Revista
Mediterránea
de Comunicación
Mediterranean Journal of Communication ISSN 1989-872X

Dr. José-Luis ROJAS-TORRIJOS
University of Sevilla. Spain. jlrojas@us.es

Dr. Jesús GARCÍA-CEPERO
University of Sevilla. Spain. jesusgcepero@gmail.com

Perception of sports data journalism among heavy users. Case study: predictive model during the 2018 Football World Cup in El País

Percepción del periodismo deportivo de datos entre usuarios habituales. Estudio de caso del modelo predictivo de El País para el Mundial de Fútbol de 2018

Dates | Received: 13/01/2020 **-** Reviewed: 20/03/2020 - In press: 16/03/2020 - Published: 01/07/2020

Abstract

In recent years, sports and, most notably, football coverage has become a breeding ground for data journalism. The vast number of statistics collected from scheduled matches have helped data analytics techniques, based on applied mathematics, to be expanded in current sports journalism. This article examines the statistical model in predicting results developed by El País for the first time to enhance its coverage of a mega sporting event: the 2018 FIFA World Cup held in Russia. This research also analyses the level of acceptance and understanding of this advanced statistical method among heavy users of sports content. To this end, a semi-structured questionnaire involving both open-ended and closed questions was conducted to know the views of both Sports Journalism students and sports reporters from twelve Spanish media outlets. The results reveal that the news values of applying advanced statistics to report on probabilities in a sports tournament encounter a reluctant attitude and an uneven level of understanding among professionals and students. Despite this, data journalism is mainly perceived as a huge opportunity to diversify the agenda and improve the quality of sports coverage.

Keywords

Data journalism; data visualisation; football; infographics; sports journalism; statistics

Resumen

El periodismo de datos ha hallado en los últimos años un terreno abonado para su expansión en las coberturas deportivas, en particular las futbolísticas. El sustrato estadístico de la competición y su carácter cíclico han favorecido el desarrollo de esta nueva modalidad de análisis con datos que se apoya en matemática aplicada para expandir el periodismo deportivo actual. Este artículo profundiza en el estudio de la aplicación, por primera vez en el diario El País, de un modelo matemático de predicción de resultados en una gran cobertura como el Mundial de fútbol de Rusia en 2018. Asimismo, evalúa el grado de aceptación y comprensión de esta metodología de estadística avanzada entre los consumidores más habituales de este tipo de información a partir de cuestionarios semiestructurados a estudiantes universitarios de Periodismo Deportivo y periodistas deportivos de medios españoles. Los resultados del estudio ponen de manifiesto que el uso de estadística avanzada para informar de probabilidades en un torneo aún encuentra una difícil aceptación periodística y un desigual grado de entendimiento entre la audiencia. Pese a ello, el periodismo de datos es percibido mayoritariamente como una gran posibilidad para mejorar la diversidad y la calidad de las coberturas deportivas.

Palabras clave

Estadística; fútbol; infografía; periodismo de datos; periodismo deportivo; visualización de datos

Rojas-Torrijos, J. L. & García-Cepero, J. (2020). Perception of sports data journalism among heavy users. Case study: predictive model during the 2018 Football World Cup in El País. *Revista Mediterránea de Comunicación/Mediterranean Journal of Communication, 11*(2), 295-310. https://www.doi.org/10.14198/MEDCOM2020.11.2.8

1. Introduction

In recent years, data journalism has become one of the main categories of work in digital media when visually representing stories and finding new ways of telling them from an analysis of ordered data sets.

D*ata journalism, a term first coined by* Simon Rogers in 2008 in his *Datablog* in *The Guardian* (Knight, 2015), has gradually been implemented since then by editorial teams thanks to the accessibility to data new digital platforms have provided.

For many authors, data journalism is nothing more than the natural evolution of precision journalism (Dader 1997) while others prefer to define it as a combination of scientific research methods, journalism and the use of a computer as an essential work tool. Crucianelli described it as "a sum of known methods to which three fruits of technological innovation are added": the large volume of data which can be accessed nowadays, interactive visualization and the addition of programmer to the journalism team (2013: 107).

However, data journalism, despite being a recent phenomenon, is still at an expansive stage in which ever more editorial teams are building equipment to work with databases (Rogers, 2014). Moreover, it has not been evenly implemented by different countries. While some large chains in Europe and America such as *The Guardian*, *Financial Times*, *The New York Times* or the Argentinean *La Nación* have data and visual journalism departments, in the Spanish media, there is still scant editorial leaning in this direction, due to "the lack of tradition and training both within and outside editorial teams, as well as the little interest shown in this from those in charge " of the media (Ferreras, 2013: 130).

To some extent, the media has gradually incorporated techniques from data journalism into their digital production routines for news in order to enrich and diversify the coverage they provide. Meanwhile, Bradshaw (2011) defined this category of journalism as the sum of gathering, refining, contextualizing, combining and communicating data. Veglis and Bratsas (2017a) referred to this type of journalism as a process which consists in extracting useful information from databases, writing articles from this information and adding visualizations to articles (which are often interactive) which help readers understand the stories better.

Likewise, the trend in data journalism has entailed publishing different categories of news, which have been classified by different authors. These classifications have been made according to the level of analysis and interpretation required for working on the data so that they can be produced, but, above all, according to the structure and volume of the data included and the methodology used to present information (Veglis and Bratsas, 2017b).

So, for example, Kang (2015) categorized data projects according to their visualization and interaction potential, but, above all, in terms of their journalistic purposes: spots which provide an overview or show how a real situation has evolved over time, charts/world maps which enable the reader to extract local interest data, visual stories in which situations are compared and differences highlighted and stories with more explanatory purposes which are usually built from intersecting different sources of data or research projects.

In a process in which statistics are ever more democratic, with a greater number of databases available with open access, journalism needs to work with large volumes of information. This means data journalism" occasionally becomes curation" (Rogers, 2013: 16), since journalists needs to refine, analyse and use key data from these databases for each story and always seek those which are most pertinent to each piece of news and which can be understood by the reader.

In this expansive phase, data journalism has covered all kinds of topics, among which sports is one of those with most potential when telling and displaying stories with statistics (Segel and Heer, 2010; Silver, 2015). Sports information is one of the specialisms in which statistics play an important role and in which it is already very complicated for a match reporter to have data to hand which can clarify what has happened (Marrero-Rivera, 2010: 131).

In fact, data journalism is gradually and naturally making headway in sports contents, since competitions create a large amount of data from the results (performance, dynamics, scores...) which are also accumulative in nature in historical series which help to make advanced measurements. So, it was only a question of time, that sports journalism would make use of this mass of statistics to enrich its coverage, as stated by Arias-Robles:

> It is hard to imagine a speciality which is better adapted to data journalism than sports news, firstly, because any type of competition creates a large amount of quantifiable information; secondly, because a methodological treatment of these data enables specific cases or trends to be shown (2017: 217).

When applied to sports news, the purpose of data journalism is to provide added value and new meaning to the statistical registries of teams and sports people, as Rojas and Rivera put it:

> Rather than just gathering and showing data, data journalism is advanced statistics applied to news in which the key is to analyse the relationships between different variables in order to arrive at some data which are not usually shown by conventional statistics, using a scientific methodology (2016: 3).

Therefore, a clear line can be drawn between basic statistics (the mere exposition or listing of data and statistical registries, as what happens, for example, in live sports broadcasts) and advanced statistics, that is, the processing of statistical data by data journalism in order to go beyond basic statistics and reach new knowledge which can be shown by means of charts and visualizations.

Within this journalistic use of advanced statistics, big data is used to draw up results predictions. Here, probability models in sports coverage are aimed at extracting readings and new interpretations from a set of datificated information that already exists. This set of ordered data is already used by the media, leagues and sports institutions specialized in the storage and processing of sports statistics, such as the American *Stats LLC* or the British *Opta Sports*.

Sports results prediction models have already been used by the media such as the *Financial Times* or the website *Five Thirty Eight*, owned by *ABC News*, which has made data its trademark (Arias-Robles, 2017: 217). These publications have developed them to diversify their sports coverage, and to provide new angles and perspectives rather than just seeking to broadcast facts. In this regard, Mayer-Schönberger and Cukier hold the belief that accuracy can be sacrificed a little if the trade-off is being able to discover a general statistics behaviour trend. However, they add that "big data turn arithmetic calculations into something which is more probabilistic than accurate" (2013: 52).

Therefore, models based on advanced statistics, such as those created by *El País* for the 2018 Football World Cup, measure probabilities and make future projections from already existing data. So, rather than making mere predictions, they determine which scenarios are more probable within a competition according to a general trend and how to tell the reader about these predictions, following a methodology which characterizes data journalism.

2. Methodology

In light of this context, this article includes a case study of the probabilistic model of results from *El País* when covering the Football World Cup in Russia held between the 14th of June and 15th of July 2018 which was the first time this Spanish general-interest newspaper used advanced statistical techniques and tools to make predictions with journalistic value within its sports section.

The mathematical forecasting model from *El País* for the World Cup in Russia, created by the analyst Kiko Llaneras, was initially published on 4th of June, more than a week before the tournament started. After this initial prediction, the Madrid-based newspaper gave predictions one after another for each team as the competition went underway.

The situation which acted as a springboard to our study, once the theoretical review was complete, was really brought about by two circumstances. Firstly, the fact that data journalism is even more incipient in sports news in the Spanish media means typical consumers of sports contents still require more time to take in, understand and accept the news value of data projection in probabilistic models of results in the coverage of a large sports event. In this respect, the probabilistic models developed by *El País* came up against reluctance and an uneven degree of comprehension among consumers.

Secondly, it should be stressed that sports data journalism has great potential which, by using techniques for extracting, analysing and visualizing statistics, can provide new angles for covering competitions as well as enhanced use of graphics and multimedia resources for displaying information visually. This potential not only constitutes an incentive for sports news to move in the direction of providing accurate journalism, and hence, greater quality, but also makes contents more diverse and attractive for the most typical consumers. This is exactly what *El País* has been doing with this modus operandi in the coverage of a large sports event.

In light of all this, the objectives of this study are as follows:

1. To examine the degree to which data journalism has developed to date in the speciality of sports journalism in Spain.

2. To analyse the journalistic value of the results prediction models from advanced statistics on the coverage of a large sports event.

3. To specifically study the probabilistic model from *El País* in its coverage of the Football World Cup in Russia in 2018, and compare it with other similar models used by other international media in sports coverage.

4. To find out the impact and degree of comprehension of the techniques used in data journalism among the main groups of typical consumers of this type of news, as well as students of Sports Journalism and the professionals which work in the media and sports sections in Spain.

5. To broaden reflection on its innovative nature, contributions and potential from a data analysis viewpoint for the future of sports journalism.

In an attempt to respond to the research objectives, qualitative research was made based on the case study of the probabilistic model developed by *El País* in its news coverage of the Football World Cup in Russia. In this regard, not only was this coverage analysed by means of different predictions for each match published in the on-line edition of this newspaper throughout the tournament, but also with the view of the media itself. So, an in-depth interview was made on the 9th of September 2018 with the analyst Kiko Llaneras, author of the mathematical model which is the subject of this study, in order to gain an insight into the methodology he developed and his own assessment of the journalistic experience referred to when the championship had finished.

In the second phase of the research, semi-structured questionnaires were drawn up which were aimed not only at students of Sports Journalism, but also at journalists on the sports editorial teams for Spanish printed and digital media, both general-interest and specialized ones. Therefore, in line with that set out by Rodríguez-Gómez, Gil-Flores and García-Jiménez (1999: 73) on how to design qualitative research, the sampling when selecting informants followed intentional, rather than random criteria in order to deal with the attitudes and perceptions of two types of target audience on the phenomenon which is the subject of this study.

Moreover, considering the questionnaire as a fundamental instrument for obtaining data and studying attitudes in terms of a problem (Igartua and Humanes, 2009: 94-95), with the results obtained from these sets of questions, the aim was to compare statements, opinions and assessments from two qualified target groups which are typical consumers of sports information.

The objectives of the initial questionnaire was firstly, to check the general feeling future journalists had about statistics and how these related to sports journalism, and, secondly, their assessment of the probabilistic model developed by *El País* for Russia 2018. In total, 52 valid responses were collected, all of which were from third-year Journalism students at the Faculty of Communication at the University of Seville and enrolled in the Sports Journalism option. The questionnaire was passed to the students online with the tool, Google Forms, on the 4th of April 2019.

Subsequently, a similar questionnaire was sent by email to 16 sports journalists from 12 different media. The responses were collected between the 11th and 21st of October 2019 thanks to a follow-up task sent to smartphones (Whatsapp) and social networks (direct messages on Twitter). The journalists participating were: Carles Vila (*Mundo Deportivo*), Javier Sánchez (El Mundo), Víctor García (*El Confidencial*), Ignacio Labarga and Alberto Benítez (*Marca*), Antonio Medina (*Estadio Deportivo*), Ignacio Delgado and Álvaro Ramírez (*El Desmarque*), Juan Luis Rodríguez Cudeiro (*El País*), Pablo Salvago (*Diario de Sevilla*), Mateo González (*ABC*), Jorge Fernández Maldonado (*As*), Eduardo Casado (*20 Minutos*) and Enrique Julián Gómez, Cristina Caparrós and Borja Pardo (*Sphera Sports*).

Although both questionnaires were very similar, and most questions were the same, they were designed ad hoc bearing in mind the characteristics of both types of target and based on the questioning from the theoretical review. Given the exploratory and qualitative nature of the research, in both situations closed questions, which are easier to measure from the reduced number of response options, were combined with more open-ended questions so that the participants could explain their personal experiences, as well as their perceptions and assessments of the topic this study is concerned with.

The questionnaire for the students consisted in six questions, all referring to their evaluation of the works published by key media for data journalism, except for the first question, which was more introductory, whose purpose was to measure student interest and acceptance of statistics in general. Meanwhile, the set of questions sent to the journalists covered seven questions. Out of these, four were similar to those posed to the students, while another two were rephrased to find out how professionals valued the news in works based on prediction models, and their thoughts on implementing data journalism in the sports editorial departments in Spain. Lastly, in the seventh question, the journalists were given some results from

the questionnaire, which were previously made to the students, for them to interpret what they meant and, in this way, points of view from different types of target audience were dealt with.

3. Results

This qualitative research primarily concerns an explanation and analysis of the mathematical prediction model of *El País* as an example of innovative sports data journalism. This model was studied by comparing the updates published by this media on its website throughout the competition with the results which were finally produced, and the explanation the tool author, Kiko Llaneras, himself gave in an interview.

3.1. Explanation of the model

The mathematical prediction model developed by *El País* for the World Cup in Russia was shown to readers on the 4th of June 2018. That is, ten days before the competition began. The first article was entitled, "Who will win the World Cup? This is how we made predictions at EL País" (1) an initial prediction was provided and a detailed explanation of the methodology used for creating this model was given.

In that initial explanation, *El País* showed the model had three fundamental parts: 1) a ranking which measured the strength of each national team; 2) a statistical model which estimated the possible results for each match; and 3) a simulator for the competition. According to Llaneras (2018) in a later article, the points ranking was based on the Elo points system, inspired by its use in chess since the 1950´s and which has been used to predict football results for some years (Hvattum and Arntzen, 2010). The Elo system consists in calculating the relative skill of a player and his probability of victory from results against opponents (Gickman and Jones 1999). In this way, the results for each national team, goals scored and conceded ("Elo expected") and their data was taken into account.

As Llaneras (2018) explained in an article published on completion of the championship, the ranking was based on the data for expected goals for over 200 matches in which national teams have fought since 2017, which was provided by the specialist company in sports statistics, *Opta Sports*, while to gauge the value of each team, data from 352 clubs and 800 players were used.

In the interview made for this research, the data analyst from *El País* claims that the model was inspired by academic works and publications from other media with similar methodologies such as *Five Thirty Eight* or The *Financial Times*. Among these, there is a reference to a study made in Germany for estimating the results of the matches in the UEFA European Championship in 2016 from a goals distribution formula named Poisson (Arroyo, Bravo, Llinás and Muñoz, 2014; Groll, Kneib, Mayr and Schauberger, 2016), in which the field factor is taken into account. That is, whether the team is playing at home, away or on a neutral pitch.

Likewise, points out that they chose to predict goals rather than victories directly "because there are two advantages in doing this: it can better explain the positions in the groups stage and it helps to predict if matches will go into extra time" (2018). Despite the studies which suggest that the models which predict goals underestimate the draws which really occur (Dixon and Coles, 1997: 266-267), Llaneras believes this bias is far less marked in national team tournaments than between clubs.

He remarked that when gauging this model, a database was used with 17,000 matches in which national teams participated and for calculating the probabilities throughout different phases of the tournament around 10,000 match simulations were made.

After this initial forecast, *El País* published different predictions for every team as the competition went underway (2). However, the competition results did not always match those predictions, which led to criticisms of the model. For example, Germany, one of the favourites was knocked out in the first round of the World Cup, despite having an almost 90% chance of being classified for the knock-out phase of the final; or Spain which had an over 84% chance of passing to the quarter-finals, was defeated in the round of 16 against Russia.

These discrepancies between the probabilities published and the actual results during the World Cup in Russia led to a debate on the journalistic value of the data and how ideal it was for publishing information based on predictions and estimations from statistical data accumulated in coverages of large sports events. Moreover, the degree to which the average, and even specialized, reader understood determined analysis and visualizations based on advanced statistics when these are not typical in the media, was questioned, and therefore, when consumers were not used to interpreting and assessing them appropriately.

3.2. Questionnaires

In order to check to what extent the *El País* model specifically and data journalism in sports news in general is accepted and understood by readers, semi-structured questionnaires were drawn up which were aimed not only at students of Sports Journalism, but also at journalists working in the sports editorial teams of Spanish printed and digital media, both general-interest and specialized ones.

3.2.1. Value and significance of statistical data

In the questionnaires, an initial question was posed that was common both to students and journalists in order to measure the degree to which those participating in this study accepted sports statistics in general and to check the significance they gave them as contents, whether these were journalistic or not. The first reading that could be extracted was that sports statistics aroused great interest: 78.8% of university students claimed they liked or were curious about checking data for a sports event while watching it live or after having watched it, and with media professionals, this figure reached 81.25%. Likewise, the participants (61.5% and 68.75%, respectively) recognised that the statistical data was especially attractive for the average spectator, consumer or reader, which may explain why the media are using them increasingly more.

However, this consideration does not mean the results are entirely positive in terms of the value and significance both groups give to the statistical data. According to the responses, they are not a tool by which sports events can be explained in themselves or which can optimally define the features of a sports person. Hence, only 25% of students and 12.5% of the journalists felt it was sufficient to look at the most important statistics of a match to gain a very approximate idea of what has happened without having seen it. However, a small percentage (30.8% and 12.5%, respectively) claimed that it is possible to know what a player is like simply by looking at his statistical record for passes, shots, victories, etc. That is, from this viewpoint, statistics do not cover or explain everything, but they may reveal interesting information.

3.2.2. The role of data in sports journalism

The last point in this introductory section of the questionnaire is connected with the relationship between statistics and sports journalism: What role do the data have in sports news journalism? Although opinions vary a great deal about the use of data in these types of contents, there seems to be one fundamental premise: 88.5% of the Journalism students and 87.5% of the media professionals agree that data in sports journalism contents are very useful for supplementing and enriching basic information. Likewise, a large proportion of the participants in the study, 67.3% of future journalists and 56.25% of those who actually work in this profession, agree that a report or preview are always incomplete if they do not include important statistics.

3.2.3. Acceptance and comprehension of data journalism

Although there seems to be some consensus as regards the importance of statistical sustenance within sports news, there is some discrepancy when it comes to understanding and accepting data journalism as a new modus operandi when covering events. This was made patent in the specific probabilistic model drawn up by *El País* for the 2018 Football World Cup (Image 1), a work which, due to its originality and specific features, received both praise and criticism. The results gleaned from the questionnaire made with this research corroborate this. Firstly, up to 19.2% of students considered the model to be error-prone on the basis of the differences in the predicted results and those that finally appeared in the tournament. Meanwhile, the other 80.8% found the study the newspaper made interesting. However, within this latter percentage, 34.6% understood it as a serious piece of journalism as opposed to 46.2% who claimed that this use of data was not just for journalistic purposes.

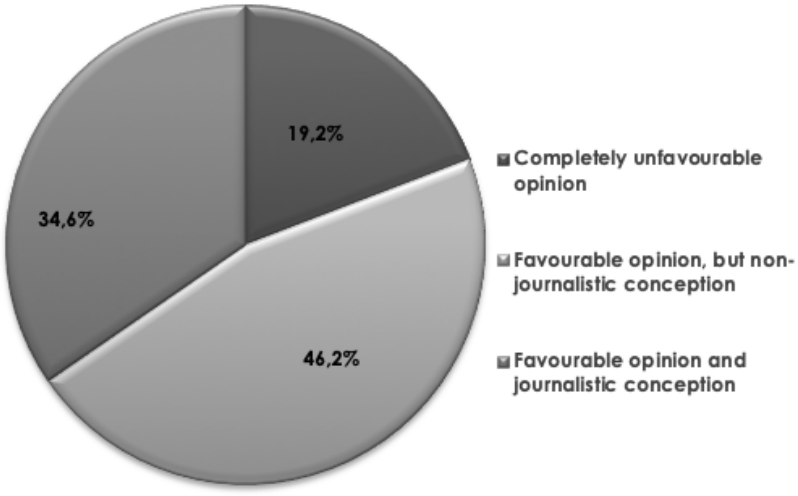
Image 1: Predictions in the preview of the World Cup

SELECCION	OCTAVOS	CUARTOS	SEMIFINAL	FINAL	GANAR
Brasil	90,3%	63,6%	45,5%	29,3%	17,9%
Alemania	89,1%	60,2%	41,9%	25,5%	15,6%
España	84,1%	65,9%	42,6%	25,6%	15,3%
Argentina	85,6%	58,8%	34,8%	19,9%	11,3%
Portugal	75,0%	53,4%	30,9%	16,0%	7,9%
Francia	79,5%	46,1%	25,5%	12,9%	6,2%
Inglaterra	80,8%	50,8%	23,6%	12,1%	5,3%
Bélgica	79,3%	49,1%	23,1%	11,2%	4,7%
Colombia	74,6%	40,8%	17,4%	7,6%	3,3%

Source: Graph published by *El País* on the 4th of June 2018.

The *El País* prediction model surprised students. For instance, only 36.5% knew that data could be used to try to predict future events. However, this innovative way of using statistics in journalism is much more popular among professionals, 62.5% of whom claimed to be aware of it. However, the debate once again hinged on whether this use was for journalistic purposes or not, and this probabilistic model is a great example because neither Journalism students nor journalists reached agreement in this respect.

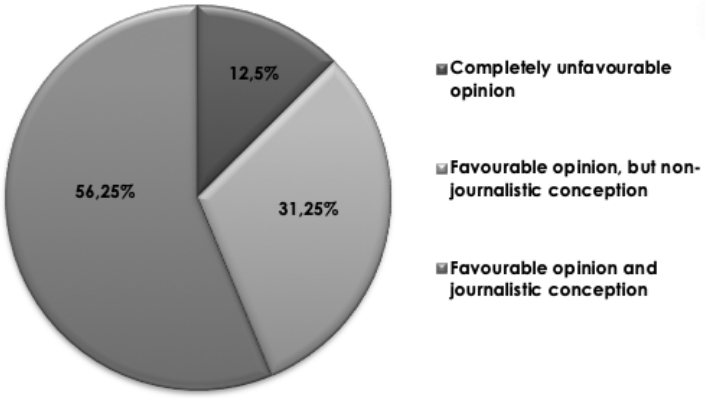
Chart 1: Evaluation of the students from the *El País predictive model.*



Source: prepared by the author

As seen in Charts 1 and 2, 56.25% of the professionals, and only 34.6% of students considered it to be a serious journalistic study, while 31.25% of professionals and 46.2% of students thought it was attractive despite being beyond the confines of strict journalism. The remaining 12.5% of journalists and 19.2% of students saw the model as useless and of no value.

Chart 2: Evaluation of the journalists from the *El País predictive model.*



Source: prepared by the author

3.2.4. The journalistic value of results prediction

But, why did some journalists claim it definitely constituted journalism and others were not convinced that it fully adhered to journalistic codes and procedures, regardless of the degree to which the predicted results were fulfilled? Some of the professionals whose opinions were gathered in this research insisted there were limitations to the use of advanced statistics for informing and explaining events from a journalistic point of view.

For example, Mateo González, chief sports editor at *ABC de Sevilla*, claimed "advanced statistics is fine for noting trends and player-performance in a team at any given time, but daring to make a real prediction about the future is too risky and perhaps not in the best interests of journalism", which would even make the model "border on the world of sports bets". Also, Carles Vila, editor of *Mundo Deportivo*, mentions betting houses and considers the methodology used in the *El País* model to be similar to the one used to establish stakes, so this model "could be used as a previous step to being informed about a result from another occasion, and is informative in that regard, but does not help much to give an overall explanation for the competition".

Likewise, Víctor García (*El Confidencial*) and Borja Pardo (*Sphera Sports*) agree that when working with data, this may provide excellent added value to the information, but they should never lay the foundations for it. Lastly, we turn to the view of Javier Sánchez, journalist at *El Mundo*, who sees it is only valuable in terms of providing a ranking for FIFA or a similar one, and that "data may be interesting for analysis and of slight interest to the reporter, but not for prediction", which is not really the work of a journalist.

Defending the model from all this criticism (which might be thought reasonable, considering how new this is in Spanish journalism), Kiko Llaneras (2018), head of the project, claimed that once the World Cup had finished, it was truly reliable, because, for example, events which had a probability of between 0 and 10% of occurring, only occurred 3% of the time, while events which had a probability of over 75% occurred 86% of the time.

In this way, rather than expecting the results to be confirmed to see if they match the model predictions, what the debate is really concerned with is ascertaining the journalistic value these types of editorial approaches have. Figures do actually communicate (in this case, equality, uncertainty...) and moreover, they are important for dissemination purposes, since they attempt to explain a sports event in a different light, with data and by crossing variables.

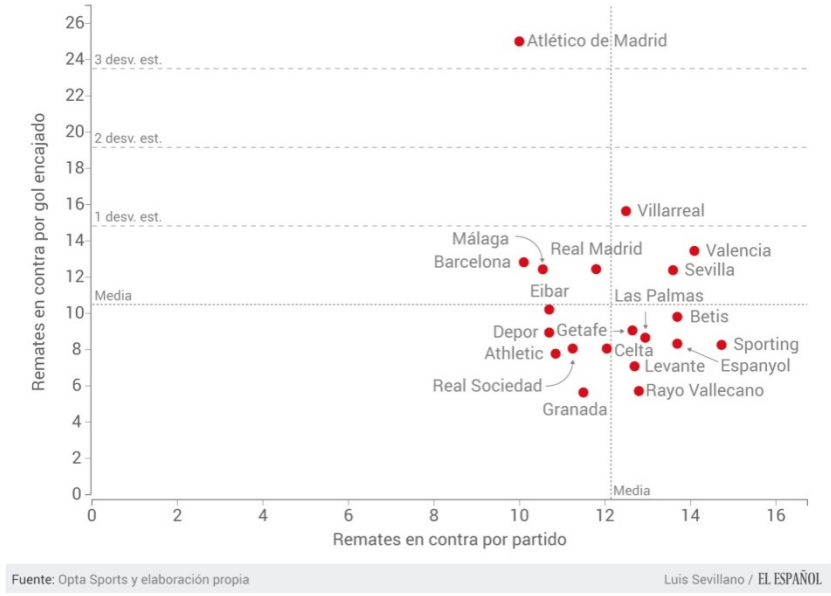
According to Llaneras, journalists must not become obsessed with finding newsworthy events, but also explain them from another point of view while entertaining their readers. The truth is that many of the professionals who participated in this research seem to embrace that notion. Pablo Salvago, journalist at *Diario de Sevilla*, considers"advanced statistics has become one more tool for creating valid and even true news", which makes the *El País* study a priori, "rigorous and objective". Also Cristina Caparrós, at *Sphera Sports*, praises its attractiveness and diligence, stating "it is not an opinion or procedure based on simple data, but, rather, a process which encompasses a range of factors in its methodology".

3.2.5. A second example of data journalism: control chart

To better gauge the degree of difficulty of the *El País* model, a second example of sports data journalism was attached to the questionnaire, which also gave rise to a range of different opinions. This was a chart which appeared in a short article by *El Español* in January 2016 (Image 2) which analysed the solid defence of Atlético de Madrid compared with other defences in the Spanish football league (3), crossing the shots against per match and shots against per goal scored for the different teams in the competition.

It was interesting to see how time passed for one of the most innovative data visualizations presented to date by this digital media, which in turn, was one of the first in Spain to commit to sports data journalism (Rojas-Torrijos and Rivera, 2016: 180). Therefore, there was an assessment of whether what was disruptive a few years ago (the chart becomes the focal point of the explanation, providing multiple readings from a combination of the variables and the text was minimalised as if it were a mere caption), is now better understood.

Image 2: Control chart



Source: Chart published by *El País* on the 20th of January 2016.

However, the results from the questionnaire clearly showed that this type of analysis and sports data visualization are not yet fully understood, even among the most typical users. Thus, Journalism students assessed the chart with an average difficulty of 3.9 on a scale from 1 to 5 with up to 50% allocating a difficulty of 4 and even 23.1% gave it the maximum difficulty (5). Yet more striking was the opinion of the professionals, as the average assessment of the difficulty they gave for this visualization was 3.75 (very similar to that given by the students), and up to 62.5% gave it a value of 4 or above.

In order to check if the chart was really so complex as it seemed, the students were asked what, to their minds, was the message or main idea behind this visualization. Indeed, crossing the statistics, it might be concluded that Atlético de Madrid was the team which was most difficult to score a goal against, which speaks wonders about its defence strategy because two things happen at the same time: firstly, the team receives very few attack shots per match, and, secondly, it has to receive many shots before its opponents manage to score a goal.

Having said that, only 7.7% of students extracted this as the main idea, while 17.3% came close to giving the ideal response, since they clearly saw the red and white team had a solid defence, but were unable to get to the bottom of the matter. Also, 34.6% of students were unable to draw any clear conclusions from the chart, openly acknowledging that they did not know how to interpret it, while 7.7% tried to solve it, but did so erroneously, giving some responses that were totally incorrect. Finally, out of the remaining 32.8% of

respondents, rather vague responses were given, with which it could not be ascertained whether they had grasped the essential message of the chart or not, as some students just showed knowledge about the general idea (shots against received by the teams in the League) while others read it in alternative ways.

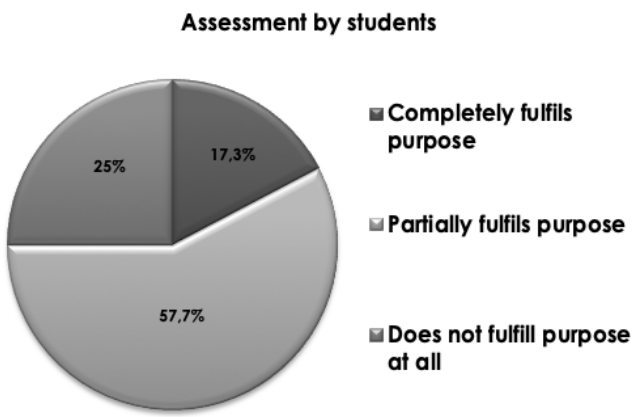
3.2.6. More information on data journalism for gaining an insight

In light of these results, it was interesting to ask the professionals about the response given by the students and what this may have been due to. To be specific, the journalists were asked if the fact that most Journalism students were unable to interpret the chart was linked to shortcomings in their training or whether it was because the chart was too complex. Surprisingly, up to 56.25% of the journalists tended to indicate the latter option, as stressed, for example, by Jorge Fernández Maldonado, sports editor at *AS*. He claimed that "it is up to journalists and the media to ensure that their final product is understood by most of their recipients" and that "if most of them do not understand it, then an alternative must be sought". As for, Javier Sánchez (*El Mundo*), he was convinced that although almost any reader could understand it, to do so, they had to stop and carefully examine it, so the effort required "time it is highly likely readers are not going to dedicate to it". Logically, then, the solution would be to try to simplify the chart or seek another method of displaying data, thereby adapting their level to their potential readers, as Víctor García (*El Confidencial*) adds.

Meanwhile, some professionals thought the problem was mainly due to training. Enrique Julián Gómez, at *Sphera Sports*, believed that in the faculties "little training is given to students in statistics and mathematics, in general", and this is something which, in his opinion, is worrying for journalists because "statistics is an ever-more fundamental discipline for showing and interpreting reality and is therefore essential to journalism. Therefore, "although it may be a complex chart for the average reader, a journalist must have a perfect and clear understanding of this type of statistical information".

Concerning these shortcomings in training, Carles Vila commented on "the lack of experience the students have in handling data", which implies that it is still not typical for journalists to work on this specialism from an early stage. Finally, there were users who gave the same importance to the chart complexity problem as to possible shortcomings in training for future journalists, without putting any special emphasis on either of these two factors which are the main causes of this lack of understanding.

Chart 3: Assessment of the functionality of the *El Español chart (I)*



Source: prepared by the author

Along with the degree of perceived difficulty, another point that was analysed was the function of the chart. That is, whether the two participating groups considered data visualization fulfilled its purpose of clearly representing the intended idea (the solidity of the defence of the two teams). Also, in this case, there was a wide range of opinions.

As shown in Chart 3, only 17.3% of students thought the visualization completely fulfilled this purpose. They understood that a defence which enables fewer goal opportunities is a more solid one; while 57.7% of them believed it was useful because it could explain a variable of the game related to defence work, but it did not totally fulfil the initial purpose of digital media; and according to 25% of them, the chart was no good for that purpose, as in their view, just because a team receives fewer shots does not necessarily mean they defend better.

Chart 4: Assessment of the functionality of the *El Español chart (II)*

Source: prepared by the author

As for the professionals, their percentages were 25%, 68.75% and 6.25% respectively (see Chart 4), which shows us that the vast majority considered this data work to be interesting, although they remained unconvinced as they did not think the idea was represented in the most appropriate way.

3.2.7. The future of sports data journalism

The last section of this research consisted in an open-ended question expressly made to professionals in order to know whether, at present, they see the establishment of data journalism as a work technique for the editorial teams of Spanish media as viable.

While, there were testimonials, like that from Antonio Medina (*Estadio Deportivo*), who believed that "data journalism is already very established in the editorial departments at least those for sports", and that even "more news comes from it than statement-based journalism", the vast majority of professionals disagreed: they considered data journalism in Spain had still not made that leap forward and that its use in the media was still too marginal for it to be considered as significant.

However, regarding the future of Spanish data journalism, some of the views expressed showed optimism. