we chose to perform a sign test based on the successive differences $D_i = X_{S,i} - X_{W,i}$, where $X_{S,i}$ corresponds to the measure in summer of year i, and $X_{W,i}$ corresponds to the measure in winter of year i. We collect all possible D_i, for all individuals and perform the sign test on this large data set. This is possible because the sign test is very robust, and not influenced by the inter-variability between individuals. All that matters for its application is that the variables $1_{D_i>0}$ are iid, which is certainly true once we admit that a difference of 6 months is enough for the independency. Under the null hypothesis, the period of the year has no influence on the biomarker, the distribution of the number of positive D_i follows a binomial distribution $B(N, 0.5)$ (N being the total number of D_i for all the individuals).

The sign tests give very small P-values for hemoglobin, erythrocytes and hematocrit, meaning that there is a significant difference between summer and winter for these three biomarkers (see Table 1). Therefore, only the ferritin and serum iron are eligible for Methods 1 and 2. For the other markers, the third method can be applied.

For the two markers ferritin and serum iron, we see that the empirical distribution of T_n (Method 1) is close to the theoretical one under H_0 (see Figure of Supplementary material available at *Biostatistics*

Table 2. *Number of abnormal series for each biomarker and method ($\alpha = 5\%$)*

	Ferritin	Serum ion	Red blood cells	Hemoglobin	Hematocrit	Total
Method 1						
Usual (n)	715	695	—	—	—	1410
Unusual (n)	42	62	—	—	—	104
Unusual (%)	5.55	8.19	—	—	—	6.87
(Multivariate)						
Usual (n)	661		—	—	—	661
Unusual (n)	49		—	—	—	49
Unusual (%)	6.90		—	—	—	6.9
Method 2						
Usual (n)	674	707	—	—	—	1381
Unusual (n)	83	50	—	—	—	133
Unusual (%)	10.96	6.61	—	—	—	8.78
Method 3						
Usual (n)	—	—	577	588	593	1758
Unusual (n)	—	—	41	38	35	114
Unusual (%)	—	—	6.63	6.07	5.57	6.09

online). This confirms that the quantiles of the theoretical distribution of T_n can be used to detect abnormal values on our dataset.

4.3. *Results*

The three methods are used on the longitudinal follow-up of five biological markers collected from elite soccer players. Some series are log-transformed to ensure that the variables were normally distributed (see Table 1), and we have carefully taken care of the seasonality (detected by the sign test with a very small P-value, see Table 1 and Section 4.2). As stated previously, it is not clear whether the hematocrit marker is impacted by the seasonal change but our results show a seasonal variability for hemoglobin, erythrocytes, and hematocrit (see Table 1). Hence, we use Method 3 for these three markers, although these variations are possibly linked to the geographical trip of individuals due to training or competition reasons. According to these preliminary analyses, we apply Method 1 (and it's multivariate extension) and Method 2 to the ferritin and the serum iron, and Method 3 to erythrocytes, hemoglobin, and hematocrit. We keep the individuals with at least 3 values per season for the Method 3. The frequency of abnormal series detected by the procedures are reported in Table 2 for a level $\alpha = 5\%$.

Most of the results are not too far from the expected false-discovery rate: between 5% and 6.7% for Ferritin/Method 1, Serum iron/Method 2, and all three others biomarkers for Method 3; 8.19% for Serum Iron/Method 1. However, the percentage of abnormal subseries for Ferritin/Method 2 is close to 11%, hence significantly different from the expected level $\alpha = 5\%$. We cannot directly assert that these differences are related to a pathological behavior; for instance it could be explained by a slight departure from the Gaussian distribution (7.05% of the Ferritin series were detected to be non-normal by the Shapiro test at level 5%). Thus, further analyses are required to shed light on these series. We find the same percentage of abnormal values for the multivariate version of Method 1 as for the two univariate procedures (6.9% in both cases). In Figure 3 below, we give some examples of abnormal observations detected by the three methods, for different biomarkers.

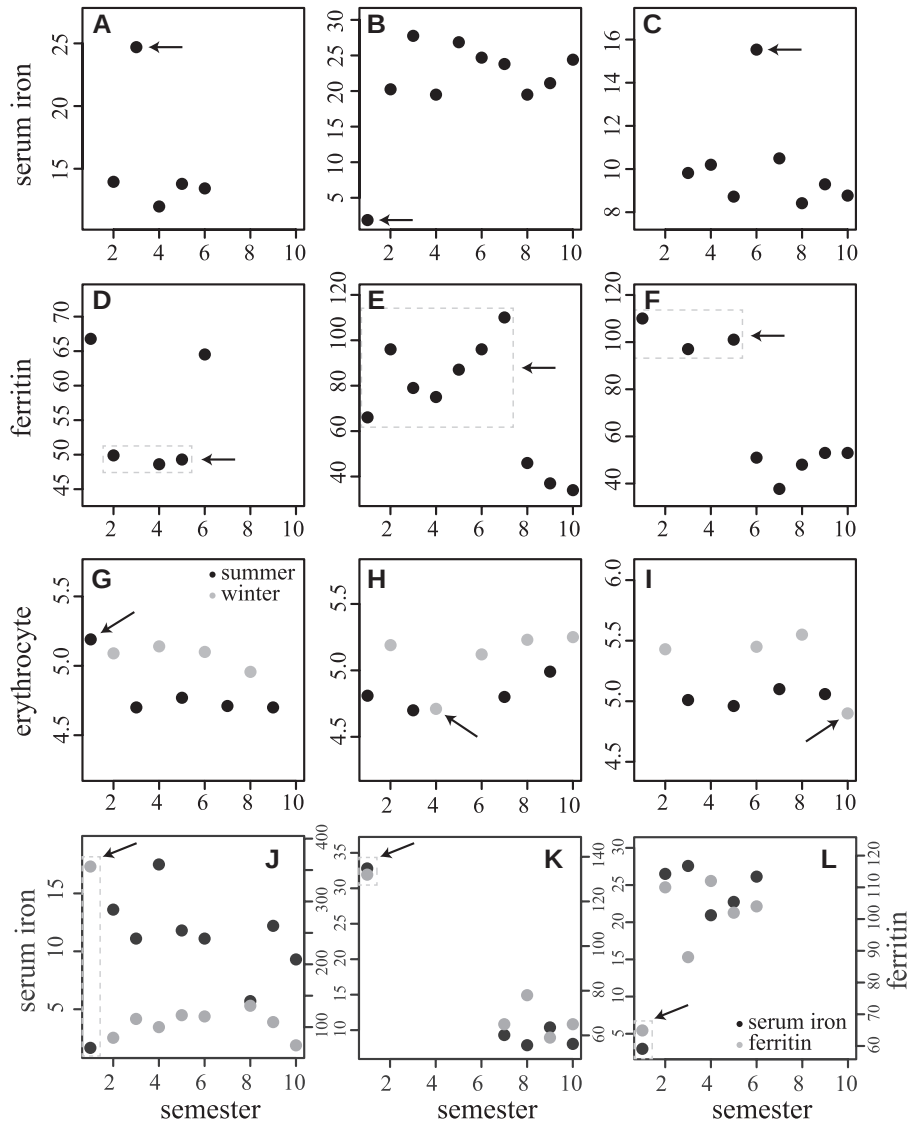


Fig. 3. Examples of results for the Method 1 (A–C), Method 2 (D–F), and Method 3 (G–I) in different individuals for the serum iron, ferritin and erythrocyte values. Panels (J–L) show some results for the multivariate version of Method 1.

5. DISCUSSION

This article introduces three Z-scores-based methods for detecting abnormal values in a series of biological measures from an individual's sample. They allow for assessing the individual baseline while taking into account the seasonal changes that alter the values of biomarkers. The multivariate approach is also developed in order to avoid multi-test issues and to take care of the possible correlations between biomarkers. Finally, we make extensive simulations to obtain the power-curve of our procedures for certain points of the alternative hypothesis. This power-curve could be used to compare our procedures to other methods.

Unfortunately, such power curves are not available in the article by Sottas *and others* (2007), and it is somewhat unclear how they could be produced. Our methods are tested on the follow-up of elite athletes and we found the results in accordance with the expected false-discovery rate, in most of the cases. Our methods can be extended to the Gaussian linear model, but the quantile tables have to be computed for each new design. A possible alternative would be to compute an approximate *p*-value by using Monte Carlo simulations.

The longitudinal values of athlete's hematological markers were reported to differ from those of the general population. Several previous works have shown that biological values are affected by the player's position, level, ethnic origin, and age among other parameters. As a result, Malcovati *and others* (2003) suggested focusing on intra-variability only. Sharpe *and others* (2006) and Sottas *and others* (2007) introduced two interesting approaches to identify outliers in longitudinal follow-ups. However, Sharpe *and others* (2006) assume that the variance of the random variables is the same for all individuals, which is not relevant in many cases. The method proposed by Sottas *and others* (2007) is convincing and provides good results, but the need of an extra population makes it rather complicated to use (a large relevant population has to be used for each different biomarker). In contrast, our methods are particularly simple, and can be tabulated, and then applied to all biomarkers, provided the distribution is Gaussian (up to a known deterministic transformation).

Bejder *and others* (2016) showed that acute hyperhydration reduces the sensitivity of the Bayes APB procedure for usual blood markers, and this could affect the sensitivity of our methods as well. However, Lobigs *and others* (2017a) proposed new plasma volume biomarkers that should help to correct this effect.

To conclude this section, let us discuss the relevance of a hierarchical Bayes procedure in our context. The following are elements of discussion:

- First, for a Bayes procedure, one needs to choose an *a priori* distribution on the parameters (μ, σ^2).This can be done by using an extra population, as in Sottas *and others* (2007), who fit a parametric model to the parameters. If this extra population is not available, but there are many individuals in the study (as in our soccer players database), one can select a subpopulation to perform this first estimation step. Note however, the methods described in the present article can be applied to one single individual without the help of an extra population.

- Once we know the *a priori* distribution, we have access via the Bayes rule to the global distribution of the variables (in the Bayesian framework, the distribution $N(\mu, \sigma^2)$ is the conditional distribution of the variables X_i given (μ, σ^2)). Hence, we can test if the observations are abnormal with respect to this distribution. However, it is important to point out that this is not what we have in mind: we want to test if, for a given individual, some observations are abnormal; this can be done efficiently only if the characteristics of the individuals are taken into account. If we use the global distribution of the variables, we will primarily detect individuals that are different from the rest of the population (because the observations of the bio-markers of these individuals will very likely be abnormal with respect to the global distribution). But we are not interested to know if an individual is different from the rest of the population (it is irrelevant for professional athletes).

- Sottas *and others* (2007) used an intermediate approach. Thanks to the Bayes rule again, they compute the conditional distribution of X_n given $X_1, \ldots, X_{n-1}$, and they use the quantiles of this distribution to see if the observation x_n is abnormal or not. This procedure is satisfactory (even if it requires an extra population for each biomarker to determine an appropriate *a priori* distribution). It is similar to our Method 0, but it does not answer the question: is there at least one abnormal value in the whole sequence of measurements of a single individual? One way to answer is to perform a sequential procedure, testing if x_2 is abnormal given x_1, if x_3 is abnormal given x_1, x_2,

and so on . . . but this leads to a non-trivial multi-test problem, as already explained at the beginning of Section 2.2.

To summarize, our methods allow for a single-detection rule, thus avoiding multi-test issues that would occur by testing if each new value is abnormal. In the univariate case, Method 1 can be applied with at least three observations per individual. This is of course a (rather mild) restriction of the method but, on another hand, it can be applied to any sequence of independent Gaussian outcomes, without the help of an extra cohort to determine an appropriate *a priori* distribution.

6. CONCLUSION

We propose three methods to detect abnormal values or subseries in longitudinal follow-ups. These methods are simple and easy to implement, and their significance levels can be easily computed via Monte Carlo simulations. They can be applied to independent and normally distributed random variables, for series with at least three observations. This last point does not seem to be a strong limitation, since today's follow-ups often include more than three observations, and the sampling rate is increasing over the years. We apply these methods to a large dataset, and we detect some abnormal values (some of them being false positives). Additional investigations should be carried out by the medical staff in order to shed light on abnormal follow-ups. The proposed methods could be improved to include additional periodic environmental effects, such as training at altitude. Practical applications include detecting abnormal values in elite athletes' follow-ups for clinical or anti-doping purposes. It would also be useful in detecting pathological values in clinical follow-ups of patients based on individual rather than population-based thresholds.

SUPPLEMENTARY MATERIAL

Supplementary material is available at http://biostatistics.oxfordjournals.org.

ACKNOWLEDGMENTS

This article is dedicated to the work and commitment of Pierre Rochcongar (1947–2016). We deeply thank the Association des Médecins de Club de Football Professionnel represented by Hervé Caumont (MD) and Emmanuel Orhant (MD). We also thank Lydie Jorda and Jean Lagrange for the preliminary analyses and Stacey Johnson for reviewing the manuscript. *Conflict of Interest*: None declared.

FUNDING

The biomarker collections and analyses were supported by the Ligue de Football Professionnel.

REFERENCES

BANFI, G., LOMBARDI, G., COLOMBINI, A. AND LIPPI, G. (2011). Analytical variability in sport hematology: its importance in an antidoping setting. *Clinical Chemistry and Laboratory Medicine* **49**, 779–782.

BEJDER, J., HOFFMANN, M. F., ASHENDEN, M., NORDSBORG, N. B., KARSTOFT, K. AND MØRKEBERG, J. (2016). Acute hyperhydration reduces athlete biological passport off-hr score. *Scandinavian Journal of Medicine & Science in Sports* **26**, 338–347.

CUSTER, E. M., FINCH, C. A., SOBEL, R. E. AND ZETTNER, A. (1995). Population norms for serum ferritin. *The Journal of Laboratory and Clinical Medicine* **126**, 88–94.

DAMSGAARD, R., MUNCH, T., MORKEBERG, J., MORTENSEN, S. P. AND GONZALEZ-ALONSO, J. (2006). Effects of blood withdrawal and reinfusion on biomarkers of erythropoiesis in humans: implications for anti-doping strategies. *Haematologica* **91**, 1006–1008.

EATON, M. L. (1983). *Multivariate Statistics: A Vector Space Approach*. New York, NY: John Wiley & Sons, Inc., p.512.

EGGER, F., MEYER, T. AND HECKSTEDEN, A. (2016). Interindividual variation in the relationship of different intensity markers—a challenge for targeted training prescriptions. *PLoS One* **11**, e0165010.

EMBRECHTS, P., KLÜPPELBERG, C. AND MIKOSCH, T. (2013). *Modelling Extremal Events: For Insurance and Finance*, Volume 33. Berlin Heidelberg New-York: Springer Science & Business Media.

GASULL, A., JOLIS, M. AND UTZET, F. (2015a). On the norming constants for normal maxima. *Journal of Mathematical Analysis and Applications* **422**, 376–396.

GASULL, A., LÓPEZ-SALCEDO, J. A. AND UTZET, F. (2015b). Maxima of gamma random variables and other weibull-like distributions and the lambert W function. *Test* **24**, 714–733.

HALL, P. (1979). On the rate of convergence of normal extremes. *Journal of Applied Probability* **16**, 433–439.

HECKSTEDEN, A., PITSCH, W., JULIAN, R., PFEIFFER, M., KELLMANN, M., FERRAUTI, A. AND MEYER, T. (2016a). A new method to individualize monitoring of muscle recovery in athletes. *International Journal of Sports Physiology and Performance*, 1–25.

HECKSTEDEN, A., SKORSKI, S., SCHWINDLING, S., HAMMES, D., PFEIFFER, M., KELLMANN, M., FERRAUTI, A. AND MEYER, T. (2016b). Blood-borne markers of fatigue in competitive athletes–results from simulated training camps. *PloS One* **11**, e0148810.

HEISTERBERG, M. F., FAHRENKRUG, J., KRUSTRUP, P., STORSKOV, A., KJÆR, M. AND ANDERSEN, J. L. (2013). Extensive monitoring through multiple blood samples in professional soccer players. *The Journal of Strength & Conditioning Research* **27**, 1260–1271.

JULIAN, R., MEYER, T., FULLAGAR, H. H. K., SKORSKI, S., PFEIFFER, M., KELLMANN, M., FERRAUTI, A. AND HECKSTEDEN, A. (2017). Individual patterns in blood-borne indicators of fatigueâŁ”trait or chance. *Journal of Strength and Conditioning Research* **31**, 608–619.

LOBIGS, L. M., SOTTAS, P.-E., BOURDON, P. C., NIKOLOVSKI, Z., EL-GINGO, M., VARAMENTI, E., PEELING, P., DAWSON, B. AND SCHUMACHER, Y. O. (2017a). A step towards removing plasma volume variance from the athlete's biological passport: the use of biomarkers to describe vascular volumes from a simple blood test. *Drug Testing and Analysis*.

LOBIGS, L. M., SOTTAS, P.-E., BOURDON, P. C., NIKOLOVSKI, Z., EL-GINGO, M., VARAMENTI, E., PEELING, P., DAWSON, B. AND SCHUMACHER, Y. O. (2017b). The use of biomarkers to describe plasma-, red cell-, and blood volume from a simple blood test. *American Journal of Hematology* **92**, 62–67.

LOMBARDI, G., RICCI, C. AND BANFI, G. (2011). Effects of winter swimming on haematological parameters. *Biochemia Medica* **21**, 71–78.

MALCOVATI, L., PASCUTTO, C. AND CAZZOLA, M. (2003). Hematologic passport for athletes competing in endurance sports: a feasibility study. *Haematologica* **88**, 570–581.

MALM, C. B., KHOO, N. S., GRANLUND, I., LINDSTEDT, E. AND HULT, A. (2016). Autologous doping with cryopreserved red blood cells–effects on physical performance and detection by multivariate statistics. *PloS One* **11**, e0156157.

MEYER, T. AND MEISTER, S. (2011). Routine blood parameters in elite soccer players. *International Journal of Sports Medicine* **32**, 875–881.

NIKOLAIDIS, P., ZIV, G., LIDOR, R. AND ARNON, M. (2014). Inter-individual variability in soccer players of different age groups playing different positions. *Journal of Human Kinetics* **40**, 213–225.

PARISOTTO, R., GORE, C. J., EMSLIE, K. R., ASHENDEN, M. J., BRUGNARA, C., HOWE, C., MARTIN, D. T., TROUT, G. J. AND HAHN, A. G. (2000). A novel method utilising markers of altered erythropoiesis for the detection of recombinant human erythropoietin abuse in athletes. *Haematologica* **85**, 564–572.

POTTGIESSER, T., SOTTAS, P.-E., ECHTELER, T., ROBINSON, N., UMHAU, M. AND SCHUMACHER, Y. O. (2011). Detection of autologous blood doping with adaptively evaluated biomarkers of doping: a longitudinal blinded study. *Transfusion* **51**, 1707–1715.

ROBINSON, N., SOTTAS, P.-E., MANGIN, P. AND SAUGY, M. (2007). Bayesian detection of abnormal hematological values to introduce a no-start rule for heterogeneous populations of athletes. *Haematologica* **92**, 1143–1144.

SALLET, P., BRUNET-GUEDJ, E., MORNEX, R. AND BAVEREL, G. (2008). Study of a new indirect method based on absolute norms of variation to detect autologous blood transfusion. *International Journal of Hematology* **88**, 362.

SAUGY, M., LUNDBY, C. AND ROBINSON, N. (2014). Monitoring of biological markers indicative of doping: the athlete biological passport. *British Journal of Sports Medicine* **48**, 827–832.

SHARPE, K., ASHENDEN, M. J. AND SCHUMACHER, Y. O. (2006). A third generation approach to detect erythropoietin abuse in athletes. *Haematologica* **91**, 356–363.

SHARPE, K., HOPKINS, W., EMSLIE, K. R., HOWE, C., TROUT, G. J., KAZLAUSKAS, R., ASHENDEN, M. J., GORE, C. J., PARISOTTO, R. AND HAHN, A. G. (2002). Development of reference ranges in elite athletes for markers of altered erythropoiesis. *Haematologica* **87**, 1248–1257.

SILVA, J. R., REBELO, A., MARQUES, F., PEREIRA, L., SEABRA, A., ASCENSÃO, A. AND MAGALHÃES, J. (2013). Biochemical impact of soccer: an analysis of hormonal, muscle damage, and redox markers during the season. *Applied Physiology, Nutrition, and Metabolism* **39**, 432–438.

SOTTAS, P.-E., BAUME, N., SAUDAN, C., SCHWEIZER, C., KAMBER, M. AND SAUGY, M. (2007). Bayesian detection of abnormal values in longitudinal biomarkers with an application to t/e ratio. *Biostatistics* **8**, 285.

SOTTAS, P.-E., ROBINSON, N., RABIN, O. AND SAUGY, M. (2011). The athlete biological passport. *Clinical Chemistry* **57**, 969–976.

THIRUP, P. (2003). Haematocrit. *Sports Medicine* **33**, 231–243.

TUFTS, D. A., HAAS, J. D., BEARD, J. L. AND SPIELVOGEL, H. (1985). Distribution of hemoglobin and functional consequences of anemia in adult males at high altitude. *The American Journal of Clinical Nutrition* **42**, 1–11.

ZORZOLI, M., PIPE, A., GARNIER, P. Y., VOUILLAMOZ, M. AND DVORAK, J. (2014). Practical experience with the implementation of an athlete's biological profile in athletics, cycling, football and swimming. *British Journal of Sports Medicine* **48**, 862–866.

ZORZOLI, M. AND ROSSI, F. (2010). Implementation of the biological passport: the experience of the international cycling union. *Drug Testing and Analysis* **2**, 542–547.

[*Received February 8, 2017; revised July 28, 2017; accepted for publication August 21, 2017*]

c) Apports et perspectives

Comme évoqué précédemment, les marqueurs biologiques présentent une variabilité interindividuelle supérieure à la variabilité intra-individuelle (53,54). Cela implique que pour détecter une variation individuelle au sein d'une série longitudinale, il ne faut utiliser que des paramètres individualisés. Des paramètres populationnels ne peuvent être utilisés pour déterminer l'atypicité d'une valeur individuelle par rapport au reste de la série personnelle. Se fondant sur ce premier constat, nos méthodes prennent en compte uniquement les variations intra-individuelles optimisant le pouvoir discriminant de l'analyse par rapport à une méthode utilisant des informations populationnelles. Par ailleurs, certains biomarqueurs fluctuent selon des variations saisonnières. Une de nos méthodes permet de prendre en compte ces fluctuations, élément indispensable au bon déroulement d'une procédure de détection d'atypicité, indépendamment de ces variations cycliques. Aujourd'hui, le suivi longitudinal concerne le prélèvement systématique de nombreux biomarqueurs différents sur le même individu. Certains d'entre eux étant corrélés, nous avons développé une approche multivariée permettant de prendre en compte les possibles corrélations entre biomarqueurs afin d'éviter les problèmes de modification du risque de première espèce lié aux multi-tests. Nous avons aussi la possibilité de calculer les courbes de puissance relatives aux différentes méthodes. De telles courbes permettent la comparaison à d'autres méthodes de détection.

Nos différentes méthodes de détection trouvent plusieurs applications dans le domaine du sport. Comme précisé précédemment, le développement de ces outils a été initié par la volonté de la Fédération Française de Football de protéger la santé de ses athlètes de haut niveau. Dans ce sens, ces outils permettent de contrôler les valeurs issues du monitoring de différents marqueurs biologiques (hématologiques, stéroïdiens) mis en place par les staffs médicaux. Associées au monitoring systématique, ces méthodes nouvelles facilitent la décision médicale. Une application pour la détection indirecte de comportements dopants est aussi possible.

Outre la détection de valeurs atypiques nécessaires à la surveillance biomédicale des athlètes, ces méthodes peuvent être utilisées dans d'autres champs d'investigation. L'optimisation de la performance en fait partie. La fatigue et la récupération sont en effet des éléments essentiels pris en compte par les staffs pour la mise en place des plannings d'entrainement. Or certains marqueurs biologiques sont utiles pour l'évaluation des niveaux de fatigue centrale ou de

récupération musculaire. L'analyse du suivi de marqueurs spécifiques grâce à ces méthodes nouvelles pourrait apporter aux staffs techniques une information supplémentaire permettant un meilleur ajustement de la charge d'entrainement.

Le monitoring des marqueurs biologiques des athlètes est aussi un élément qui s'intègre parfaitement dans un programme global de prévention des blessures. Alerté par la détection d'une atypicité, les staffs pourraient investiguer l'état de forme de l'athlète et prévenir un potentiel risque de blessure.

Enfin, les modes de vie actuels dans nos sociétés voient la prévalence des affections longues durées (ALD) augmenter dangereusement (87). Les patients atteints d'ALD sont régulièrement suivis pour contrôler l'évolution de leurs pathologies. Une forme de contrôle est l'analyse régulière d'un certains nombres de biomarqueurs : hémoglobine glyquée dans le diabète de type 2, taux de cholestérol, thyréostimuline (TSH) en cas d'hyperthyroïdie et de paramètres cliniques : tension artérielle, indice de masse corporelle. Concernant les paramètres sanguins, les valeurs des patients sont comparées à des normes définies a priori et il n'existe pas à notre connaissance une analyse de l'évolution de ces paramètres sanguins sur le long terme bien que la volonté d'individualiser le dépistage de certaines pathologies est de plus en plus présente (88). L'application de nos méthodes de détection d'anormalité sur ces suivis permettrait d'aider le médecin dans sa prise de décision sur l'évolution des traitements à prescrire. Un tel suivi chez des patients à risques car ayant des prédispositions génétiques ou un style de vie particulièrement sédentaire serait très pertinent dans un objectif de prévention.

3) Etude 2 : Correction du test de Mann-Whitney pour des variables corrélées.

a) Introduction

La recherche dans le champ des sciences du sport se heurte à une problématique inhérente au haut niveau : le nombre d'individus à analyser est faible. En effet les sportifs de haut niveau se situent à l'extrémité de la distribution de la population générale, que ce soit à propos de leurs caractéristiques physiques, physiologiques ou leurs performances (6,8). Ainsi, les études mises en place peinent à regrouper assez d'individus et à compléter les conditions de validité des tests paramétriques usuels. Des statisticiens se sont penchés sur cette problématique et offrent des outils aux chercheurs en sciences du sport pour adapter leurs analyses. C'est le cas Will G Hopkins qui en collaboration avec des chercheurs en sciences du sport adapte et

propose des outils issus des problématiques rencontrés sur le terrain et les laboratoires par les professionnels du sport (89–94). Des articles critiquent ces méthodes proposées par Hopkins et alertent les scientifiques du sport de problèmes liés à leurs applications (95–97). L'inférence bayésienne est évidemment une des alternatives existant aux problématiques statistiques rencontrées dans la recherche en sport de haut de niveau. Alternative rencontrant de plus en plus d'intérêt dans ce domaine (98,99). Les *sport scientists* utilisent une batterie d'outils statistiques au quotidien allant des statistiques descriptives basiques à l'application de statistiques inférentielles (100). Les données manipulées peuvent être quantitatives continues ou discrètes, qualitatives ordonnées ou non. Les paramètres de dispersion et de position sont aussi largement utilisés dans la compréhension des variables observées (101–104). Ce constat introduit le fait que les techniques statistiques utilisées au quotidien par les chercheurs des différents domaines du sport sont étroitement liées à celles utilisées en santé publique. Il existe en effet une réelle transposition des questions de recherche entre le sport de haut niveau et la santé publique : l'évaluation de l'effet d'un traitement et celui d'un protocole d'entrainement, la détection de variation de marqueurs biologiques, la prévention de maladies et de blessures, la mise en évidence de facteurs de risques. Des échanges de méthodes statistiques entre le monde de la santé publique et du sport de haut niveau amènent au développement de nouveaux outils. C'est le cas d'une étude sur les causes de mortalité et la longévité réalisée sur une cohorte d'Olympiens français. Inscrite dans un contexte de recherche épidémiologique mettant en application des modèles d'analyses de survie classiquement utilisés en recherche biomédicale et estimant l'effet de telles ou telles pathologies sur le nombre d'années de vies perdues, cette étude a abouti à la publication d'une nouvelle méthode d'analyse permettant d'estimer les années de vie gagnées. Ici, nous nous intéressons à un autre outil statistique très utilisé en recherche biomédicale : le test de Mann-Whitney (aussi appelé test de Wilcoxon à deux échantillons).

b) The Mann–Whitney U-Statistic for α-Dependent Sequences (105–107,107–122)

ISSN 1066-5307, Mathematical Methods of Statistics, 2017, Vol. 26, No. 2, pp. 111–133. © Allerton Press, Inc., 2017.

The Mann–Whitney U-Statistic for α-Dependent Sequences

J. Dedecker[1]* and G. Saulière[2]**

[1] *Univ. Paris Descartes, Sorbonne Paris Cité, Paris, France*

[2] *Univ. Paris Sud, Inst. de Recherche bioMéd. et d'Epidém. du Sport (IRMES) Inst. Natl. du Sport de l'Expertise et de la Performance (INSEP), Paris, France*

Received November 23, 2016; in final form, April 25, 2016

Abstract—We give the asymptotic behavior of the Mann–Whitney U-statistic for two independent stationary sequences. The result applies to a large class of short-range dependent sequences, including many nonmixing processes in the sense of Rosenblatt [17]. We also give some partial results in the long-range dependent case, and we investigate other related questions. Based on the theoretical results, we propose some simple corrections of the usual tests for stochastic domination; next we simulate different (nonmixing) stationary processes to see that the corrected tests perform well.

Keywords: U-statistic, weak dependence, stationary sequences.

2000 Mathematics Subject Classification: 62G30, 60F05.

DOI: 10.3103/S1066530717020028

1. INTRODUCTION AND MAIN RESULT

Let $(X_i)_{i\in\mathbb{Z}}$ and $(Y_i)_{i\in\mathbb{Z}}$ be two independent stationary sequences of real-valued random variables. Let $m = m_n$ be a sequence of positive integers such that $m_n \to \infty$ as $n \to \infty$. The Mann–Whitney U-statistic may be defined as

$$U_n = U_{n,m} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} \mathbf{1}_{X_i<Y_j}.$$

We shall also use the notation $\pi = \mathbb{P}(X_1 < Y_1)$, $H_Y(t) = \mathbb{P}(Y_1 > t)$ and $G_X(t) = \mathbb{P}(X_1 < t)$.

It is well known that if $(X_i)_{i\in\mathbb{Z}}$ and $(Y_i)_{i\in\mathbb{Z}}$ are two independent sequences of independent and identically distributed (iid) random variables, then U_n may be used to test $\pi = 1/2$ against $\pi \neq 1/2$ (or $\pi > 1/2$, $\pi < 1/2$). When $\pi > 1/2$ we shall say that Y stochastically dominates X in a weak sense. This property of weak domination is true, for instance, if Y stochastically dominates X, that is if $H_Y(t) \geq H_X(t)$ for any real t, with a strict inequality for some t_0. But it also holds in many other situations, for instance, if X and Y are two Gaussian random variables with $\mathbb{E}(X) < \mathbb{E}(Y)$, whatever the variances of X and Y.

In this paper, we study the asymptotic behavior of U_n under some conditions on the α-dependence coefficients of the sequences $(X_i)_{i\in\mathbb{Z}}$ and $(Y_i)_{i\in\mathbb{Z}}$. Let us recall the definition of these coefficients.

Definition 1.1. Let $(X_i)_{i\in\mathbb{Z}}$ be a strictly stationary sequence, and let $\mathcal{F}_0 = \sigma(X_k, k \leq 0)$. For any nonnegative integer n, let

$$\alpha_{1,\mathbf{X}}(n) = \sup_{x\in\mathbb{R}} \|\mathbb{E}(\mathbf{1}_{X_n\leq x} \mid \mathcal{F}_0) - F(x)\|_1,$$

where F is the distribution function of X_0.

