Document generated on 08/02/2024 8:22 p.m.

Criminologie

Fraudes, violences et autres comportements déviants dans le sport professionnel et olympique

Opportunités et limites des sources ouvertes en ligne comme moyen de renseignement

Fraud, violence and other deviant behaviors in professional and Olympic sport

Opportunities and constraints of online open sources as channel for intelligence

Fraudes, violencias y otras conductas desviadas en el deporte profesional e olímpico

Oportunidades y limites de las fuentes abiertas en línea como medio de información

Stefano Caneppele, Ph. D., Giulia Cinaglia, Christina Sperrer and Fiona Langlois

Volume 52, Number 2, Fall 2019

La criminologie de l'information : état des lieux et perspectives

URI: https://id.erudit.org/iderudit/1065858ar
DOI: https://doi.org/10.7202/1065858ar

See table of contents

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0316-0041 (print)
1492-1367 (digital)

Explore this journal

Cite this article

Caneppele, S., Cinaglia, G., Sperrer, C. & Langlois, F. (2019). Fraudes, violences et autres comportements déviants dans le sport professionnel et olympique : opportunités et limites des sources ouvertes en ligne comme moyen de renseignement. *Criminologie*, *52*(2), 110–140. https://doi.org/10.7202/1065858ar

Article abstract

Online open source media are increasingly being used as tools for intelligence analysis. This contribution explores how they can be used in the field of deviance in professional and Olympic sport. Our study considered on the field fraud (doping, game manipulation, eligibility fraud) and off the field fraud (corruption), on the field and off the field violence (hooliganism and terrorism), and other detrimental conducts by sportsmen whether off or on the field of play. To explore the magnitude of these phenomena, we performed an online open source analysis in 2016 (in English) and identified 775 cases reported by online media, with frauds accounting for 85 % of all cases. Specifically, doping, followed by match-fixing are the most frequent cases. In total, 87 different countries were involved, but some sub regions, notably Eastern Europe (17.4 %), East Africa (16.1 %) and Australia and New Zealand (10.2 %)) concentrate most of the observed cases. In terms of temporal patterns, concentration of cases in August and November for fraud, and in June and July for violence may suggest a seasonality of deviant behaviors in sport. By looking at biases linked to language and types of source, further research may contribute to systematic monitoring of the presence of these phenomena around the world.

Tous droits réservés © Les Presses de l'Université de Montréal, 2019

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

https://apropos.erudit.org/en/users/policy-on-use/

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

https://www.erudit.org/en/

Fraudes, violences et autres comportements déviants dans le sport professionnel et olympique

Opportunités et limites des sources ouvertes en ligne comme moyen de renseignement

Stefano Caneppele, Ph. D.[1,2]
Professeur
École des sciences criminelles, Université de Lausanne
stefano.caneppele@unil.ch

Giulia Cinaglia
Candidate au doctorat
École des sciences criminelles, Université de Lausanne
giulia.cinaglia@unil.ch

Christina Sperrer
Consultante forensique
Deloitte GmbH
christinasperrer@gmx.at

Fiona Langlois
Candidate au doctorat
École des sciences criminelles, Université de Lausanne
fiona.langlois@unil.ch

RÉSUMÉ • *Les sources ouvertes en ligne sont de plus en plus utilisées comme outils de renseignement. Cette contribution explore la manière dont elles peuvent être utilisées pour étudier la déviance dans le sport professionnel et olympique. L'étude a considéré les fraudes*

1. École des sciences criminelles, Université de Lausanne, ESC – Sorge – BCH, CH – 1015 Lausanne, Suisse.

2. Les auteurs souhaitent remercier Mme Sara Aniello pour son aide précieuse dans le processus d'extraction des données d'*EMM Newsbrief*. Contribution des auteurs : conceptualisation, S. Caneppele, G. Cinaglia, F. Langlois et C. Sperrer ; méthodologie, S. Caneppele., G. Cinaglia et C. Sperrer ; récolte des données, C. Sperrer et S. Caneppele ; analyse formelle, S. Caneppele ; rédaction originale, F. Langlois, S. Caneppele et C. Sperrer ; rédaction, révision et correction, S. Caneppele, G. Cinaglia et F. Langlois ; supervision, S. Caneppele.

perpétrées sur le terrain (le dopage, la manipulation de match et les fraudes à l'éligibilité) et en dehors (la corruption), les violences sur et en dehors du terrain de jeu (le hooliganisme et le terrorisme), ainsi que les autres comportements préjudiciables commis par les sportifs, que ce soit sur le terrain ou en dehors. Un dispositif de veille en ligne a été mis en place pour la récolte d'articles publiés en 2016 en anglais. Des 775 cas qui ont été relevés, les fraudes (surtout les cas de dopage et de manipulations de matchs) comptaient à elles seules pour 85 % des cas. Au total, 87 pays sont impliqués dans cette étude, bien que certaines sous-régions – l'Europe de l'Est (17,4 %), l'Afrique de l'Est (16,1 %), ainsi que l'Australie et la Nouvelle-Zélande (10,2 %) – se démarquent par une concentration plus élevée de cas. Quant aux patterns temporels, des distributions sont concentrées surtout aux mois d'août et novembre pour les fraudes, et aux mois de juin et juillet pour les violences. Ces tendances peuvent laisser supposer une saisonnalité des comportements déviants. En se concentrant sur les biais liés au langage et aux types de sources, des recherches plus approfondies sur le sujet pourraient contribuer à une surveillance systématique de la représentation médiatique de ces phénomènes dans le monde.

MOTS CLÉS • *Dopage, fraudes sportives, violences, corruption, sport, sources ouvertes, renseignement.*

Introduction

« *Citius, altius, fortius* »[3]. La devise olympique, attribuée au père des Jeux olympiques modernes, Pierre de Coubertin, décrit bien la manière par laquelle la compétition sportive a été relancée à la fin du 19e siècle. L'expression latine évoquait à la fois la tradition helléniste – très en vogue à la fin des années 1800 –, la solennité d'un événement promouvant la compétition entre plusieurs pays, ainsi que le respect impératif des règles des Jeux par les athlètes. Cette idéalisation de l'olympisme a sûrement contribué à la création du mythe qui a permis au Comité international olympique (CIO) de prospérer pendant plusieurs décennies et d'ériger les manifestations à cinq anneaux en événements mondiaux. Cependant, l'idée du sport comme symbole d'autonomie, d'honnêteté, de santé et d'égalité (*fairness*) a été remise en question à plusieurs reprises. D'une part, les historiens affirment aujourd'hui que la déviance dans la compétition sportive n'est pas un phénomène récent (Gill, 2015). D'autre part, depuis les années 2000, de nombreux scandales ont émergé, touchant les disciplines les plus populaires, les fédérations et les manifestations sportives (par exemple le football et la FIFA [Fédération internationale de football association], l'athlétisme et l'IAAF [Association internationale des fédérations d'athlétisme], le tennis et l'ATP [Association of Tennis Professionals] ou les Jeux olym-

3. « Plus vite, plus haut, plus fort ».

piques et le CIO). Ces affaires ont généré une forte réaction au sein de l'opinion publique, entraînant chez les médias un intérêt marqué pour les comportements des athlètes ou des dirigeants sportifs, remettant en question cette image d'un sport pur (Rowe et Ohl, 2006). Par conséquent, les organisations sportives ont réalisé l'urgence de réagir contre une série de comportements potentiellement dangereux, notamment pour la réputation de la discipline (comportant une perte de sponsorisations, et donc, des fonds). Dernièrement, plusieurs institutions sportives ont essayé de résumer cette nécessité de conformité sous le concept d'intégrité du sport (Australian Sports Commission, 2016 ; Barbary, Dyer et Frandano, 2011 ; Godkin, 2013 ; McNamee, 2013 ; SportAccord, 2011 ; Treagus, Cover et Beasley 2011). D'un point de vue criminologique, cette résurgence de l'intérêt médiatique pour un phénomène mondial (c.-à-d. la déviance dans le domaine du sport[4]) offre la possibilité de tester les opportunités et les limites liées à l'utilisation des sources ouvertes comme outils de veille sur le phénomène de la déviance sportive. En effet, les sources ouvertes jouent un rôle informatif concernant les comportements déviants qui atteignent une visibilité publique grâce à leur caractère extraordinaire (Chibnall, 1977 ; Jewkes, 2004). Les médias deviennent dès lors intéressants car ces conduites, bien que très discutées mondialement, possèdent un tel niveau de spécificité qu'elles ne sont pas définies dans des catégories univoques de statistiques policières. Les médias permettent alors de pallier le chiffre noir des données officielles.

Par sources ouvertes sont notamment entendus les médias, les magazines, la radio, la télévision, autrement dit toute source de données accessible en tout temps et par tout individu. Cette démarche, applicable à un ensemble hétéroclite de sujets, est désormais utilisée sous le nom de « veille » pour la collecte et l'analyse de données relatives aux phénomènes les plus variés (Serrano, 2014). Des activités de recherche ont précédemment été entreprises, notamment sur le terrorisme (Choi, Ko, Kim et Kim, 2014), sur l'utilisation des sources ouvertes comme moyen d'analyse de crimes et de prévention de potentielles futures infractions (Jayaweera *et al.*, 2015) ou sur la vente de biens volés sur Internet (Aniello et Caneppele, 2018). En 2011, Gorse et Chadwick ont publié un rapport sur des cas de « corruption dans le sport », collectés, totalement ou en partie, sur des

4. Dans le cadre de cette étude, le terme « sport » englobe le sport professionnel et olympique.

sources ouvertes. Le concept de « corruption dans le sport » couvrait quatre typologies d'affaires : le dopage, les manipulations de matchs[5] en vue d'obtenir un avantage financier (par les paris sportifs) ou sportif (non-relégation à des divisions inférieures) et l'utilisation abusive d'informations privilégiées à des fins de paris. Dans cette étude, 2089 affaires avaient été détectées entre 2000 et 2010, dont 95,6 % impliquaient le dopage, 4,31 % concernaient les matchs truqués – pour des avantages sportifs (2,73 %) et financiers (1,58 %) – et 1,63 % portaient sur l'utilisation abusive d'informations privilégiées.

Les comportements déviants dans le monde du sport à intégrer dans le processus de veille sur Internet

Comparativement aux travaux de Gorse et Chadwick (2011), cette contribution vise à couvrir un plus large éventail de comportements déviants. La recherche, inspirée en partie de la littérature existante (voir par exemple les travaux de Barbary *et al*., 2011 ; Caneppele, Cinaglia et Langlois, 2019 ; Carpenter, 2012 ; Chappelet, 2017 ; Hemphill et Wilson-Evered, 2016 ; Moller, 2010), classe les comportements à surveiller en trois grandes catégories : la fraude, la violence et les autres comportements préjudiciables (Tableau 1). Chaque catégorie est soumise à une classification ultérieure. Des distinctions sont ainsi faites entre les comportements susceptibles d'influencer directement le résultat des compétitions sportives (sur le terrain) et les comportements hors du terrain. Ces derniers incluent toute action affectant l'organisation, la gestion ou le bon déroulement des événements sportifs, ainsi que la conduite attendue par les membres des institutions sportives.

Fraude. Cette catégorie inclut l'ensemble des comportements qui entraînent une altération des résultats sportifs ou qui affectent les décisions des institutions sportives. Lorsque l'on s'intéresse aux conduites déviantes dans le domaine sportif (discussions publiques ou recherches scientifiques), les deux phénomènes principalement abordés sont le dopage et les matchs truqués (Carpenter, 2012 ; Hemphill et Wilson-Evered, 2016 ; Moller, 2010). Le dopage est généralement défini comme l'utilisation de substances et de méthodes interdites susceptibles d'améliorer la performance sportive d'un athlète (comme les stéroïdes anabo-

5. Les termes « trucage de match », « manipulation de match » et « manipulation de compétition sportive » sont ici utilisés comme synonymes.

TABLEAU 1

Catégories des comportements sélectionnées pour la veille en ligne

Fraudes sur le terrain	**Fraudes hors du terrain**
FST1 : Moyens qui, biologiquement ou technologiquement, participent à l'amélioration des prestations sportives des athlètes (dopage biologique et dopage technologique)	**FHT1 :** Fraudes visant l'adjudication des événements sportifs, l'attribution des contrats liés à leur déroulement, ainsi que l'attribution des travaux de construction des infrastructures (corruption)
FST2 : Fraudes pour lesquelles les athlètes sont amenés à modifier leurs performances sportives en vue d'obtenir des avantages économiques ou sportifs (manipulations des matchs liées ou non aux paris sportifs)	**FHT2 :** Fraude dans l'organisation et la gestion des institutions sportives (notamment la corruption et la malversation des fonds)
FST3 : Fraudes autorisant les athlètes à prendre part aux compétitions, contrairement aux règles de participation applicables (fraude de scolarité, fraude à l'éligibilité)	
Violences sur le terrain	**Violences hors du terrain**
VST1 : Violences et comportements préjudiciables des joueurs, des entraîneurs, des arbitres ou du personnel, pendant le déroulement de matchs (le racisme, par exemple)	**VHT1 :** Violences des supporteurs lors des manifestations sportives (le hooliganisme, par exemple)
	VHT2 : Violences politiques lors des manifestations sportives (le terrorisme, par exemple)
Autres conduites préjudiciables sur le terrain	**Autres conduites préjudiciables en dehors du terrain**
ACPST1	**ACPHT1**

lisants ou la retransfusion sanguine) (Hemphill et Wilson-Evered, 2016). Moller (2010) ajoute à cette définition le dopage génétique, qui consiste à injecter du matériel génétique dans le muscle. L'Agence mondiale antidopage (AMA) fournit une définition plus large de ce qui constitue le dopage, ne se limitant pas aux résultats positifs d'une analyse biologique d'un athlète (World Anti-Doping Agency, 2017). Selon Moller (2010), outre le recours aux substances dopantes, l'amélioration des performances sportives peut également être liée à des technologies qu'il qualifie de « technologie illégale d'amélioration des performances » ou de « dopage technologique » (comme l'utilisation en natation de maillots de bain à faible frottement). Une autre fraude visant

à diriger les résultats sportifs est le trucage de match, défini comme un acte intentionnel ou une intrusion dans la compétition dans le but d'influencer le résultat final ou certains moments de la rencontre. Par conséquent, le déroulement ou l'issue de l'événement n'est plus méconnu. Cette pratique procure un avantage indu pour soi-même ou pour autrui (Council of Europe, 2014). On distingue deux types de matchs truqués : les matchs truqués dans un but purement sportif et ceux motivés par des paris sportifs (Gorse et Chadwick, 2010 ; Hill, 2015 ; Moller, 2010). Les manipulations de match motivées par le sport reposent sur une collusion, « comme des accords bilatéraux entre les parties impliquées » (Schulenkorf et Frawley, 2017, p. 66, notre traduction). Enfin, parmi les fraudes sur le terrain se trouve la fraude à l'éligibilité. Les athlètes concernés par ce type de conduite mentent délibérément sur leur sexe, leur âge, leur nationalité ou leur formation scolaire (Barbary *et al.*, 2011). Des fraudes à l'éligibilité liées à l'âge des athlètes sont souvent découvertes. En mentant délibérément sur son âge, un athlète va rivaliser avec des athlètes plus âgés ou plus jeunes que lui, disposant ainsi potentiellement d'un avantage physique. La naturalisation des athlètes étrangers est une autre problématique. En effet, bien que la remise d'un passeport à un sportif étranger ne soit pas une pratique illégale, elle est parfois considérée comme allant à l'encontre de l'esprit du sport (Brown, 2017). Enfin, la fraude à l'éligibilité peut aussi concerner la formation scolaire des athlètes étudiants. Cette problématique est principalement observée aux États-Unis. Typiquement, des athlètes de haut niveau inscrits dans les universités reçoivent des bourses pour pratiquer un sport, mais ne suivent pas de cours. Pour tromper le système, les participants changent les notes, soumettent des textes plagiés ou créent de fausses attestations d'examens (Dalton, 2016).

Quant aux fraudes commises en dehors du terrain, elles sont typiquement liées aux situations de corruption entre au moins deux individus. La corruption est définie par Brooks, Aleem et Button (2013) comme un acte consistant à « offrir, promettre, donner, accepter ou solliciter un avantage en tant qu'incitation à une action illégale, contraire à l'éthique ou à un abus de confiance » (p. 5, notre traduction). Elle peut notamment impliquer le versement et l'acceptation de pots-de-vin souvent liés à des cadeaux ou autres avantages, à l'attribution des événements sportifs, de droits de diffusion et à la nomination à des postes importants au sein d'organismes sportifs (Barbary *et al.*, 2011).

Violence. La violence de la part du public ou des amateurs de sport peut survenir aussi bien à l'intérieur qu'à l'extérieur du lieu de l'événement. Ce type de conduite est couramment appelé « hooliganisme » (Milojević, Simonović, Janković, Otašević et Turanjanin, 2013). Cette terminologie indique le « comportement destructeur d'individus et de groupes de personnes qui ne respectent pas les normes et les lois sociales et qui vandalisent leur environnement » (Milojević *et al.,* 2013, p. 14, notre traduction). Même si le football est fréquemment sujet au hooliganisme, d'autres sports comme le basketball, le handball, le rugby et le hockey sont également touchés (Milojević *et al.*, 2013). Sur le terrain, certains comportements violents des athlètes ou des membres du personnel, notamment par l'adoption de propos racistes, peuvent également susciter des tensions hors terrain.

Autres conduites préjudiciables. Avant, pendant et après une compétition, les athlètes représentent leur équipe et leur discipline sportive. Une mauvaise conduite, tant sur le terrain qu'en dehors, porte atteinte non seulement à leur propre réputation, mais également à celle de leur équipe, du sponsor, de la marque qu'ils représentent et du sport en général. L'exposition des athlètes sur les réseaux sociaux (en particulier les athlètes « vedettes ») amplifie cet effet (Han et Dodds, 2013 ; Sanderson et Browning, 2013). De nombreux cas de sportifs adoptant de mauvaises conduites ont été observés, qu'il s'agisse de consommation d'alcool ou de drogue, ou d'abus sexuels (Barbary *et al.,* 2011), affectant ainsi l'image de l'athlète et du sport qu'il représente. Généralement, pour réglementer leur comportement, les contrats des joueurs et les accords de sponsoring imposent des clauses morales, stipulant ce qui est autorisé et ce qui ne l'est pas (Jonson, Lynch et Adair, 2013).

Objectifs et questions de recherche

Cette contribution vise à comprendre dans quelle mesure une méthode de veille sur Internet est applicable pour l'analyse des fraudes, des violences et des autres conduites préjudiciables dans le sport. Plus précisément, l'étude – de type exploratoire – a été diligentée par la question suivante : quels types de fraudes, violences, et conduites préjudiciables sont rapportés par les médias en ligne ?

Pour cibler les analyses, cette question a été décomposée en quatre sous-questions. Des hypothèses ont été posées pour chacune d'entre elles.

1) D'après les sources à disposition, quelle est la fréquence d'apparition des comportements déviants retenus ? Est-ce qu'un comportement est rapporté dans les médias plus souvent que les autres ? L'hypothèse posée consiste à penser que les conduites sont rapportées dans les médias avec une certaine fréquence, et que certaines apparaissent plus régulièrement que d'autres.
2) Grâce aux sources à disposition, peut-on identifier les disciplines sportives les plus touchées par des cas de fraudes, de violences et d'autres conduites préjudiciables ? L'hypothèse est que les sports les plus populaires (football, athlétisme et tennis) sont les plus exposés médiatiquement. De ce fait, les conduites déviantes détectées dans ces disciplines sont plus à même d'être portées à la connaissance du public.
3) En considérant les sources à disposition, peut-on établir des patterns géographiques pour certains types de comportements ? L'hypothèse consiste à dire que les cas ne surviennent pas de manière aléatoire, mais suivent une distribution géographique, indiquant des spécificités régionales par rapport aux fraudes, aux violences et aux autres conduites préjudiciables.
4) À partir des sources à disposition, peut-on identifier des patterns temporels, en s'intéressant spécifiquement à la distribution mensuelle des cas ? L'hypothèse consiste à dire que les cas ne surviennent pas de manière aléatoire, mais suivent une distribution temporelle, indiquant une saisonnalité des fraudes, des violences et des autres conduites préjudiciables dans le monde du sport.

Méthodologie

Pour la définition d'un système de veille sur Internet concernant les fraudes, les violences et les autres comportements déviants dans le sport, il est nécessaire de faire une distinction entre les passages méthodologiques suivants : a) la collecte du corpus des articles en ligne pertinents ; b) la construction de la base de données ; et c) la définition de la stratégie d'analyse des données.

Collecte des articles. Pour la construction du corpus d'articles traitant des cas de fraudes, de violences et autres comportements déviants, il a été nécessaire de suivre cinq étapes (Tableau 2).

TABLEAU 2

Description des étapes suivies dans la collecte des articles en ligne

Étapes	Description de l'activité
1	Définition des typologies des fraudes, des violences et autres comportements déviants, objets de la veille
2	Sélection des mots clés associés à chaque typologie
3	Détermination de la plateforme agrégeant les sources ouvertes en ligne
4	Définition de la langue de recherche et de la période de veille
5	Collecte des articles par mots clés, par un système d'extraction automatique
6	Tri des données

Le premier passage a porté sur la définition des typologies de comportements faisant l'objet de la veille (étape 1). Sur la base de la littérature existante, cette recherche a déterminé trois macro-typologies (fraudes, violences et autres comportements préjudiciables) en opérant une distinction entre « comportements sur le terrain » – à savoir qui produisent un effet direct sur le résultat sportif – et « comportements hors du terrain » – c'est-à-dire les autres conduites (Tableau 1). Lors de la deuxième étape, des mots clés ont été sélectionnés pour chaque typologie. Ces termes ont été choisis sur la base des mots précédemment repérés dans les articles traitant des fraudes, des violences et autres comportements déviants dans le sport (Tableau 3).

L'agrégateur *The European Media Monitor (EMM) Newsbrief*[6] est la plateforme qui a été utilisée dans cette étude (étape 3). Développée par le Centre commun de recherche de la Commission européenne, cette plateforme emmagasine des sites d'information en 70 langues. Elle permet d'effectuer des recherches par mots clés par critère de langue(s), par pays et par période temporelle. En tant qu'agrégateur, *EMM Newsbrief* ne conserve pas les articles, mais renvoie le lecteur aux liens URL, qui correspondent aux critères de recherche insérés. À noter toutefois que certains articles peuvent ne plus être accessibles, soit parce que le contenu n'est plus disponible, soit parce que le lien est devenu inactif.

6. Accessible au lien : https://emm.newsbrief.eu/

TABLEAU 3

Mots clés utilisés pour l'extraction d'articles dans *EMM Newsbrief*

Fraudes	**FST1 :** Dopage biologique et dopage technologique	Exact : *doping* et au moins : sport Exact : *performance enhancing substances* et pas : *doping*
	FST2 : Manipulations de matchs	Exact : *match-fixing* Exact : *spot-fixing*
	FST3 : Fraude à l'éligibilité	Exact : *academic fraud* et au moins : *sport*, *athletes* Exact : *sex testing* et au moins : *sport*, *athletes* Exact : *age cheating* et au moins : *sport*, *athletes*, *football*
	FHT1, FHT2 : Corruption	Exact : *corruption in sports* Exact : *bribery* et au moins : *sport* Exact : *sports corruption* et au moins : *sport*, *athletes* et pas : *match-fixing*
Violences	**VST1 :** Violences sur le terrain	Exact : *racist behaviour* et au moins : *sport*, *athletes* Exact : *misbehaviour* et au moins : *sport*, *athletes* Exact : *conduct detrimental* et au moins : *sport*, *athletes*
	VHT1 : Violences des supporteurs	Exact : *fan violence* et au moins : *sport*, *football* Exact : *hooliganism* et au moins : *sport*, *athletes* et pas : *match-fixing*, *doping*
	VHT2 : Terrorisme	Exact : *terror attack* et : *sport event*
Autres conduites préjudiciables	**ACPST1 :** Autres conduites sur le terrain	Exact : *misbehaviour* et au moins : *sport*, *athletes* Exact : *conduct detrimental* et au moins : *sport*, *athletes*
	ACPHT : Autres conduites hors du terrain	

Dès que les mots clés ont été définis et l'agrégateur sélectionné, la langue de recherche et la période de veille ont été spécifiées (étape 4). Pour cette recherche, seule la langue anglaise a été retenue, en raison de sa nature de *lingua franca*. En effet, le but était de tester le potentiel du système de veille avec la langue la plus utilisée. Concernant la période de la veille, l'ensemble des articles publiés en 2016 ont été pris en considération, sachant que les affaires apparues avant cette date, mais relatées en 2016 ont également été incluses. Choisir une seule année s'explique par la volonté d'assurer l'efficacité et l'efficience de l'analyse des données collectées. Finalement, pour gérer les processus de requêtes

par mots clés, un script Python a été développé, afin de relever automatiquement les articles pertinents (étape 5). Pour chaque résultat, le script a extrait le titre de l'article, sa date de publication et son lien URL. Ces informations – collectées en 2017 – ont été automatiquement générées dans des fichiers.csv (un fichier par mot clé). Au final, 40 617 articles ont été collectés. Les fichiers bruts ont ensuite été triés (étape 6) et les doublons ont été supprimés. À la suite de cela, le corpus des données comprenait 22 666 articles. Seul 1,3 % (N = 295) de ces articles a été retenu. Les autres ont été exclus soit parce qu'ils traitaient d'un même sujet (57,4 %, N = 13 005), qu'ils n'étaient pas pertinents[7] (30,5 %, N = 6924), qu'ils abordaient un problème de manière trop générale, ne traitant pas de cas spécifiques (10,5 %, N = 2381), soit parce que le lien URL n'était plus actif ou le contenu n'était plus disponible (0,3 %, N = 61). Cette procédure de tri est illustrée au Tableau 4.

Construction de la base de données

Le corpus des articles retenus a été utilisé pour la construction d'une base de données sur les fraudes, les violences et les autres conduites préjudiciables sur le terrain et en dehors dans le sport. L'unité d'analyse est principalement la personne physique (par exemple un athlète) ou morale (par exemple une fédération). Néanmoins, pour les cas de violence en dehors du terrain, l'unité d'analyse est l'événement sportif concerné. De ce fait, il est possible d'obtenir plusieurs cas à partir d'un même article. En résumé, à partir des 295 articles retenus, 775 cas ont été inclus dans la base de données. Pour chaque cas, 12 variables ont été définies (Tableau 5).

Stratégie d'analyse des données

Cette étude adopte une approche quantitative, basée sur des analyses descriptives bivariées, avec deux composantes ultérieures, l'une géographique et l'autre temporelle. La stratégie d'analyse a été diligentée par les quatre sous-questions précédemment formulées.

Pour répondre à la sous-question 1, relative à la fréquence d'apparition des comportements déviants, des analyses bivariées ont été effectuées afin de comparer la distribution des phénomènes selon la « Macro-catégorie » (variable 4) et la « Catégorie » (variable 5).

7. Aussi définis comme « faux positifs ».

TABLEAU 4

Résultats du tri des articles extraits automatiquement, après la suppression des doublons (N = 22 666)

	Articles Retenus	Articles Exclus				Total
		Même sujet	*Non pertinent*	*Non spécifique*	*Lien inactif ou contenu inaccessible*	
Dopage	140	9 124	2 604	641	29	12 538
Proportion (%)[1]	*1,1 %*	*72,8 %*	*20,8 %*	*5,1 %*	*0,2 %*	*100,0 %*
Manipulations de matchs	61	3 002	531	1 573	28	5 195
Proportion (%)	*1,2 %*	*57,8 %*	*10,2 %*	*30,3 %*	*0,5 %*	*100,0 %*
Fraudes à l'éligibilité	19	102	38	7	4	170
Proportion (%)	*11,2 %*	*60,0 %*	*22,4 %*	*4,1 %*	*2,4 %*	*100,0 %*
Corruption	13	457	420	60	0	950
Proportion (%)	*1,4 %*	*48,1 %*	*44,2 %*	*6,3 %*	*0,0 %*	*100,0 %*
Violences sur le terrain et en dehors	51	293	3 305	91	0	3 740
Proportion (%)	*1,4 %*	*7,8 %*	*88,4 %*	*2,4 %*	*0,0 %*	*100,0 %*
Autres conduites préjudiciables sur le terrain et en dehors	11	27	26	9	0	73
Proportion (%)	*15,1 %*	*37,0 %*	*35,6 %*	*12,3 %*	*0,0 %*	*100,0 %*
Total	295	13 005	6 924	2 381	61	22 666
Proportion (%)	*1,3 %*	*57,4 %*	*30,5 %*	*10,5 %*	*0,3 %*	*100,0 %*

1. Les proportions correspondent au ratio (n/N) entre les articles retenus ou exclus (n) sur l'ensemble des articles collectés pour une catégorie donnée (dopage, manipulations de matchs, etc.) (N).

Quant à la sous-question 2, concernant les disciplines les plus touchées par les conduites préjudiciables, des analyses bivariées ont été réalisées entre les variables « Sport » (variable 9) et « Catégorie ».

En ce qui concerne la sous-question 3, portant sur la distribution géographique des cas, des analyses bivariées ont été conduites en croisant les variables 6 et 7 (respectivement « Continent » et « Sous-région continentale ») avec la variable 5. Cette analyse a conduit à la réalisation d'un diagramme alluvial, permettant ainsi de visualiser le flux de distribution géographique par catégories. Une visualisation globale de l'ensemble des comportements recensés dans la base de

TABLEAU 5

Présentation des variables retenues pour la construction de la base de données

Variable	Nom	Type de variable	Modalités
1	ID	Numérique	Entre 1 et 775
2	Date	Date	Entre le 01.01.2016 et le 31.12.2016
3	Lien du site	Texte	www
4	Macro-catégorie	Texte	Fraudes / violences / autres conduites préjudiciables
5	Catégorie	Texte	Dopage / Manipulations de matchs / Fraude à l'éligibilité / Corruption / Violence sur le terrain et en dehors (violence des supporteurs et terrorisme) / Autres conduites préjudiciables sur le terrain et en dehors
6	Continent	Texte	Afrique / Amérique / Asie / Europe / Océanie
7	Sous-région continentale	Texte	Selon les standards M49 des Nations unies https://unstats.un.org/unsd/methodology/m49/
8	Pays	Texte	Tous les pays du monde
9	Sport	Texte	Types de sports
10	Type d'unité d'analyse	Texte	Personne physique / Personne morale/ Événement sportif
11	Rôle/Position	Texte	Par exemple : athlète, entraîneur, arbitre, dirigeant
12	Commentaires	Texte	Champ optionnel pour insérer des commentaires sur le type de conduite, l'éventuelle sanction et l'institution sportive qui a décidé la sanction

données a également été réalisée. Afin de standardiser les résultats, le nombre de cas par pays a été normalisé par le nombre d'athlètes olympiques ayant participé aux Jeux olympiques de Rio (2016). Cet indicateur a été choisi en postulant que la proportion d'athlètes représentant un pays donné aux Jeux olympiques est représentative de la proportion d'athlètes professionnels au sein de ce même pays. Dans cette optique, les taux ont tout d'abord été normalisés sur une échelle allant de 0 à 100, puis les valeurs ont été transformées selon une échelle logarithmique (log-10) pour harmoniser la distribution.

En regard de la sous-question 4, relative à la distribution d'apparition mensuelle des cas, deux analyses bivariées ont été effectuées. L'une associant les variables « Date » (variable 2) et « Macro-catégorie » (variable 4), et l'autre, plus spécifique, associant les variables 2 et 5, respectivement « Date » et « Catégorie ».

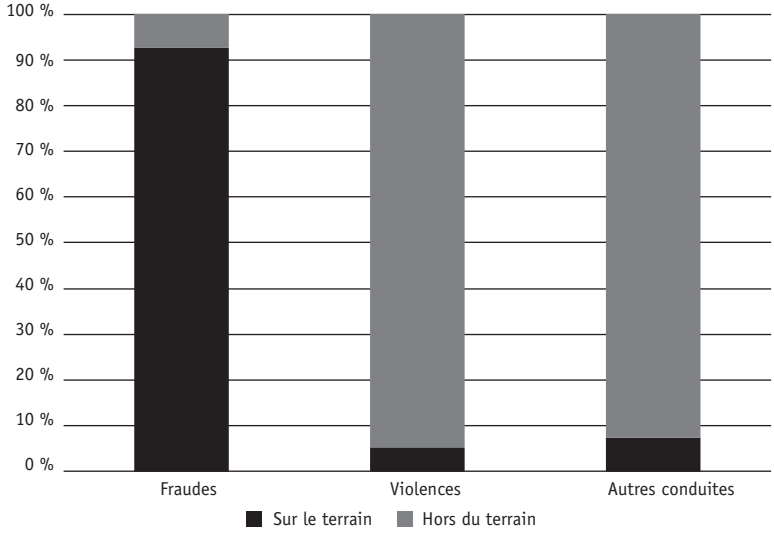
Résultats

Fréquence d'apparition des cas de comportements déviants dans les médias et popularité des disciplines touchées

D'après les analyses, il ressort que la macro-catégorie « Fraudes » regroupe 85 % des cas. Contrairement aux deux autres typologies de conduite, les fraudes se manifestent principalement sur le terrain (Figure 1). Les cas de violence, quant à eux, concernent majoritairement des actes de hooliganisme perpétrés par des supporteurs, le football étant la discipline la plus touchée. Enfin, la participation des athlètes à des paris sportifs (pratique habituellement interdite par plusieurs fédérations) constitue le comportement principalement observé dans la macro-catégorie « Autres conduites préjudiciables ».

FIGURE 1

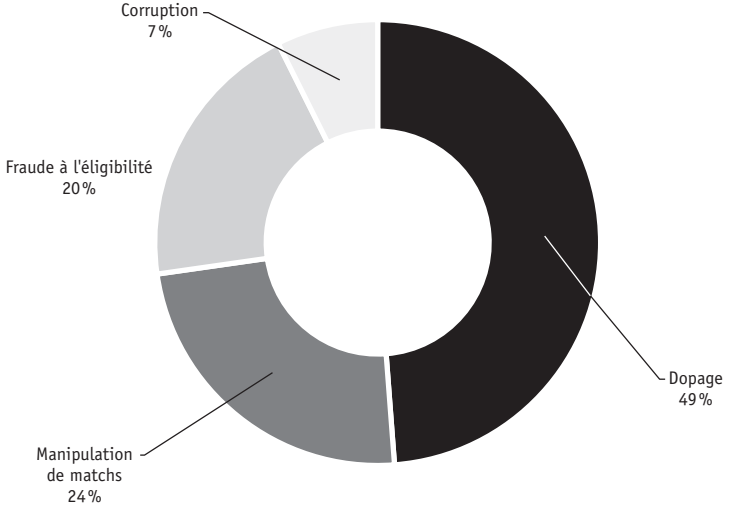
Distribution des conduites adoptées sur le terrain et en dehors. Fraudes (N = 664), violences (N = 57) et autres conduites préjudiciables (N = 54)



En analysant plus en détail les typologies des fraudes, près de la moitié des cas concerne le dopage (49 %), suivi des cas de manipulations de matchs (24 %), de fraude à l'éligibilité (20 %) et de corruption (7 %) (*cf.* Figure 2). Ces résultats semblent donc appuyer la première hypothèse posée.

FIGURE 2

Cas de fraudes par typologie (N = 664)

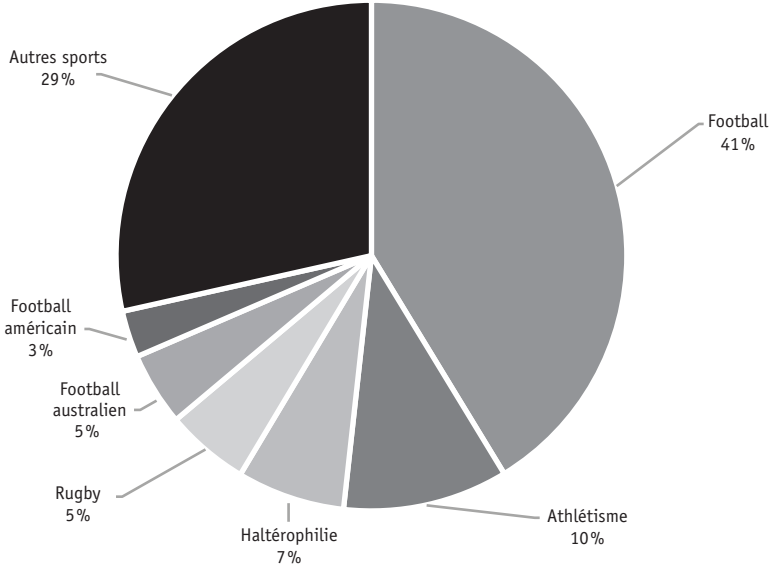


Dans l'ensemble, 37 disciplines sportives sont concernées par les cas contenus dans la base de données. Comme l'illustre la Figure 3, globalement le football est le sport qui regroupe la majorité des cas (41 %), suivi de l'athlétisme (10 %), de l'haltérophilie (7 %), du rugby et du football australien (5 %), et du football américain (3 %). Les résultats semblent donc corroborer la deuxième hypothèse posée. Cependant, des variations sont observées selon les zones géographiques étudiées (Tableau 6), susceptibles de refléter la popularité d'un sport dans une région donnée. Par exemple, en Afrique, le football, l'athlétisme et le rugby concentrent 91,7 % des cas recensés sur ce continent. En Amérique, les trois disciplines principalement touchées sont le football américain, le football, tandis que la troisième catégorie correspond principalement aux cas liés à l'organisation des Jeux de Rio. Ensemble, ils constituent 58,4 % des cas recensés sur ce continent. En Asie, le football, l'haltérophilie (diffusée surtout dans les pays asiatiques de l'ex-bloc soviétique)

et le baseball représentent 64,2 % des cas. En Europe, 58,4 % des cas concernent le football, l'athlétisme et l'haltérophilie. Finalement, en Océanie, c'est le football australien, suivi du rugby et du football qui rassemblent 87,2 % des cas. En général, sur l'ensemble des continents, les cas liés au football sont largement prédominants, exception faite pour le dopage, où l'athlétisme, l'haltérophilie et le football australien sont responsables de près de la moitié des cas recensés (48,5 %, N = 157).

FIGURE 3

Répartition en pourcentage par discipline sportive des fraudes, violences et autres comportements préjudiciables (N = 775)



Distributions géographique et temporelle des cas recensés dans les médias

Concernant la distribution géographique, les résultats indiquent que 87 pays sont touchés par des cas de conduite préjudiciable dans le sport. Les analyses ont ensuite été détaillées selon les sous-régions concernées (Tableau 7 et Figure 4). Il apparaît ainsi qu'une grande partie des cas détectés sont concentrés en Europe de l'Est (17,4 %), en Afrique de l'Est (16,1 %) et dans la région Australie–Nouvelle-Zélande (10,2 %), corroborant de ce fait la troisième sous-hypothèse de cette étude.

TABLEAU 6

Répartition des cas recensés par continent. Mise en exergue des trois disciplines les plus touchées selon la région géographique concernée

Continent	Nombre total de cas	Nombre de disciplines concernées	Principales disciplines touchées	Proportion de cas touchant les trois disciplines principalement visées [%]
Afrique	217	11	Football Athlétisme Rugby	91,7
Amérique	89	16	Football américain Football Jeux olympiques	58,4
Asie	120	15	Football Haltérophilie Baseball	64,2
Europe	255	23	Football Athlétisme Haltérophilie	58,4
Océanie	94	9	Football australien Rugby Football	87,2
Monde	775	37	Football Athlétisme Haltérophilie	58,6

Comme l'illustre la Figure 1, les fraudes perpétrées sur le terrain (dopage, manipulations des matchs et fraudes à l'éligibilité) représentent les comportements déviants les plus récurrents dans la base de données. De ce fait, seule cette macro-catégorie a été considérée lors de l'analyse de la distribution géographique des cas. D'après les résultats, il semblerait qu'il existe des patterns géographiques liés aux catégories de comportements, comme le démontre le diagramme alluvial présenté à la Figure 4. Par exemple, les manipulations de matchs semblent toucher principalement l'Asie, l'Afrique et l'Europe. De même, les cas de fraudes à l'éligibilité sont concentrés en Afrique, et plus spécifiquement en Afrique de l'Est. Enfin, les cas de dopage ont été majoritairement observés en Europe de l'Est.

TABLEAU 7

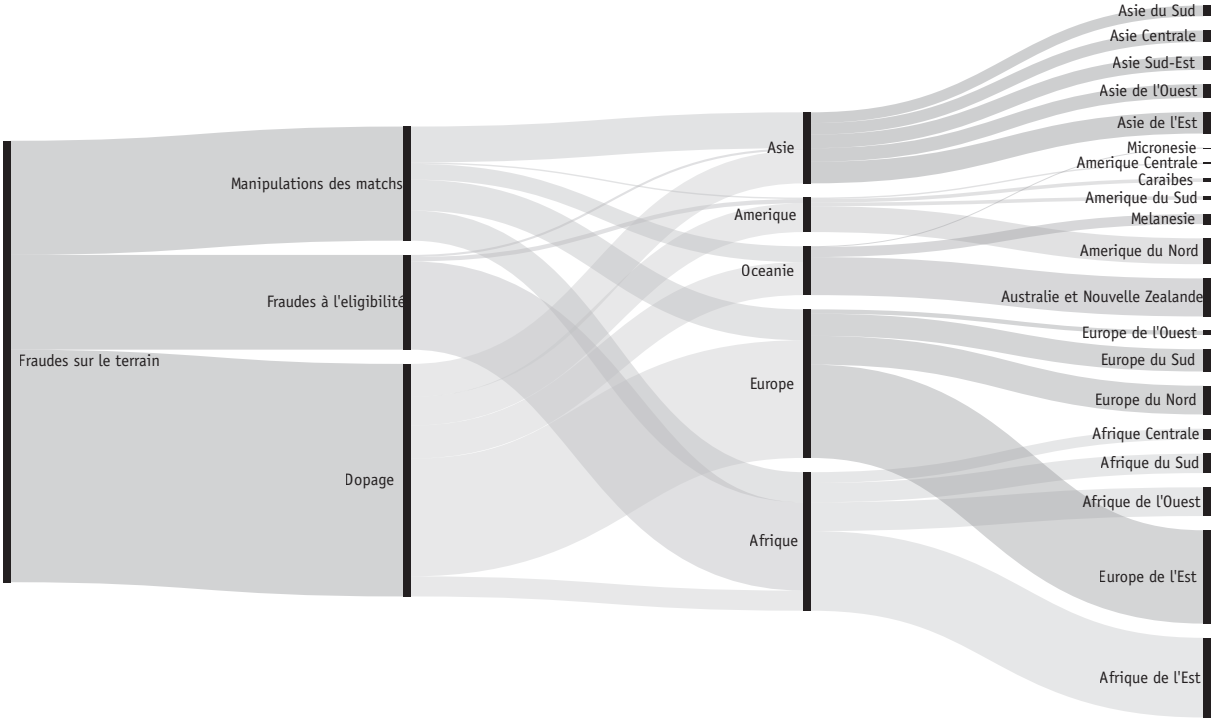
Répartition des cas (N = 775) par continent et par sous-région, prévalence sur le total des cas concernés

Continent	Sous-région	Nombre de cas détectés	Proportion de cas détectés par sous-région [%]
Afrique	Afrique centrale	15	1,9
Afrique	Afrique de l'Est	125	16,1
Afrique	Afrique de l'Ouest	44	5,7
Afrique	Afrique du Nord	4	0,5
Afrique	Afrique du Sud	29	3,7
Amérique	Amérique centrale	2	0,3
Amérique	Amérique du Nord	51	6,6
Amérique	Amérique du Sud	26	3,4
Amérique	Caraïbes	10	1,3
Asie	Asie centrale	16	2,1
Asie	Asie de l'Est	38	4,9
Asie	Asie de l'Ouest	24	3,1
Asie	Asie du Sud-Est	24	3,1
Asie	Asie du Sud	18	2,3
Europe	Europe de l'Est	135	17,4
Europe	Europe du Nord	62	8,0
Europe	Europe du Sud	38	4,9
Europe	Europe de l'Ouest	20	2,6
Océanie	Australie et Nouvelle-Zélande	79	10,2
Océanie	Mélanésie	14	1,8
Océanie	Micronésie	1	0,1

De plus, les analyses réalisées ont permis de produire une carte mondiale sur la distribution des cas retenus dans la base de données (Figure 5). Il en ressort que la majorité des cas sont situés au Kenya, en Russie, en Australie, aux États-Unis, au Royaume-Uni, au Nigéria et en Afrique du Sud. Les résultats, normalisés selon la méthodologie précédemment explicitée, montrent qu'il existe des concentrations de

FIGURE 4

Diagramme alluvial représentant les fraudes sur le terrain selon le continent et la sous-région concernée (N = 664)



cas dans des régions spécifiques, tels que le cluster des pays de l'ex-bloc soviétique et celui des pays d'Afrique de l'Est et du Sud, dont le Kenya, l'Afrique du Sud, l'Ouganda et la Tanzanie font partie.

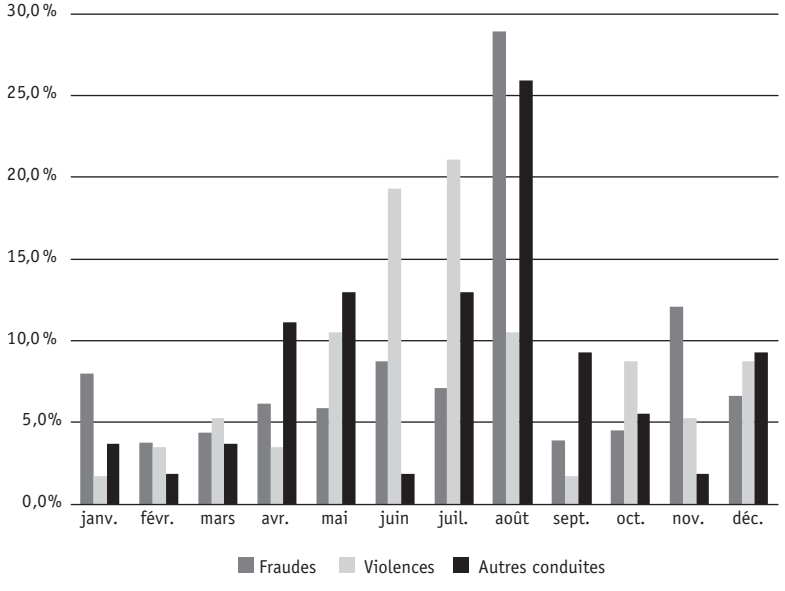
FIGURE 5

Carte mondiale représentant les taux de fraudes, de violences et autres comportements préjudiciables apparus dans les médias en 2016, en fonction du pays concerné. Taux normalisés par nombre d'athlètes olympiques (valeur logarithmique)

Quant à la dimension temporelle, des distributions de cas concentrés aux mois d'août et de novembre concernant les fraudes, aux mois de juin et de juillet pour les violences, et au mois d'août pour les autres conduites préjudiciables permettent d'établir l'existence d'un pattern mensuel (Figure 6). Cette observation semble confirmer l'hypothèse relative à la dernière sous-question.

FIGURE 6

Distribution mensuelle des cas de fraudes (N = 664), de violences (N = 57) et autres comportements préjudiciables (N = 54), en pourcentage



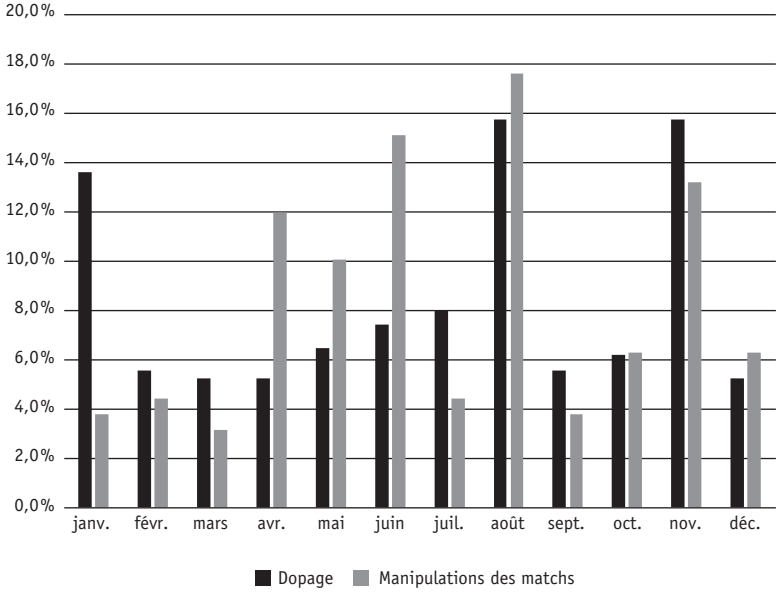
De plus, deux tendances distinctes se dégagent lorsqu'on analyse de plus près la distribution mensuelle des deux fraudes sur le terrain les plus fréquemment détectées. Les mois de janvier, d'août et de novembre se démarquent pour une prépondérance de cas de dopage, alors que les manipulations de matchs ont été plus fréquemment rapportées aux mois d'avril, de juin, d'août et de novembre.

Discussion

Cette étude s'est concentrée sur les fraudes, les violences et les autres comportements déviants dans le sport professionnel et olympique. La stratégie de collecte de données s'est basée sur un processus de veille à partir de sources ouvertes accessibles en ligne (c.-à-d. le recensement automatique des articles pertinents sur l'année 2016). Le principal enjeu de ce travail portait sur le tri des données, et, plus spécifiquement, sur l'élimination des doublons et sur l'identification et la gestion des faux positifs (c.-à-d. tout article contenant au moins un des mots clés, mais

FIGURE 7

Distribution mensuelle des cas de dopage (N = 324) et de manipulations de matchs (N = 159), en pourcentage



non pertinent pour la recherche). En effet, à la suite du processus de tri, seul 1,3 % des articles extraits a été retenu pour les analyses (soit 295 articles sur 22 666, à partir desquels 775 cas ont été dégagés). Les deux phénomènes recensant le moins de faux positifs sont les fraudes à l'éligibilité et les autres comportements préjudiciables (respectivement, 11,2 % et 15,1 % des articles extraits finalement retenus). La spécificité des comportements associés à ces catégories (par exemple la fraude de scolarité) et le peu d'intérêt médiatique qu'ils suscitent peuvent expliquer ces résultats.

À l'inverse, les mots « corruption » et « dopage » sont plus fréquemment utilisés dans le langage courant, ce qui génère une proportion plus importante de faux positifs. Associer différemment les mots clés pourrait contribuer à l'amélioration du processus, assurant alors une meilleure pertinence des articles extraits.

Cette recherche visait à déterminer s'il existe dans le sport des conduites déviantes plus fréquemment abordées dans les médias, et si

des patterns (géographiques et temporels) peuvent être dégagés. Parmi les macro-catégories considérées, les analyses révèlent que les fraudes sportives sont les sujets suscitant le plus grand intérêt médiatique. En effet, le dopage, les manipulations de matchs et la fraude à l'éligibilité concernaient 664 cas. Le football est le sport principalement touché par les trucages de matchs et par les fraudes à l'éligibilité, tandis que l'athlétisme et l'haltérophilie sont les disciplines les plus visées par le dopage. Ces derniers résultats concordent avec les statistiques officielles (World Anti-Doping Agency, 2017) et avec les recherches scientifiques existantes. En effet, Alaranta *et al.* (2006) expliquent que les sports d'endurance et de puissance sont généralement les plus touchés par le fléau du dopage. Néanmoins, il est important de mentionner que 2016 a été une année particulièrement prolifique en matière de fraudes sportives sur le terrain, avec deux grands scandales. Le premier, mis en évidence dans le rapport McLaren (2016), concernait la possible existence d'un système de dopage étatique organisé en Russie. Le second, quant à lui, portait sur la suspicion de matchs de tennis truqués, ayant conduit à la nomination d'une commission indépendante par les autorités régulant le tennis mondial[8]. Ces deux scandales ont sans doute contribué au grand nombre d'articles collectés sur ces thématiques. D'autres événements ont eu également un impact sur le corpus d'articles extraits, comme le dispositif mis en place par le CIO en 2016 pour retester les échantillons des athlètes ayant participé aux Jeux olympiques de 2008, de 2012 et de 2014 (CIO, 2016). Ce processus a conduit à plusieurs disqualifications, fortement discutées dans les médias. Un mécanisme similaire a été observé pour les matchs truqués, grâce à la forte pression exercée sur les institutions sportives pour agir contre cette problématique émergente et à l'intérêt médiatique renforcé par les récents scandales. Enfin, la question de la corruption au sein des institutions sportives de football a fait l'objet d'une enquête du Département de la Justice des États-Unis en 2015, dont les conséquences ont été ressenties sur plusieurs années (U.S. Department of Justice, 2015). Toutes ces situations peuvent donc expliquer pourquoi le dopage, les manipulations de matchs et la corruption sont parmi les comportements le plus fréquemment observés. La fraude à l'éligibilité, quant à elle, semble être un comportement visible à échelle régionale. Par exemple, la fraude liée à l'âge est principalement observée en Afrique, où les pays

8. *Cf.* http://tennisirp.com/

ne disposent pas nécessairement de bases de données fiables permettant de vérifier l'âge réel des individus, une lacune facilitant l'apparition de ce type de fraude (Cryer, 2014). De même, dans cette étude, la fraude de scolarité a été observée seulement aux États-Unis, favorisée par le fonctionnement du système universitaire local. L'année 2016 a également connu deux grands événements sportifs : le Championnat européen de football (organisé en France) et les Jeux olympiques d'été à Rio. Le premier a été source de comportements violents de la part des supporteurs, tandis que lors du deuxième, certains comportements préjudiciables d'athlètes ont été rapportés. Le cas le plus emblématique concerne celui des nageurs américains, qui avaient prétendu avoir été braqués par de faux policiers à Rio (Agence France-Presse, 2016).

En s'intéressant à la répartition de cas par discipline, les résultats corroborent l'hypothèse que les sports les plus populaires, notamment le football, sont aussi ceux qui suscitent le plus d'intérêt (aussi bien chez le public que dans les médias). Ceci est d'autant plus vrai lorsque des comportements déviants apparaissent dans ces sports. Cependant, il est intéressant de constater que les disciplines les plus touchées semblent varier entre les continents, selon les traditions sportives locales (par exemple, en Amérique le football américain et le football sont les disciplines les plus visées, alors qu'en Australie il s'agit plutôt du football australien et du rugby). L'existence de patterns continentaux et sous-régionaux a été confirmée par le diagramme alluvial. Ce dernier illustre notamment l'impact qu'a eu le scandale de dopage en Russie, indiquant l'Europe de l'Est comme étant la sous-région européenne la plus touchée par ce type de comportement. Pareillement, l'Afrique de l'Est ressort comme étant la sous-région la plus exposée à la fraude à l'éligibilité.

En outre, la présente étude propose une vue d'ensemble de la répartition des fraudes, des violences et des autres comportements préjudiciables à l'échelle mondiale. Cette carte permet notamment de présenter deux zones transrégionales (l'ex-bloc soviétique et la partie de l'Afrique du Sud et de l'Est) comme les plus exposées aux conduites déviantes dans le sport. Normaliser le nombre de cas par le nombre d'athlètes ayant participé aux Jeux olympiques d'été 2016 se veut une tentative de standardisation des valeurs absolues pour harmoniser les résultats. À noter toutefois que la détection de cas dans un certain pays n'est pas directement corrélée à l'occurrence effective des cas. En effet, l'efficacité des moyens de lutte mis en place par certains pays influencera

directement le type et la quantité de cas détectés, les exposant davantage médiatiquement.

Concernant les distributions temporelles, une hypothèse peut être posée quant à l'existence de cycles de déviance sportive. Par exemple, les matchs truqués sont principalement observés lors des fins de saisons sportives en Europe (avril-juin) et en août, période durant laquelle se disputent des matchs amicaux. Les violences, quant à elles, semblent liées aux grandes manifestations sportives de football. Enfin, les comportements préjudiciables semblent plutôt associés aux Jeux olympiques, ce qui peut s'expliquer par le rassemblement d'un grand nombre d'athlètes dans un même lieu, pendant plusieurs semaines. À noter que la détection de cas dans un certain pays n'est pas directement corrélée à l'occurrence effective des cas. En effet, l'efficacité des moyens de lutte mis en place par certains pays influencera directement le type et la quantité de cas détectés, les exposant davantage médiatiquement.

Les résultats de cette étude indiquent que plusieurs disciplines sportives sont touchées par le dopage, alors que le nombre est plus faible concernant le trucage de matchs. Ils semblent donc concorder avec les conclusions de Gorse et Chadwick (2011), estimant que les affaires de dopage sont plus courantes que les affaires de trucage de matchs. Deux explications sont possibles : le dopage serait réellement plus fréquent que le trucage de matchs, ou bien les méthodes sont plus efficaces pour détecter des cas de dopage. En effet, les athlètes dopés sont principalement identifiés par des tests antidopage et, pour seulement une petite partie des cas, des enquêtes sont nécessaires. En revanche, dans les cas de manipulations de matchs motivés par des paris sportifs, les alertes de paris suspects ne suffisent pas à elles seules à démontrer qu'un trucage de match a effectivement eu lieu. Des enquêtes doivent donc être menées et peuvent s'étendre sur une longue période temporelle, sans pour autant garantir la résolution de l'ensemble des cas. De plus, à ce jour, le système d'antidopage a une portée mondiale, qui – par le biais d'accords internationaux – pose plus d'obligations de transparence que pour d'autres comportements. Une différence entre les résultats de cette étude et ceux obtenus par Gorse et Chadwick (2011) concerne le nombre de manipulations de match perpétrées pour obtenir un avantage sportif. Dans cette étude, aucun incident n'a été détecté, alors qu'entre 2001 et 2010, les deux auteurs en avaient relevé 24. Cette discordance pourrait s'expliquer par le choix des sources, une problématique qui sera discutée plus tard comme limite dans cette étude. Néanmoins, les

affaires de matchs truqués motivés par le sport semblent plus difficiles à reconnaître, puisqu'elles ne sont pas liées au marché des paris et, par conséquent, il n'existe pas forcément de moyens de preuves suffisants pour démontrer qu'un match a été manipulé.

Finalement, avec 775 cas observés sur une année contre les 2089 cas recensés sur 10 ans pour Gorse et Chadwick (2011), cette étude suggère que le volume des comportements déviants semble être plus important aujourd'hui. Ces résultats peuvent signifier que le nombre de cas a réellement augmenté, que la tolérance envers ces comportements a diminué, que des méthodes plus efficaces ont été mises en place pour les détecter et/ou que les médias relatent plus d'événements qui sont propres à cette thématique (Marchetti et Dargelos, 2000 ; Ohl, 2000).

Pour conclure, cette étude ne prétend pas pouvoir déterminer la gravité d'un comportement suivant son nombre d'occurrences. En effet, un seul cas de corruption peut avoir d'importantes retombées, impliquant de nombreuses organisations sportives, ainsi qu'un vaste réseau (potentiellement criminel) de personnes mondialement connectées.

Limites

Cette étude comporte des limites. Tout d'abord, le panel de comportements retenus n'est pas exhaustif. En effet, il existe d'autres conduites préjudiciables qui pourraient être prises en considération, comme les agressions sexuelles par exemple. La représentativité des cas abordés par les sources choisies constitue une deuxième limite. En effet, les faits relatés ne correspondent que partiellement à la réalité. Plus particulièrement, les médias ne rapportent pas l'ensemble des cas avérés (certains étant classés et/ou non accessibles au public), ni parfois ceux ayant fait l'objet de soupçons ou qui ne sont pas encore connus. De plus, lorsque des affaires sont dévoilées, il n'est pas rare que certains journalistes décident de ne pas les traiter, que ça soit par manque d'intérêt ou selon un choix arbitraire. Par exemple, un épisode isolé de racisme peut passer inaperçu, alors que l'implication d'un officiel dans une affaire de corruption rendra cette dernière plus intéressante, et elle sera donc plus vraisemblablement relatée. Une limite ultérieure concerne l'engouement médiatique que suscitent certains cas en fonction de la gravité perçue du comportement en cause. Dans cette étude, le dopage (biologique ou technologique) et les trucages de matchs sont des thématiques abordées dans 483 cas

relatés par les médias, alors que seuls 111 traitent de l'ensemble des cas de violence et de comportements préjudiciables. La dernière limite est liée au choix des sources utilisées lors de la collecte de données. En effet, la sélection effectuée pourrait avoir impacté la détection de cas au sein de certaines disciplines et dans certains pays. De ce fait, l'absence de cas doit plutôt être considérée comme étant la conséquence des paramètres choisis (telles que la langue ou la période temporelle) ou des limites intrinsèques à la plateforme utilisée.

À l'aune de ces limites, l'interprétation des résultats doit se faire prudemment, puisque les cas détectés ne représentent pas forcément l'amplitude du phénomène étudié. De plus, l'organisation de grandes manifestations sportives – comme les Jeux olympiques ou la Coupe du monde – aura généralement tendance à amplifier le nombre de cas relevé au sein du pays hôte. Par conséquent, aucune corrélation ne peut être établie entre le nombre de cas touchant un sport ou un pays donné et le niveau de déviance.

Conclusion

Cette étude s'est intéressée aux conduites déviantes dans le sport. Grâce à un dispositif de veille basé sur l'analyse du contenu de sources ouvertes en ligne, les types de fraudes, de violences et de conduites préjudiciables relatés en 2016 ont été analysés. Les distributions géographiques et temporelles de la survenance de ces comportements selon les disciplines touchées ont également été étudiées. Cette étape a permis de dégager des patterns.

À partir du corpus de 295 articles extraits automatiquement, 775 cas ont été relevés dans différents pays et disciplines sportives. Au total, 87 pays sont impliqués, bien que la plupart des cas concernent des pays européens et africains. Quant aux disciplines sportives, 48 % des cas rapportés concernent le football, l'athlétisme et l'haltérophilie. Le dopage et les manipulations de matchs sont les comportements les plus fréquemment observés.

Les résultats obtenus dans ce travail laissent entendre que les sources ouvertes peuvent être employées pour étudier les fraudes, les violences et les autres comportements déviants dans le sport professionnel et olympique. Ainsi, la mise en place d'une veille opérationnelle sur plusieurs années pourrait permettre de détecter des variations dans les patterns, que ça soit sur le plan de l'évolution temporelle ou de la répartition géographique des cas.

Des recherches futures pourraient se consacrer au développement de méthodes de détection de cas sur une période temporelle plus large, en intégrant plusieurs langues, ce qui permettrait d'analyser le phénomène de manière plus complète et plus précise. Il existe néanmoins d'autres aspects qu'il convient de considérer : l'accessibilité et la complétude des données (c.-à-d. la couverture médiatique des cas, le type d'information disponible sur chaque cas selon la source utilisée), leur granularité (c.-à-d. le niveau de détail) et la fiabilité de l'information.

Cette contribution constitue une première tentative dans l'appréhension de conduites déviantes dans le sport par les articles parus dans les médias. Des recherches ultérieures sont néanmoins nécessaires afin de déterminer dans quelle mesure la veille en ligne basée sur les sources ouvertes peut constituer un outil de compréhension d'un tel phénomène mondial.

Références

Agence France-Presse. (2016, 19 août). JO-2016 : le nageur américain Ryan Lochte « s'excuse » pour sa vraie-fausse agression. *Le Point*. Repéré à https://www.lepoint.fr/sport/jo-2016-le-nageur-americain-ryan-lochte-s-excuse-pour-sa-vraie-fausse-agression-19-08-2016-2062484_26.php

Alaranta, A., Alaranta, H., Holmila, J., Palmu, P., Pietilä, K. et Helenius, I. (2006). Self-reported attitudes of elite athletes towards doping : Differences between type of sport. *International Journal of Sports Medicine*, *27*(10), 842-846.

Aniello, S. et Caneppele, S. (2018). Selling stolen goods on the online markets : An explorative study. *Global Crime*, *19*(1), 42-62.

Australian Sports Commission. (2016). What is sport integrity ? Repéré à https://www.sportaus.gov.au/integrity_in_sport

Barbary, V., Dyer, D. et Frandano, P. (2011). *Guarding the Game. Preserving the Integrity of Sport.* Londres, Royaume-Uni : Monitor Quest.

Brooks, G., Aleem, A. et Button, M. (2013). *Fraud, corruption and sport.* New York, NY : Palgrave Macmillan.

Brown, A. (2017, 9 août). Documentary alleges African athlete trafficking & doping. *The Sports Integrity Initiative.* Repéré à https://www.sportsintegrityinitiative.com/documentary-alleges-african-athlete-trafficking-doping/

Caneppele, S., Cinaglia, G. et Langlois, F. (2019). Corruption, opportunity and sporting integrity. Dans E. Bayle (dir.), *Le système olympique. Passé, présent et futur* (p. 289-298). Lausanne, Suisse : Presses polytechniques et universitaires romandes.

Carpenter, K. (2012). Match-fixing—The biggest threat to sport in the 21st century ? *International Sports Law Review*, *2*(1), 13-24.

Chappelet, J.-L. (2017). Pour un cadre global de lutte contre la corruption dans le sport. *Jurisport*, *181*(12), 19-22.

Chibnall, S. (1977). *Law-and-order news : An analysis of crime reporting in the British press.* Londres, Royaume-Uni : Tavistock.

Choi, D., Ko, B., Kim, H. et Kim, P. (2014). Text analysis for detecting terrorism-related articles on the web. *Journal of Network and Computer Applications, 38*, 16-21.

Comité international olympique. (2016, 17 mai). Le CIO prend des mesures décisives pour protéger les athlètes intègres – Les athlètes dopés de Beijing, Londres et Sotchi dans le collimateur. Repéré à https://www.olympic.org/fr/news/the-ioc-takes-decisive-action-to-protect-the-clean-athletes-doped-athletes-from-beijing-london-and-sochi-all-targeted

Council of Europe. (2014). Council of Europe convention on the manipulation of sports competitions. *Council of Europe Treaty Series 215*. Strasbourg, France : Council of Europe.

Cryer, A. (2014, 13 février). Why the problem of age fraud is ´rampant´ in African football. *BBC Sport.* Repéré à https://www.bbc.com/sport/football/26174252

Dalton, N. (2016, 19 septembre). Why does athletic academic fraud happen ? Repéré à https://www.pressreleasepoint.com/why-does-athletic-academic-fraud-happen

Gill, N. S. (2015). Cheating during the ancient Olympics. Instances of bribery and cheating at the Ancient Olympics. *ThoughtCo.* Repéré à https://www.thoughtco.com/cheating-during-the-ancient-olympics-120134

Godkin, A. (2013). *National integrity of sport unit. Update on sports integrity.* Australian Government. National Integrity of Sport Unite. Repéré à https://ors.sa.gov.au/__data/assets/pdf_file/0004/148522/National-Sport-Integrity-Unit-Presentation.pdf

Gorse, S. et Chadwick, S. (2010). Conceptualising corruption in sport : Implications for sponsorship programmes. *The European Business Review, 2010*, 40-45.

Gorse, S. et Chadwick, S. (2011). *The prevalence of corruption in international sport : A statistical analysis.* Centre for the International Business of Sport. Coventry, Royaume-Uni.

Han, P. et Dodds, M. A. (2013). Can a sport organization monitor its employees' and athletes' use of social media ? *Journal of Physical Education, Recreation & Dance*, *84*(3), 9-12. https://doi.org/10.1080/07303084.2013.763668

Hemphill, D. et Wilson-Evered, E. (2016). *Sport integrity readiness kit.* Melbourne, Australie : Victoria University.

Hill, D. (2015). Jumping into fixing. *Trends in Organized Crime*, *18*(3), 212-228.

Jayaweera, I., Sajeewa, C., Liyanage, S., Wijewardane, T., Perera, I. et Wijayasiri, A. (2015). Crime analytics : Analysis of crimes through newspaper articles. *Moratuwa Engineering Research Conference (MERCon)*, 277-282.

Jewkes, Y. (2004). *Media representations of the causes of crime.* The Centre for Crime and Justice Studies. Repéré à https://www.crimeandjustice.org.uk/sites/crimeandjustice.org.uk/files/09627250408553594.pdf

Jonson, P. T., Lynch, S. et Adair, D. (2013). The contractual and ethical duty for a professional athlete to be an exemplary role model : Bringing the sport and sportsperson into unreasonable and unfair disrepute. *Australian and New Zealand Sports Law Journal*, *8*(1), 55-88.

Marchetti, D. et Dargelos, B. (2000). Les 'professionnels' de l'information sportive. Entre exigences professionnelles et contraintes économiques. *Regards Sociologiques, Association Regards Sociologiques*, 67-87.

McLaren, R. H. (2016). *The independant person report.* Repéré à https://www.wada-ama.org/sites/default/files/resources/files/20160718_ip_report_newfinal.pdf

McNamee, M. (2013). The integrity of sport : Unregulated gambling, match fixing and corruption. *Sport, Ethics and Philosophy*, *7*(2), 173-174.

Milojević, S., Simonović, B., Janković, B., Otašević, B. et Turanjanin, V. (2013). *Youth and hooliganism at sports events.* OSCE Mission to Serbia. Belgrade, Serbie. Repéré à https://www.files.ethz.ch/isn/180014/English.pdf

Moller, K. (2010). *Examination of threats to the integrity of sports.* Oxford Research.

Ohl, F. (2000). Le journalisme sportif, une production sous influence : l'exemple de la presse quotidienne régionale. *Regards Sociologiques*, (20), 89106.

Rowe, D. et Ohl, F. (2006). Sports et médias. Dans F. Ohl (dir.), *Sociologie du sport. Perspectives internationales et mondialisation* (p. 65-84). Paris, France : Presses Universitaires de France.

Sanderson, J. et Browning, B. (2013). Training versus monitoring : A qualitative examination of athletic department practices regarding student-athletes and Twitter. *Qualitative Research Reports in Communication, 14*(1), 10511. https://doi.org/10.1080/17459435.2013.835348

Schulenkorf, N. et Frawley, S. (2017). *Critical issues in global sport management.* Londres, Royaume-Uni : Routledge.

Serrano, L. (2014). *Vers une capitalisation des connaissances orientée utilisateur : extraction et structuration automatiques de l'information issue de sources ouvertes* (Thèse de doctorat inédite). Université de Caen Basse-Normandie.

SportAccord. (2011). *Model rules on sports integrity in relation to sports betting for all international sports federations and organisations.* Lausanne, Suisse.

Treagus, M., Cover, R. et Beasley, C. (2011). *Integrity in sport literature review.* Adelaïde, Australie : Université d'Adélaïde.

U.S. Department of Justice. (2015, 27 mai). Nine FIFA Officials and Five Corporate Executives Indicted for Racketeering Conspiracy and Corruption. Repéré à https://www.justice.gov/opa/pr/nine-fifa-officials-and-five-corporate-executives-indicted-racketeering-conspiracy-and

World Anti-Doping Agency. (2017). *2015 anti-doping rule violations (ADRVs) report.* Repéré à https://www.wada-ama.org/sites/default/files/resources/files/2015_adrvs_report_web_release_0.pdf

Fraud, violence and other deviant behaviors in professional and Olympic sport: opportunities and constraints of online open sources as channel for intelligence

ABSTRACT • *Online open source media are increasingly being used as tools for intelligence analysis. This contribution explores how they can be used in the field of deviance in professional and Olympic sport. Our study considered on the field fraud (doping, game manipulation, eligibility fraud) and off the field fraud (corruption), on the field and off the field violence (hooliganism and terrorism), and other detrimental conducts by sportsmen whether off or on the field of play. To explore the magnitude of these phenomena, we performed an online open source analysis in 2016 (in English) and identified 775 cases reported by online media, with frauds accounting for 85% of all cases. Specifically, doping, followed by match-fixing are the most frequent cases. In total, 87 different countries were involved, but some sub regions, notably Eastern Europe (17.4%), East Africa (16.1%) and Australia and New Zealand (10.2%)) concentrate most of the observed cases. In terms of temporal patterns, concentration of cases in August and November for fraud, and in June and July for violence may suggest a seasonality of deviant behaviors in sport. By looking at biases linked to language and types of source, further research may contribute to systematic monitoring of the presence of these phenomena around the world.*

KEYWORDS • *Doping, sport fraude, violence, corruption, sport, open sources, intelligence.*

Fraudes, violencias y otras conductas desviadas en el deporte profesional e olímpico. Oportunidades y limites de las fuentes abiertas en línea como medio de información

RESUMEN • *Las fuentes abiertas en línea son utilizadas cada vez más como mecanismos de inteligencia. Nuestro estudio explora de que manera podrían utilizarse las fuentes abiertas para estudiar la desviación en el deporte profesional y olímpico. En el estudio se han considerado los fraudes cometidos en el campo (dopaje, manipulación de los partidos, fraude a la admisibilidad) y fuera del campo (corrupción), la violencia en el campo de juego y fuera de este (gamberrismo y terrorismo) y también otros comportamientos perjudiciales por parte de los deportistas, tanto en el campo como fuera. Un mecanismo de vigilia en línea fue utilizado para recoger los artículos publicados en ingles en el año 2016. 775 casos fueron identificados, los fraudes (sobre todo los casos de dopaje y manipulación de partidos) representando por si solos 85% de los casos. En total, 87 países estuvieron implicados en este estudio, aunque unas subregiones – Europa del Este (17,4%), Africa del Este (16,1%) y Australia y Nueva Zelanda (10,2%) – se destacan por una concentración mas elevada de casos. Sobre los* patterns *temporales, hay distribuciones concentradas sobre todo en los meses de agosto y noviembre para los fraudes, y los meses de junio y julio para las violencias. Esas tendencias pueden significar una estacionalidad de los comportamientos perjudiciales. Atacándose a algunos de los sesgos relacionados con el lenguaje y los tipos de fuentes, investigaciones más profundas sobre el tema podrían contribuir a una vigilancia sistemática de la representación mediática de estos fenómenos en el mundo.*

PALABRAS CLAV • *Dopaje, fraude deportivo, violencia, corrupción, deporte, fuentes abiertas, inteligencia.*