*E-mail: jerome.dedecker@parisdescartes.fr

**E-mail: Guillaume.SAULIERE@insep.fr

We shall also always require that the sequences are 2-ergodic in the following sense.

Definition 1.2. A strictly stationary sequence $(X_i)_{i\in\mathbb{Z}}$ is 2-ergodic if, for any bounded measurable function f from $\mathbb{R}^2$ to $\mathbb{R}$ and any nonnegative integer k,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} f(X_i, X_{i+k}) = \mathbb{E}(f(X_0, X_k)) \quad \text{almost surely.} \tag{1.1}$$

Note that the almost sure limit in (1.1) always exists, by the ergodic theorem. The property of 2-ergodicity means only that this limit is constant. This property is weaker than usual ergodicity, which would imply the same property for any bounded function f from $\mathbb{R}^\ell$ to $\mathbb{R}$, $\ell \in \mathbb{N} - \{0\}$.

Our main result is the following theorem.

Theorem 1.1. *Let $(X_i)_{i\in\mathbb{Z}}$ and $(Y_i)_{i\in\mathbb{Z}}$ be two independent stationary 2-ergodic sequences of real-valued random variables. Assume that*

$$\sum_{k=1}^{\infty}\alpha_{1,\mathbf{X}}(k) < \infty \quad \textit{and} \quad \sum_{k=1}^{\infty}\alpha_{1,\mathbf{Y}}(k) < \infty. \tag{1.2}$$

Let (m_n) be a sequence of positive integers such that

$$\lim_{n\to\infty}\frac{n}{m_n} = c \quad \textit{for some } c \in [0,\infty). \tag{1.3}$$

Then the random variables $\sqrt{n}(U_n - \pi)$ converge in distribution to $\mathcal{N}(0, V)$ with

$$\begin{aligned} V = {} & \operatorname{Var}(H_Y(X_0)) + 2\sum_{k=1}^{\infty}\operatorname{Cov}(H_Y(X_0), H_Y(X_k)) \\ & + c\Big(\operatorname{Var}(G_X(Y_0)) + 2\sum_{k=1}^{\infty}\operatorname{Cov}(G_X(Y_0), G_X(Y_k))\Big). \end{aligned} \tag{1.4}$$

Note that the result of Theorem 1.1 has been obtained by Serfling [18] under the more restrictive condition

$$\sum_{k\geq 0}\alpha(k) < \infty \quad \text{and} \quad \alpha(k) = O\left(\frac{1}{k\log k}\right) \tag{1.5}$$

on the sequences $(X_i)_{i\in\mathbb{Z}}$ and $(Y_i)_{i\in\mathbb{Z}}$, where the $\alpha(k)$'s are the usual α-mixing coefficients of Rosenblatt [17].

The notion of α-dependence that we consider in the present paper is much weaker than α-mixing in the sense of Rosenblatt [17]. Some properties of these α-dependence coefficients are stated in the book by Rio [16], and many examples are given in the two papers [7], [8]. Let us recall one of these examples. Let

$$X_i = \sum_{k\geq 0} a_k \varepsilon_{i-k}, \tag{1.6}$$

where $(a_k)_{k\geq 0} \in \ell^2$ and the sequence $(\varepsilon_i)_{i\in\mathbb{Z}}$ is a sequence of independent and identically distributed real-valued random variables with mean zero and finite variance. If the cumulative distribution function F of X_0 is γ-Hölder for some $\gamma \in (0, 1]$, then

$$\alpha_{1,\mathbf{X}}(n) \leq C\Big(\sum_{k\geq n} a_k^2\Big)^{\frac{\gamma}{\gamma+2}}$$

for some positive constant C (this follows from the computations in Section 6.1 of [8]). In particular, if a_k is geometrically decreasing, then so is $\alpha_{1,\mathbf{X}}(n)$ (whatever the index γ).

Note that, without extra assumptions on the distribution of ε_0, such linear processes have no reasons to be mixing in the sense of Rosenblatt. For instance, it is well known that the linear process

$$X_i = \sum_{k\geq 0} \frac{\varepsilon_{i-k}}{2^{k+1}}, \qquad \text{where} \quad \mathbb{P}(\varepsilon_1 = -1/2) = \mathbb{P}(\varepsilon_1 = 1/2) = 1/2,$$

is not α-mixing (see for instance [1]). By contrast, for this particular example, $\alpha_{1,\mathbf{X}}(n)$ is geometrically decreasing.

To conclude, let us also mention the paper [5], where it is proved that the coefficients $\alpha_{1,\mathbf{X}}(n)$ can be computed for a large class of dynamical systems on the unit interval (see Subsection 6.2 of the present paper for more details).

Remark 1.1. Other types of dependence are considered by Dewan and Prakasa Rao [13] and Dehling and Fried [11]. The paper [13] deals with the U-statistic U_n for two independent samples of associated random variables. The paper [11] deals with general U-statistics of functions of β-mixing sequences, for either two independent samples, or for two adjacent samples from the same time series (see Subsection 5.2 for an extension of our results to this more difficult context). The class of functions of β-mixing sequences contains also many examples, with a large intersection with the class of α-dependent sequences. The advantage of our approach is that it leads to conditions that are optimal in a precise sense (see Section 3). On the other hand, it seems more difficult to deal with general U-statistics in our context (some kind of regularity is needed, for instance we could certainly deal with U-statistics based on functions which have bounded variation in each coordinate).

Remark 1.2. Let us comment on the rate of convergence in Theorem 1.1. Berry–Esseen type estimates seem difficult to obtain in our dependence context, but one can give precise rates of convergence with respect to the minimal Wasserstein metric of order 1. Let $\alpha_{2,\mathbf{X}}(n)$ and $\alpha_{2,\mathbf{Y}}(n)$ be the coefficients defined in Definition 2.1 below, and let Λ_1 be the set of Lipschitz functions such that $|f(x) - f(y)| \leq |x - y|$ for any reals x, y. Combining the decomposition (4.2) and Proposition 4.1 below with Theorem 2.2 in [10] (see also Theorem 5.1 and its proof in [9]), one can prove that

- If (1.3) holds, and $\sum_{k>0} k\,\alpha_{2,\mathbf{X}}(k) < \infty$ and $\sum_{k>0} k\,\alpha_{2,\mathbf{Y}}(k) < \infty$, then

$$\sup_{f\in\Lambda_1} \left|\mathbb{E}\big(f(\sqrt{n}(U_n - \pi))\big) - \mathbb{E}(f(N_V))\right| = O\left(\frac{\log n}{\sqrt{n}}\right),$$

  where N_V is a Gaussian random variable with mean 0 and variance V.

- If (1.3) holds, and $\alpha_{2,\mathbf{X}}(k) = O(k^{-1-\delta})$ and $\alpha_{2,\mathbf{Y}}(k) = O(k^{-1-\delta})$ for some $\delta \in (0,1)$, then

$$\sup_{f\in\Lambda_1} \left|\mathbb{E}\big(f(\sqrt{n}(U_n - \pi))\big) - \mathbb{E}(f(N_V))\right| = O\left(\frac{1}{n^{\delta/2}}\right).$$

Note that there is no "gap" between these results and Theorem 1.1 in the sense that we get a rate of convergence in the central limit theorem as soon as $\alpha_{2,\mathbf{X}}(k) = O(k^{-1-\delta})$ and $\alpha_{2,\mathbf{Y}}(k) = O(k^{-1-\delta})$ for some $\delta > 0$. Moreover, this rate of convergence is that of the iid case (up to the $\log n$ term) as soon as $\delta > 1$. To be complete, it is clear from the proof of Theorem 5.1 in [9] that the constants involved in these rates of convergence depend only on the two sequences $(\alpha_{2,\mathbf{X}}(k))$, $(\alpha_{2,\mathbf{Y}}(k))$, on the two series involved in the definition of V, and on the constant c in (1.3).

2. TESTING THE STOCHASTIC DOMINATION FROM TWO INDEPENDENT TIME SERIES

The statistic $\sqrt{n}(U_n - 1/2)$ cannot be used to test $H_0\colon \pi = 1/2$ versus one of the possible alternatives because under H_0 the asymptotic distribution of $\sqrt{n}(U_n - 1/2)$ depends on the unknown quantity V defined in (1.4). In the next Proposition, we propose an estimator of V under some slightly stronger conditions on the dependence coefficients.

Definition 2.1. Let $(X_i)_{i\in\mathbb{Z}}$ be a strictly stationary sequence of real-valued random variables, and let $\mathcal{F}_0=\sigma(X_k,k\le 0)$. Let $f_x(z)=\mathbf{1}_{z\le x}-F(x)$, where F is the cumulative distribution function of X_0. For any nonnegative integer n, let

$$\alpha_{2,\mathbf{X}}(n)=\max\Big\{\alpha_{1,\mathbf{X}}(n),\sup_{x,y\in\mathbb{R},j\ge i\ge n}\big\|\mathbb{E}\big(f_x(X_i)f_y(X_j)|\mathcal{F}_0\big)-\mathbb{E}\big(f_x(X_i)f_y(X_j)\big)\big\|_1\Big\}.$$

Proposition 2.1. *Let $(X_i)_{i\in\mathbb{Z}}$ and $(Y_i)_{i\in\mathbb{Z}}$ be two independent stationary sequences, and let (m_n) be a sequence of positive integers satisfying* (1.3)*. Assume that*

$$\sum_{k=1}^{\infty}\alpha_{2,\mathbf{X}}(k)<\infty\quad\textit{and}\quad\sum_{k=1}^{\infty}\alpha_{2,\mathbf{Y}}(k)<\infty. \tag{2.1}$$

Let

$$H_m(t)=\frac{1}{m}\sum_{j=1}^{m}\mathbf{1}_{Y_j>t}\quad\textit{and}\quad G_n(t)=\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{X_i<t},$$

and

$$\hat\gamma_X(k)=\frac{1}{n}\sum_{i=1}^{n-k}(H_m(X_i)-\bar H_m)(H_m(X_{i+k})-\bar H_m),$$

$$\hat\gamma_Y(\ell)=\frac{1}{m}\sum_{j=1}^{m-\ell}(G_n(Y_j)-\bar G_n)(G_n(Y_{j+\ell})-\bar G_n),$$

where $n\bar H_m=\sum_{i=1}^n H_m(X_i)$ and $m\bar G_n=\sum_{j=1}^m G_n(Y_j)$. Let also (a_n) and (b_n) be two sequences of positive integers tending to infinity as n tends to infinity such that $a_n=o(\sqrt{n})$ and $b_n=o(\sqrt{m_n})$. Then

$$V_n=\hat\gamma_X(0)+2\sum_{k=1}^{a_n}\hat\gamma_X(k)+\frac{n}{m_n}\Big(\hat\gamma_Y(0)+2\sum_{\ell=1}^{b_n}\hat\gamma_Y(\ell)\Big)$$

converges in $\mathbb{L}^2$ to the quantity V defined in (1.4)*.*

Combining Theorem 1.1 and Proposition 2.1, we obtain that, under $H_0\colon \pi=1/2$, if $V>0$, the random variables

$$T_n=\frac{\sqrt{n}(U_n-1/2)}{\sqrt{\max\{V_n,0\}}} \tag{2.2}$$

converge in distribution to $\mathcal{N}(0,1)$.

Remark 2.1. The condition (2.1) is not restrictive: in all the natural examples we know, the coefficients $\alpha_{2,\mathbf{X}}(k)$ behave as $\alpha_{1,\mathbf{X}}(k)$ (this is the case for instance for the linear process (1.6)). Note that if we assume (2.1) instead of (1.2), the ergodicity is no longer required in Theorem 1.1. Finally, following the proof of Theorem 4.1 in [2], p. 89, one can also build a consistent estimator of V under the condition

$$\sum_{k=1}^{\infty}\sqrt{\frac{\alpha_{1,\mathbf{X}}(k)}{k}}<\infty\quad\text{and}\quad\sum_{k=1}^{\infty}\sqrt{\frac{\alpha_{1,\mathbf{Y}}(k)}{k}}<\infty, \tag{2.3}$$

which is not directly comparable to (2.1) (for instance, the first part of (2.1) is satisfied if $\alpha_{2,\mathbf{X}}(k)=O(k^{-1}(\log k)^{-a})$ for some $a>1$, while the first part of (2.3) requires $\alpha_{1,\mathbf{X}}(k)=O(k^{-1}(\log k)^{-a})$ for some $a>2$).

Remark 2.2. The choice of the two sequences (a_n) and (b_n) is a delicate matter. If the coefficients $\alpha_{2,\mathbf{X}}(k)$ decrease very quickly, then a_n should increase very slowly (it suffices to take $a_n \equiv 0$ in the iid setting). On the contrary, if $\alpha_{2,\mathbf{X}}(k) = O(k^{-1}(\log k)^{-a})$ for some $a > 1$, then the terms in the covariance series have no reason to be small, and one should take a_n close to $\sqrt{n}$ to estimate many of these covariance terms. A data-driven criterion for choosing a_n and b_n is an interesting (but probably difficult) question, which is far beyond the scope of the present paper.

However, from a practical point of view, there is an easy way to proceed: one can plot the estimated covariances $\hat{\gamma}_X(k)$'s and choose a_n (not too large) in such a way that

$$\hat{\gamma}_X(0) + 2\sum_{k=1}^{a_n} \hat{\gamma}_X(k)$$

should represent an important part of the covariance series

$$\mathrm{Var}(H_Y(X_0)) + 2\sum_{k=1}^{\infty} \mathrm{Cov}(H_Y(X_0), H_Y(X_k)).$$

Similarly, one may choose b_n from the graph of the $\hat{\gamma}_Y(\ell)$'s. As we shall see in the simulations (Section 6), if the decays of the true covariances $\gamma_X(k)$ and $\gamma_Y(\ell)$ are not too slow, this provides an easy and reasonable choice for a_n and b_n.

3. LONG-RANGE DEPENDENCE

In this section, we shall see that condition (1.2) cannot be weakened for the validity of Theorem 1.1. More precisely, we shall give some examples where (1.2) fails, and the asymptotic distribution of $a_n(U_n - \pi)$ (if any) can be non-Gaussian.

Let us begin with the boundary case, where

$$\alpha_{2,\mathbf{X}}(k) = O((k+1)^{-1}) \quad \text{and} \quad \alpha_{2,\mathbf{Y}}(k) = O((k+1)^{-1}). \tag{3.1}$$

In that case, one can prove the following result.

Proposition 3.1. *Let $(X_i)_{i\in\mathbb{Z}}$ and $(Y_i)_{i\in\mathbb{Z}}$ be two independent stationary sequences. Assume that (3.1) is satisfied and let (m_n) be a sequence a positive integers satisfying (1.3). Then*

1. *For any $r \in (2,4)$, there exists a positive constant C_r such that, for any $x > 0$,*

$$\limsup_{n\to\infty} \mathbb{P}\big(\sqrt{n/\log n}(U_n - \pi) > x\big) \le \frac{C_r}{x^r}.$$

2. *There are some examples of stationary Markov chains $(X_i)_{i\in\mathbb{Z}}$ and $(Y_i)_{i\in\mathbb{Z}}$ for which the sequence $\sqrt{n/\log n}(U_n - \pi)$ converges in distribution to a nondegenerate normal distribution.*

We consider now the case where

$$\sum_{k=1}^{\infty} \alpha_{1,\mathbf{X}}(k) < \infty \quad \text{and} \quad \alpha_{1,\mathbf{Y}}(k) = O((k+1)^{-(p-1)}) \quad \text{for some } p \in (1,2). \tag{3.2}$$

This means that $(X_i)_{i\in\mathbb{Z}}$ is short-range dependent, while $(Y_i)_{i\in\mathbb{Z}}$ is possibly long-range dependent. In that case, we shall prove the following result.

Proposition 3.2. *Let $(X_i)_{i\in\mathbb{Z}}$ and $(Y_i)_{i\in\mathbb{Z}}$ be two independent stationary sequences. Assume that (3.2) is satisfied for some $p \in (1,2)$ and that $(X_i)_{i\in\mathbb{Z}}$ is 2-ergodic. Let (m_n) be a sequence of positive integers such that*

$$\lim_{n\to\infty} \frac{n}{m_n^{2(p-1)/p}} = c \quad \textit{for some } c \in [0,\infty). \tag{3.3}$$

Then

1. *There exists a positive constant C_p such that, for any $x > 0$,*

$$\limsup_{n\to\infty} \mathbb{P}\big(\sqrt{n}(U_n - \pi) > x\big) \le \frac{C_p}{x^p}.$$

2. *There are some examples of stationary Markov chains $(X_i)_{i\in\mathbb{Z}}$ and $(Y_i)_{i\in\mathbb{Z}}$ for which the sequence $\sqrt{n}(U_n - \pi)$ converges in distribution to $Z + \sqrt{c}S$, where Z is independent of S, Z is a Gaussian random variable, and S is a stable random variable of order p.*

Remark 3.1. Note that we do not deal with the case where the two samples are long-range dependent; in that case the degenerate U-statistic from the Hoeffding decomposition (see (4.2)) is no longer negligible, which gives rise to other possible limits. See the paper [12] for a treatment of this difficult question in the case of functions of Gaussian processes.

4. PROOFS

4.1. Proof of Theorem 1.1

The main ingredient of the proof is the following proposition.

Proposition 4.1. *Let $(X_i)_{i\in\mathbb{Z}}$ and $(Y_i)_{i\in\mathbb{Z}}$ be two independent stationary sequences. Let also*

$$f(x,y) = \mathbf{1}_{x<y} - H_Y(x) - G_X(y) + \pi. \tag{4.1}$$

Then

$$\mathbb{E}\Big(\Big(\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} f(X_i, Y_j)\Big)^2\Big) \le \frac{16}{mn}\sum_{i=0}^{n-1}\sum_{j=0}^{m-1} \min\{\alpha_{1,\mathbf{X}}(i), \alpha_{1,\mathbf{Y}}(j)\}.$$

Let us admit this proposition for a while, and see how it can be used to prove the theorem. Starting from (4.1), we easily see that

$$\sqrt{n}(U_n - \pi) = \frac{\sqrt{n}}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} f(X_i, Y_j) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(H_Y(X_i) - \pi) + \frac{\sqrt{n}}{m}\sum_{j=1}^{m}(G_X(Y_j) - \pi). \tag{4.2}$$

This is the well-known Hoeffding decomposition of U-statistics; the first term on the right-hand side of (4.2) is called a *degenerate* U-statistic, and is handled *via* Proposition 4.1. Indeed, it follows easily from this proposition that

$$n\mathbb{E}\Big(\Big(\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} f(X_i, Y_j)\Big)^2\Big) \le \frac{16}{m}\sum_{i=0}^{n-1}(i+1)\alpha_{1,\mathbf{X}}(i) + \frac{16}{m}\sum_{j=0}^{m-1}(j+1)\alpha_{1,\mathbf{Y}}(j). \tag{4.3}$$

Since (1.2) holds, then $\alpha_{1,\mathbf{X}}(n) = o(n^{-1})$ and $\alpha_{1,\mathbf{Y}}(n) = o(n^{-1})$ (because these coefficients decrease) and by Cesaro's lemma

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n}(i+1)\alpha_{1,\mathbf{X}}(i) = 0 \quad \text{and} \quad \lim_{m\to\infty}\frac{1}{m}\sum_{j=0}^{m}(j+1)\alpha_{1,\mathbf{Y}}(j) = 0.$$

From (4.3) and Condition (1.3), we infer that

$$\frac{\sqrt{n}}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} f(X_i, Y_j) \quad \text{converges to 0 in } \mathbb{L}^2 \text{ as } n \to \infty. \tag{4.4}$$

It remains to control the two last terms on the right-hand side of (4.2). We first note that these two terms are independent, so it is enough to prove the convergence in distribution for both terms. The functions

H_Y and G_X being monotonic, the central limit theorem for stationary and 2-ergodic α-dependent sequences (see for instance [5]) implies that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(H_Y(X_i)-\pi) \quad \text{converges in distribution to } \mathcal{N}(0,v_1), \text{ where}$$

$$v_1 = \mathrm{Var}(H_Y(X_0)) + 2\sum_{k=1}^{\infty}\mathrm{Cov}\left(H_Y(X_0), H_Y(X_k)\right), \tag{4.5}$$

and

$$\frac{1}{\sqrt{m}}\sum_{j=1}^{m}(G_X(Y_j)-\pi) \quad \text{converges in distribution to } \mathcal{N}(0,v_2), \text{ where}$$

$$v_2 = \mathrm{Var}(G_X(Y_0)) + 2\sum_{k=1}^{\infty}\mathrm{Cov}\left(G_X(Y_0), G_X(Y_k)\right). \tag{4.6}$$

Combining (4.2), (4.4), (4.5) and (4.6), the proof of Theorem 1.1 is complete.

It remains to prove Proposition 4.1. We start with the elementary inequality

$$\mathbb{E}\Big(\Big(\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}f(X_i,Y_j)\Big)^2\Big) \le \frac{4}{n^2m^2}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=i}^{n}\sum_{\ell=j}^{m}\mathbb{E}(f(X_i,Y_j)f(X_k,Y_\ell)). \tag{4.7}$$

To control the term $\mathbb{E}(f(X_i,Y_j)f(X_k,Y_\ell))$, we first work conditionally on $\mathbf{Y}=(Y_i)_{i\in\mathbb{Z}}$. Note first that, since $\mathbf{Y}$ is independent of $\mathbf{X}$, $\mathbb{E}(f(X_i,Y_j)|\mathbf{Y})=0$. Since $|f(x,y)|\le 2$,

$$\left|\mathbb{E}(f(X_i,Y_j)f(X_k,Y_l) \mid \mathbf{Y}=\mathbf{y})\right| \le 2\|\mathbb{E}(f(X_k,y_\ell)\mid\mathcal{F}_i)\|_1, \tag{4.8}$$

where $\mathcal{F}_i=\sigma(X_k, k\le i)$. Now, by definition of f,

$$\begin{aligned}\|\mathbb{E}(f(X_k,y_\ell)\mid\mathcal{F}_i)\|_1 &\le \|\mathbb{E}(\mathbf{1}_{X_k<y_\ell}\mid\mathcal{F}_i)-G_X(y_\ell)\|_1 + \|\mathbb{E}(H_Y(X_k)\mid\mathcal{F}_i)-\pi\|_1\\ &\le 2\alpha_{1,\mathbf{X}}(k-i).\end{aligned} \tag{4.9}$$

From (4.8) and (4.9), it follows that

$$|\mathbb{E}(f(X_i,Y_j)f(X_k,Y_\ell))| \le \|\mathbb{E}(f(X_i,Y_j)f(X_k,Y_\ell)\mid\mathbf{Y})\|_1 \le 4\alpha_{1,\mathbf{X}}(k-i). \tag{4.10}$$

In the same way

$$|\mathbb{E}(f(X_i,Y_j)f(X_k,Y_\ell))| \le 4\alpha_{1,\mathbf{Y}}(\ell-j). \tag{4.11}$$

Proposition 4.1 follows from (4.7), (4.10) and (4.11).

4.2. Proof of Proposition 2.1

We proceed in two steps.

Step 1. Let

$$\gamma_X^*(k)=\frac{1}{n}\sum_{i=1}^{n-k}(H_Y(X_i)-\pi)(H_Y(X_{i+k})-\pi), \quad \gamma_Y^*(\ell)=\frac{1}{m}\sum_{j=1}^{m-\ell}(G_X(Y_j)-\pi)(G_X(Y_{i+k})-\pi),$$

and note that

$$\begin{aligned}\mathbb{E}(\gamma_X^*(k)) &= \frac{n-k}{n}\mathrm{Cov}(H_Y(X_0),H_Y(X_k))\\ \mathbb{E}(\gamma_Y^*(\ell)) &= \frac{n-k}{n}\mathrm{Cov}(G_X(Y_0),G_X(Y_\ell)).\end{aligned} \tag{4.12}$$

In this first step, we shall prove that

$$V_n^* = \gamma_X^*(0) + 2\sum_{k=1}^{a_n} \gamma_X^*(k) + \frac{n}{m_n}\Big(\gamma_Y^*(0) + 2\sum_{\ell=1}^{b_n} \gamma_Y^*(\ell)\Big)$$

converges in $\mathbb{L}^2$ to V. Clearly, taking into account (4.12) and the fact that a_n, b_n tend to infinity as $n \to \infty$, it suffices to show that $\mathrm{Var}(V_n^*)$ converges to 0 as $n \to \infty$. We only deal with the second term in the expression of V_n^*, the other ones may be handled in the same way. Letting $Z_i = H_Y(X_i) - \pi$, and $\gamma_k = \mathrm{Cov}(H_Y(X_0), H_Y(X_k))$ we have

$$\mathrm{Var}\Big(\sum_{k=1}^{a_n} \gamma_X^*(k)\Big) = \frac{1}{n^2}\sum_{k=1}^{a_n}\sum_{\ell=1}^{a_n}\sum_{i=1}^{n-k}\sum_{j=1}^{n-\ell} \mathbb{E}\big((Z_iZ_{i+k} - \gamma_k)(Z_jZ_{j+\ell} - \gamma_\ell)\big). \tag{4.13}$$

We will examine several cases.

On $\Gamma_1 = \{j \geq i + k\}$,

$$|\mathbb{E}((Z_iZ_{i+k} - \gamma_k)(Z_jZ_{j+\ell} - \gamma_\ell))| \leq 2\alpha_{2,\mathbf{X}}(j - i - k).$$

Consequently

$$\begin{aligned}\frac{1}{n^2}\sum_{k=1}^{a_n}\sum_{\ell=1}^{a_n}\sum_{i=1}^{n-k}\sum_{j=1}^{n-\ell} &\big|\mathbb{E}((Z_iZ_{i+k} - \gamma_k)(Z_jZ_{j+\ell} - \gamma_\ell))\big|\mathbf{1}_{\Gamma_1} \\ &\leq \frac{2}{n^2}\sum_{k=1}^{a_n}\sum_{\ell=1}^{a_n}\sum_{i=1}^{n}\Big(\sum_{j\geq 0}\alpha_{2,\mathbf{X}}(j)\Big) \leq \frac{Ca_n^2}{n}\end{aligned} \tag{4.14}$$

for some positive constant C.

On $\Gamma_2 = \{i \leq j < i + k\} \cap \{i + k \leq j + \ell\}$,

$$|\mathbb{E}((Z_iZ_{i+k} - \gamma_k)(Z_jZ_{j+\ell} - \gamma_\ell))| = |\mathbb{E}((Z_iZ_{i+k} - \gamma_k)Z_jZ_{j+\ell})| \leq 2\alpha_{1,\mathbf{X}}(j + \ell - i - k).$$

Consequently

$$\begin{aligned}\frac{1}{n^2}\sum_{k=1}^{a_n}\sum_{\ell=1}^{a_n}\sum_{i=1}^{n-k}\sum_{j=1}^{n-\ell} &|\mathbb{E}((Z_iZ_{i+k} - \gamma_k)(Z_jZ_{j+\ell} - \gamma_\ell))|\mathbf{1}_{\Gamma_2} \\ &\leq \frac{2}{n^2}\sum_{k=1}^{a_n}\sum_{\ell=1}^{a_n}\sum_{i=1}^{n}\Big(\sum_{j\geq 0}\alpha_{1,\mathbf{X}}(j)\Big) \leq \frac{Ca_n^2}{n}.\end{aligned} \tag{4.15}$$

On $\Gamma_3 = \{i \leq j\} \cap \{i + k > j + \ell\}$,

$$\big|\mathbb{E}\big((Z_iZ_{i+k} - \gamma_k)(Z_jZ_{j+\ell} - \gamma_\ell)\big)\big| = |\mathbb{E}((Z_jZ_{j+\ell} - \gamma_\ell)Z_iZ_{i+k})| \leq 2\alpha_{1,\mathbf{X}}(i + k - j - \ell).$$

Consequently

$$\begin{aligned}\frac{1}{n^2}\sum_{k=1}^{a_n}\sum_{\ell=1}^{a_n}\sum_{i=1}^{n-k}\sum_{j=1}^{n-\ell} &\big|\mathbb{E}\big((Z_iZ_{i+k} - \gamma_k)(Z_jZ_{j+\ell} - \gamma_\ell)\big)\big|\mathbf{1}_{\Gamma_3} \\ &\leq \frac{2}{n^2}\sum_{k=1}^{a_n}\sum_{\ell=1}^{a_n}\sum_{j=1}^{n}\Big(\sum_{i\geq 0}\alpha_{1,\mathbf{X}}(i)\Big) \leq \frac{Ca_n^2}{n}.\end{aligned} \tag{4.16}$$

From (4.14), (4.15) and (4.16), setting $\Gamma = \{i \leq j\} = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3$, one gets

$$\frac{1}{n^2}\sum_{k=1}^{a_n}\sum_{\ell=1}^{a_n}\sum_{i=1}^{n-k}\sum_{j=1}^{n-\ell} \big|\mathbb{E}\big((Z_iZ_{i+k} - \gamma_k)(Z_jZ_{j+\ell} - \gamma_\ell)\big)\big|\mathbf{1}_{\Gamma} \leq \frac{3Ca_n^2}{n}.$$

Of course, interchanging i and j, the same is true on $\Gamma^c = \{i > j\}$, and finally

$$\frac{1}{n^2}\sum_{k=1}^{a_n}\sum_{\ell=1}^{a_n}\sum_{i=1}^{n-k}\sum_{j=1}^{n-\ell}\left|\mathbb{E}\big((Z_iZ_{i+k}-\gamma_k)(Z_jZ_{j+\ell}-\gamma_\ell)\big)\right| \le \frac{6Ca_n^2}{n}. \tag{4.17}$$

From (4.13), (4.17) and the fact that $a_n = o(\sqrt{n})$, we infer that $\sum_{k=1}^{a_n}\gamma_X^*(k)$ converges in $\mathbb{L}^2$ to $\sum_{k>0}\mathrm{Cov}(H_Y(X_0), H_Y(X_k))$. Since the other terms in the definition of V_n^* can be handled in the same way, it follows that V_n^* converges in $\mathbb{L}^2$ to V.

Step 2. In this second step, in the expression of $\gamma_X^*(k)$, we replace $H_Y(X_i)$ by $H_m(X_i)$, $H_Y(X_{i+k})$ by $H_m(X_{i+k})$, and π by $\bar{H}_m$. Similarly, in the expression of $\gamma_Y^*(\ell)$, we replace $G_X(Y_j)$ by $G_n(Y_j)$, $G_X(Y_{j+\ell})$ by $G_n(Y_{j+\ell})$, and π by $\bar{G}_n$. If we can prove that these replacements in the expression of V_n^* are negligible in $\mathbb{L}^2$, the conclusion of Proposition 2.1 will follow.

Let

$$\gamma_{1,X}(k) = \frac{1}{n}\sum_{i=1}^{n-k}(H_m(X_i)-\pi)(H_Y(X_{i+k})-\pi).$$

Then

$$\|\gamma_X^*(k)-\gamma_{1,X}(k)\|_2 \le \|H_m(X_i)-H_Y(X_i)\|_2 = \left(\int \|H_m(x)-H_Y(x)\|_2^2\mathbb{P}_X\,(dx)\right)^{1/2}, \tag{4.18}$$

the last equality being true because $(X_i)_{i\in\mathbb{Z}}$ and $(Y_i)_{i\in\mathbb{Z}}$ are independent. Now,

$$\begin{aligned}\|H_m(x)-H_Y(x)\|_2^2 &\le \frac{1}{m}\Big(\mathrm{Var}(\mathbf{1}_{Y_0>x}) + 2\sum_{k=1}^{n-1}\big|\,\mathrm{Cov}(\mathbf{1}_{Y_0>x},\mathbf{1}_{Y_k>x})\big|\Big)\\ &\le \frac{2\sum_{k=0}^{n-1}\alpha_{1,\mathbf{Y}}(k)}{m}.\end{aligned} \tag{4.19}$$

From (4.18) and (4.19), we infer that there exists a positive constant C_1 such that

$$\Big\|\gamma_X^*(0)-\gamma_{1,X}(0)+2\sum_{k=1}^{a_n}(\gamma_X^*(k)-\gamma_{1,X}(k))\Big\|_2 \le \frac{C_1a_n}{\sqrt{m}}.$$

By condition (1.3) and the fact that $a_n = o(\sqrt{n})$, this last quantity tends to 0 as $n\to\infty$.

Let now

$$\gamma_{2,X}(k) = \frac{1}{n}\sum_{i=1}^{n-k}(H_m(X_i)-\bar{H}_m)(H_Y(X_{i+k})-\pi).$$

Then

$$\|\gamma_{1,X}(k)-\gamma_{2,X}(k)\|_2 \le \Big\|\bar{H}_m-\frac{1}{n}\sum_{i=1}^{n}H_Y(X_i)\Big\|_2 + \Big\|\frac{1}{n}\sum_{i=1}^{n}H_Y(X_i)-\pi\Big\|_2. \tag{4.20}$$

Proceeding as in (4.19) for the first term on the right-hand side of (4.20) and noting that

$$\Big\|\frac{1}{n}\sum_{i=1}^{n}H_Y(X_i)-\pi\Big\|_2^2 \le \frac{2\sum_{k=0}^{n-1}\alpha_{1,\mathbf{X}}(k)}{n},$$

we get that there exists a positive constant C_2 such that

$$\Big\|\gamma_{1,X}(0)-\gamma_{2,X}(0)+2\sum_{k=1}^{a_n}(\gamma_{1,X}(k)-\gamma_{2,X}(k))\Big\|_2 \le \frac{C_2a_n}{\sqrt{m}}+\frac{C_2a_n}{\sqrt{n}}.$$

Again, this last quantity tends to 0 as $n\to\infty$.

Finally, all the other replacements may be done in the same way and lead to negligible contributions in $\mathbb{L}^2$. Hence

$$\lim_{n\to\infty} \|V_n - V_n^*\|_2 = 0$$

and the result follows from Step 1.

4.3. Proof of Proposition 3.1

We start with (4.2) again, but with the normalization $\sqrt{n/\log n}$ instead of $\sqrt{n}$. Since (3.1) holds,

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n}(i+1)\alpha_{1,\mathbf{X}}(i) < \infty \quad \text{and} \quad \limsup_{m\to\infty} \frac{1}{m}\sum_{j=0}^{m}(j+1)\alpha_{1,\mathbf{Y}}(j) < \infty.$$

From Inequality (4.3) and Condition (1.3), we infer that

$$\frac{\sqrt{n}}{nm\sqrt{\log n}}\sum_{i=1}^{n}\sum_{j=1}^{m} f(X_i, Y_j) \quad \text{converges to 0 in } \mathbb{L}^2 \text{ as } n \to \infty. \tag{4.21}$$

Let us prove Item 1. From (4.2) and (4.21),

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{P}\big(\sqrt{n/\log n}(U_n - \pi) > x\big) &\le \limsup_{n\to\infty} \mathbb{P}\Big(\sum_{i=1}^{n}(H_Y(X_i) - \pi) > x\sqrt{n\log n}/2\Big) \\ &\quad + \limsup_{n\to\infty} \mathbb{P}\Big(\frac{n}{m}\sum_{j=1}^{m}(G_X(Y_i) - \pi) > x\sqrt{n\log n}/2\Big). \end{aligned} \tag{4.22}$$

Since (3.1) holds, it follows from Proposition 5.1 in [6] that for any $r \in (2,4)$ there exists a positive constant $c_{1,r}$ such that for any $x > 0$

$$\limsup_{n\to\infty} \mathbb{P}\Big(\sum_{i=1}^{n}(H_Y(X_i) - \pi) > x\sqrt{n\log n}/2\Big) \le \frac{c_{1,r}}{x^r}.$$

Since (1.3) holds, the same is true for the second term on the right-hand side of (4.22) (for a positive constant $c_{2,r}$), and Item 1 follows.

Let us prove Item 2. We use a result by Gouëzel [14] about intermittent maps. Let $\theta_{1/2}$ be the map described in Subsection 6.2 of the present paper. As explained in Subsection 6.2, there exists a unique absolutely continuous $\theta_{1/2}$-invariant measure $\nu_{1/2}$ on $[0,1]$, and (see the comments on the case $\alpha = 1/2$ in Section 1.3 of Gouëzel's paper), on the probability space $([0,1], \nu_{1/2})$,

$$\frac{1}{\sqrt{n\log n}}\sum_{k=1}^{n}\big(H_Y(\theta_{1/2}^k) - \nu_{1/2}(H_Y)\big)$$

converges in distribution to a nondegenerate normal distribution provided $H_Y(0) \neq \nu(H_Y)$ (which is true for instance if Y has distribution $\nu_{1/2}$, since in that case $H_Y(0) = 1$ and $\nu(H_Y) = 1/2$). As fully explained in Subsection 6.2, there exists a stationary Markov chain $(X_i)_{i\in\mathbb{Z}}$ such that the distribution of $(X_1, \dots, X_n)$ is the same as $(\theta_{1/2}^n, \dots, \theta_{1/2})$ on the probability space $([0,1], \nu_{1/2})$. Take $(Y_i)_{i\in\mathbb{Z}}$ an independent copy of the chain $(X_i)_{i\in\mathbb{Z}}$. For such chains, both

$$\frac{1}{\sqrt{n\log n}}\sum_{i=1}^{n}(H_Y(X_i) - \pi) \quad \text{and} \quad \frac{1}{\sqrt{m\log m}}\sum_{j=1}^{m}(G_X(Y_j) - \pi)$$

converge to the same nondegenerate normal distribution (note that $\pi = 1/2$ in that case). Moreover, as recalled in Subsection 6.2, there exist two positive constants A, B such that

$$\frac{A}{k+1} \le \alpha_{2,\mathbf{X}}(k) = \alpha_{2,\mathbf{Y}}(k) \le \frac{B}{k+1}.$$

This completes the proof of Item 2.

4.4. Proof of Proposition 3.2

We start with (4.2) again. At the end of the proof, we shall prove that if (3.2) and (3.3) are satisfied, then (4.4) holds.

Let us now prove Item 1. From (4.2) and (4.4),

$$\limsup_{n\to\infty}\mathbb{P}\big(\sqrt{n}(U_n-\pi)>x\big)\le\limsup_{n\to\infty}\mathbb{P}\Big(\sum_{i=1}^{n}(H_Y(X_i)-\pi)>x\sqrt{n}/2\Big)$$
$$+\limsup_{n\to\infty}\mathbb{P}\bigg(\frac{n}{m}\sum_{j=1}^{m}(G_X(Y_i)-\pi)>x\sqrt{n}/2\bigg). \tag{4.23}$$

Since the first part of (3.2) is satisfied, it follows from the central limit theorem for stationary and 2-ergodic α-dependent sequences that

$$\lim_{n\to\infty}\mathbb{P}\Big(\sum_{i=1}^{n}(H_Y(X_i)-\pi)>x\sqrt{n}/2\Big)=\mathbb{P}(Z>x/2), \tag{4.24}$$

where Z is a Gaussian random variable. From Condition (3.2) and Remark 3.3 in [6], we infer that if (3.3) holds, then there exists a positive constant κ_p such that, for any $x>0$,

$$\limsup_{n\to\infty}\mathbb{P}\bigg(\frac{n}{m}\sum_{j=1}^{m}(G_X(Y_i)-\pi)>x\sqrt{n}/2\bigg)\le\frac{\kappa_p}{x^p}. \tag{4.25}$$

Item 1 follows from (4.23), (4.24) and (4.25).

Let us prove Item 2. For the sequence $(X_i)_{i\in\mathbb{Z}}$, take a sequence of iid random variables uniformly distributed over $[0,1]$. To find the sequence $(Y_i)_{i\in\mathbb{Z}}$ we use again a result by Gouëzel [14] about intermittent maps. Let $\theta_{1/p}$ be the map described in Subsection 6.2 of the present paper. As explained is Subsection 6.2, there exists a unique absolutely continuous $\theta_{1/p}$-invariant measure $\nu_{1/p}$ on $[0,1]$, and (see Theorem 1.3 in [14]), on the probability space $([0,1],\nu_{1/p})$,

$$\frac{1}{m^{1/p}}\sum_{k=1}^{m}\big(G_X(\theta_{1/p}^k)-\nu_{1/p}(G_X)\big)$$

converges in distribution to a nondegenerate stable distribution of order p provided $G_X(0)\neq\nu(G_X)$ (this condition is satisfied here because $G_X(0)=0$ and $\nu_{1/p}(G_X)>0$). As fully explained in Subsection 6.2, there exists a stationary Markov chain $(Y_i)_{i\in\mathbb{Z}}$ such that the distribution of $(Y_1,\dots,Y_n)$ is the same as $(\theta_{1/p}^n,\dots,\theta_{1/p})$ on the probability space $([0,1],\nu_{1/p})$. For such a chain,

$$\frac{1}{m^{1/p}}\sum_{j=1}^{m}(G_X(Y_j)-\pi)$$

converges in distribution to a nondegenerate stable distribution of order p. Moreover, as recalled in Subsection 6.2, there exist two positive constants A,B such that

$$\frac{A}{(k+1)^{p-1}}\le\alpha_{1,\mathbf{Y}}(k)\le\frac{B}{(k+1)^{p-1}}.$$

This completes the proof of Item 2.

It remains to prove (4.4). From Proposition 4.1 we get that

$$n\mathbb{E}\bigg(\bigg(\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}f(X_i,Y_j)\bigg)^2\bigg)\le\sum_{i=0}^{\infty}\bigg(\frac{16}{m}\sum_{j=0}^{m-1}\min\{\alpha_{1,\mathbf{X}}(i),\alpha_{1,\mathbf{Y}}(j)\}\bigg). \tag{4.26}$$

Note that

$$\frac{16}{m}\sum_{j=0}^{m-1}\min\{\alpha_{1,\mathbf{X}}(i),\alpha_{1,\mathbf{Y}}(j)\}\le 16\alpha_{1,\mathbf{X}}(i), \tag{4.27}$$

and that, for any positive integer i, since $\alpha_{1,\mathbf{Y}}(j)\to 0$ as $j\to\infty$,

$$\lim_{m\to\infty}\frac{16}{m}\sum_{j=0}^{m-1}\min\{\alpha_{1,\mathbf{X}}(i),\alpha_{1,\mathbf{Y}}(j)\}=0, \tag{4.28}$$

by Cesaro's lemma. Since $\sum_{i>0}\alpha_{1,\mathbf{X}}(i)<\infty$, it follows from (4.26), (4.27), (4.28), and the dominated convergence theorem that

$$\lim_{n\to\infty} n\mathbb{E}\left(\left(\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}f(X_i,Y_j)\right)^2\right)=0,$$

proving (4.4).

5. OTHER RELATED RESULTS

5.1. Testing the Stochastic Domination by a Known Distribution

We briefly describe here a procedure for testing the (weak) domination by a known distribution μ. Let $H(t)=\mu((t,\infty))$, and let $(X_i)_{i\in\mathbb{Z}}$ be a stationary 2-ergodic sequence of real-valued random variables. Let $\pi=\mathbb{E}(H(X_0))$ and define

$$\bar{H}_n=\frac{1}{n}\sum_{i=1}^{n}H(X_i).$$

If

$$\sum_{k=1}^{\infty}\alpha_{1,\mathbf{X}}(k)<\infty, \tag{5.1}$$

then the random variables $\sqrt{n}(\bar{H}_n-\pi)$ converge in distribution to $\mathcal{N}(0,V_1)$, where

$$V_1=\mathrm{Var}(H(X_0))+2\sum_{k=1}^{\infty}\mathrm{Cov}(H(X_0),H(X_k)). \tag{5.2}$$

Now, we want to test $H_0\colon \pi=1/2$ (no relation of weak domination between μ and $\mathbb{P}_X$) against one of the possible alternatives: $H_1\colon \pi\neq 1/2$ (one of the two distributions weakly dominates the other one), $H_1\colon \pi>1/2$ (μ weakly dominates $\mathbb{P}_X$), $H_1\colon \pi<1/2$ ($\mathbb{P}_X$ weakly dominates μ).

Again, one cannot use directly the statistics $\sqrt{n}(\bar{H}_n-1/2)$ to test $H_0\colon \pi=1/2$ versus one of the possible alternatives because under H_0 the asymptotic distribution of $\sqrt{n}(\bar{H}_n-1/2)$ depends of the unknown quantity V_1 defined in (5.2).

To estimate V_1, we take

$$V_{1,n}=\hat{\gamma}(0)+2\sum_{k=1}^{a_n}\hat{\gamma}(k)$$

for some sequence (a_n) of positive integers tending to infinity as n goes to infinity and such that $a_n=o(\sqrt{n})$, where

$$\hat{\gamma}(k)=\frac{1}{n}\sum_{i=1}^{n-k}(H(X_i)-\bar{H}_n)(H(X_{i+k})-\bar{H}_n).$$

Now, as in Proposition 2.1, if (5.1) holds for $\alpha_{2,\mathbf{X}}(k)$ instead of $\alpha_{1,\mathbf{X}}(k)$, then v_n converges in $\mathbb{L}^2$ to v. It follows that under $H_0\colon \pi = 1/2$, if $v > 0$, the random variables

$$T'_n = \frac{\sqrt{n}(\bar{H}_n - 1/2)}{\sqrt{\max\{V_{1,n}, 0\}}}$$

converge in distribution to $\mathcal{N}(0,1)$.

5.2. The Case of Two Adjacent Samples from the Same Time Series

Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary sequence of real-valued random variables. Let $m = m_n$ be a sequence of positive integers such that $m_n \to \infty$ as $n \to \infty$. We consider here the U-statistic

$$U_n = U_{n,m} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=n+1}^{n+m} \mathbf{1}_{X_i<Y_j},$$

where $Y_j = f(X_j)$, f being a monotonic function. This statistic can be used to test if there is a change in the time series after the time n, and more precisely if Y_{n+1} weakly dominates X_1. Let $\pi = \mathbb{P}(X_1 < f(X_1^*))$, where X_1^* is an independent copy of X_1. Let also $H_Y(t) = \mathbb{P}(f(X_1) > t)$ and $G_X(t) = \mathbb{P}(X_1 < t)$.

Finding the asymptotic behavior of U_n is a slightly more difficult problem than for two independent samples because we cannot use the independence to control the degenerate U-statistic as in Proposition 4.1. However it can be done by using a more restrictive coefficient than $\alpha_{2,\mathbf{X}}$.

Definition 5.1. Let $(X_i)_{i\in\mathbb{Z}}$ be a strictly stationary sequence of real-valued random variables, and let $\mathcal{F}_0 = \sigma(X_k, k \le 0)$. Let P be the law of X_0 and $P_{(X_i,X_j)}$ be the law of (X_i, X_j). Let $P_{X_k|\mathcal{F}_0}$ be the conditional distribution of X_k given $\mathcal{F}_0$, and let $P_{(X_i,X_j)|\mathcal{F}_0}$ be the conditional distribution of (X_i, X_j) given $\mathcal{F}_0$. Define the random variables

$$b(k) = \sup_{x\in\mathbb{R}} \big|P_{X_k|\mathcal{F}_0}(f_x)\big|,$$
$$b(i,j) = \sup_{(x,y)\in\mathbb{R}^2} \big|P_{(X_i,X_j)|\mathcal{F}_0}(f_x \otimes f_y) - P_{(X_i,X_j)}(f_x \otimes f_y)\big|,$$

where the function f_x has been defined in Definition 2.1. Define now the coefficients

$$\beta_{1,\mathbf{X}}(k) = \mathbb{E}(b(k)) \quad \text{and} \quad \beta_{2,\mathbf{X}}(k) = \max\big\{\tilde{\beta}_{1,\mathbf{X}}(k), \sup_{i>j\ge k} \mathbb{E}((b(i,j)))\big\}.$$

Of course, by definition of the coefficients, we always have that $\alpha_{2,\mathbf{X}}(k) \le \beta_{2,\mathbf{X}}(k)$. The coefficients $\beta_{2,\mathbf{X}}$ are weaker than the usual β-mixing coefficients of $(X_i)_{i\in\mathbb{Z}}$. Many examples of nonmixing processes for which $\beta_{2,\mathbf{X}}$ can be computed are given in [8]. For all the examples that we mention in the present paper, the coefficient $\beta_{2,\mathbf{X}}(k)$ is of the same order as the coefficient $\alpha_{2,\mathbf{X}}(k)$ (see [4] for the example of intermittent maps described in Subsection 6.2).

Theorem 5.1. *Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary sequence of real-valued random variables. Assume that*

$$\sum_{k=1}^{\infty} \beta_{2,\mathbf{X}}(k) < \infty. \tag{5.3}$$

Let (m_n) be a sequence of positive integers satisfying (1.3)*. Then $\sqrt{n}(U_n - \pi)$ converge in distribution to $\mathcal{N}(0, V)$ with*

$$V = \mathrm{Var}(H_Y(X_0)) + 2\sum_{k=1}^{\infty} \mathrm{Cov}(H_Y(X_0), H_Y(X_k))$$
$$+ c\Big(\mathrm{Var}(G_X(f(X_0))) + 2\sum_{k=1}^{\infty} \mathrm{Cov}(G_X(f(X_0)), G_X(f(X_k)))\Big). \tag{5.4}$$

For the estimation of V, one can prove the following analogue of Proposition 2.1.

Proposition 5.1. *Let $(X_i)_{i\in\mathbb{Z}}$ be an independent stationary sequence, and let $Y_j = f(X_j)$, where f is a monotone function. Let (m_n) be a sequence of positive integers satisfying* (1.3)*. Assume that* (5.3) *is satisfied. Let*

$$H_m(t) = \frac{1}{m}\sum_{j=n+1}^{n+m} \mathbf{1}_{Y_j>t} \quad \text{and} \quad G_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i<t},$$

and

$$\hat{\gamma}_X(k) = \frac{1}{n}\sum_{i=1}^{n-k}(H_m(X_i) - \bar{H}_m)(H_m(X_{i+k}) - \bar{H}_m),$$

$$\hat{\gamma}_Y(\ell) = \frac{1}{m}\sum_{j=n+1}^{n+m-\ell}(G_n(Y_j) - \bar{G}_n)(G_n(Y_{j+\ell}) - \bar{G}_n),$$

where $n\bar{H}_m = \sum_{i=1}^{n} H_m(X_i)$ and $m\bar{G}_n = \sum_{j=n+1}^{n+m} G_n(Y_j)$. Let also (a_n) and (b_n) be two sequences of positive integers tending to infinity as n tends to infinity such that $a_n = o(\sqrt{n}/\log(n))$ and $b_n = o(\sqrt{m_n}/\log(m_n))$. Then

$$V_n = \hat{\gamma}_X(0) + 2\sum_{k=1}^{a_n}\hat{\gamma}_X(k) + \frac{n}{m_n}\Big(\hat{\gamma}_Y(0) + 2\sum_{\ell=1}^{b_n}\hat{\gamma}_Y(\ell)\Big)$$

converges in $\mathbb{L}^2$ to the quantity V defined in (5.4)*.*

Combining Theorem 5.1 and Proposition 5.1, we obtain that under $H_0\colon \pi = 1/2$, if $V > 0$, the random variables

$$T_n = \frac{\sqrt{n}(U_n - 1/2)}{\sqrt{\max\{V_n, 0\}}} \tag{5.5}$$

converge in distribution to $\mathcal{N}(0,1)$.

The proof of Proposition 5.1 follows the same lines as that of Proposition 2.1, so we do not give the details here. The main change concerns Step 2, Inequalities (4.18) and (4.19), where we used independence between the samples. Here instead, one can use the fact that

$$\big\|\sup_{x\in\mathbb{R}}|H_n(x) - H(x)|\big\|_2^2 \le C(\log(m))^2\frac{\sum_{k=0}^{m-1}\beta_{1,\mathbf{X}}(k)}{m}.$$

This can be easily proved by following the proof of Proposition 7.1 in Rio [16] up to (7.8) and by using the computations in the proof of Theorem 2.1 in [3].

The proof of Theorem 5.1 also follows that of Theorem 1.1; the only nontrivial change concerns Proposition 4.1. Here, we have to deal with

$$L_n = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=n+1}^{n+m} f(X_i, Y_j),$$

where f is defined in (4.1). We want to prove that $n\mathbb{E}(L_n^2)$ tends to 0 as $n\to\infty$. As in the proof of Proposition 4.1, we start with the elementary inequality

$$\mathbb{E}(L_n^2) \le \frac{4}{n^2m^2}\sum_{i=1}^{n}\sum_{j=n+1}^{n+m}\sum_{k=i}^{n}\sum_{\ell=j}^{n+m}\mathbb{E}(f(X_i,Y_j)f(X_k,Y_\ell)). \tag{5.6}$$

Let $(X_i^*)_{i\in\mathbb{Z}}$ be an independent copy of $(X_i^*)_{i\in\mathbb{Z}}$ and $Y_i^* = f(X_i^*)$, and write

$$\mathbb{E}\big(f(X_i,Y_j)f(X_k,Y_\ell)\big) = \mathbb{E}\big(f(X_i,Y_j)f(X_k,Y_\ell)\big) - \mathbb{E}\big(f(X_i,Y_j^*)f(X_k,Y_\ell^*)\big) + \mathbb{E}\big(f(X_i,Y_j^*)f(X_k,Y_\ell^*)\big). \tag{5.7}$$

Now, using the independence, the second term on the right-hand side of (5.7) can be handled exactly as in Proposition 4.1. It remains to deal with the first term in the right-hand side of (5.7). Note that

$$\begin{aligned}\big|\mathbb{E}\big(f(X_i,Y_j)f(X_k,Y_\ell)\big) - \mathbb{E}\big(f(X_i,Y_j^*)f(X_k,Y_\ell^*)\big)\big| \\ \le \int \big|\mathbb{E}\big(f(x_i,Y_j)f(x_k,Y_\ell) \mid X_i = x_i, X_k = x_k\big) \\ - \mathbb{E}\big(f(x_i,Y_j)f(x_k,Y_\ell)\big)\big| P_{(X_i,X_k)}(dx_i,dx_k).\end{aligned} \tag{5.8}$$

For any fixed x, the function $y \to h_x(y) = f(x,y)$ is a bounded variation function whose variation is bounded by 2, and $\mathbb{E}(h_x(X_i)) = 0$. Proceeding as in Lemma 1 of [7], one can then prove that

$$\begin{aligned}\int \big|\mathbb{E}\big(f(x_i,Y_j)f(x_k,Y_\ell) \mid X_i = x_i, X_k = x_k\big) - \mathbb{E}\big(f(x_i,Y_j)f(x_k,Y_\ell)\big)\big| P_{(X_i,X_k)}(dx_i,dx_k) \\ \le 4\beta_{2,\mathbf{X}}(j-k).\end{aligned} \tag{5.9}$$

Considering separately the two cases $j-k > n^{1/4}$ and $j-k \le n^{1/4}$, we get the upper bound

$$\begin{aligned}\frac{1}{n^2m^2}\sum_{i=1}^{n}\sum_{j=n+1}^{n+m}\sum_{k=i}^{n}\sum_{\ell=j}^{n+m}\big|\mathbb{E}\big(f(X_i,Y_j)f(X_k,Y_\ell)\big) - \mathbb{E}\big(f(X_i,Y_j^*)f(X_k,Y_\ell^*)\big)\big| \\ \le \frac{4}{\sqrt{n}m} + \frac{4}{m}\sum_{k>n^{1/4}}\beta_{2,\mathbf{X}}(k)\end{aligned} \tag{5.10}$$

from which we easily derive that $n\mathbb{E}(L_n^2)$ tends to 0 as $n \to \infty$ (since (1.3) and (5.3) are satisfied).

6. SIMULATIONS

6.1. Example 1: Functions of an Auto-Regressive Process

In this section, we first simulate $(Z_1,\dots,Z_n)$, according to the simple AR(1) equation

$$Z_{k+1} = \frac{1}{2}(Z_k + \varepsilon_{k+1}),$$

where Z_1 is uniformly distributed over $[0,1]$, and $(\varepsilon_i)_{i\ge2}$ is a sequence of iid random variables with distribution $\mathcal{B}(1/2)$ independent of Z_1.

One can check that the transition kernel of the chain $(Z_i)_{i\ge1}$ is

$$K(f)(x) = \frac{1}{2}\left(f\left(\frac{x}{2}\right) + f\left(\frac{x+1}{2}\right)\right),$$

and that the uniform distribution on $[0,1]$ is the unique invariant distribution by K. Hence, the chain $(Z_i)_{i\ge1}$ is strictly stationary.

It is well known that the chain $(Z_i)_{i\ge1}$ is not α-mixing in the sense of Rosenblatt [17] (see for instance [1]). However, one can prove that the coefficients $\alpha_{2,\mathbf{X}}$ of the chain $(Z_i)_{i\ge1}$ are such that

$$\alpha_{2,\mathbf{Z}}(k) \le 2^{-k}$$

(and the same upper bound is true for $\beta_{2,\mathbf{Z}}(k)$, see for instance Section 6.1 in [8]).

Let now Q_{μ,σ^2} be the inverse of the cumulative distribution function of the law $\mathcal{N}(\mu,\sigma^2)$. Let then

$$X_i = Q_{\mu_1,\sigma_1^2}(Z_i).$$

The sequence $(X_i)_{i\le 1}$ is also a stationary Markov chain (as an invertible function of a stationary Markov chain), and one can easily check that $\alpha_{2,\mathbf{X}}(k) = \alpha_{2,\mathbf{Z}}(k)$. By construction, X_i is $\mathcal{N}(\mu, \sigma^2)$-distributed, but the sequence $(X_i)_{i\ge 1}$ is not a Gaussian process (otherwise it would be mixing in the sense of Rosenblatt).

Next, we simulate iid random variables $(Y_1, \dots, Y_m)$ with common distribution $\mathcal{N}(\mu_2, \sigma_2^2)$.

Note that, if $\mu_1 = \mu_2$, the hypothesis $H_0 \colon \pi = 1/2$ is satisfied and the random variables T_n defined in (2.2) converge in distribution to $\mathcal{N}(0, 1)$.

We shall study the behavior of T_n for $\sigma_1^2 = 4$, $\sigma_2^2 = 1$ and different choices of μ_1, μ_2, a_n and b_n. As explained in Section 2, the quantities a_n and b_n may be chosen by analyzing the graph of the estimated auto-covariances $\hat{\gamma}_X(k)$ and $\hat{\gamma}_Y(\ell)$. For this examples, these graphs suggest a choice of $a_n = 3$ or 4 and $b_n = 0$ (see Fig. 1).

We shall compute T_n for different choices of (n, m) (from $(150, 100)$ to $(750, 500)$ with $n/m = 1.5$). We estimate the three quantities $\mathrm{Var}(T_n)$, $\mathbb{P}(T_n > 1.645)$ and $\mathbb{P}(|T_n| > 1.96)$ via a classical Monte-Carlo procedure, by averaging over $N = 2000$ independent trials.

If $\mu_1 = \mu_2$, we say that $\mathbb{P}(T_n > 1.645)$ is level 1 (the level of the one-sided test) and $\mathbb{P}(|T_n| > 1.96)$ is level 2 (the level of the two-sided test). If a_n, b_n are well chosen, the estimated variance should be close to 1 and the estimated levels 1 and 2 should be close to 0.05.

If $\mu_1 \neq \mu_2$, we say that $\mathbb{P}(T_n > 1.645)$ is power 1 (the power of the one-sided test) and $\mathbb{P}(|T_n| > 1.96)$ is power 2 (the power of the two-sided test).

• Case $\mu_1 = \mu_2 = 0$ and $a_n = 0$, $b_n = 0$ (no correction):

$n; m$	150 ; 100	300 ; 200	450 ; 300	600 ; 400	750 ; 500
Estimated variance	2.13	2.199	2.118	2.112	2.273
Estimated level 1	0.129	0.145	0.121	0.132	0.14
Estimated level 2	0.177	0.194	0.18	0.18	0.191

Here, since $a_n = 0$, we do not estimate any of the covariance terms $\gamma_X(\ell)$; hence, we compute T_n as if both series $(X_i)_{i\ge 1}$ and $(Y_j)_{j\ge 1}$ were uncorrelated. The result is that the estimated variance is too large (around 2.1 whatever (n, m)) and so are the estimated levels 1 and 2 (around 0.13 for level 1 and 0.18 for level 2, whatever (n, m)). This means that the-one sided or two-sided tests build with T_n will reject the null hypothesis too often.

• Case $\mu_1 = \mu_2 = 0$ and $a_n = 3$, $b_n = 0$:

$n; m$	150 ; 100	300 ; 200	450 ; 300	600 ; 400	750 ; 500
Estimated variance	1.227	1.148	1.172	1.119	1.018
Estimated level 1	0.071	0.061	0.065	0.06	0.046
Estimated level 2	0.074	0.069	0.071	0.066	0.054

As suggested by the graphs of the estimated auto-covariances (see Figure 1), the choice $a_n = 3$, $b_n = 0$ should give a reasonable estimation of V. This is indeed the case: the estimated variance is now around 1.1, the estimated level 1 is around 0.06 and the estimated level 2 around 0.07.

• Case $\mu_1 = \mu_2 = 0$ and $a_n = 4$, $b_n = 0$:

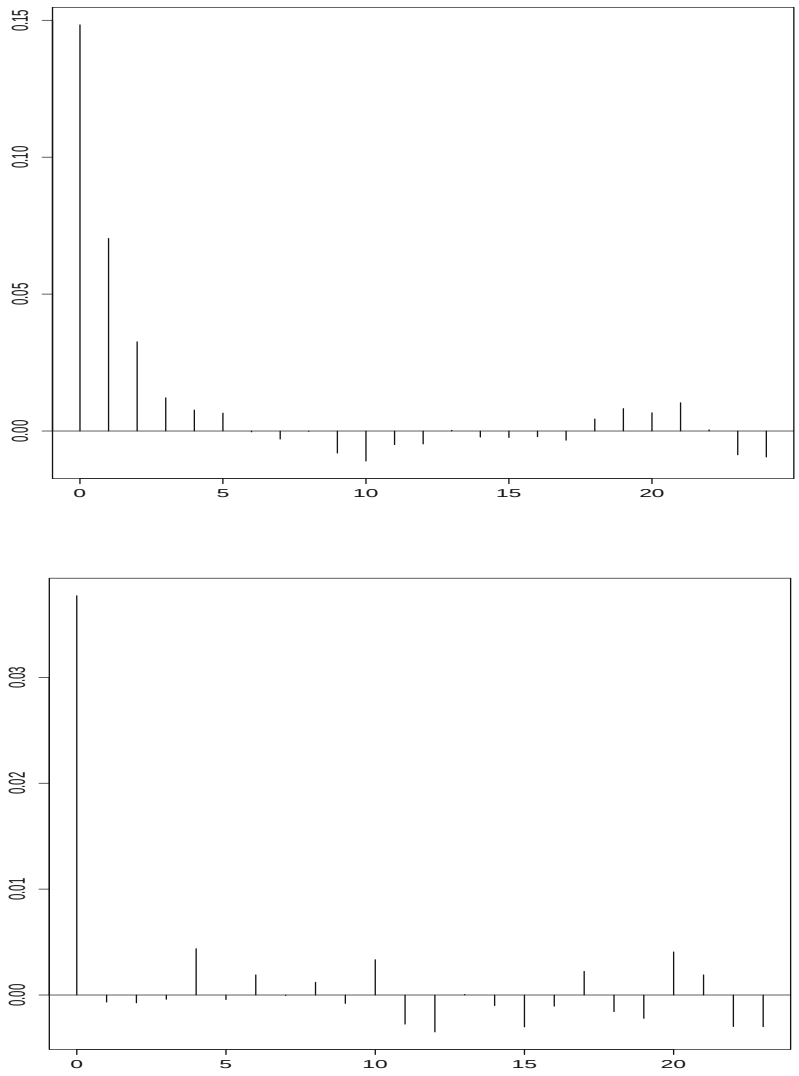


Fig. 1. Estimation of the $\gamma_X(k)$'s (top) and $\gamma_Y(\ell)$'s (bottom) for Example 1, with $\mu_1 = \mu_2 = 0$, $n = 300$ and $m = 200$.

$n; m$	150 ; 100	300 ; 200	450 ; 300	600 ; 400	750 ; 500
Estimated variance	1.151	1.11	1.121	1.03	0.987
Estimated level 1	0.06	0.05	0.058	0.049	0.042
Estimated level 2	0.062	0.067	0.06	0.057	0.046

Here, we see that the choice $a_n = 4$, $b_n = 0$ works well also, and seems even slightly better than the choice $a_n = 3, b_n = 0$.

• Case $\mu_1 = 0$, $\mu_2 = 0.2$ and $a_n = 4$, $b_n = 0$:

In this example, H_0 is not satisfied, and $\pi > 1/2$. However, μ_2 is close to μ_1, and since $\sigma_1 = 2$ and $\sigma_2 = 1$, Y_1 does not stochastically dominate X_1 (there is only stochastic domination in a weak sense). We shall see how the two tests are able to see this weak domination, by giving some estimation of the powers.

$n; m$	150 ; 100	300 ; 200	450 ; 300	600 ; 400	750 ; 500
Estimated variance	1.229	1.136	1.107	1.066	1.059
Estimated power 1	0.185	0.276	0.352	0.4	0.476
Estimated power 2	0.13	0.187	0.25	0.291	0.344

As one can see, the powers of the one sided and two-sided tests are always greater than 0.05, as expected. Still as expected, these powers increase with the size of the samples. For $n = 750, m = 500$, the power of the one-sided test is around 0.47 and that of the two-sided test is around 0.34.

To be complete, we also describe the case where $\sigma_1 = 1$ and $\sigma_2 = 2$.

• Case $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 = 0.2$, $\sigma_2 = 2$, and $a_n = 4$, $b_n = 0$:

In that case the quantity $\pi > 1/2$ is the same as in the preceding paragraph, but now Y stochastically dominates X (in the strong sense) since $H_Y(t) > H_X(t)$ for any real t.

$n; m$	150 ; 100	300 ; 200	450 ; 300	600 ; 400	750 ; 500
Estimated variance	1.034	1.123	1.086	1.008	1.103
Estimated power 1	0.211	0.307	0.407	0.448	0.509
Estimated power 2	0.139	0.217	0.294	0.331	0.401

As one can see by comparing the last two tables, for the same $\pi > 1/2$, the two tests seem more powerful when there is a strict stochastic domination than when there is only a weak stochastic domination.

6.2. Example 2: Intermittent Maps

For γ in $]0, 1[$, we consider the intermittent map θ_γ from $[0, 1]$ to $[0, 1]$, introduced by Liverani, Saussol and Vaienti [15]:

$$\theta_\gamma(x) = \begin{cases} x(1 + 2^\gamma x^\gamma) & \text{if } x \in [0, 1/2[\\ 2x - 1 & \text{if } x \in [1/2, 1]. \end{cases} \tag{6.1}$$

It follows from [15] that there exists a unique absolutely continuous θ_γ-invariant probability measure ν_γ, with density h_γ.

Let us briefly describe the Markov chain associated with θ_γ and its properties. Let first K_γ be the Perron–Frobenius operator of θ_γ with respect to ν_γ defined as follows: for any functions u, v in $\mathbb{L}^2([0, 1], \nu_\gamma)$

$$\nu_\gamma(u \cdot v \circ \theta_\gamma) = \nu_\gamma(K_\gamma(u) \cdot v). \tag{6.2}$$

Relation (6.2) states that K_γ is the adjoint operator of the isometry $U\colon u \mapsto u \circ \theta_\gamma$ acting on $\mathbb{L}^2([0, 1], \nu_\gamma)$. It is easy to see that the operator K_γ is a transition kernel, and that ν_γ is invariant by K_γ. Let now $(\xi_i)_{i\geq 1}$ be a stationary Markov chain with invariant measure ν_γ and transition kernel K_γ. It is well known that on the probability space $([0, 1], \nu_\gamma)$ the random vector $(\theta_\gamma, \theta_\gamma^2, \ldots, \theta_\gamma^n)$ is distributed as $(\xi_n, \xi_{n-1}, \ldots, \xi_1)$. Now it is proved in [5] that there exist two positive constants A, B such that

$$\frac{A}{(n+1)^{(1-\gamma)/\gamma}} \leq \alpha_{2,\boldsymbol{\xi}}(n) \leq \frac{B}{(n+1)^{(1-\gamma)/\gamma}}$$

(this is also true for the coefficient $\beta_{2,\boldsymbol{\xi}}(n)$, see [4]).

Moreover, the chain $(\xi_i)_{i\geq 1}$ is not α-mixing in the sense of Rosenblatt [17].

Let $X_i(x) = \theta^i_{\gamma_1}(x)$ and $Y_j(y) = \theta^j_{\gamma_2}(y)$. On the probability space $([0,1]\times[0,1], \nu_{\gamma_1}\otimes\nu_{\gamma_2})$ the two sequences $(X_i)_{i\geq 1}$ and $(Y_j)_{j\geq 1}$ are independent.

Note that, if $\gamma_1 = \gamma_2$, the hypothesis $H_0\colon \pi = 1/2$ is satisfied and the random variables T_n defined in (2.2) converge in distribution to $\mathcal{N}(0,1)$.

We shall study the behavior of T_n for various choices of γ_1, γ_2, a_n and b_n. As explained in Section 2, the quantities a_n and b_n may be chosen by analyzing the graph of the estimated auto-covariances $\hat{\gamma}_X(k)$ and $\hat{\gamma}_Y(\ell)$.

As in Subsection 6.1, we shall compute T_n for various choices of (n,m) (from $(150,100)$ to $(750,500)$ with $n/m = 1.5$). We estimate the three quantities $\mathrm{Var}(T_n)$, $\mathbb{P}(T_n > 1.645)$ and $\mathbb{P}(|T_n| > 1.96)$ via a classical Monte-Carlo procedure, by averaging over $N = 2000$ independent trials.

If $\gamma_1 = \gamma_2$, we say that $\mathbb{P}(T_n > 1.645)$ is level 1 (the level of the one-sided test) and $\mathbb{P}(|T_n| > 1.96)$ is level 2 (the level of the two-sided test). If a_n, b_n are well chosen, the estimated variance should be close to 1 and the estimated levels 1 and 2 should be close to 0.05.

If $\gamma_1 \neq \gamma_2$, we say that $\mathbb{P}(T_n > 1.645)$ is power 1 (the power of the one-sided test) and $\mathbb{P}(|T_n| > 1.96)$ is power 2 (the power of the two-sided test).

Note that we cannot simulate exactly $(X_i)_{i\geq 1}$ and $(Y_j)_{j\geq 1}$, since the distribution $\nu_{1/4}$ is unknown. We start by picking X_1 and Y_1 independently over $[0, 0.05]$ (in order to start near the neutral fixed point 0) and then generate $X_i = \theta^{i-1}_{\gamma_1}(X_1)$ and $Y_j = \theta^{j-1}_{\gamma_2}(Y_1)$ for $i > 1, j > 1$. We shall see that the tests are robust to this (slight) lack of stationarity.

• Case $\gamma_1 = \gamma_2 = 1/4$ and $a_n = 0$, $b_n = 0$ (no correction):

$n; m$	150 ; 100	300 ; 200	450 ; 300	600 ; 400	750 ; 500
Estimated variance	5.172	4.658	4.858	4.764	4.751
Estimated level 1	0.204	0.196	0.19	0.206	0.218
Estimated level 2	0.358	0.359	0.361	0.361	0.37

Here again, since $a_n = b_n = 0$, we compute T_n as if both series $(X_i)_{i\geq 1}$ and $(Y_j)_{j\geq 1}$ were uncorrelated. It leads to a disastrous result with estimated variance around 4.8, estimated level 1 around 0.2, and estimated level 2 around 0.36. This means that the one-sided or two-sided tests build with T_n will reject the null hypothesis much too often.

• Case $\gamma_1 = \gamma_2 = 1/4$ and $a_n = 4$, $b_n = 4$:

$n; m$	150 ; 100	300 ; 200	450 ; 300	600 ; 400	750 ; 500
Estimated variance	1.352	1.308	1.329	1.324	1.28
Estimated level 1	0.057	0.0645	0.062	0.069	0.067
Estimated level 2	0.089	0.082	0.087	0.087	0.075

As suggested by the graphs of the estimated auto-covariances (see Fig. 2), the choice $a_n = 4$, $b_n = 4$ should give a reasonable estimation of V. We see that the procedure is much better than if no correction is made: the estimated variance is now around 1.3, the estimated level 1 is around 0.065 and the estimated level 2 is around 0.085.

• Case $\gamma_1 = \gamma_2 = 1/4$ and $a_n = 5$, $b_n = 5$:

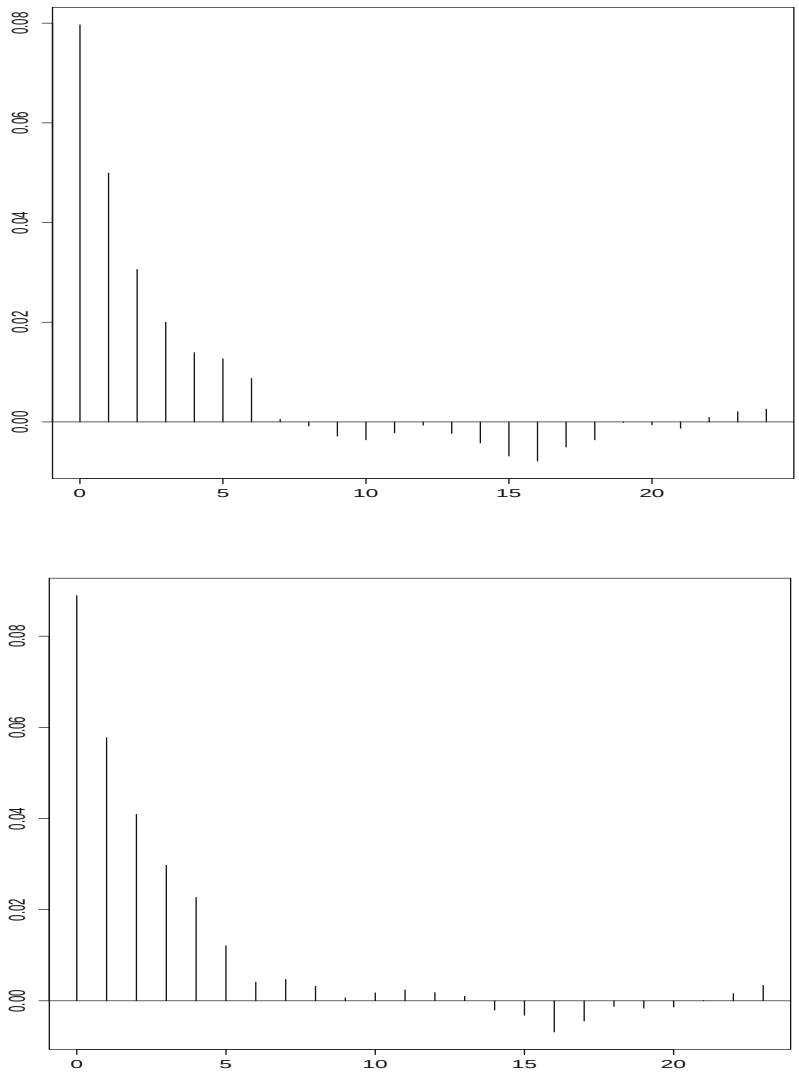


Fig. 2. Estimation of the $\gamma_X(k)$'s (top) and $\gamma_Y(\ell)$'s (bottom) for Example 2, with $\gamma_1 = \gamma_2 = 1/4$, $n = 300$ and $m = 200$.

$n; m$	150 ; 100	300 ; 200	450 ; 300	600 ; 400	750 ; 500
Estimated variance	1.349	1.249	1.164	1.181	1.185
Estimated level 1	0.057	0.062	0.055	0.055	0.063
Estimated level 2	0.084	0.077	0.068	0.071	0.07

For $a_n = 5$, $b_n = 5$, the results seem slightly better than for $a_n = 4$, $b_n = 4$. Even for large samples, the estimated level 2 is still larger than the expected level: around 0.07 instead of 0.05 (but 0.07 has to be compared to 0.36 which is the estimated level 2 without correction). Recall that the coefficients $\alpha_{2,\mathbf{X}}(n)$ and $\alpha_{2,\mathbf{Y}}(n)$ are exactly of order $n^{-1/3}$, which is quite slow. The procedure can certainly be improved for larger n, m by estimating more covariance terms (in accordance with Proposition 2.1).

• Case $\gamma_1 = 0.25, \gamma_2 = 0.1$ and $a_n = 5$, $b_n = 4$.

We first estimate the quantity $\pi = \nu_{\gamma_1} \otimes \nu_{\gamma_2}(\{(x,y)\colon x < y\})$ via a realization of

$$\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\mathbf{1}_{X_i<Y_j}.$$

For $n = m = 30000$, this gives an estimate around 0.529 for π meaning that Y weakly dominates X. Let us see how the tests based on T_n can detect this departure from $H_0\colon \pi = 1/2$.

Based on the estimated covariances, a reasonable choice is $a_n = 5, b_n = 4$. This choice leads to the following estimated powers (based on $N = 2000$ independent trials).

$n; m$	150 ; 100	300 ; 200	450 ; 300	600 ; 400	750 ; 500
Estimated variance	1.24	1.2	1.2	1.23	1.117
Estimated power 1	0.11	0.165	0.185	0.226	0.241
Estimated power 2	0.091	0.12	0.13	0.158	0.17

As one can see, the powers of the one-sided and two-sided tests are always greater than 0.05, as expected. Still as expected, these powers increase with the size of the samples. For $n = 750, m = 500$, the power of the one-sided test is around 0.24 and that of the two-sided test is around 0.17. This has to be compared with the estimated levels when T_n is centered at 0.529 instead of 0.5 (recall that 0.529 is the previous estimate of π on very large samples): in that case, for $n = 750, m = 500$ and $a_n = 5, b_n = 4$, one gets an estimated level 1 equal to 0.059, and an estimated level 2 equal to 0.068, which are both close to 0.05 as expected.

6.3. Intermittent maps: the case of two adjacent samples

Here, we shall illustrate the result of Subsection 5.2. As in Subsection 6.2, we first simulate X_1 according to the uniform distribution over $[0, 0.05]$, and then generate $X_i = \theta_\gamma^{i-1}(Z_1)$, but now for the indexes $i \in \{2, \dots, n+m\}$. For $j \in \{n+1, \dots, n+m\}$, we take $Y_j = f(X_j)$ for a given monotone function f.

In our first experiment, we simply take $f = \mathrm{Id}$, so that $Y_j = X_j$. Hence the assumption $H_0\colon \pi = 1/2$ is satisfied.

For the second experiment, we take $f(x) = x^{4/5}$, so that $f(X_1) > X_1$ almost surely. Hence $\pi > 1/2$.

Let T_n be defined in (5.5). We follow the same scheme as in Subsection 6.2, and we estimate the three quantities $\mathrm{Var}(T_n)$, $\mathbb{P}(T_n > 1.645)$ and $\mathbb{P}(|T_n| > 1.96)$ via a classical Monte-Carlo procedure, by averaging over $N = 2000$ independent trials.

• Case $\gamma = 1/4$, $f = \mathrm{Id}$ and $a_n = 0$, $b_n = 0$ (no correction):

$n; m$	150 ; 100	300 ; 200	450 ; 300	600 ; 400	750 ; 500
Estimated variance	4.851	4.908	4.551	4.926	4.85
Estimated level 1	0.294	0.276	0.252	0.256	0.251
Estimated level 2	0.368	0.373	0.354	0.366	0.362

Here, as in the corresponding two-sample case (see the table for $\gamma_1 = \gamma_2 = 1/4$), the uncorrected tests lead to a disastrous result, with an estimated level 1 around 0.25 and an estimated level 2 around 0.36.

• Case $\gamma = 1/4$, $f = \mathrm{Id}$ and $a_n = 5$, $b_n = 5$:

$n; m$	150 ; 100	300 ; 200	450 ; 300	600 ; 400	750 ; 500
Estimated variance	1.369	1.302	1.272	1.205	1.171
Estimated level 1	0.114	0.097	0.086	0.081	0.075
Estimated level 2	0.109	0.096	0.079	0.072	0.066

The corrected tests with the choice $a_n = b_n = 5$ give a much better result than the uncorrected ones. One can observe, however, that the estimated levels are still too high ($\geq 8\%$) for moderately large samples (up to $n = 450, m = 300$). The results seem less good than in the corresponding two-sample case (see the table for $\gamma_1 = \gamma_2 = 1/4$), which is of course not surprising.

• Case $\gamma = 1/4$, $f(x) = x^{4/5}$ and $a_n = 5$, $b_n = 5$:

$n; m$	150 ; 100	300 ; 200	450 ; 300	600 ; 400	750 ; 500
Estimated variance	1.596	1.448	1.334	1.311	1.451
Estimated power 1	0.321	0.385	0.451	0.49	0.548
Estimated power 2	0.242	0.293	0.34	0.383	0.442

As one can see, the powers of the one-sided and two-sided tests are always greater than 0.05, as expected. Still as expected, these powers increase with the size of the samples. For $n = 750, m = 500$, the power of the one-sided test is around 0.55 and that of the two-sided test is around 0.44.

ACKNOWLEDGMENTS

We thank Herold Dehling for helpful comments on a preliminary version of the paper.

REFERENCES

1. R. C. Bradley, "Basic Properties of Strong Mixing Conditions", in *Dependence in Probability and Statistics. A Survey of Recent Results. Oberwolfach, 1985*, Ed. by E. Eberlein and M. S. Taqqu (Birkhäuser, 1986), pp. 165–192.
2. S. Dédé, *Théorèmes limites fonctionnels et estimation de la densité spectrale pour des suites stationnaires*, PhD thesis (Univ. Pierre et Marie Curie, 2009). https://tel.archives-ouvertes.fr/tel-00440850
3. J. Dedecker, "An Empirical Central Limit Theorem for Intermittent Maps", Probab. Theory Rel. Fields **148**, 177–195 (2010).
4. J. Dedecker, H. Dehling, and M. S. Taqqu, "Weak Convergence of the Empirical Process of Intermittent Maps in $\mathbb{L}^2$ under Long-Range Dependence", Stochastics and Dynamics **15** (2015), 29 pages.
5. J. Dedecker, S. Gouëzel and F. Merlevède, "Some Almost Sure Results for Unbounded Functions of Intermittent Maps and Their Associated Markov Chains", Ann. Inst. Henri Poincaré Probab. Statist. **46**, 796–821 (2010).
6. J. Dedecker and F. Merlevède, "A Deviation Bound for α-Dependent Sequences with Applications to Intermittent Maps", Stochastics and Dynamics **17** (2017), 27 pages.
7. J. Dedecker and C. Prieur, "New Dependence Coefficients. Examples and Applications to Statistics", Probab. Theory Rel. Fields **132**, 203–236 (2005).
8. J. Dedecker and C. Prieur, "An Empirical Central Limit Theorem for Dependent Sequences", Stochastic Process. Appl. **117**, 121–142 (2007).
9. J. Dedecker and C. Prieur, "Some Unbounded Functions of Intermittent Maps for Which the Central Limit Theorem Holds", ALEA Lat. Am. J. Probab. Math. Statist. **5**, 29–45 (2009).
10. J. Dedecker and E. Rio, "On Mean Central Limit Theorems for Stationary Sequences", Ann. Inst. Henri Poincaré Probab. Statist. **44**, 693–726 (2008).
11. H. Dehling and R. Fried, "Asymptotic Distribution of Two-Sample Empirical U-Quantiles with Applications to Robust Tests for Shifts in Location", J. Multivar. Anal. **105**, 124–170 (2012).

12. H. Dehling, A. Rooch, and M. S. Taqqu, "Nonparametric Change-Point Tests for Long-Range Dependent Data", Scand. J. Statist. **40**, 153–173 (2013).
13. I. Dewan and B. L. S. Prakasa Rao, "Mann–Whitney Test for Associated Sequences", Ann. Inst. Statist. Math. **55**, 111–119 (2003).
14. S. Gouëzel, "Central Limit Theorems and Stable Laws for Intermittent Maps", Probab. Theory Rel. Fields **128**, 82–122 (2004).
15. C. Liverani, B. Saussol and S. Vaienti, "A Probabilistic Approach to Intermittency", Ergodic Theory Dynam. Systems **19**, 671–685 (1999).
16. E. Rio, *Théorie asymptotique des processus aléatoires faiblement dépendants*, in *Mathématiques et Applications* (Springer-Verlag, 2000), Vol. 31.
17. M. Rosenblatt, "A Central Limit Theorem and a Strong Mixing Condition", Proc. Nat. Acad. Sci. USA **42**, 43–47 (1956).
18. R. J. Serfling, "The Wilcoxon Two-Sample Statistic on Strongly Mixing Processes", Ann. Math. Statist. **39**, 1202–1209 (1968).

c) Apports et perspectives

L'application typique du test de Mann-Whitney en recherche biomédicale concerne l'administration en double aveugle de deux types de traitements différents à deux groupes de patients choisis aléatoirement. Chez chaque patient des deux groupes est ensuite relevée une donnée quantitative (le taux de cholestérol par exemple). La question sera alors s'il existe des différences significatives entre les groupes, et plus précisément si les réalisations de la variable quantitative ont tendance à être plus élevées dans un groupe que dans l'autre. On teste donc la domination stochastique des variables issues de deux échantillons indépendants. Les conditions de validité du test de Mann-Whitney sont modestes : il s'applique dès que les observations sont issues de deux échantillons de variables quantitatives continues et indépendantes et identiquement distribuées (i.i.d), sans supposer que la loi des variables appartient à une famille paramétrique identifiée.

Cependant, les problématiques rencontrées dans le contexte propre au sport de haut niveau peuvent rapidement mettre en défaut ces conditions de validité. Nous l'avons vu précédemment, le suivi individuel dans le temps du sportif est le meilleur moyen pour rendre compte de sa progression et permettre d'ajuster son environnement pour le rendre propice à sa performance. Dans ce cas, nous sommes confrontés à des données longitudinales individuelles, c'est le cas par exemple si nous souhaitons comparer les performances de deux athlètes suivis sur un an. L'hypothèse que les variables sont i.i.d dans chaque échantillon n'est alors plus respectée.

Dans l'article « *The Mann-Whitney U-statistic for alpha-dependent sequences* », nous proposons une méthodologie pour corriger le test de Mann-Whitney lorsque les deux séries d'observations proviennent de variables non i.i.d, mais stationnaires et « à mémoire courte » (ce qui implique que les corrélations entre les variables de chaque série décroissent suffisamment vite avec le temps).

Nous avons proposé la correction du test de Mann-Whitney permettant son application à une problématique récurrente dans le sport de haut niveau, à savoir la comparaison de variables corrélées. La justification de cette correction est asymptotique. Nous montrons dans un premier temps que si les variables sont corrélées, la loi limite de la statistique de Mann-Whitney converge vers une loi normale dont la variance s'écrit en fonction des coefficients de corrélation des deux suites de variables uniformément distribuées formées à partir des

variables de départ. La décroissance des corrélations ne suffit en général pas à assurer la convergence en loi normale, d'autres conditions sont nécessaires. Ainsi, nous avons utilisé des coefficients de mélange qui mesurent la dépendance temporelle en contrôlant les quantités $\left|Cov\left(f(X_0), g(X_k)\right)\right|$ uniformément sur des classes de fonctions. De telles mesures de la dépendance ont été introduites par Rosenblatt(115) qui a proposé un contrôle uniforme des covariances pour toutes les fonctions f et g bornées en valeur absolue par 1 : c'est la notion d'*alpha* mélange fort au sens de Rosenblatt. Par la suite, de nombreux ont montré que ces coefficients étaient quelques peu restrictifs et excluaient même certains processus Markoviens classiques. En 2005, Dedecker et Prieur (111) ont proposé des coefficients dits « de mélange faible » en considérant un contrôle uniforme des variances pour toutes les fonctions bornées en valeur absolue par 1, et toutes les fonctions g qui sont des fonctions indicatrices d'intervalles. Ces auteurs ont montré que les coefficients faibles ainsi définis conservaient la plupart des bonnes propriétés des coefficients de Rosenblatt et permettaient de traiter un nombre considérables de processus supplémentaires. Il est même possible de calculer ces coefficients pour des processus déterministes ayant un comportement quasi-aléatoire (*cf.* les transformations intermittentes définies en section 6.2 de l'article). Notre résultat principal (Théorème 1.1 de l'article) est un théorème imite central pour la U-statistique de Mann-Whitney obtenu sous des conditions de mélange faible, don nous montrons l'optimalité dans la Section 3 de l'article.

Dans un second temps, nous montrons comment estimer cette variance de façon consistante, de sorte que la statistique de test corrigée (c'est à dire re-normalisée par l'estimateur de l'écart-type) converge en loi vers la loi normale centrée et réduite. On peut donc ensuite en déduire une zone de rejet basée sur les quantiles de la loi normale, de sorte que le niveau asymptotique du test soit bien celui attendu. Les performances de ce test sont ensuite validées sur différents jeux de simulation, qui montrent en particulier que la procédure fonctionne encore lorsque que l'on s'approche du cas de la « longue mémoire », pourvu que le nombre d'observations dans chaque groupe soit suffisant. Les simulations montrent aussi que, dans les cas standard ou la décroissance des corrélations est très rapide, le test corrigé est valide pour de petites tailles d'échantillons.

D'un point de vue méthodologique et pratique, nous proposons deux graphiques de « résidus » représentant les estimateurs des coefficients de corrélation des variables uniformes formées à

partir des variables de départ. Ces graphiques permettent d'identifier si une correction du test est nécessaire ou non, et si oui, de choisir le nombre d'estimateurs de ces corrélations nécessaire pour estimer correctement le terme de variance asymptotique. Là encore, cette méthodologie est validée par les simulations.

Cette correction du test de Mann-Whitney peut trouver des applications spécifiques dans le domaine du sport. Comparer des paramètres chez deux athlètes sur une période définie, que ce soit des performances dans le cadre de la comparaison de protocoles d'entrainement, des capacités physiques dans le cadre d'observations de réponses à un protocole de ré-athlétisation suite à une blessure ou bien dans un contexte de détection.

4) Conclusions générales

L'individualisation du suivi du sportif est indispensable à sa préparation à performer au meilleur niveau. La mise en place de ce suivi concerne différents membres du staff entourant l'athlète qui se coordonnent pour prendre en compte l'ensemble des paramètres en considération. Ainsi sont analysés conjointement les performances de l'athlète et son état de santé. Etat de santé qui doit être optimisé pour permettre à l'athlète d'encaisser les charges d'entrainement nécessaires à sa préparation physique et technique (123). Cette optimisation de la santé et de la performance résulte d'une dynamique, d'échanges entre les médecins, entraineurs et préparateurs physiques. Les informations fournies et reçues permettent d'ajuster au quotidien les charges d'entrainement et les soins de l'athlète pour optimiser sa récupération et sa qualité d'entrainement. Or ces informations échangées entre les membres du staff sont basées sur l'interprétation de données complexes comme celles présentées dans notre première étude. C'est dans l'interprétation de ces données que les outils biostatistiques peuvent aider le sport de haut niveau. Comme illustré à travers ces deux premières études, le développement d'outils et leurs applications adaptées pour répondre à un problème donné fourni des informations claires et interprétables au staff, enrichissant le contexte et facilitant la prise de décisions.

Ces outils appliqués dans notre cas à l'individualisation du suivi de sportifs participant à des sports collectifs sont transposables aux sports individuels. L'évolution des paramètres sanguins chez les cyclistes est une thématique très développée au sein de la fédération française de cyclisme et les méthodes présentées dans notre première étude peuvent évidemment s'appliquer. La prise en compte des problématiques spécifiques au cyclisme

comme les conditions d'entrainement à haute altitude permettrait de nourrir nos réflexions et d'implémenter de nouvelles extensions méthodologiques toujours dans l'idée d'optimiser le suivi de l'état de santé des sportifs.

L'individualisation de la performance est aussi indispensable à sa compréhension. Alors que l'évolution des performances au cours d'une vie est bien connue (7), un des intérêts majeur des fédérations et clubs sportifs est la détection chez les jeunes. Cette détection réalisée à travers des outils spécifiques, des tests physiques et techniques mais aussi et surtout grâce à l'expertise des entraineurs est assistée par des méthodologies statistiques rendant compte de l'évolution des capacités des sportifs. La correction du test de Mann-Whitney proposée dans l'étude 2 appliquée aux données issues de ces tests de détection réalisés permettrait de comparer l'évolution de deux jeunes sportifs sur une période donnée. Les informations fournies pourraient ainsi alimenter la réflexion des formateurs, leur permettre de cibler des problèmes, d'adapter les entrainements chez des jeunes en moins bonne réussite.

Comme abordé plus haut, la quantification des charges d'entrainements et de compétitions des sportifs est une problématique émanant directement du terrain et faisant l'objet de nombreuses études (123). C'est la dynamique de cette charge, son interprétation dans le temps qui intéresse principalement les chercheurs au regard de la description des associations de charge avec l'émergence de bonne ou mauvaise performance ou l'apparition de blessure (124). Ces charges, mesurées pour chaque athlète permettent d'affiner le suivi individuel afin de fournir des retours aux athlètes et aux entraineurs. Dans l'objectif de prévenir la survenue de blessures, des modèles de quantification de charge sont mis en place dans le sport de haut niveau et sont complétés par la prise en compte d'autres facteurs comme le niveau de fatigue déclaré. Ces outils de quantification permettent d'observer au mieux les cinétiques des charges et sont en constante évolution méthodologique pour répondre au mieux aux problématiques rencontrées sur le terrain. C'est le cas de cette étude qui se propose d'apporter à un questionnement ciblé: la difficulté de maintenir un suivi régulier et chronique des athlètes de haut niveau en raison de leurs emploi du temps (compétitions, stages internationaux, sélections nationales, blessures,…), un outil de test statistique pratique fondé sur un travail collaboratif intégrant les sciences du sport, la réalité du terrain et les biostatistiques. Les données produites par ces méthodologies de suivi de la charge sont longitudinales et doivent être observées individuellement, ce qui offre la perspective d'adapter les méthodes développées dans cette partie à ce type de suivi. C'est par un regard croisé de

multiples indicateurs et sous le prisme des réponses individuelles analysées par les méthodes statistiques appropriées que le monitoring prend sens (125).

5) Transition

Nous avons abordé dans cette première partie l'importance accordée par les staffs aux données issues de l'individualisation du suivi des sportifs dans les sports collectifs. A travers deux études et l'évocation d'intéressantes perspectives fut mis en avant comment les biostatistiques peuvent et doivent accompagner les staffs dans la compréhension et l'interprétation de ces données. L'individualisation est déterminante dans les sports collectifs où les capacités physiques et techniques requises diffèrent grandement d'un poste à l'autre. Nous allons maintenant appréhender la dimension collective de la performance. Comment la gestion d'une équipe, d'un groupe peut optimiser sa performance ? Comment l'utilisation parcimonieuse ou le manque de turnover au sein d'un groupe peut impacter la survenue de blessures ? Ces deux questions sont traitées dans la partie suivante.

IV. Partie 2 : Application d'analyses statistiques dans le sport : *le collectif comme moteur de la performance*

1) Introduction générale

Nous avons vu en première parti l'importance de l'individualisation du suivi des sportifs de haut niveau dans l'optimisation de leur performance et la minimisation de l'exposition aux risques de blessures. Ce monitoring personnalisé est indispensable à la compréhension de la performance et des mécanismes de survenue des blessures. Cependant, dans les sports collectifs, la gestion du collectif est aussi un élément déterminant. Cette gestion va impacter les performances globales de l'équipe mais aussi les performances individuelles ainsi que la survenue de blessures au sein du groupe (126). Corollairement, les blessures subies au sein d'un effectif impactent négativement sa performance (126–128). Les blessures sont aussi la première cause d'indisponibilité des joueurs à l'entrainement et en mach (129). Cette indisponibilité forcée limite les choix de l'entraineur pour constituer son équipe et entrave par conséquent la création d'un collectif fort au cours du temps. Bengtsson et al. souligne en 2013 qu'il existe trop peu d'études analysant les relations entre les rotations des joueurs au sein de l'effectif au cours du temps et la performance de l'équipe (130). Nous avons réalisé deux études sur le rugby à 15 dans le but d'identifier ces relations entre expérience collective et performance.

2) Etude 3 : Efficacité collective au sein du XV de France: une histoire de sélections et de temps.

a) Introduction

La performance au rugby dépend en partie des capacités individuelles des joueurs sélectionnés (131–133). Leurs aptitudes physiques de force et de vitesse ainsi que leur maitrise technique et tactique du jeu sont déterminants dans la réussite de l'équipe (134,135). Le résultat du match est aussi largement influencé par les synergies physiques et stratégiques de l'équipe (135,136,136). Des individus qui se connaissent, qui partagent un passé commun sont plus aptes à évoluer efficacement qu'un groupe qui se découvre. Les principales études décrivant la performance d'un collectif analysent la somme des performances individuelles, ont une approche très qualitative de l'évaluation de l'expérience collective (ressenti des joueurs entre eux) ou utilisent la somme d'expérience individuelle (16) comme définition de

l’expérience du groupe. Notre volonté est de mettre en place de nouveaux outils quantitatifs d’évaluation de l’expérience collective, plus particulièrement au sein du XV de France.

b) Collective effectiveness in the XV de France: selections and time matter. (7,16,37,38,40–43,89,93,132,137–160)

Volume 17 Issue 6 July 2017 ISSN 1746-1391

European Journal of Sport Science

European Journal of Sport Science
The Official Journal of the European College of Sport Science

Taylor & Francis
Taylor & Francis Group

Routledge

ISSN: 1746-1391 (Print) 1536-7290 (Online) Journal homepage: http://www.tandfonline.com/loi/tejs20

Collective effectiveness in the XV de France: selections and time matter

Adrien Sedeaud, Guillaume Saulière, Laurie-Anne Marquet, Scott Del Vecchio, Avner Bar-Hen & Jean-François Toussaint

To cite this article: Adrien Sedeaud, Guillaume Saulière, Laurie-Anne Marquet, Scott Del Vecchio, Avner Bar-Hen & Jean-François Toussaint (2017) Collective effectiveness in the XV de France: selections and time matter, European Journal of Sport Science, 17:6, 656-664, DOI: 10.1080/17461391.2017.1291742

To link to this article: https://doi.org/10.1080/17461391.2017.1291742

European Journal of Sport Science, 2017
Vol. 17, No. 6, 656–664, https://doi.org/10.1080/17461391.2017.1291742

ORIGINAL ARTICLE

Collective effectiveness in the *XV de France:* selections and time matter

ADRIEN SEDEAUD[1,2], GUILLAUME SAULIÈRE[1,2], LAURIE-ANNE MARQUET [3,4], SCOTT DEL VECCHIO[1,2], AVNER BAR-HEN[5], & JEAN-FRANÇOIS TOUSSAINT[1,2,6]

[1]*IRMES, Institut de Recherche bioMédicale et d'Epidémiologie du Sport, INSEP, Paris, France;* [2]*Université Paris-Sud Orsay, Île-de-France, France;* [3]*Laboratoire Sport, Expertise et Performance, INSEP, Paris, France;* [4]*Laboratoire Motricité Humaine, Education, Sport, Santé, Université Nice Sophia-Antipolis, Nice, France;* [5]*MAP5, UFR de Mathématiques et Informatique, Université Paris Descartes, Paris, France &* [6]*CIMS, Paris, France*

Abstract
The aim of this study was to quantify the impact of selections and shared selections in the rugby union. Players' names, positions, and number of selections were collected for all XV de France's games (1906–2014). Every team's percentage of renewal of workforce was calculated for backs and forwards. During the 1987–2014 period, all second row forwards (locks), halfbacks, and centres' shared selections (number of times when two players have competed together) were recreated. The Best vs. Rest method was applied to these remodelled dyads. They were analysed and compared with surrounding teammates as well as opponents. Head coaches similarly change their workforce for upcoming matches after winning or losing (around 30%), but losing teams renew significantly more positions in their line-ups. The recreated halfbacks, locks, and centres reveal a common pattern. Whether victorious or not, the 'renewed couples' victory percentage will congregate towards the XV de France's victory percentage. For all the best recreated couples, the cumulated number of selections for forwards' is always higher than the ones part of less efficient teams: 231.3 ± 80 vs. 212.9 ± 91 selections for locks' teammates (Effect sizes (ES) small, possibly positive, 54.8%). In best recreated couples, number 8's are significantly more experienced than their counterparts in less efficient pairs (ES small, likely positive, 76.3%). The XV de France's collective effectiveness relies on a balance between stability and workforce renewal, which allows the building of specific position interactions and builds on experimented forwards packs. Selections and shared selections are serious collective performance parameters associated with performance.

Keywords: *Rugby union, collective performance, shared selections, sharing experience, teamwork*

Introduction

Rugby performance depends on energy resources, physical qualities such as strength, speed, and power (Cunniffe, Proctor, Baker, & Davies, 2009; Lacome, Piscione, Hager, & Bourdin, 2014; Mellalieu, Trewartha, & Stokes, 2008) as well as technical and tactical skills that work together to bring out team effectiveness (Hendricks, Matthews, Roode, & Lambert, 2014; Hendricks, Roode, Matthews, & Lambert, 2013; Hendricks, Till, Brown, & Jones, 2016; Sewry, Lambert, Roode, Matthews, & Hendricks, 2015). The plurality of these parameters continues to benefit the collective effectiveness with success (i.e. the ability to play well together), which is usually assessed with experience. In rugby, experience is a performance indicator that has been studied and used through different approaches, like skill measures at different levels. For example, in rugby, match performance depends on tackle ability (Gabbett, 2005, 2008; Sewry et al., 2015; Wheeler & Sayers, 2009), and a player's tackle ability relates to playing experience (Gabbett & Ryan, 2009). A greater playing experience is associated with a better tackling technique, (Gabbett & Ryan, 2009). This outcome is in accordance with the considerable practice required for improvements in tackling performance (Gabbett & Ryan, 2009). This important practice time is needed not only for an individual player, but for the team's cohesiveness as well. It is likely that the collective effectiveness responds in a

Correspondence: Adrien Sedeaud, INSEP (Institut National du Sport, de l'Expertise et de la Performance) – IRMES (Institut de Recherche bio-Médicale et d'Épidémiologie du Sport), 11, Avenue du Tremblay, 75012 Paris, France. E-mail: adrien.sedeaud@insep.fr

similar way. Indeed, collective experience gathered from previous Rugby World Cups (RWC) has been shown to be a performance factor for forwards (Sedeaud et al., 2012). It was shown that teams that won more RWC games comprised of forwards with more shared experience of the RWC (Sedeaud et al., 2012). These findings are consistent with previous studies which have found that international players may be selected because of greater skill and experience (Gabbett & Ryan, 2009; Hendricks & Lambert, 2014; Walsh, Young, Hill, Kittredge, & Horn, 2007). Another parameter affected by experience is ball-carrying technique while sprinting. Indeed, the more experienced players sprint faster with the ball under one arm and in both hands than their less experienced counterparts, whereas experienced and novice players do not differ between their sprint time while running without the ball (Walsh et al., 2007). However, most studies assess individual experience by the number of games played or other cumulated athletes characteristics, but rarely in a collective way (Shearer, 2015).

Most of the previous mentioned studies are done at the individual level. In other team sports, team cohesion, confidence, and collective efficacy are also linked to performance (Heuzé, Raimbault, & Fontayne, 2006; Stajkovic, Lee, & Nyberg, 2009). Players involved in a larger number of games do not only have better individual results, but are also considered more efficient and more involved in building group cohesiveness and team performance (Fransen et al., 2014; Heuzé et al., 2006; Leo, Sanchez-Miguel, Sanchez-Oliva, Amado, & Garcia-Calvo, 2013). These studies focused on interviews with players concerning their perception of their own actions, their teammates' actions, and their own impact on the group's performance. However, it was rarely done in a systematic view of shared comprehension or with a multiyear scope. Duch, Waitzman, and Amaral (2010) quantified the impact of team members on the performance using methods inspired by social network analysis. Their approach was based on the concept that the composition of teams determines their odds of success (Guimerà, Uzzi, Spiro, & Amaral, 2005; Wuchty, Jones, & Uzzi, 2007), especially the players' interplay contributions. In rugby teams, cohesion dimensions make up a significant proportion of the variance in the collective efficacy scores (Kozub & McDonnell, 2000). Interplay between teammates is crucial in collective sports, especially when executing strategy and tactic (Hendricks et al., 2013; Sewry et al., 2015) such as during the set piece (i.e. scrum, line-out). In this context, turnovers have progressively become more decisive due to defences being disorganized. In this case, the ability to read and act on opportunities together is critical to create overlap.

Throughout all phases of play, rugby is the result of collective interactions, either in defence (placement, interaction, tackles, and support during rucks, forage, mauls …) or in offence (placement, runs, lures, fixing, support, timing …). Game experience has been linked to performance and success in rugby union, but it has most often been recorded individually or by summing individual performances (Shearer, 2015). Collective effectiveness as a synergic construction (through shared playing time) has not been analysed over long periods of time. The purpose of this study was to quantify the impact of selections and shared selections on the game's result.

Methods

Data collection

After obtaining the approval of the Institutional Ethics Committee, individual characteristics (name, position, starters or substitutes, date of first selection, selection during a match) for each international game were collected from the first *XV de France* official match (1 January 1906) to the final match of the 2014 Six Nations tournament. All data of games played by the *XV de France* were collected, including the date, opponent, score, match venue, type of match: test match, V then VI Nations, World Cup (RWC) from the official French national team's website: www.ffr.fr. These data were derived from the longitudinal follow-up of players, with measurements done by *XV de France*'s team physician. Since 1906, the analysis includes 702 games detailed in 11,832 international selections for 1054 different rugby players.

Selections trend

Over the entire period (108 years), mean numbers of selections of the team were computed. Then, the mean numbers of selections of the team were analysed by match result. Specific analysis was conducted during the 1987–2014 period, reflecting the arrival of the RWC and professional rugby (since 1995).

Workforce renewal

During this period, the percentage of renewal of workforce by team (declined in forwards and backs) was calculated in function of game results (win or loss). This percentage was computed as the proportion of players involved in two consecutive matches.

Shared selections

In order to identify success factors for the best second row forwards (locks), halfbacks, and centres during the 1987–2014 period, all shared selections (as starters) for each dyads were computed game by game. Shared selection represents the number of times when two players have already competed together. For locks, halfbacks, and centres, percentages of never renewed players after a loss or a win, were significantly higher than their counterparts. For these players, we redrew all the historical shared selections as a 'couple' based upon the game results. A threshold of four games shared by couples was set for historical recreated shared selections. In order to compare locks, centres, and halfbacks, two groups were developed: best and rest. The 'best vs. rest' method divided the previous groups in two by the median of the percentage of victory (percentage derived at the end of the dyads''career'). The 'best' group represents dyads with the highest percentage of victory (above the median), whereas the 'rest' depicts couples with a lower percentage (below the median).

Overall selections

Relative to these best and rest recreated halfbacks, locks, and centres, an analysis of their opponents and the numbers of selections by the players who make up the teams are proportionally drawn. All 15 players forming teams surrounding halfbacks, centres, and locks are recreated by circles (where the diameter is proportional to the number of mean selections by position), which are drawn on a rugby field.

Statistical analyses

Data were reported as mean ± standard deviation. The normality of the distribution of each variable was tested using a Shapiro test. Mean numbers of selections of the team for victories versus defeats and 'best' against 'rest' groups comparisons were compared by Student's *t*-tests. Percentages of renewal of workforce were compared by Kruskall–Wallis test. The Effect sizes (ES) were assessed with magnitude-based inferential approach, using standardization to define magnitude thresholds. ES were classified as trivial (< 0.2), small (> 0.2–0.6), moderate (> 0.6–1.2), large (> 1.2–2.0), and very large (> 2.0–4.0) (Batterham & Hopkins, 2006; Winter, Abt, & Nevill, 2014). Number of selection and workforce renewal differences were assessed qualitatively as follows: < 1%, almost certainly not; 1–5%, very unlikely; 5–25%, probably not; 25–75%, possibly; 75–97.5%, likely; 97.5–99%, very likely; > 99%, almost certain (Hopkins, Marshall, Batterham, & Hanin, 2009). Frequency of facing opponents was compared by χ^2 test. The level of significance was set at $p < .05$.

Results

Selections trend

During the twentieth century (1906–2014), the number of games per year and the mean numbers of selections of the team consequently increased: from 1.3 selections for players in 1906 to 22.5 in 2014. Throughout this period, there was a *trivial* difference between the number of selections per player in winning teams and losing teams 19.6 ± 18.7 vs. 16.3 ± 18.0 selections, respectively (trivial, ES = 0.18, possibly trivial, 70,6%). Since 1987, players who have participated in the RWC have a significantly higher number of selections than players who have participated in other international competitions (tournaments: six Nations, summer and autumn tournaments): 33 ± 22.3 vs. 25.4 ± 21.5 selections, respectively (small, ES = 0.35, very likely positive, 95.3%).

Workforce renewal

The percentages of workforce renewal for forwards and backs, after victorious games are 33 ± 26.4 and 32.3 ± 25%, respectively, and are not significantly different after a lost game: 35.2 ± 27.1 and 35.1 ± 26.5%, respectively (the effect for forwards is trivial, ES = 0.08, likely trivial, 82.5% and the effect for backs is trivial, ES = 0.11, likely trivial, 76.6%). By contrast, teams who lose have a significantly higher workforce renewal rate. In winning teams only, the forwards' group changed by 29.8 ± 26.9%, whereas in losing teams, their counterparts underwent a change of 41.5 ± 24.4% (Figure 1). In victorious teams only, the backs' group changed by 29.5 ± 25.3%, whereas in losing teams, their counterparts underwent a change of 40.8 ± 24.7% (Figure 1) (for both forwards and backs, the effect is small, ES = 0.45, very likely positive, 98.5%). Workforce renewal does not determine the outcome of a winning or losing team in previous games, but a higher workforce renewal rate is associated with the loss of the game.

Shared selections

Figure 2 illustrates all shared selections by halfbacks, centres, and locks, and their respective victory percentage during the 1987–2014 period (left side). A common pattern between halfbacks, centres, and

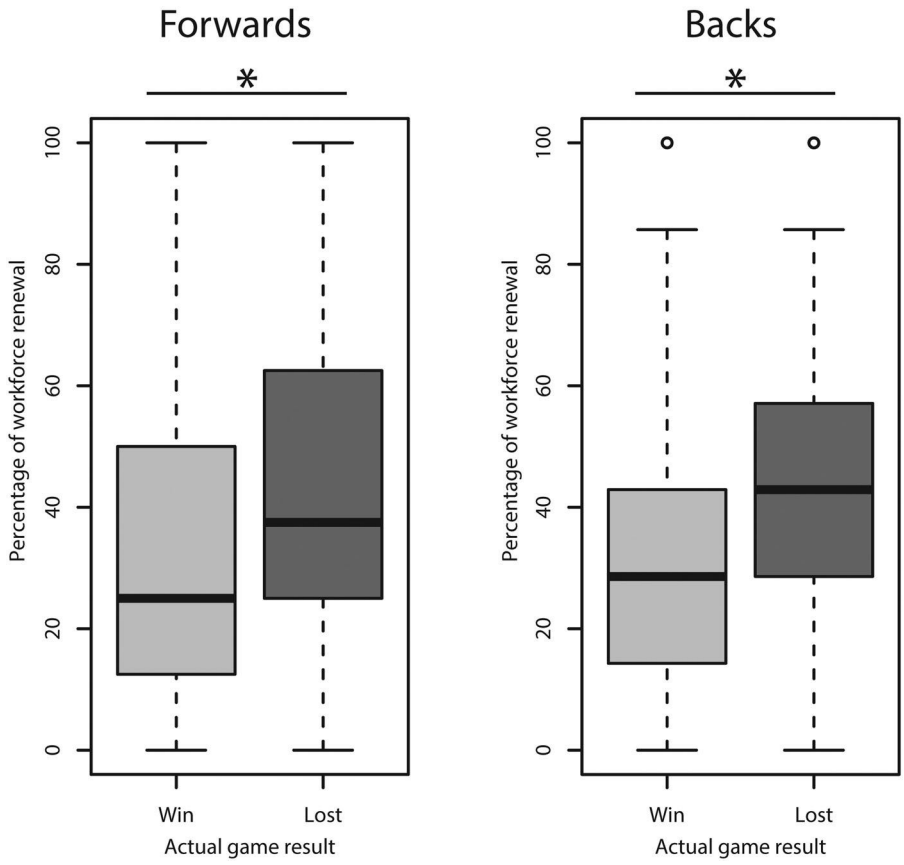


Figure 1. Percentage of turnovers for forwards and backs based on the actual game result. *$p < .05$ between game results.

locks is revealed: whether the match was won or not, renewed couples contribute to the *XV de France*'s overall victory percentage (Figure 2). However 'best vs. rest' classification indicates that the majority of the 'best' begins with a win, except for the Para-Trinh-Duc (first match the 23 February 2008 against England, 6 Nations) and Accoceberry-Deylaud (first match the 26 June 1994 against New Zealand, Test match) halfbacks and Mermoz-Fritz centres (first match the 27 June 2009 against Australia, Test match), which begin with one or two defeats. Over time, a convergence phenomenon allowed all recreated pairs to move towards the team's average winning percentage: 62.4% (for 1987–2014 period). In Figure 2, the right side indicates the best and rest reconstructed halfbacks, locks, and centres' opponents. There are no significant differences between opponent distributions from the Tri nations, six nations or other teams, for the three recreated pairs in the best and rest groups.

Overall selections

For every best recreated team, the number of selections for forwards is always higher than the rest of the team's: 223.6 ± 91.3 vs. 213.3 ± 100 selections (the effect is trivial, ES = 0.11, possibly trivial, 64.8%) for forwards surrounding halfbacks, 240.6 ± 96.3 vs. 229.9 ± 81 selections (the effect is trivial, ES = 0.12, possibly trivial, 70.8%) for locks' teammates and 231.3 ± 80 vs. 212.9 ± 91 selections (the effect is small, ES = 0.22, possibly positive, 54.8%) for those playing with centres. Among the best teams, left loosehead props and eighthmen surround halfbacks with significantly more selections than those in less efficient teams, 23.1 ± 18.8 vs. 17.7 ± 14.1 selections (the effect is small, ES = 0.3, likely positive, 75.2%) and 30 ± 24.1 vs. 22.6 ± 21.1 selections (the effect is small, ES = 0.3, likely positive, 76.3%), respectively (Figure 3). Recreated locks among the best teams consist of a number 4 with more selections than locks in rest teams (50.2 ± 37.1 vs. 29.7 ± 20 selections, the effect is moderate, ES = 0.7, most likely positive, 99.8%). Best recreated locks are surrounded by blindside flankers, and eighthmen who add up a significantly higher number of selections than their counterparts in rest teams, 29.5 ± 16.9 vs. 24.6 ± 17.3 selections (the effect is small, ES = 0.3, possibly positive, 74.2%) and 32.3 ± 26 vs. 26.1 ± 21 selections (the effect is

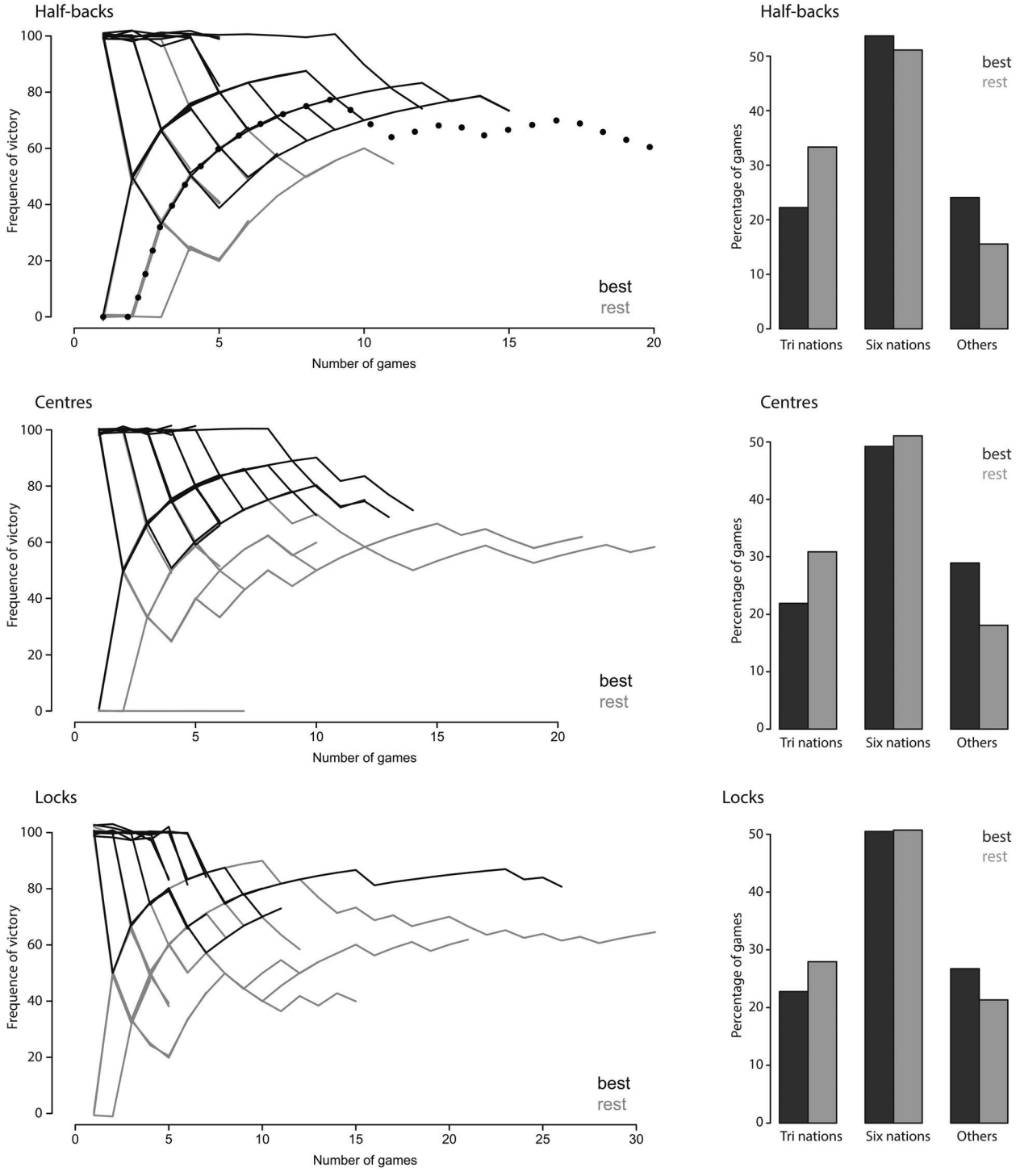


Figure 2. To the left: the frequency of victory changes based on shared selection of halfbacks, centres and locks. For example, the halfbacks' figure expresses in a dotted line the recreated para-Trinh-duc association, which began with two defeats (the first two points at 0% of frequency of victory) and followed with seven wins. Each line corresponds to a recreated association for one dyad of a halfback, centre and lock during the 1987–2014 period. To the right: the percentage of games played against tri nations teams (New Zealand, South Africa and Australia), six nations teams (England, Ireland, wales, scotland and Italy) and other squads of the best and rest recreated association (for best and rest halfbacks, centres and locks from the top of the figure downwards).

small, ES = 0.3, possibly positive, 69.8%), respectively (Figure 3). Centres in best teams are surrounded by higher selected number 4's (40.8 ± 28.3 vs. 24.1 ± 11.7 selections, the effect is moderate, ES = 0.6, most likely positive, 99.6%). In rest teams, these players are surrounded by fly-halves,

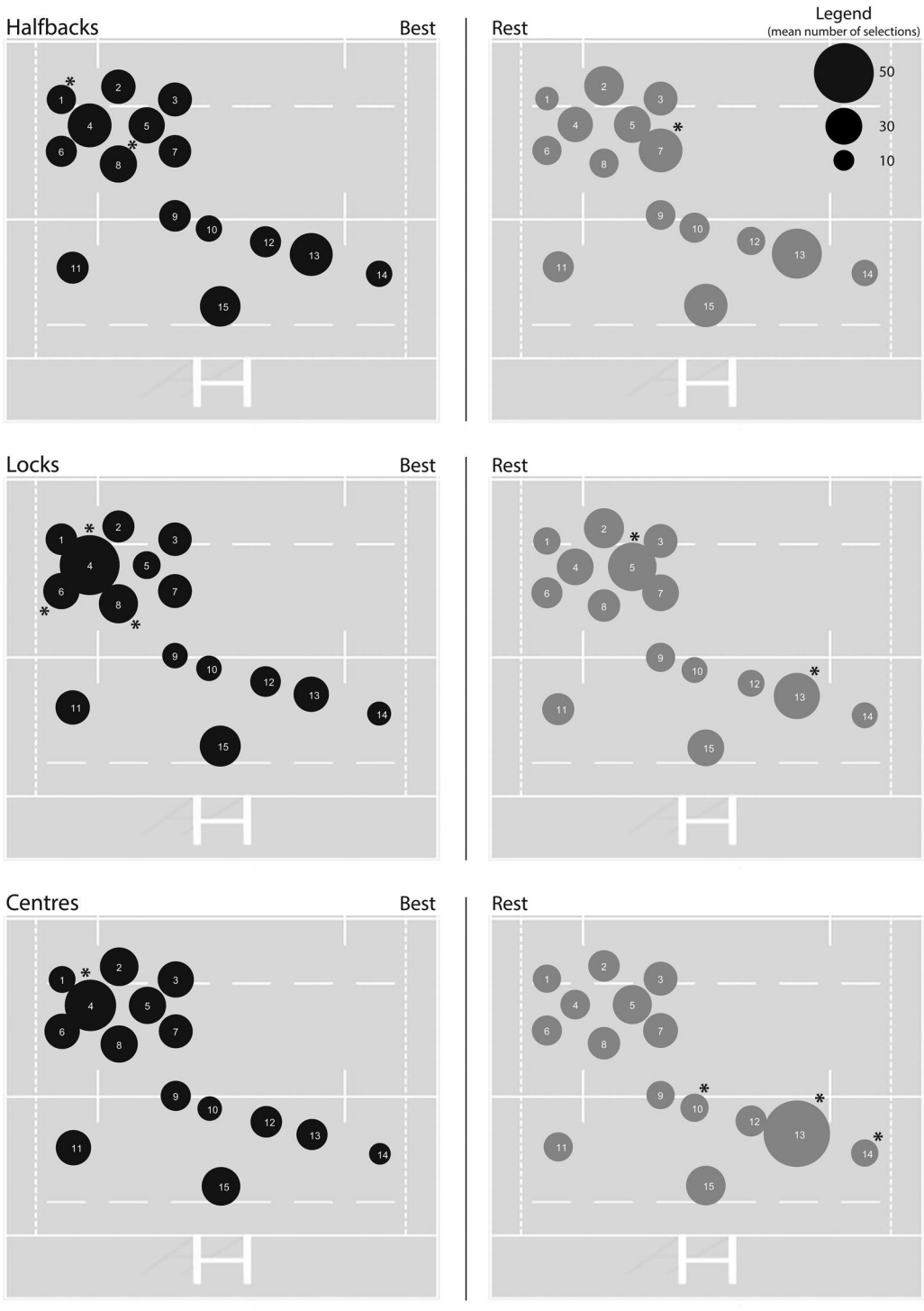


Figure 3. The comparison of players' selections by position surrounded recreated halfbacks, locks and centres pairs in best and rest teams. $^{*}p < .05$ between best and rest team players of the same position. Circle diameters depict the number of mean selections by position. Their positions and respective numbers are depicted as a standard rugby union team organization.

outside centres, and right wings with significantly more selections than their counterparts in best teams, 22.8 ± 16.4 vs. 18.4 ± 12.8 selections (the effect is small, ES = 0.3, likely positive, 76.8%), 56.7 ± 35.1 vs. 24.1 ± 19.7 selections (the effect is large, ES = 1.2, likely positive, 90.8%) and 22.2 ± 18.4 vs. 16.1 ± 11.9 selections (the effect is small, ES = 0.4, likely positive, 90.7%), respectively.

Discussion

To our knowledge, the present study is the first to quantify the effect of an individual's selections number and specifically shared selections on rugby union game results.

Selections trend and XV de France history

During the 1906–2014 period, the number of total players' selections in victorious teams is significantly greater than those in defeated teams. Head coaches select more experienced players to participate in RWC. Indeed, picking the 'toughest' players on the international scene is coherent with a greater playing experience that correlates with better tackling technique (Gabbett & Ryan, 2009) and skill (Gabbett & Ryan, 2009; Hendricks & Lambert, 2014; Walsh et al., 2007).

Workforce renewal and consequences

Whether teams win or lose, head coaches change around 30% of their workforce. The fact that they stabilize 70% of players is in accordance with studies, which found that a previous performance was a positive predictor of collective efficacy at the group level (Myers, Payment, & Feltz, 2004; Watson, Chemers, & Preiser, 2001). Nonetheless, teams who lose games have a significantly higher workforce renewal rate: around 40% for all teams. These results probably underline the optimum rate of workforce renewal, which seems to depend on team size (Whitfield, 2008). Indeed, small groups draw benefit from some workforce renewal (Whitfield, 2008), but can only do so in stable conditions and line-up remaining unchanged (Palla, Barabási, & Vicsek, 2007). According to Whitfield, it is recommended to have a 'healthy mix of rookies and veterans', which this study showed 70% incumbents (2008). These results are also consistent with previous results showing that collective experience of the forwards' pack increased with the final ranking (Sedeaud et al., 2012). These results express the relationship and trade-off between stability and renewal.

Shared selections and recreated effectiveness

Teams show more than the total of their individuals; they also exhibit the impact of players' experience hidden in the team performance; consequently, sharing experience reveals a new kind of parameter. Such an indicator may evaluate the interactions among members, beyond the players' individual characteristics (Bourbousson, Poizat, Saury, & Seve, 2010). Over time, a convergence phenomenon for all recreated pairs emerges: when shared selections increase, the percentage of victories tends to increase as well. Time is required for the construction of collective effectiveness, especially for the most exposed positions (halfback, centre, and locks). Indeed, effective interpersonal coordination is based on pre-shared elements and common organization (Lim & Klein, 2006). In rugby, identifying how quantitative measures of an individual's sharing game may impact team performance is crucial, as this sport requires both individual and collective work to win duels and contests. The strategic common benchmarks (during static phase, movement or fixing) are also better acquired during shared training and consequently reinvest during the game (such as to act and react according to teammates) in response to the opponents' moves. During turnover state, it is imperative to create uncertainty among opponents, but still remain understandable to one's partners. These adjustments require a higher rate of games shared. When players break, their teammates need to be efficient at supporting them and making clean rucks to ensure continuity of progression. This kind of cohesion requires the creation of collective references and team ability to anticipate. Playing together for a long time appears to be the key point in bringing out these situational adaptations.

Overall selections and surrounding teammates

Individual features such as personal selections are quantified to rebuild the kind of team that borders recreated dyads (halfbacks, centres, and locks). Indeed, teams' composition is involved in their odds of success (Guimerà et al., 2005; Wuchty et al., 2007). For all best reconstructed teams, dyads are surrounded by higher selected forwards (particularly number 1 and 8 for halfbacks, numbers 4, 6, and 8 for locks and 4 for centres), like in RWC teams, in which the percentage of forwards participating at the previous RWC increased with the final ranking attained (Sedeaud et al., 2012). Collective effectiveness combines not only physical and tactical individual skills, but also a strong complicity, acquired over the years. Collectively adapting to opponents' attacks and providing a common effort require a shared knowledge and a combined action, especially for forwards. These collective investments are the forwards' cornerstone actions, such as providing support in breakdowns, synchronization during line-outs, maul for placement, collective push, and orientation in scrums (Sedeaud et al., 2012). Best vs. Rest analyses do

not show significant differences in opponents: best teams are not better because they play less successful teams. Moreover, numbers 4, 6, and, 8 which surround the locks, have the highest number of selections. This result is consistent with the fact that these teammates permanently interact and provide common effort either on offence or defence or during breakdown. Eighthmen and openside flankers are not only the first forwards in a breakdown to secure the ball, but are also good at reading the opposition's attacking lines, a task that requires strong coordination with their teammates and experience. Centres with the lowest winning percentages play on teams with the numbers 10, 13, and 14, who have a higher number of selections. This result also questions the backs' role specificity and qualities: speed, change of direction, agility, high handling skills, with line break, off-loads pass, and outflanking capabilities. Indeed, speed and change of direction capacities require athletes of a younger age (Berthelot et al., 2012). Surrounding teammates' selections underline the fact that a team of experts is not necessarily an expert team (Bourbousson et al., 2010; Eccles & Tenenbaum, 2004). Teams require coordination that rely on shared knowledge and action (Eccles & Tenenbaum, 2004).

This study establishes one indicator among others of Pierre Villepreux's thoughts: 'The reality of rugby is not managing an isolated player but the optimal management of a complex system in which it operates' (1993).

Perspectives and limitations

Similar to Duch et al, who provided an objective quantification of individual and team performance (2010), this study provides an indicator of playing together through shared and surrounded selection of a rugby union team. This is the first step towards a relevant and systemic indicator of personal and collective experience, which are both connected through dyads. To develop such a complete set of indicators, it will be necessary to measure all team members' shared selections (including substitutes) and incorporate game actions (pass, try, penalty, tackle …). Another limitation of this work lies in the fact that neither the players' age nor injuries are taken into account. This indicator do not integrate shared matches during the year by players of same club. Network dynamics is another scope to explore and apply to team construction. Finally, quantifying selections primarily underlines the head coach's choices rather than collective building. Questions to be studied will address how and how long the process of building collective effectiveness develops.

Conclusion

This study is the first to quantify the effect of selections number and shared selections on rugby union game results. The *XV de France*'s collective effectiveness relies on a balance between stability and workforce renewal. Time is required for building interactions within specific positions such as halfbacks. Moreover, the best halfbacks, locks, and centres are surrounded by more experimented forwards. The number of shared selections is an important quantifiable parameter to access collective effectiveness and impact collective performance.

Acknowledgements

The authors wish to thank INSEP teams for their full support. The results of the current study do not constitute endorsement of the product by the authors or the journal.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Batterham, A. M., & Hopkins, W. G. (2006). Making meaningful inferences about magnitudes. *International Journal of Sports Physiology and Performance*, *1*(1), 50–57.

Berthelot, G., Len, S., Hellard, P., Tafflet, M., Guillaume, M., Vollmer, J.-C., … Toussaint, J.-F. (2012). Exponential growth combined with exponential decline explains lifetime performance evolution in individual and human species. *Age (Dordrecht, Netherlands)*, *34*(4), 1001–1009. doi:10.1007/s11357-011-9274-9

Bourbousson, J., Poizat, G., Saury, J., & Seve, C. (2010). Team coordination in basketball: Description of the cognitive connections among teammates. *Journal of Applied Sport Psychology*, *22*(2), 150–166.

Cunniffe, B., Proctor, W., Baker, J. S., & Davies, B. (2009). An evaluation of the physiological demands of elite rugby union using Global Positioning System tracking software. *Journal of Strength and Conditioning Research / National Strength & Conditioning Association*, *23*(4), 1195–1203. doi:10.1519/JSC.0b013e3181a3928b

Duch, J., Waitzman, J. S., & Amaral, L. A. N. (2010). Quantifying the performance of individual players in a team activity. *PloS One*, *5*(6), e10937. doi:10.1371/journal.pone.0010937

Eccles, D. W., Tenenbaum, G., & others. (2004). Why an expert team is more than a team of experts: A social-cognitive conceptualization of team coordination and communication in sport. *Journal of Sport and Exercise Psychology*, *26*(4), 542–560.

Fransen, K., Decroos, S., Vanbeselaere, N., Vande Broek, G., De Cuyper, B., Vanroy, J., & Boen, F. (2014). Is team confidence the key to success? The reciprocal relation between collective efficacy, team outcome confidence, and perceptions of team performance during soccer games. *Journal of Sports Sciences*, 1–13. doi:10.1080/02640414.2014.942689

Gabbett, T. J. (2005). Science of rugby league football: A review. *Journal of Sports Sciences*, *23*(9), 961–976. doi:10.1080/02640410400023381

Gabbett, T. J. (2008). Influence of fatigue on tackling technique in rugby league players. *Journal of Strength and Conditioning Research/National Strength & Conditioning Association*, *22*(2), 625–632. doi:10.1519/JSC.0b013e3181635a6a

Gabbett, T., & Ryan, P. (2009). Tackling technique, injury risk, and playing performance in high-performance collision sport athletes. *International Journal of Sports Science and Coaching*, *4*(4), 521–533.

Guimerà, R., Uzzi, B., Spiro, J., & Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science (New York, N.Y.)*, *308*(5722), 697–702. doi:10.1126/science.1106340

Hendricks, S., & Lambert, M. I. (2014). Theoretical model describing the relationship between the number of tackles in which a player engages, tackle injury risk and tackle performance. *Journal of Sports Science & Medicine*, *13*(3), 715–717.

Hendricks, S., Matthews, B., Roode, B., & Lambert, M. (2014). Tackler characteristics associated with tackle performance in rugby union. *European Journal of Sport Science*, *14*(8), 753–762. doi:10.1080/17461391.2014.905982

Hendricks, S., Roode, B., Matthews, B., & Lambert, M. (2013). Defensive strategies in rugby union. *Perceptual and Motor Skills*, *117*(1), 65–87. doi:10.2466/30.25.PMS.117x17z6

Hendricks, S., Till, K., Brown, J. C., & Jones, B. (2016). Rugby union needs a contact skill-training programme. *British Journal of Sports Medicine*. doi:10.1136/bjsports-2016-096347

Heuzé, J.-P., Raimbault, N., & Fontayne, P. (2006). Relationships between cohesion, collective efficacy and performance in professional basketball teams: An examination of mediating effects. *Journal of Sports Sciences*, *24*(1), 59–68. doi:10.1080/02640410500127736

Hopkins, W. G., Marshall, S. W., Batterham, A. M., & Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine and Science in Sports and Exercise*, *41*(1), 3–13. doi:10.1249/MSS.0b013e31818cb278

Kozub, S., & McDonnell, J. (2000). Exploring the relationship between cohesion and collective efficacy in rugby teams. *Journal of Sport Behaviour*, *23*(2), 120–129.

Lacome, M., Piscione, J., Hager, J.-P., & Bourdin, M. (2014). A new approach to quantifying physical demand in rugby union. *Journal of Sports Sciences*, *32*(3), 290–300. doi:10.1080/02640414.2013.823225

Leo, F. M., Sanchez-Miguel, P. A., Sanchez-Oliva, D., Amado, D., & Garcia-Calvo, T. (2013). Analysis of cohesion and collective efficacy profiles for the performance of soccer players. *Journal of Human Kinetics*, *39*, 221–229. https://doi.org/10.2478/hukin-2013-0085

Lim, B.-C., & Klein, K. J. (2006). Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behavior*, *27*(4), 403–418.

Mellalieu, S., Trewartha, G., & Stokes, K. (2008). Science and rugby union. *Journal of Sports Sciences*, *26*(8), 791–794. doi:10.1080/02640410701819099

Myers, N. D., Payment, C. A., & Feltz, D. L. (2004). Reciprocal relationships between collective efficacy and team performance in women's Ice Hockey. *Group Dynamics: Theory, Research, and Practice*, *8*(3), 182–195.

Palla, G., Barabási, A.-L., & Vicsek, T. (2007). Quantifying social group evolution. *Nature*, *446*(7136), 664–667. doi:10.1038/nature05670

Sedeaud, A., Marc, A., Schipman, J., Tafflet, M., Hager, J.-P., & Toussaint, J.-F. (2012). How they won Rugby World cup through height, mass and collective experience. *British Journal of Sports Medicine*, *46*(8), 580–584. doi:10.1136/bjsports-2011-090506

Sewry, N., Lambert, M., Roode, B., Matthews, B., & Hendricks, S. (2015). The relationship between playing situation, defence and tackle technique in rugby union. *International Journal of Sports Science & Coaching*, *10*(6), 1115–1128. doi:10.1260/1747-9541.10.6.1115

Shearer, D. A. (2015). Collective efficacy at the Rugby World Cup 2015 – The role of imagery and observation. *European Journal of Sport Science*, *15*(6), 530–535. doi:10.1080/17461391.2015.1034787

Stajkovic, A. D., Lee, D., & Nyberg, A. J. (2009). Collective efficacy, group potency, and group performance: Meta-analyses of their relationships, and test of a mediation model. *The Journal of Applied Psychology*, *94*(3), 814–828. doi:10.1037/a0015659

Villepreux, P. (1993). Formation au rugby de movement. Toulouse: Editions Cépadus.

Walsh, M., Young, B., Hill, B., Kittredge, K., & Horn, T. (2007). The effect of ball-carrying technique and experience on sprinting in rugby union. *Journal of Sports Sciences*, *25*(2), 185–192. doi:10.1080/02640410600630571

Watson, C. B., Chemers, M. M., & Preiser, N. (2001). Collective efficacy: A multilevel analysis. *Personality and Social Psychology Bulletin*, *27*(8), 1057–1068.

Wheeler, K., & Sayers, M. (2009). Contact skills predicting tackle-breaks in rugby union. *International Journal of Sports Science & Coaching*, *4*(4), 535–544. doi:10.1260/174795409790291420

Whitfield, J. (2008). Collaboration: Group theory. *Nature*, *455*(7214), 720–723. doi:10.1038/455720a

Winter, E. M., Abt, G. A., & Nevill, A. M. (2014). Metrics of meaningfulness as opposed to sleights of significance. *Journal of Sports Sciences*, *32*(10), 901–902. doi:10.1080/02640414.2014.895118

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science (New York, N.Y.)*, *316*(5827), 1036–1039. doi:10.1126/science.1136099

c) Apports et perspectives

Cette étude introduit de nouvelles perspectives d'analyse de l'expérience collective au sein d'une équipe de rugby, d'une analyse globale de la rotation des joueurs à une analyse plus fine des relations entre différents groupes de joueurs.

L'analyse du turn-over, conséquence des choix du sélectionneur, d'un match à l'autre montre que les victoires sont associées à un turn-over de 30% en moyenne par rapport au match précédent contre environ 40% pour les défaites. Ce résultat montre l'impact brut de la gestion de l'effectif d'un match à l'autre. Modifier brutalement son effectif semble avoir un impact négatif sur la performance de l'équipe. En conséquence, nous avons introduit la notion d'expérience partagée par plusieurs joueurs participant à des taches complémentaires sur le terrain. Cette notion permet de prendre en compte l'aspect évolutif du temps de jeu commun de paires de joueurs au cours de leurs carrières internationales. Cette idée d'expérience partagée est estimée par les reconstructions de l'ensemble des sélections partagées entre toutes les paires de centres, les deuxièmes lignes et les charnières. Au cours du temps, un phénomène de convergence pour toutes les paires recrées émerge : le pourcentage de victoires des paires augmentent avec leur nombre de sélections partagées.

Ainsi, l'efficacité d'une équipe ne repose pas seulement sur la somme d'individualités mais sur la création de références collectives qui demande du temps. Partager du temps de jeu, que l'équipe gagne ou perde, semble être un point clef pour faire émerger des adaptations situationnelles. Cette étude est la première à retracer historiquement l'impact des sélections des coéquipiers et les sélections partagées entre les joueurs sur le résultat d'un match. Cependant n'ont été analysées que les sélections partagées entre deux joueurs de l'effectif. C'est une première étape vers la construction d'un indicateur estimant l'expérience collective partagée par l'ensemble des joueurs de l'effectif. Dans la continuité de cette logique, créer un indicateur de partage de temps de jeu collectif qui serait représentatif d'un nombre de matches cumulés et partagés par l'ensemble des joueurs présent sur le terrain introduirait une nouvelle mesure cohérente.

3) Etude 4: Quantification la performance collective au rugby à 15.

a) Introduction

La performance dans les sports collectifs ne résulte donc pas de la somme d'individualités mais de la gestion complexe des interactions qu'il existe entre les joueurs. Notre première étude sur le XV de France a montré l'importance de la quantification de l'expérience collective dans la compréhension de sa performance. L'étude qui suit présente un double objectif : estimer l'expérience collective d'une équipe et son évolution au cours du temps, puis mesurer son impact sur les performances de l'équipe. Ici, l'expérience collective d'un groupe sélectionné à un match *t* est définie comme la somme de l'ensemble des sélections partagées par l'ensemble des joueurs composant le groupe. Cet indicateur appelé *Cumulative Shared Selections* (CSS), sélections cumulées partagées en français, résume intelligiblement la complexité des interactions créées entre les joueurs au cours du temps par le temps de jeu disputé en commun. Par sa définition, il évolue d'un match à l'autre, reflétant la stratégie de gestion de l'effectif mise en place par le sélectionneur. L'objectif est d'analyser l'évolution de cet indicateur d'expérience collective et de mesurer son impact sur la performance des meilleures équipes internationales de rugby à XV. Fort des conclusions de l'étude précédente, nous supposons qu'un lien fort relie l'histoire vécue par un effectif et sa performance.

b) Quantifying collective performance in rugby union

i. Introduction

'The reality of rugby is not managing an isolated player but the optimal management of a complex system in which it operates'. Echoing to Pierre Villepreux's thought, many studies have shown the importance of investigating cohesion and collective efficacy factors as key performance indicators in team sports (137,146,151,156,161,162). In rugby union, may be more than in any other team sport, no significant effort can be performed without the help and cooperation of every player. In this context of cooperation, a team's performance can be described both by its productivity and by the sum of its players' abilities. Confronting this state with Pierre Villepreux's thought, the main goal of a manager will be to make his team's performance greater than the simple sum of the parts. Getting more than the simple sum of the parts is described by Oliver (163) as filling a gap. Filling a gap between individual and teammate fit is defined as the complementarity of skills, as a function of how the team is constructed.

To do so, the manager needs to create cohesion and collective efficacy, the two have been proven determinants of performance. Qualitative aspects of collective experience on individual and collective performance have been explored. Heuzé (151) established that teammate's from more cohesive teams share stronger beliefs in their team. This enhances individuals' perception of collective efficacy thereby improving cohesion. In parallel, Marcos (161) added that improving team cohesion in basketball ameliorates a teammate's perception of efficacy. Moreover, Leo (164) stated that players with higher cohesion and collective efficacy profiles belonged to the best teams. Generating synergy through a coordinated effort which allows each member to maximize their strengths and minimize their weakness (165) is a long-term process (162). It requires coordination based on shared knowledge and action (160) that teammates can only achieve through time spent together and traditional methods of team-building (140). Thus, impact of cohesion and collective efficacy gathered through collective experience has been widely qualitatively correlated with improvement of team performance.

Fewer studies have led to quantify player profiles and their potential relation on performance. Individual experience among the collective is a determinant for the team to perform. Previous studies have found that rugby union international players may be selected because of greater skill and experience (43,154,159). Furthermore, the sum of individual experience among the

squad has also been reviewed. Sedeaud et al (16) showed that teams including forwards with previous World Cup experience perform better. Moreover, specific analysis on collective effectiveness of XV de France showed that more experienced forwards surround the best halfbacks, locks and centres (162). More studies explored how shared collective experience could affect team performance. Kahane (166) showed that in Major League Baseball, turnover of players from one season to another had a significantly negative impact on attendance when games are played at home. This negative effect leads ultimately to a decreasing likelihood of winning for the home team (167). Impact of workforce renewal rate on results has also been analysed in rugby union and soccer. A high rate of turnover in the XV de France between two consecutive games is associated with the loss of the game (a mean squad turnover of around 30% is observed in victorious games and 41% for those lost) (162). Similar results have been found in soccer by Carling (126). The best results of a soccer team over 5 seasons are obtained when the fewest number of players are alternated over the season. The team won the national championship with the same 10 players involved in at least 75% of the total minutes played during the season. During the other seasons, less than 6 players shared this characteristic of high play-time. This was possibly due to higher player availability and low injury incidence, which had a significant influence on team success (168). Those results imply that squad management strategies directly foster team performance.

The impact of team experience on team performance has been analysed through qualitative estimators such as cohesion and collective efficacy, and then quantified by indicators summing individual experiences or assessing a global estimation of the group experience.

The purpose of this study was to quantify team experience and its evolution through time, based on cumulative shared selections of players. Then, to assess its impact on team performance. As specified by Shearer (140), the prestige and the four-year cycle between two Rugby World Cup events make this competition an ideal field of investigation. We assume that a strong link exists between a club's history and its performance. That the greater the experience, the better the group's performance will be.

ii. Methods

Data collection

After obtaining the approval of the Institutional Ethics Committee, scoresheets of all games involving at least one of all 10 nations participating at the Rugby Championship (*Argentina, Australia, New Zealand and South Africa*) and the Six Nations Championship (*England, France, Ireland, Italy, Scotland and Wales*) were collected from the end of the 1999 Rugby World Cup (RWC) up to the 2015 RWC final. Date of the game, competition, points scored and conceded, names and positions (forwards and backs) of players involved as starters were recorded for each international game. Data were collected from an up-to-date rugby news and statistics website: www.statbunker.com. The site is updated after every game, and has a comprehensive list of statistics, covering all aspects of rugby union. Additional to the scoresheet data, World Rugby Ranking points of each nation were recorded weekly from its introduction at the beginning of 2004 up to the last game covered by our study period. Ranking points were collected on the World Rugby website: www.worldrugby.org.

Team experience indicator: Cumulative shared selections

First, we need to quantify team experience at a given game. A single indicator based on the set-lists of the starters selected by the head coach is produced. For each game, number of selections that each player has shared with the others during the previous games are computed and added up. This calculation quantifies the Cumulative Shared Selections (CSS) of a team at a given game. The number of CSS is computed over the entire period (1999-2015).

Then we aim to analyse the evolution of the team experience indicator through time. As implied by Shearer (140), we choose to divide our study period into sequences of 4 years, each sequence corresponds to the period between two consecutive RWCs. Mean number of CSS is computed by nation for each sequence.

Slope, intercept, ranking points and victory

The behaviour of the team experience indicator through time is analysed. Slopes and intercepts from linear regression between CSS's evolution and games played are computed and compared for each sequence and nation. The slope between two RWCs reflects how the CSS evolved in preparation for the following RWC. The intercept is used to for analysis adjustment over the basal team experience at the beginning of a RWC cycle.

Both World Rugby Ranking points and percentage of victories are used to estimate national team performance. From the same perspective of analysing evolution of the team experience, mean ranking points and victory percentage are computed by nation for each 4 year time block.

Statistical analysis

Data were reported as mean ± standard deviation. Multiple linear regression analysis was used for establishing the potential associations between team performance and team experience indicators which were entered as dependent variable. Results were considered significant at $p < 0.05$. All statistical analyses were performed using R (version 3.3.2; The R Foundation for Statistical Computing, Vienna, Austria).

Team experience estimated by the number of CSS computed in the first years of our study period is biased and does not take into account all previous games shared by the players. Therefore, specific analysis is conducted over three 4-year sequences between 2004 and 2015.

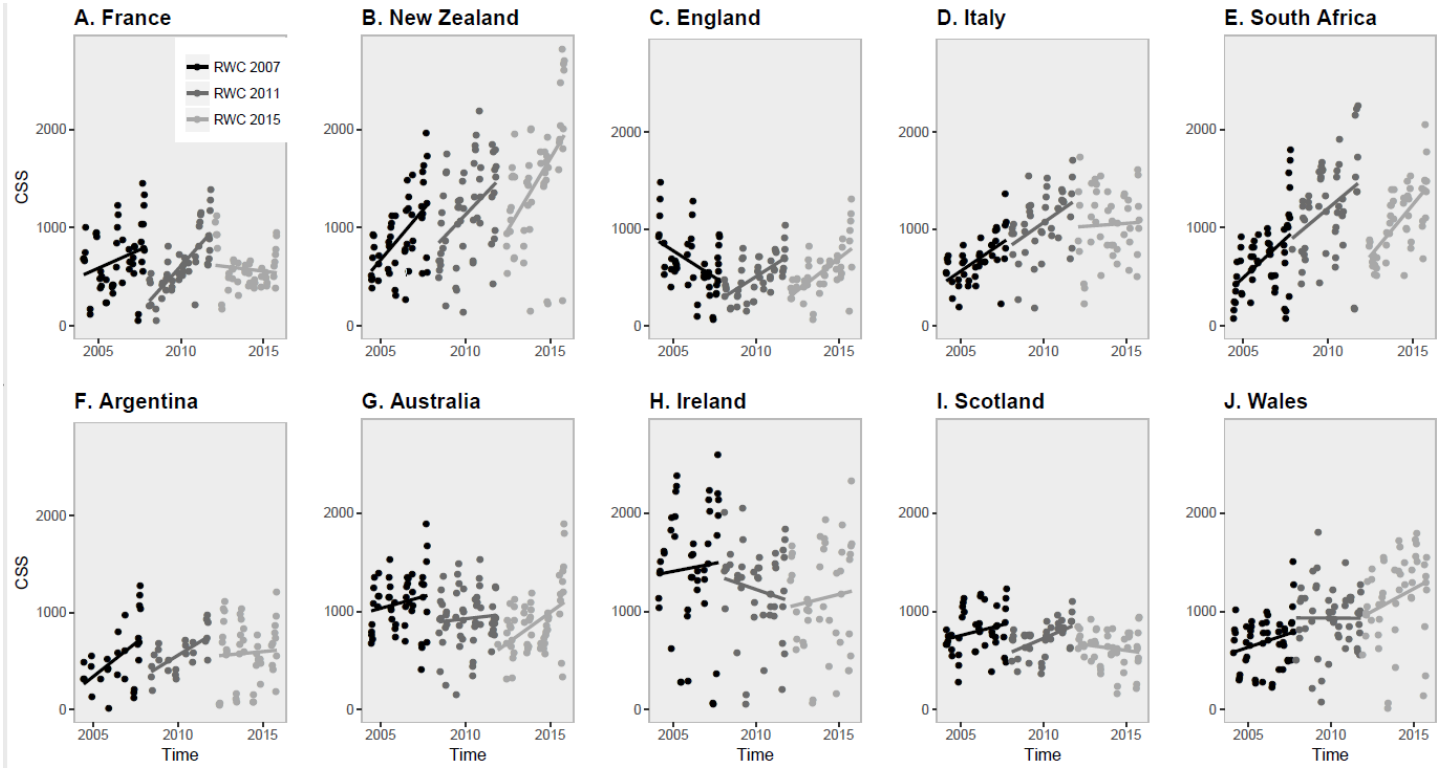


Figure 1: CSS evolutions over the 2007, 2011 and 2015 RWC.

iii. Results

Variable descriptions

The study period covers three sequences of 4 years between the 2007, 2011 and 2015 RWCs. Mean number of games played by the 10 nations over a sequence is 45.9 ± 6.9, involving an average of 62.1 ± 7.6 different players. Regarding the experience indicators, mean number of CSS over a sequence is 860.4 ± 284.9 and it evolves, on average, with a slope of 7.9 ± 8.1 and an intercept of 677.0 ± 287.5. Regarding performance indicators, the ten nations completed the sequences with mean ranking points of 82.21 ± 5.59 and a mean victory percentage of 55.27 ± 18.95.

Multiple linear regressions

In multiple linear regressions, CSS slope is found to be significantly associated with both ranking points (*p* Value = 0.042) and victory percentage (*p* Value = 0.001), adjusted on the CSS intercept (Table 1).

Table 1: Multiple linear regression results. Effects of team experience indicators over team performance.

	Coefficient (β)	SE	95% CI	*p* Value
Ranking Points				
CSS Slope	0.36	0.16	0.074 to 0.643	0.042
CSS Intercept	0.005	0.005	-0.003 to 0.013	0.330
Victory Percentage				
CSS Slope	1.78	0.45	0.992 to 2.568	0.001
CSS Intercept	0.02	0.013	-0.005 to 0.04	0.193

Example: France VS New Zealand

Looking at France's trend of CSS, the slopes leading to the first two world cups are 7.21 and 18.20 resulting in a 4th and 2nd place ranking, respectively (Table 2). Then, the slope of the last RWC period is -1.47 leading to a defeat in the quarter final.

In the meantime, New Zealand's evolution of CSS shows a constant increase with an intercept of 504.52 during the first sequence and one of 876.95 during the last one. During the first sequence, mean number of CSS was 911.17 and 1451.26 during the last one. New Zealand suffered a defeat in the quarter finals with a slope of 16.59 in the 2007 RWC but then won the next two world cups with a slope of 14.39 and 21.27.

Other nations

Figure 1 illustrates CSS evolutions over the 2007, 2011 and 2015 RWC for the ten teams involved in four and six nations. In 2007, the RWC winner displayed a slope of 12.99 (Table 2). During the entire 12 years of follow-up, only Wales and New Zealand share a continuous progression of their CSS (Figure 1) and an uninterrupted improvement of victory percentage and ranking score (Table 2).

Table 2: Descriptions of the variables from the 10 nations over the 2007, 2011 and 2015 RWC.

Nation	RWC	CSS	Slope	Intercept	Ranking Points	Victory Percentage	RWC Result
	2007	911.17 ± 403.48	16.59 [9.61 ; 23.60]	504.52 [307.71 ; 701.32]	92.59 ± 1.50	87.5	quarter finalist
New Zealand	2011	1147.60 ± 488.61	14.39 [6.98 ; 21.80]	744.67 [506.22 ; 983.13]	91.39 ± 1.49	83.64	winner
	2015	1451.26 ± 632.65	21.27 [11.43 ; 31.11]	876.95 [571.53 ; 1182.36]	92.70 ± 1.06	94.12	winner
	2007	675.49 ± 324.91	7.21 [0.88 ; 13.54]	495.33 [313.55 ; 677.11]	85.18 ± 1.26	70.83	semi finalist
France	2011	627.04 ± 304.52	18.20 [13.78 ; 22.62]	208.42 [91.74 ; 325.10]	82.02 ± 1.91	60,00	finalist
	2015	577.60 ± 206.58	-1.47 [-6.29 ; 3.35]	611.42 [484.21 ; 738.62]	81.58 ± 1.66	46.51	quarter finalist
	2007	645.06 ± 309.81	-9.03 [-15.25 ; -2.81]	861.73 [690.28 ; 1033.18]	83.39 ± 4.78	44.68	finalist
England	2011	515.53 ± 212.82	10.6 [6.83 ; 14.37]	271.83 [172.25 ; 371.42]	82.33 ± 1.5	54.55	quarter finalist
	2015	561.18 ± 259.9	12.65 [7.97 ; 17.33]	270.23 [146.62 ; 393.84]	84.24 ± 1.57	63.64	other
	2007	679.85 ± 238.3	11.55 [6.3 ; 16.79]	437.35 [310.9 ; 563.81]	72.81 ± 1.44	30,00	other
Italy	2011	1061.93 ± 328.79	11.72 [4.01 ; 19.42]	810.04 [619.93 ; 1000.16]	73 ± 1.08	21.43	other
	2015	1045.64 ± 358.9	1.07 [-7.33 ; 9.47]	1021.04 [799.23 ; 1242.84]	73.31 ± 2.01	24.44	other
	2007	684.96 ± 386.84	12.99 [6.97 ; 19.01]	334.24 [147.4 ; 521.08]	85.74 ± 1.97	67.31	winner
South Africa	2011	1194.94 ± 483.66	12.72 [3.52 ; 21.92]	876.93 [612.55 ; 1141.32]	88.54 ± 2.07	63.27	quarter finalist
	2015	1060.6 ± 356.13	16.97 [11.34 ; 22.59]	644.94 [486.61 ; 803.27]	86.91 ± 2.17	69.57	semi finalist
	2007	543.71 ± 331.87	24.1 [11.06 ; 37.14]	194.28 [-22.22 ; 410.78]	79.29 ± 3.53	64.29	semi finalist
Argentina	2011	575.88 ± 196.92	18.34 [9.07 ; 27.6]	346.66 [214.24 ; 479.08]	81.01 ± 2.83	37.5	quarter finalist
	2015	585 ± 319.17	2.44 [-4.51 ; 9.39]	526.34 [334.73 ; 717.95]	77.72 ± 2.58	28.26	semi finalist
	2007	1085.35 ± 293.76	4.41 [-1.48 ; 10.3]	975.04 [805.79 ; 1144.28]	86.53 ± 1.48	60.42	quarter finalist
Australia	2011	927.15 ± 296.89	2.22 [-2.85 ; 7.29]	864.91 [701.76 ; 1028.06]	85.86 ± 1.17	59.26	semi finalist
	2015	874.39 ± 320.51	9.13 [4.53 ; 13.73]	609.64 [456.34 ; 762.95]	86.28 ± 1.81	60,00	finalist
	2007	1441.59 ± 646.01	4.49 [-11.11 ; 20.09]	1340.61 [937.62 ; 1743.61]	82.27 ± 1.56	61.36	other
Ireland	2011	1217.73 ± 457.29	-4.16 [-14.79 ; 6.47]	1313.41 [1032.61 ; 1594.22]	81.01 ± 2.3	56.82	quarter finalist
	2015	1133.61 ± 543.61	5.74 [-7.31 ; 18.8]	1004.39 [667.06 ; 1341.71]	82.08 ± 2.75	61.9	quarter finalist
	2007	796.95 ± 231.49	5.43 [-0.35 ; 11.2]	680.22 [537.69 ; 822.76]	75.34 ± 1.31	33.33	quarter finalist
Scotland	2011	724.92 ± 171.38	8.33 [4.16 ; 12.5]	558.28 [462.65 ; 653.92]	76.98 ± 1.87	42.11	other
	2015	624.79 ± 186.72	-1.76 [-5.68 ; 2.17]	667.83 [557.39 ; 778.26]	76.49 ± 1.22	37.5	quarter finalist
	2007	693.21 ± 269.52	5.35 [-0.14 ; 10.84]	562.16 [407.68 ; 716.63]	78.2 ± 2.48	45.65	other
Wales	2011	931.72 ± 332.22	0.33 [-6.29 ; 6.94]	923.35 [729.58 ; 1117.12]	79.43 ± 1.89	51.02	semi finalist
	2015	1135.6 ± 439.76	8.76 [-0.19 ; 17.71]	921.03 [669.05 ; 1173]	81.65 ± 2.17	54.17	quarter finalist

iv. Discussion

This study is the first to create a key collective performance indicator and reveal its association with ranking and victory.

We have investigated the four-year cumulative shared selections evolution of the national rugby teams involved in four and six nations between consecutive RWCs from 2004 to 2015. A general indicator quantifying collective experience was computed. Its impact on rugby union game results was assessed. Previous work introduced the idea that shared selections are highly involved in rugby union teams' performance and that time is needed to let collective effectiveness emerge (162). Sedeaud et al. concentrated their investigations on shared experiences between pairs of players (halfbacks, locks or centers). Investigating shared experiences between all players reveals a new way to test these relationships.

Team cohesion and collective efficacy are widely used to qualify collective experience (151,156). These qualities are acquired through time spent by the players with each other (140) : CSS is proposed as a global statistical tool to quantify all the experiences shared between the players during successive games. By definition, this indicator changes after each game. Then it is a natural candidate to analyse team dynamic construction over a defined period. Slopes and intercepts of CSS evolution over four-year cycles preceding RWCs put into numbers the strategies established by the managers to create their teams.

With regards to the CSS trends, its slopes and intercepts, both intra- and inter-nation variability appear wide. Positive and negative slopes can be observed for the same team from one four-year cycle to the next. For example, England presents opposite CSS evolutions leading up to the 2007, 2011 and 2015 RWCs with a negative slope followed by two positive ones. Those slopes are associated with a 10 percent increase in victories per four-year sequence (from 44% to 64%). On the other hand, South Africa seems to apply a different strategy for two successive RWCs, with two consecutive positive slopes (increasing team building), with the second one starting precisely where the first one ends, implying some continuity in the policy of player selection. Such a continuity observed for the 2011 RWC four-year cycle could be explained by the confidence in the players after their success in the 2007 RWC. Nonetheless, a continuous improvement in CSS during a long period is paradoxical. It implies the general aging of the group which in reality is naturally topped out by the physical demands. Mean ages of RWC players of 25 years for backs and 27 for forwards (16) reflect such a reality. The construction of a collective workforce that shares

consequent playing time also faces the necessary youth physical prerequisites. Indeed, specific positions in rugby union demand repeated sprints, changes of direction, capacities that require young athletes (7,12).

Concerning multiple linear results exploring effects of experience indicators over performance ones, CSS slope is found to significantly impact both ranking points and victory percentage while CSS intercept does not. This is an interesting result, which implies that improving CSS (i.e. international games played together) over the entire four-year period prior to the RWC is more important that initial collective experience garnered in the past. In other words, progression is a key variable rather than the starting point. These results are consistent with previous conclusions showing that the same halfbacks, locks or centres selected over time, obtained at the end of their common career, a winning percentage similar to the team's average (162).

The issues confronting the observations of France and New Zealand illustrates these findings. The rendez-vous of the RWC every 4 year allows to analyse how teams are built while preparing for such competitions (140). For the best team, optimizing shared selections is a suitable strategy, but one that takes time and should be considered to have some limits. Furthermore, high intercepts are more often associated with negative or gentle slopes: Australia, England in 2007, Ireland during the entire 12 years, and Italy in 2015.

This reveals that the first year after an RCW may be crucial, that early choices of head coaches may be a primary determinant for the rest of the four-year period (e.g. keeping attention to the incorporation process, mixing experienced players with young ones). New Zealand parallels these trends with a slight increase of intercepts, which results in a brilliant mix between players. This mix brings together experienced players, indispensable to the tactical group cohesion and optimal operation and progressively incorporates young players at the peak of their physical performance.

Since professionalization in 1995, rugby union match activities have changed with an increase of passes, tackles, rucks, tries, effective time of play but a reduction of scrums and mean participation time per player (169). Actual considerations to prevent injuries tends to reduce this participation (170).

Dwindling effective time of play coupled with the fact that nations compete about only 11 games per year, force teams to capitalize on every opportunity to develop collective experience. Effective time of play is about 44% of overall match time (39) during which

players share specific actions according to their positions. Loose, locks and front row forwards are involved on average in 12.9, 15.0 and 10.9 rucks and 23.4, 21.4 and 21.7 scrums, respectively (171). Backs have a greater involvement in passes and high speed runs. They make about 11 passes per game and cover about 851 meters at high speed (36). Those are relative small numbers of events and fewer events lead to a try, a penalty kick or a drop. Players need to have perfect knowledge of each other in order to achieve synchronization allowing them to rightly form and clear rucks, to complete decisive passes at high speed or to perform offloads.

A positive CSS slope is correlated with a good performance as it reflects an increase in the time of play, accumulation of passes, coordinated high-speed runs, and succession of rucks in competitive environments. Some studies demonstrated that synergy of running line velocity, position on field between ball carrier and support player in non-ecological 2 vs 1 situations, or 4 vs 2 + 2 are imperative (172). However, these studies focused only on "parts of a complex system" which constantly evolves between teammates and opponents. Indeed, athletes individual performances, athletes synchronization by dyads or small groups and team performance are neither proportional nor linear (173) and we need elements of understanding in ecological situations. The CSS provides a scoring parameter based on shared time of play during matches, which is consequently involved in such an ecological system.

v. Perspectives and limitations

The aim of this study was to quantify time spent together by the players and test whether a strong collective experience was correlated with better collective performance. Taking into account the fact that a squad can be composed of some players that play in the same club might be relevant. Groll (174) showed in soccer that too many players coming from the same clubs negatively impacts the national selection's performance because of a lack of diversity (to many players scattered in too few clubs). In addition, it leads to a lack of knowledge of foreign game features. Future studies must investigate the impact of this factor in rugby union for comparison with results found by Groll in soccer. Comparisons between TOP 14 French players who come from many different clubs, with the New Zealand players involved in Super Rugby playing primarily for only 5 franchises could be relevant in order to identify different efficiency collective patterns.

Social network analysis techniques are increasingly applied to analyse behaviours within teams and inter-player interactions during games (25,175–178). Bipartite structure of complex networks as described in Guillaume & Latapy 2006 (179) might be an interesting tool to deeply scope the evolution of players' interactions through time and successive games, and might reveal individual or specific positions influencing the entire team evolution.

Due to a lack of information, we were not able to take into account the impact of injuries on CSS evolution. The negative effect on team performance of players unavailability caused by injuries has been shown in soccer (126,128,129,168). As reported by Bengtsson (130), the influence of team rotation strategies on a) team performance and b) injury rates remains unclear. While our results lift the veil on the first relationship, further research must help to define the second one. Studying CSS impact on a team performance during a championship with large injury incidences (such as in rugby union) might help to better understand inherent relations between squad management, team performance and player injury.

vi. Conclusion

For the first time, a single estimator (CSS) allows to relate the evolution of a team's experience and its performance through time. This study reveals the potential of this indicator. It would be captivating to transpose this methodology to other team sports and other competition formats, such as a championship. Investigating more precisely the links between players through social network analyses would also make it possible to discretise relationships and detect key individuals or groups of individuals.

National managers need to create the most competitive squad for each competition. They have to decide what to do with the time and few games to play that are given to them. The number of cumulative shared selections is a parameter that could help them in the decision making process.

c) Apports et perspectives

L'objectif de cette étude était de quantifier le temps passé entre les différents joueurs composant un collectif à travers un nouvel indicateur : le nombre de CSS, puis de tester si cette expérience partagée impactait la performance de l'équipe.

Dans notre étude, nous avons ciblé comme objectif sportif de « performer » à la coupe du monde de rugby qui a lieu tous les 4 ans. Le but du sélectionneur est alors d'utiliser les quatre années précédant la compétition pour trouver, tester, confirmer ses joueurs afin d'avoir un effectif roder, qui se connait parfaitement, qui pourra s'adapter à toutes les situations. La création d'une expérience collective est donc un processus dynamique, une entreprise longue et sinueuse. Observer les pentes d'évolution des CSS sur des périodes de 4 ans entre chaque coupe du monde a permis de bien rendre compte des stratégies mises en place par les sélectionneurs. Une pente positive, reflétant une augmentation du temps de jeu commun, une accumulation du nombre de passes tentées, ratées et réussies, du nombre de courses coordonnées est corrélée à de meilleures performances qu'une évolution négative ou moins prononcée. Nous avons développé un indicateur unique (CSS) permettant de rendre compte à un match *t* de l'expérience collective d'une équipe alignée et d'éclairer sur la gestion du groupe à travers l'analyse de son évolution au cours du temps. Une gestion favorisant la création d'expérience collective est corrélée positivement à de meilleures performances.

4) Perspectives générales

A travers ces deux études portant sur le rugby à 15, une spécifique au XV de France dans un premier temps puis la seconde élargie au dix meilleures équipes internationales, nous avons montré l'intérêt de l'analyse de l'expérience collectif d'un groupe dans la compréhension de sa performance.

a) Analyse des réseaux sociaux

Pour réaliser ces analyses, nous nous sommes inspirés des idées et hypothèses issues d'études sociologiques traitant largement le sujet avec des outils qualitatifs (137,146,151,156,161). Les outils utilisés dans l'analyse des réseaux sociaux (*Social Network Analysis*, *SNA*) ont également été source d'inspiration (25,175–178). En effet l'indicateur *CSS* considère les relations sociales existant entre les joueurs, une équipe étant dans notre cas un réseau social composé de liens (les sélections) entre des nœuds (les joueurs). Les méthodes de *SNA* sont de

plus en plus utilisées pour étudier les comportements des joueurs d'une même équipe et avec ceux de l'équipe adverse au cours d'un match. Dans ce contexte, les nœuds sont les joueurs et les interactions peuvent être des passes, des contacts, des coordonnées de déplacement. Concernant notre problématique d'observation et d'estimation de l'expérience collective, la structure des réseaux bipartis permettrait d'analyser l'aspect dynamique de l'évolution des sélections communes entre les joueurs d'un match à l'autre (179).

L'indicateur *CSS* fourni une information macroscopique : l'expérience collective d'une équipe est estimée à un match à un instant *t*. La volonté est de compléter cette information macroscopique par une analyse microscopique plus fine des relations entre les joueurs. Et par la même occasion affiner l'analyse de la dynamique de construction d'une équipe à travers le temps. Utilisant les données issues de l'étude 4 avec le même focus sur les quatre années précédant les coupes du monde de rugby à 15, nous avons déterminé les sélections partagées entre les différents joueurs année par année. A partir de la reconstruction de ces relations, nous avons construit un réseau social par année illustrant les interactions créées par l'entraineur. Chaque nœud du réseau représente un joueur et chaque lien entre deux points, par son épaisseur, le nombre de matches communs aux deux joueurs sur la période considérée. L'ensemble des joueurs utilisés pendant les quatre ans précédant la coupe du monde sont catégorisés selon leur participation ou non à la compétition finale, selon le choix ou non de l'entraineur. La *Figure 2* ci-dessous représente les réseaux créés année par année par Philippe Saint-André entre les 73 joueurs qu'il a convoqués au cours de son mandat à la tête du XV de France. Les points roses correspondent aux joueurs participant à la coupe du monde 2015, reliés par des liens roses. Les points bleus correspondent aux joueurs testés mais non retenus pour la coupe du monde, ils sont reliés entre eux et avec les joueurs de la liste définitive par des liens gris.

Cette représentation dynamique des réseaux permet de visualiser la stratégie mise en place par le sélectionneur. Les liens roses rendant compte de l'expérience collective créée au cours des années, les liens gris matérialisant les tests de l'entraineur. La première année, la quasi-totalité des joueurs testés sont bleus et ne rejouerons plus par la suite. Puis d'année en année on observe la proportion de joueurs roses augmenter, résultant de la certitude de l'entraineur dans ses choix, impliquant la création d'un vrai collectif, la capitalisation du temps que les joueurs passent ensemble à chaque rassemblement international. Malgré la persistance jusqu'à la veille de la coupe du monde d'incertitudes, de joueurs néo sélectionnés, d'ultimes tests, conséquences possibles de blessures et d'indisponibilités forcées.

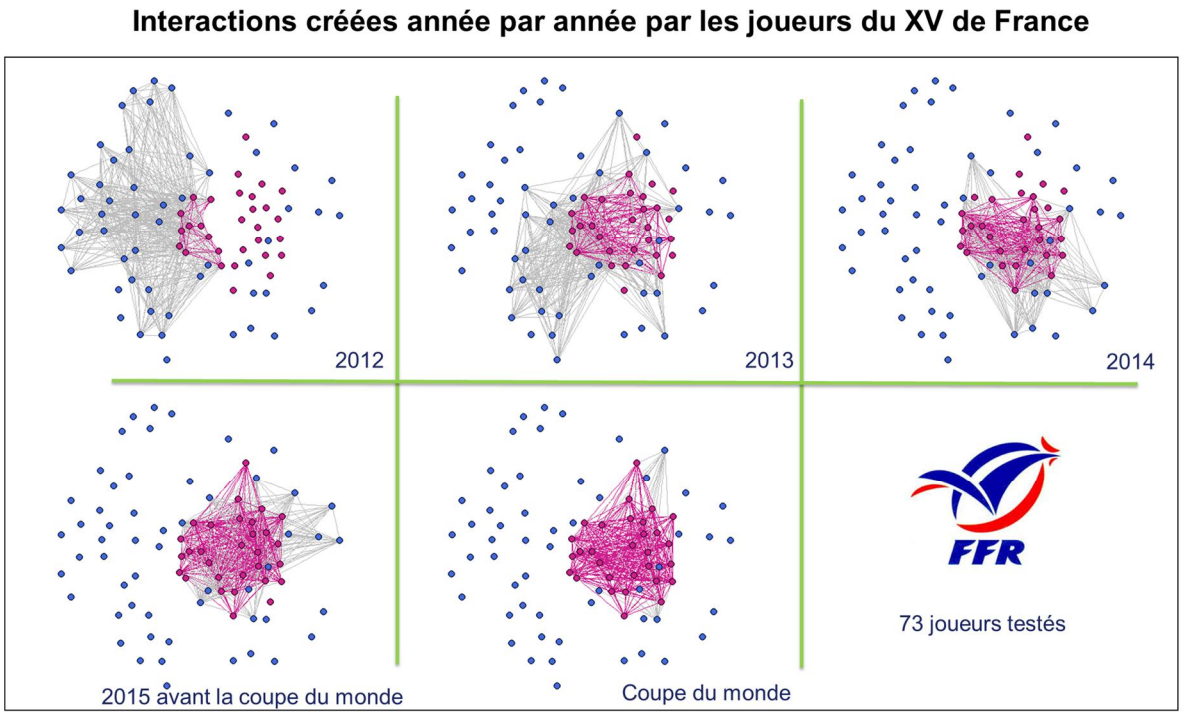


Figure 2: *Evolution des liens créés année par année au sein du XV de France au cours des 4 ans précédant la coupe du monde 2015. En rose sont les joueurs retenus par le sélectionneur pour participer à la compétition et les liens créés entre eux. La largeur du lien est proportionnelle aux nombres de sélections partagées par deux joueurs sur la période.*

Les propriétés dynamiques de ces types de réseaux sociaux offrent des possibilités nouvelles dans l'interprétation des performances d'un collectif dans une grande compétition internationale. Ces résultats ne sont qu'exploratoires et nécessitent des analyses plus poussées. Mais l'on peut déjà sentir l'intérêt de cette science des réseaux pour la compréhension de la création d'expérience collective au sein d'un groupe. L'application de ces outils à d'autres formats de compétitions comme un championnat fermé avec un nombre de joueurs fixe et limité, dont les blessures et absences impacteraient plus significativement le groupe qu'une sélection nationale, est une perspective prometteuse.

b) Prévention des risques de blessures

Ces deux études menées sur le rugby à 15 représentent un exemple d'appropriation par le sport de méthodes d'analyses utilisées dans d'autre domaine, ici, l'analyse des réseaux sociaux. Appropriation utile dans un premier temps à l'optimisation de la performance d'un collectif. Cependant nous avons comme perspective l'approfondissement de ces analyses et

l'application à la minimisation du risque de blessure des joueurs. Option peu envisageable avec les données traitées dans cette partie puisque les joueurs en sélections nationales ne jouent qu'une dizaine de match au plus par année. L'idée ici est de contextualiser l'importance de l'évolution des *CSS* sur la performance du groupe en prenant en compte la survenue des blessures au sein de l'effectif. Pour se faire, nous souhaitons par la suite transposer nos méthodes à l'analyse d'équipes participant à un championnat. Ce format de compétition avec un nombre de joueurs sélectionnables fixe du début à la fin de la saison ainsi qu'un grand nombre de matches répartis le long de la saison semble être le plus pertinent pour répondre à cette nouvelle problématique. En ce sens, la *Figure 3* ci-dessous décrit l'évolution du nombre de *CSS* sur deux saisons de Ligue 1, avec représentées en vert les équipes terminant la saison aux places qualificatives pour les coupes d'Europe, en bleu les équipes reléguées et en rouge les autres. On peut observer entre autre une nette rupture de ces différentes évolutions à la mi-saison pour les équipes reléguées.

Quels sont les mécanismes derrière ces différentes évolutions ? Comment la gestion du groupe mise en place par l'entraineur impacte l'expérience et la performance ? Quel est le poids des blessures dans l'évolution d'un collectif ? Une gestion d'effectif augmente-elle le risque de se blesser plus qu'une autre ?

Toutes ces interrogations peuvent se résumer dans la suivante : existe-t-il une gestion d'effectif optimale permettant d'enchainer les matches au meilleur niveau tout en minimisant l'exposition aux risques de blessures liés aux calendriers chargés et aux périodes avec une fréquence de matches soutenue? L'étude épidémiologique de la survenue des blessures au sein d'effectifs au cours de plusieurs saisons sera nécessaire pour répondre à ces questions. Ce qui impliquera à nouveau l'appropriation par le monde sportif de méthodes d'analyses propre à la santé publique, comme les mesures d'incidence de blessures, les taux de guérison, la sévérité des pathologies, des modèles de Cox et de Kaplan Meier pour estimer les retours au meilleur niveau de performance ou encore différencier les récidives des blessures nouvelles.

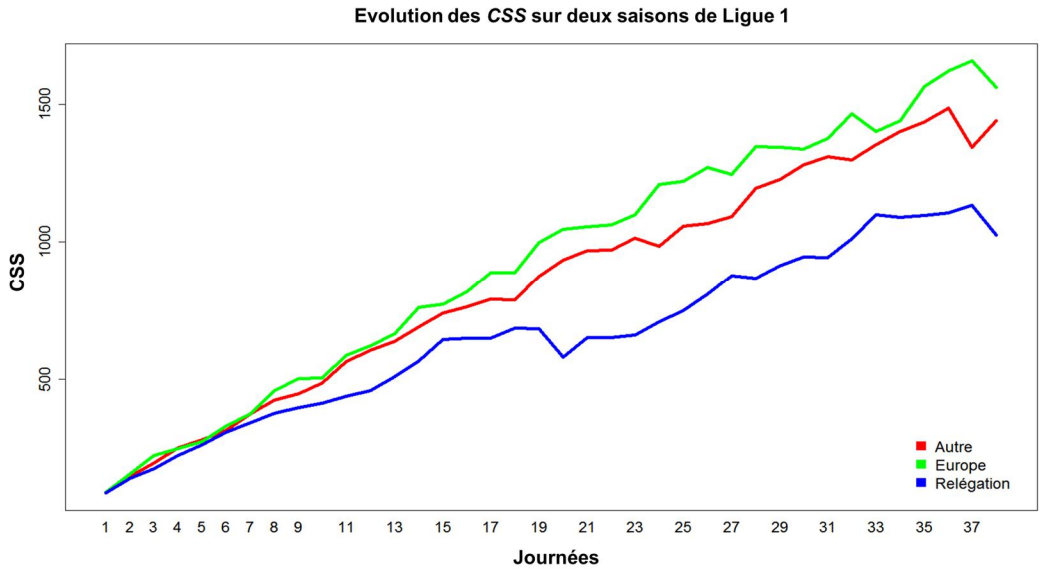


Figure 3: Evolution des CSS des équipes de Ligue 1 sur deux saisons.

5) Conclusions

La première partie de cette thèse traite le développement et l'amélioration de méthodes statistiques. Inscrites dans une réflexion globale autour de l'athlète regroupant et coordonnant les expertises de l'ensemble des corps décisionnels (techniques, tactiques, médicaux), ces méthodes optimisent l'individualisation du suivi de l'athlète, aidant ainsi à la prise de décisions. Nous l'avons vu, les différences inter individuelles sont telles dans les sports collectifs que cette individualisation est indispensable. Mais pour comprendre les déterminants de la performance d'un sport collectif, l'analyse du groupe l'est tout autant. C'est ce qu'ont montré les deux études précédentes sur le rugby, l'analyse de la gestion du groupe à travers des outils statistiques adaptés permet d'éclairer sur le résultat positif ou négatif de l'équipe.

Comprendre l'ensemble des interactions entre la gestion de l'effectif, les charges individuelles ressenties, la survenue de blessures et les performances individuelles et collectives, va nécessiter la prise en compte de la double vision individuelle et collective dans l'analyse. Le volet macroscopique estimé par les *CSS* sera complémenté par sa version microscopique à travers les *SNA*, permettant l'identification fine des liens et des individus concernés. Le niveau de lecture très fin permis par l'analyse des séries individuelles de biomarqueurs associé à l'étude épidémiologique des blessures harmoniseront la compréhension de l'état de santé des athlètes. L'observation des phénomènes inhérents aux performances et aux blessures à travers ce double spectre individuel et collectif sera indispensable à notre compréhension de ces processus.

V. Conclusions générales

A travers cette thèse a été traité l'apport des données dans la compréhension des performances et de l'état de santé des joueurs de sports collectifs. Nous l'avons évoqué, la professionnalisation du sport de haut niveau a été accompagnée d'une importante médiatisation et n'a pas échappé à l'envahissement des données. De par leur complexité, les sports collectifs ont été particulièrement intéressés par les données et ont suscités l'intérêt des producteurs de données. Cet intérêt partagé a permis aux professionnels du sport de haut de niveau de formuler des demandes problématisant les questions rencontrées sur le terrain et professionnels de la donnée de produire des données pour offrir une réponse. Les études présentées dans cette thèse illustrent certaines étapes de cet échange, de cette navette entre la question émanant du staff, les données qui y répondent et le staff qui s'approprie cette réponse pour l'inscrire dans son processus de décision.

L'accent a été mis sur deux points importants. Le premier est la nécessité d'une double analyse des sports collectifs pour optimiser leur compréhension, discutée par les différentes études présentées. La dimension individuelle d'abord, permettant d'adapter les charges, le suivi médical, les entrainements de chaque joueur. Cette individualisation de la prise en charge des joueurs optimise leurs performances et la compréhension de leur état de santé minimise les risques de blessures. Puis la dimension collective à travers laquelle les impacts de la gestion du groupe sont mis en évidence. C'est cette gestion mise en place par l'entraineur une fois qu'il possède toutes les informations individuelles disponibles qui impactera la performance de l'équipe et influera sur l'évolution de l'état de forme des joueurs. C'est donc bien la double vision individuelle et collective qui est indispensable à la bonne lecture et interprétation des évènements relatifs aux performances du groupe et aux blessures qu'il subit.

Le second est la confirmation de l'importance de l'incorporation de biostatisticiens dans les processus de développement des problématiques soulevées par les *sport scientists*, les professionnels de la santé, les préparateurs physiques, les analystes vidéo, les entraineurs, les cadres fédéraux. Une place naturelle doit être faite au sein de ces staffs à des personnes capables d'analyser les données achetées et produites pour répondre aux questions que se posent l'ensemble du staff. Ces biostatisticiens doivent être capables de comprendre le contexte sportif et scientifique des interrogations, de les traduire en tests et modèles statistiques adaptés puis de retranscrire les résultats les résultats intelligiblement et clairement afin d'aider la prise de décision.

VI. Bibliographie

1. Millington B, Millington R. “The Datafication of Everything”: Toward a Sociology of Sport and Big Data. Sociol Sport J. 2015 Jun 1;32(2):140–60.

2. Rein R, Memmert D. Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. SpringerPlus. 2016 Aug 24;5(1):1410.

3. Antero J, Saulière G, Marck A, Toussaint J-F. A Medal in the Olympics Runs in the Family: A Cohort Study of Performance Heritability in the Games History. Front Physiol [Internet]. 2018 [cited 2018 Oct 17];9. Available from: https://www.frontiersin.org/articles/10.3389/fphys.2018.01313/full?&utm_source=Email_to_authors_&utm_medium=Email&utm_content=T1_11.5e1_author&utm_campaign=Email_publication&field=&journalName=Frontiers_in_Physiology&id=400135

4. Antero-Jacquemin J, Desgorces FD, Dor F, Sedeaud A, Haïda A, LeVan P, et al. Row for Your Life: A Century of Mortality Follow-Up of French Olympic Rowers. PLOS ONE. 2014 Nov 17;9(11):e113362.

5. Antero-Jacquemin J, Rey G, Marc A, Dor F, Haïda A, Marck A, et al. Mortality in Female and Male French Olympians: A 1948-2013 Cohort Study. Am J Sports Med. 2015 Jun 1;43(6):1505–12.

6. Antero-Jacquemin J da S, Berthelot G, Marck A, Noirez P, Latouche A, Toussaint J-F. Learning From Leaders: Life-span Trends in Olympians and Supercentenarians. J Gerontol Ser A. 2015 Aug 1;70(8):944–9.

7. Berthelot G, Len S, Hellard P, Tafflet M, Guillaume M, Vollmer J-C, et al. Exponential growth combined with exponential decline explains lifetime performance evolution in individual and human species. AGE. 2012 Aug;34(4):1001–9.

8. Berthelot G, Sedeaud A, Marck A, Antero-Jacquemin J, Schipman J, Saulière G, et al. Has Athletic Performance Reached its Peak? Sports Med Auckl NZ. 2015 Sep;45(9):1263–71.

9. Haïda A, Dor F, Guillaume M, Quinquis L, Marc A, Marquet L-A, et al. Environment and scheduling effects on sprint and middle distance running performances. PloS One. 2013;8(11):e79548.

10. Hermine O, Dine G, Genty V, Marquet L-A, Fumagalli G, Tafflet M, et al. Eighty percent of French sport winners in Olympic, World and Europeans competitions have mutations in the hemochromatosis HFE gene. Biochimie. 2015 Dec;119:1–5.

11. Marc A, Sedeaud A, Guillaume M, Rizk M, Schipman J, Antero-Jacquemin J, et al. Marathon progress: demography, morphology and environment. J Sports Sci. 2014;32(6):524–32.

12. Marck A, Antero J, Berthelot G, Saulière G, Jancovici J-M, Masson-Delmotte V, et al. Are We Reaching the Limits of Homo sapiens? Front Physiol [Internet]. 2017 Oct 24 [cited 2018 May 22];8. Available from: http://journal.frontiersin.org/article/10.3389/fphys.2017.00812/full

13. Marck A, Berthelot G, Foulonneau V, Marc A, Antero-Jacquemin J, Noirez P, et al. Age-Related Changes in Locomotor Performance Reveal a Similar Pattern for Caenorhabditis elegans, Mus domesticus, Canis familiaris, Equus caballus, and Homo sapiens. J Gerontol Ser A. 2017 Apr 1;72(4):455–63.

14. Marijon E, Tafflet M, Antero-Jacquemin J, El Helou N, Berthelot G, Celermajer DS, et al. Mortality of French participants in the Tour de France (1947–2012). Eur Heart J. 2013 Oct 21;34(40):3145–50.

15. Sedeaud A, Vidalin H, Tafflet M, Marc A, Toussaint J-F. Rugby morphologies: "bigger and taller", reflects an early directional selection. J Sports Med Phys Fitness. 2013 Apr;53(2):185–91.

16. Sedeaud A, Marc A, Schipman J, Tafflet M, Hager J-P, Toussaint J-F. How they won Rugby World Cup through height, mass and collective experience. Br J Sports Med. 2012 Jun;46(8):580–4.

17. Sedeaud A, Marc A, Marck A, Dor F, Schipman J, Dorsey M, et al. BMI, a performance parameter for speed improvement. PloS One. 2014;9(2):e90183.

18. Sedeaud A, Marc A, Schipman J, Schaal K, Danial M, Guillaume M, et al. Secular trend: morphology and performance. J Sports Sci. 2014;32(12):1146–54.

19. Bromley S, Drew M, Talpey S, McIntosh A, Finch C. Collecting Health and Exposure Data in Australian Olympic Combat Sports: Feasibility Study Utilizing an Electronic System. JMIR Hum Factors. 2018 Oct 9;5(4):e27.

20. Hecksteden A, Faude O, Meyer T, Donath L. How to Construct, Conduct and Analyze an Exercise Training Study? Front Physiol. 2018;9:1007.

21. Bishop D. An applied research model for the sport sciences. Sports Med Auckl NZ. 2008;38(3):253–63.

22. Aughey RJ, Buchheit M, Garvican-Lewis LA, Roach GD, Sargent C, Billaut F, et al. Yin and yang, or peas in a pod? Individual-sport versus team-sport athletes and altitude training. Br J Sports Med. 2013 Dec;47(18):1150–4.

23. Marc A, Sedeaud A, Schipman J, Antero JJ, Saulière G, Kryger KO, et al. Geographic enrollment of the top 100 in athletics running events from 1996 to 2012. J Sports Med Phys Fitness. 2015 Dec 2;

24. Marc A, Sedeaud A, Schipman J, Saulière G, Toussaint JF. Age and Performance from 10 Seconds to a 6-Days Race. J Athl Enhanc. 2018;7(4):1–5.

25. Araújo D, Davids K. Team Synergies in Sport: Theory and Measures. Front Psychol [Internet]. 2016 Sep 21 [cited 2018 May 1];7. Available from: http://journal.frontiersin.org/Article/10.3389/fpsyg.2016.01449/abstract

26. Duarte R, Araújo D, Correia V, Davids K, Marques P, Richardson MJ. Competing together: Assessing the dynamics of team–team and player–team synchrony in professional association football. Hum Mov Sci. 2013 Aug;32(4):555–66.

27. Duarte R, Araújo D, Freire L, Folgado H, Fernandes O, Davids K. Intra- and inter-group coordination patterns reveal collective behaviors of football players near the scoring zone. Hum Mov Sci. 2012 Dec;31(6):1639–51.

28. Memmert D, Lemmink KAPM, Sampaio J. Current Approaches to Tactical Performance Analyses in Soccer Using Position Data. Sports Med. 2017 Jan 1;47(1):1–10.

29. Kite CS, Nevill A. The Predictors and Determinants of Inter-Seasonal Success in a Professional Soccer Team. J Hum Kinet. 2017 Sep;58:157–67.

30. Tenga A, Holme I, Ronglan LT, Bahr R. Effect of playing tactics on goal scoring in Norwegian professional soccer. J Sports Sci. 2010 Feb;28(3):237–44.

31. Castellano J, Casamichana D, Lago C. The Use of Match Statistics that Discriminate Between Successful and Unsuccessful Soccer Teams. J Hum Kinet. 2012 Mar;31:139–47.

32. Lago-Peñas C, Dellal A. Ball Possession Strategies in Elite Soccer According to the Evolution of the Match-Score: the Influence of Situational Variables. J Hum Kinet. 2010 Sep 30;25(1):93–100.

33. Oberstone J. Differentiating the Top English Premier League Football Clubs from the Rest of the Pack: Identifying the Keys to Success. J Quant Anal Sports [Internet]. 2009 Jan 19;5(3). Available from: https://www.degruyter.com/view/j/jqas.2009.5.3/jqas.2009.5.3.1183/jqas.2009.5.3.1183.xml

34. Harrop K, Nevill A. Performance indicators that predict success in an English professional League One soccer team. Int J Perform Anal Sport. 2014 Dec 1;14(3):907–20.

35. Coutinho D, Gonçalves B, Travassos B, Abade E, Wong DP, Sampaio J. Effects of pitch spatial references on players' positioning and physical performances during football small-sided games. J Sports Sci. 2018 Oct 11;0(0):1–7.

36. Lacome M, Piscione J, Hager J-P, Carling C. Fluctuations in running and skill-related performance in elite rugby union match-play. Eur J Sport Sci. 2017 Feb 7;17(2):132–43.

37. Lacome M, Piscione J, Hager J-P, Bourdin M. A new approach to quantifying physical demand in rugby union. J Sports Sci. 2014 02;32(3):290–300.

38. Cunniffe B, Proctor W, Baker JS, Davies B. An evaluation of the physiological demands of elite rugby union using Global Positioning System tracking software. J Strength Cond Res Natl Strength Cond Assoc. 2009 07;23(4):1195–1203.

39. Pollard BT, Turner AN, Eager R, Cunningham DJ, Cook CJ, Hogben P, et al. The ball in play demands of international rugby union. J Sci Med Sport [Internet]. 2018 Mar [cited 2018 May 1]; Available from: http://linkinghub.elsevier.com/retrieve/pii/S1440244018300677

40. Hendricks S, Matthews B, Roode B, Lambert M. Tackler characteristics associated with tackle performance in rugby union. Eur J Sport Sci. 2014 11-17;14(8):753–762.

41. Hendricks S, Roode B, Matthews B, Lambert M. Defensive Strategies in Rugby Union. Percept Mot Skills. 2013 08-01;117(1):65–87.

42. Sewry N, Lambert M, Roode B, Matthews B, Hendricks S. The Relationship Between Playing Situation, Defence and Tackle Technique in Rugby Union. Int J Sports Sci Coach. 2015 12-01;10(6):1115–1128.

43. Walsh M, Young B, Hill B, Kittredge K, Horn T. The effect of ball-carrying technique and experience on sprinting in rugby union. J Sports Sci. 2007 Jan 15;25(2):185–92.

44. Robinson N, Sottas P-E, Schumacher YO. The Athlete Biological Passport: How to Personalize Anti-Doping Testing across an Athlete's Career? Acute Top Anti-Doping. 2017;62:107–18.

45. Lombardi G, Lanteri P, Colombini A, Lippi G, Banfi G. Stability of haematological parameters and its relevance on the athlete's biological passport model. Sports Med Auckl NZ. 2011 Dec 1;41(12):1033–42.

46. Lee EC, Fragala MS, Kavouras SA, Queen RM, Pryor JL, Casa DJ. Biomarkers in Sports and Exercise: Tracking Health, Performance, and Recovery in Athletes. J Strength Cond Res. 2017 Oct;31(10):2920–37.

47. Sottas P-E, Robinson N, Rabin O, Saugy M. The athlete biological passport. Clin Chem. 2011 Jul;57(7):969–76.

48. Robinson N, Sottas P-E, Mangin P, Saugy M. Bayesian detection of abnormal hematological values to introduce a no-start rule for heterogeneous populations of athletes. Haematologica. 2007 Aug;92(8):1143–4.

49. Sottas P-E, Baume N, Saudan C, Schweizer C, Kamber M, Saugy M. Bayesian detection of abnormal values in longitudinal biomarkers with an application to T/E ratio. Biostat Oxf Engl. 2007 Apr;8(2):285–96.

50. Baume N, Saudan C, Desmarchelier A, Strahm E, Sottas P-E, Bagutti C, et al. Use of isotope ratio mass spectrometry to detect doping with oral testosterone undecanoate: inter-individual variability of 13C/12C ratio. Steroids. 2006 May;71(5):364–70.

51. Sottas P-E, Robinson N, Giraud S, Taroni F, Kamber M, Mangin P, et al. Statistical Classification of Abnormal Blood Profiles in Athletes. Int J Biostat [Internet]. 2006 Jan 20 [cited 2018 Oct 17];2(1). Available from: https://www.degruyter.com/view/j/ijb.2006.2.1/ijb.2006.2.1.1011/ijb.2006.2.1.1011.xml

52. Banfi G. Limits and pitfalls of Athlete's Biological Passport. Clin Chem Lab Med. 2011 Sep;49(9):1417–21.

53. Malcovati L, Pascutto C, Cazzola M. Hematologic passport for athletes competing in endurance sports: a feasibility study. Haematologica. 2003 May;88(5):570–81.

54. Egger F, Meyer T, Hecksteden A. Interindividual Variation in the Relationship of Different Intensity Markers—A Challenge for Targeted Training Prescriptions. PLoS ONE [Internet]. 2016 Oct 27;11(10). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5082935/

55. Spencer M, Losnegard T, Hallén J, Hopkins WG. Variability and predictability of performance times of elite cross-country skiers. Int J Sports Physiol Perform. 2014 Jan;9(1):5–11.

56. Haugen TA, Solberg PA, Foster C, Morán-Navarro R, Breitschädel F, Hopkins WG. Peak Age and Performance Progression in World-Class Track-and-Field Athletes. Int J Sports Physiol Perform. 2018 Mar 15;1–24.

57. Batterham AM, Hopkins WG. Emergence of large treatment effects from small trials. JAMA. 2013 Feb 27;309(8):768.

58. Banfi G, Lombardi G, Colombini A, Lippi G. Analytical variability in sport hematology: its importance in an antidoping setting. Clin Chem Lab Med. 2011;49(5):779–782.

59. Bejder J, Hoffmann MF, Ashenden M, Nordsborg NB, Karstoft K, Mørkeberg J. Acute hyperhydration reduces athlete biological passport OFF-hr score. Scand J Med Sci Sports. 2016 Mar;26(3):338–47.

60. Custer EM, Finch CA, Sobel RE, Zettner A. Population norms for serum ferritin. J Lab Clin Med. 1995 Jul;126(1):88–94.

61. Damsgaard R, Munch T, Mørkeberg J, Mortensen SP, González-Alonso J. Effects of blood withdrawal and reinfusion on biomarkers of erythropoiesis in humans: Implications for anti-doping strategies. Haematologica. 2006 Jul;91(7):1006–8.

62. Eaton ML. Multivariate statistics: A vector space approach. JOHN WILEY SONS INC 605 THIRD AVE N Y NY 10158 USA 1983 512. 1983;

63. Embrechts P, Klüppelberg C, Mikosch T. Modelling extremal events: for insurance and finance. Vol. 33. Springer Science & Business Media; 2013.

64. Gasull A, López-Salcedo JA, Utzet F. Maxima of Gamma random variables and other Weibull-like distributions and the Lambert$\backslash$varvec ${$W$}$ function. Test. 2015;24(4):714–733.

65. Gasull A, Jolis M, Utzet F. On the norming constants for normal maxima. J Math Anal Appl. 2015;422(1):376–396.

66. Hall P. On the Rate of Convergence of Normal Extremes. J Appl Probab. 1979;16(2):433–9.

67. Hecksteden A, Pitsch W, Julian R, Pfeiffer M, Kellmann M, Ferrauti A, et al. A New Method to Individualize Monitoring of Muscle Recovery in Athletes. Int J Sports Physiol Perform. 2016 Dec 14;1–25.

68. Hecksteden A, Skorski S, Schwindling S, Hammes D, Pfeiffer M, Kellmann M, et al. Blood-borne markers of fatigue in competitive athletes–results from simulated training camps. PloS One. 2016;11(2):e0148810.

69. Heisterberg MF, Fahrenkrug J, Krustrup P, Storskov A, Kjær M, Andersen JL. Extensive monitoring through multiple blood samples in professional soccer players. J Strength Cond Res. 2013 May;27(5):1260–71.

70. Julian R, Meyer T, Fullagar HHK, Skorski S, Pfeiffer M, Kellmann M, et al. Individual Patterns in Blood-Borne Indicators of Fatigue-Trait or Chance. J Strength Cond Res. 2017 Mar;31(3):608–19.

71. Lobigs LM, Sottas P-E, Bourdon PC, Nikolovski Z, El-Gingo M, Varamenti E, et al. The use of biomarkers to describe plasma-, red cell-, and blood volume from a simple blood test. Am J Hematol. 2017 Jan 1;92(1):62–7.

72. Lombardi G, Ricci C, Banfi G. Effects of winter swimming on haematological parameters. Biochem Medica. 2011;71–8.

73. Malm CB, Khoo NS, Granlund I, Lindstedt E, Hult A. Autologous Doping with Cryopreserved Red Blood Cells - Effects on Physical Performance and Detection by Multivariate Statistics. PloS One. 2016;11(6):e0156157.

74. Meyer T, Meister S. Routine blood parameters in elite soccer players. Int J Sports Med. 2011 Nov;32(11):875–81.

75. Nikolaidis P, Ziv G, Lidor R, Arnon M. Inter-individual Variability in Soccer Players of Different Age Groups Playing Different Positions. J Hum Kinet. 2014 Apr 9;40:213–25.

76. Parisotto R, Gore CJ, Emslie KR, Ashenden MJ, Brugnara C, Howe C, et al. A novel method utilising markers of altered erythropoiesis for the detection of recombinant human erythropoietin abuse in athletes. Haematologica. 2000 Jun;85(6):564–72.

77. Pottgiesser T, Sottas P-E, Echteler T, Robinson N, Umhau M, Schumacher YO. Detection of autologous blood doping with adaptively evaluated biomarkers of doping: a longitudinal blinded study: ADAPTIVE MODEL AND BLOOD TRANSFUSION. Transfusion (Paris). 2011 Aug;51(8):1707–15.

78. Sallet P, Brunet-Guedj E, Mornex R, Baverel G. Study of a new indirect method based on absolute norms of variation to detect autologous blood transfusion. Int J Hematol. 2008 Nov 1;88(4):362.

79. Saugy M, Lundby C, Robinson N. Monitoring of biological markers indicative of doping: the athlete biological passport. Br J Sports Med. 2014 May 1;48(10):827–32.

80. Sharpe K, Ashenden MJ, Schumacher YO. A third generation approach to detect erythropoietin abuse in athletes. Haematologica. 2006 Mar;91(3):356–63.

81. Sharpe K, Hopkins W, Emslie KR, Howe C, Trout GJ, Kazlauskas R, et al. Development of reference ranges in elite athletes for markers of altered erythropoiesis. Haematologica. 2002 Dec;87(12):1248–57.

82. Silva JR, Rebelo A, Marques F, Pereira L, Seabra A, Ascensão A, et al. Biochemical impact of soccer: an analysis of hormonal, muscle damage, and redox markers during the season. Appl Physiol Nutr Metab Physiol Appl Nutr Metab. 2014 Apr;39(4):432–8.

83. Thirup P. Haematocrit. Sports Med. 2003;33(3):231–243.

84. Tufts DA, Haas JD, Beard JL, Spielvogel H. Distribution of hemoglobin and functional consequences of anemia in adult males at high altitude. Am J Clin Nutr. 1985 Jul;42(1):1–11.

85. Zorzoli M, Pipe A, Garnier PY, Vouillamoz M, Dvorak J. Practical experience with the implementation of an athlete's biological profile in athletics, cycling, football and swimming. Br J Sports Med. 2014 May;48(10):862–6.

86. Zorzoli M, Rossi F. Implementation of the biological passport: the experience of the International Cycling Union. Drug Test Anal. 2010 Dec;2(11–12):542–7.

87. Sécurité sociale 2016 [Internet]. Cour des comptes. [cited 2018 Oct 17]. Available from: https://www.ccomptes.fr/fr/publications/securite-sociale-2016

88. Pistilli B, Mazouni C, Zingarello A, Faron M, Saghatchian M, Grynberg M, et al. Individualized Prediction of Menses Recovery After Chemotherapy for Early-stage Breast Cancer: A Nomogram Developed From UNICANCER PACS04 and PACS05 Trials. Clin Breast Cancer. 2018 Aug 23;

89. Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive statistics for studies in sports medicine and exercise science. Med Sci Sports Exerc. 2009 01;41(1):3–13.

90. Hopkins WG, Batterham AM, Impellizzeri FM, Pyne DB, Rowlands DS. Statistical perspectives: all together NOT. J Physiol. 2011 Nov 1;589(Pt 21):5327-5329-5332.

91. Batterham AM, Hopkins WG. The case for magnitude-based inference. Med Sci Sports Exerc. 2015 Apr;47(4):885.

92. Hopkins WG, Batterham AM. Error Rates, Decisive Outcomes and Publication Bias with Several Inferential Methods. Sports Med Auckl NZ. 2016 Oct;46(10):1563–73.

93. Batterham AM, Hopkins WG. Making meaningful inferences about magnitudes. Int J Sports Physiol Perform. 2006 03;1(1):50–57.

94. Hopkins WG, Hawley JA, Burke LM. Design and analysis of research on sport performance enhancement. Med Sci Sports Exerc. 1999 Mar;31(3):472–85.

95. Sainani KL. The Problem with "Magnitude-based Inference": Med Sci Sports Exerc. 2018 Oct;50(10):2166–76.

96. Welsh AH, Knight EJ. "Magnitude-based inference": a statistical review. Med Sci Sports Exerc. 2015 Apr;47(4):874–84.

97. Barker RJ, Schofield MR. Inference about magnitudes of effects. Int J Sports Physiol Perform. 2008 Dec;3(4):547–57.

98. Montagna S, Hopker J. A Bayesian Approach for the Use of Athlete Performance Data Within Anti-doping. Front Physiol [Internet]. 2018 Jul 19;9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6060570/

99. Pla R, Leroy A, Massal R, Bellami M, Kaillani F, Hellard P, et al. A bayesian approach to quantify morphological impact on performance in international elite swimmingunder. (Under review at BJSM).

100. Church JB. Basic Statistics for the Strength and Conditioning Professional. Strength Cond J. 2008 Jun;30(3):51.

101. Baumgartner TA, Jackson AS. Measurement for Evaluation in Physical Education and Exercise Science. WCB/McGraw-Hill; 1999. 536 p.

102. Pettitt RW. The standard difference score: a new statistic for evaluating strength and conditioning programs. J Strength Cond Res. 2010 Jan;24(1):287–91.

103. Turner A, Brazier J, Bishop C, Chavda S, Cree J, Read P. Data Analysis for Strength and Conditioning Coaches: Using Excel to Analyze Reliability, Differences, and Relationships. Strength Cond J. 2015 Feb;37(1):76.

104. Flanagan EP. The Effect Size Statistic—Applications for the Strength and Conditioning Coach. Strength Cond J. 2013 Oct;35(5):37.

105. Bradley RC. Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions. Probab Surv. 2005;2(0):107–44.

106. Dede S. Théorèmes limites fonctionnels et estimation de la densité spectrale pour des suites stationnaires. [Internet] [phdthesis]. Université Pierre et Marie Curie - Paris VI; 2009 [cited 2018 Oct 18]. Available from: https://tel.archives-ouvertes.fr/tel-00440850/document

107. Dedecker J, Dehling HG, Taqqu M. Weak convergence of the empirical process of intermittent maps in L2 under long-range dependence. Stoch Dyn. 2015;15(2):29pp.

108. Dedecker J, Rio E. On mean central limit theorems for stationary sequences. Ann IHP Probab Stat. 2008;44(4):693–726.

109. Dedecker J, Prieur C. Some Unbounded Functions of Intermittent Maps for Which the Central Limit Theorem Holds [Internet]. 2007 [cited 2018 Oct 18]. Available from: /paper/Some-Unbounded-Functions-of-Intermittent-Maps-for-Dedecker-Prieur/6f5c5817663a47cddc8e7e061cd1722f1a1a0f0e

110. Dedecker J. An empirical central limit theorem for intermittent maps. Probab Theory Relat Fields. 2010 Sep 1;148(1–2):177–95.

111. Dedecker J, Prieur C. New dependence coefficients. Examples and applications to statistics. Probab Theory Relat Fields. 2005 Jun 1;132(2):203–36.

112. Dedecker J, Merlevède F. A deviation bound for α-dependent sequences with applications to intermittent maps. Stoch Dyn. 2017;17(1):27pp.

113. Dedecker J, Gouëzel S, Merlevède F. Some almost sure results for unbounded functions of intermittent maps and their associated Markov chains. Ann IHP Probab Stat. 2010;46(3):796–821.

114. Dedecker J, Prieur C. An empirical central limit theorem for dependent sequences. Stoch Process Their Appl. 2007;117:121–42.

115. Rosenblatt M. A Central Limit Theorem and a Strong Mixing Condition. Proc Natl Acad Sci U S A. 1956;42(1):43–7.

116. Rio E. Théorie asymptotique des processus aléatoires faiblement dépendants [Internet]. Berlin Heidelberg: Springer-Verlag; 2000 [cited 2018 Oct 18]. (Mathématiques et Applications). Available from: //www.springer.com/la/book/9783540659792

117. Liverani C, Saussol B, Vaienti S. A probabilistic approach to intermittency. Ergod Theory Dyn Syst. 1999 Jun;19(3):671–85.

118. Gouëzel S. Central limit theorem and stable laws for intermittent maps. Probab Theory Relat Fields. 2004 Jan 1;128(1):82–122.

119. Dewan I, Rao BLSP. Mann-Whitney test for associated sequences. Ann Inst Stat Math. 2003 Mar 1;55(1):111–9.

120. Dehling H, Rooch A, Taqqu MS. Non-Parametric Change-Point Tests for Long-Range Dependent Data. Scand J Stat. 2013 Mar 1;40(1):153–73.

121. Dehling H, Fried R. Asymptotic Distribution of Two-sample Empirical U-quantiles with Applications to Robust Tests for Shifts in Location. J Multivar Anal. 2012 Feb;105(1):124–140.

122. Serfling RJ. The Wilcoxon Two-sample Statistic on Strongly Mixing processes. Ann Math Stat. 1968 Aug;39(4):1202–9.

123. Sedeaud A, Sène JM, Krantz N, Saulière G, Moussa I, Toussaint JF. L'importance de la quantification de la charge d'entraînement : exemple d'un modèle. Sci Sports. 2018 Feb;33(1):22–32.

124. Moussa I, Saulière G, Sène JM, Toussaint JF, Sedeaud A. Training load, fatigue level and injury: a short report on French high-level judokas. (Under review).

125. Colombo C, Sedeaud A. L'individualisation. (INSEP).

126. Carling C, Le Gall F, McCall A, Nédélec M, Dupont G. Squad management, injury and match performance in a professional soccer team over a championship-winning season. Eur J Sport Sci. 2015;15(7):573–82.

127. Hägglund M, Waldén M, Magnusson H, Kristenson K, Bengtsson H, Ekstrand J. Injuries affect team performance negatively in professional football: an 11-year follow-up of the UEFA Champions League injury study. Br J Sports Med. 2013 Aug;47(12):738–42.

128. Windt J, Ekstrand J, Khan KM, McCall A, Zumbo BD. Does player unavailability affect football teams' match physical outputs? A two-season study of the UEFA champions league. J Sci Med Sport. 2018 May;21(5):525–32.

129. Parry L, Drust B. Is injury the major cause of elite soccer players being unavailable to train and play during the competitive season? Phys Ther Sport. 2006 May 1;7(2):58–64.

130. Bengtsson H, Ekstrand J, Hägglund M. Muscle injury rates in professional football increase with fixture congestion: an 11-year follow-up of the UEFA Champions League injury study. Br J Sports Med. 2013 Aug;47(12):743–7.

131. Olds T. The evolution of physique in male rugby union players in the twentieth century. J Sports Sci. 2001 Apr;19(4):253–62.

132. Mellalieu S, Trewartha G, Stokes K. Science and rugby union. J Sports Sci. 2008 06;26(8):791–794.

133. Duthie GM. A framework for the physical development of elite rugby union players. Int J Sports Physiol Perform. 2006 Mar;1(1):2–13.

134. Argus CK, Gill ND, Keogh JWL, Hopkins WG, Beaven CM. Changes in strength, power, and steroid hormones during a professional rugby union competition. J Strength Cond Res. 2009 Aug;23(5):1583–92.

135. Goncalo JA, Polman E, Maslach C. Can confidence come too soon? Collective efficacy, conflict and group performance over time. Organ Behav Hum Decis Process. 2010 Sep;113(1):13–24.

136. Greenlees IA, Graydon JK, Maynard IW. The impact of collective efficacy beliefs on effort and persistence in a group task. J Sports Sci. 1999 Feb;17(2):151–8.

137. Leo FM, Sanchez-Miguel PA, Sanchez-Oliva D, Amado D, Garcia-Calvo T. Analysis of Cohesion and Collective Efficacy Profiles for the Performance of Soccer Players. J Hum Kinet. 2013 12-31;39:221–229.

138. Whitfield J. Collaboration: Group theory. Nature. 2008 10-09;455(7214):720–723.

139. Watson CB, Chemers MM, Preiser N. Collective efficacy: A multilevel analysis. Pers Soc Psychol Bull. 2001;27(8):1057–1068.

140. Shearer DA. Collective efficacy at the Rugby World Cup 2015 – The role of imagery and observation. Eur J Sport Sci. 2015 Aug 18;15(6):530–5.

141. Stajkovic AD, Lee D, Nyberg AJ. Collective efficacy, group potency, and group performance: meta-analyses of their relationships, and test of a mediation model. J Appl Psychol. 2009 05;94(3):814–828.

142. Wheeler K, Sayers M. Contact Skills Predicting Tackle-Breaks in Rugby Union. Int J Sports Sci Coach. 2009 12-01;4(4):535–544.

143. Boughatts W. Exploring the Relationship between Cohesion and Collective Efficacy in Tunisian Sports Teams: Validation of the Tunisian Version of the Collective Efficacy Questionnaire for Collective Sports. Int J Sci Cult Sport. 2017 Jan 1;5(22):46–59.

144. Villepreux P. Formation au rugby de mouvement. Editions Cépaduès; 1993.

145. Gabbett TJ. Influence of fatigue on tackling technique in rugby league players. J Strength Cond Res Natl Strength Cond Assoc. 2008 03;22(2):625–632.

146. Fransen K, Decroos S, Vanbeselaere N, Vande Broek G, De Cuyper B, Vanroy J, et al. Is team confidence the key to success? The reciprocal relation between collective efficacy, team outcome confidence, and perceptions of team performance during soccer games. J Sports Sci. 2015 Feb 7;33(3):219–31.

147. Winter EM, Abt GA, Nevill AM. Metrics of meaningfulness as opposed to sleights of significance. J Sports Sci. 2014;32(10):901–902.

148. Palla G, Barabási A-L, Vicsek T. Quantifying social group evolution. Nature. 2007 Apr;446(7136):664–7.

149. Duch J, Waitzman JS, Amaral LAN. Quantifying the performance of individual players in a team activity. PloS One. 2010;5(6):e10937.

150. Myers ND, Payment CA, Feltz DL. Reciprocal Relationships Between Collective Efficacy and Team Performance in Women's Ice Hockey. Group Dyn Theory Res Pract. 2004;8(3):182.

151. Heuzé J-P, Raimbault N, Fontayne P. Relationships between cohesion, collective efficacy and performance in professional basketball teams: An examination of mediating effects. J Sports Sci. 2006 Jan;24(1):59–68.

152. Hendricks S, Till K, Brown JC, Jones B. Rugby union needs a contact skill-training programme. Br J Sports Med. 2016 11-07;

153. Gabbett TJ. Science of rugby league football: a review. J Sports Sci. 2005 09;23(9):961–976.

154. Gabbett T, Ryan P. Tackling technique, injury risk, and playing performance in high-performance collision sport athletes. Int J Sports Sci Coach. 2009;4(4):521–533.

155. Guimerà R, Uzzi B, Spiro J, Amaral LAN. Team assembly mechanisms determine collaboration network structure and team performance. Science. 2005 04-29;308(5722):697–702.

156. Bourbousson J, Poizat G, Saury J, Seve C. Team Coordination in Basketball: Description of the Cognitive Connections Among Teammates. J Appl Sport Psychol. 2010 Apr 30;22(2):150–66.

157. Lim B-C, Klein KJ. Team mental models and team performance: a field study of the effects of team mental model similarity and accuracy. J Organ Behav. 2006;27(4):403–418.

158. Wuchty S, Jones BF, Uzzi B. The increasing dominance of teams in production of knowledge. Science. 2007 05-18;316(5827):1036–1039.

159. Hendricks S, Lambert MI. Theoretical model describing the relationship between the number of tackles in which a player engages, tackle injury risk and tackle performance. J Sports Sci Med. 2014;13(3):715.

160. Eccles DW, Tenenbaum G. Why an expert team is more than a team of experts: A social-cognitive conceptualization of team coordination and communication in sport. J Sport Exerc Psychol. 2004;26(4):542–560.

161. Marcos FML, Miguel PAS, Oliva DS, Calvo TG. Interactive effects of team cohesion on perceived efficacy in semi-professional sport. J Sports Sci Med. 2010;9(2):320.

162. Sedeaud A, Saulière G, Marquet L-A, Del Vecchio S, Bar-Hen A, Toussaint J-F. Collective effectiveness in the XV de France: selections and time matter. Eur J Sport Sci. 2017 Jul;17(6):656–64.

163. Oliver D, Fienen MN. Importance of Teammate Fit: Frescoball Example. J Quant Anal Sports [Internet]. 2009 Jan 15 [cited 2018 May 1];5(1). Available from: https://www.degruyter.com/view/j/jqas.2009.5.1/jqas.2009.5.1.1098/jqas.2009.5.1.1098.xml

164. Leo FM, Sánchez-Miguel PA, Sánchez-Oliva D, Amado D, García-Calvo T. Analysis of Cohesion and Collective Efficacy Profiles for the Performance of Soccer Players. J Hum Kinet [Internet]. 2013 Jan 1 [cited 2018 May 1];39(1). Available from: http://www.degruyter.com/view/j/hukin.2013.39.issue-1/hukin-2013-0085/hukin-2013-0085.xml

165. Jain N. Build teams to run marathons, not sprints. In: 97 things every project manager should know: Collective wisdom from the experts. O'Reilly Media, Inc.; 2009. p. 96.

166. Kahane L, Shmanske S. Team roster turnover and attendance in major league baseball. Appl Econ. 1997 Apr;29(4):425–31.

167. Smith EE, Groetzinger JD. Do Fans Matter? The Effect of Attendance on the Outcomes of Major League Baseball Games. J Quant Anal Sports [Internet]. 2010 Jan 28 [cited 2018 May 1];6(1). Available from: https://www.degruyter.com/view/j/jqas.2010.6.1/jqas.2010.6.1.1192/jqas.2010.6.1.1192.xml

168. Hägglund M, Waldén M, Magnusson H, Kristenson K, Bengtsson H, Ekstrand J. Injuries affect team performance negatively in professional football: an 11-year follow-up of the UEFA Champions League injury study. Br J Sports Med. 2013 May 2;bjsports-2013-092215.

169. Quarrie KL, Hopkins WG. Changes in player characteristics and match activities in Bledisloe Cup rugby union from 1972 to 2004. J Sports Sci. 2007 Jun;25(8):895–903.

170. Observatoire médical du rugby. 45 préconisations pour la santé des joueurs. Fédération Française de Rugby; 2018 Mar.

171. Campbell PG, Peake JM, Minett GM. The Specificity of Rugby Union Training Sessions in Preparation for Match Demands. Int J Sports Physiol Perform. 2017 Sep 5;1–23.

172. Passos P, Milho J, Button C. Quantifying synergies in two-versus-one situations in team sports: An example from Rugby Union. Behav Res Methods. 2018 Apr;50(2):620–9.

173. Seifert L, Araújo D, Komar J, Davids K. Understanding constraints on sport performance from the complexity sciences paradigm: An ecological dynamics framework. Hum Mov Sci. 2017 Dec;56:178–80.

174. Groll A, Abedieh J. Spain retains its title and sets a new record – generalized linear mixed models on European football championships. J Quant Anal Sports. 2013 Mar 30;9(1):51–66.

175. Cintia P, Coscia M, Pappalardo L. The Haka network: Evaluating rugby team performance with dynamic graph analysis. In: Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on. IEEE; 2016. p. 1095–1102.

176. Grund TU. Network structure and team performance: The case of English Premier League soccer teams. Soc Netw. 2012 Oct;34(4):682–90.

177. Pina TJ, Paulo A, Araújo D. Network Characteristics of Successful Performance in Association Football. A Study on the UEFA Champions League. Front Psychol [Internet]. 2017 Jul 13 [cited 2018 May 1];8. Available from: http://journal.frontiersin.org/article/10.3389/fpsyg.2017.01173/full

178. Ribeiro J, Silva P, Duarte R, Davids K, Garganta J. Team Sports Performance Analysed Through the Lens of Social Network Theory: Implications for Research and Practice. Sports Med. 2017 Feb 15;1–8.

179. Guillaume J-L, Latapy M. Bipartite structure of all complex networks. Inf Process Lett. 2004;90(5):215–221.

université PARIS-SACLAY | ÉCOLE DOCTORALE Santé Publique

Titre : Analyse du rapport bénéfices risques et des déterminants de la performance dans les sports collectifs.

Mots clés : sport collectif, individualisation, blessures, performance, expérience collective

Résumé : L'investissement massif du monde du sport de haut niveau par les données impacte particulièrement les sports collectifs. Les performances réalisées par la pratique d'un sport collectif sont multifactorielles et sont étroitement liées à la survenue des blessures.
Cette complexité est reflétée par la composition hétérogène des staffs techniques et des équipes de recherche (entraineurs, préparateurs physiques, médecins, kinésithérapeutes, *sport scientists*,…) travaillant à la compréhension de ces performances. De la réflexion commune des membres de ces équipes pluridisciplinaires émergent des questions et des données sont produites pour y répondre. Dans ce contexte, les biostatistiques trouvent naturellement leur place à l'interface entre les données et les problématiques posées pour aider à la prise de décision.
Les données produites par les sports collectifs nécessitent d'être observées d'un point de vue du joueur et du groupe. Nous avons dans un premier temps montré comment l'individualisation du suivi du joueur pouvait participer à la compréhension de ses performances et à optimiser la lecture de son état de santé. Ensuite nous nous sommes intéressés à l'importance de la prise en compte de la gestion du groupe dans l'explication de sa performance et de la survenue de blessures en son sein.
Les études présentées dans cette thèse illustrent certaines étapes de l'échange nécessaire entre les différentes disciplines rassemblées autour des sciences du sport et montrent comment les biostatistiques s'inscrivent dans le processus de décision des staffs.

Title : Analysis of the risk / benefit ratio and the determinants of performance in team sports.

Keywords : team sport, individualization, injuries, performance, team experience

Summary: Data has massively invested the world high-level sport and team sport are particularly concerned. Performances from team sport are multifactorial and are closely linked to the occurrence of injuries. This complexity is reflected by the heterogeneous composition of technical staff and research team (coaches, physical trainers, doctors, physiotherapists, sports scientist,…) working to understand these performances. From the joint reflection of the members of these multidisciplinary teams emerge questions and data are produced to answer them. In this context, biostatistics naturally find their place in the interface between data and issues raised to help decision-making.
Data from team sports need to be observed from both player's and group's point of view. We first showed how the individualization of the player's follow-up could participate in the understanding of his performances and optimize the reading of his health. Then we were interested in the importance of taking into account the management of the group in the explanation of its performances and the occurrence of injuries within it.
Studies presented in this thesis illustrate some stapes of the necessary exchange between the different disciplines gathered around sports sciences and show how biostatistics fit into the decision-making process of the staffs.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France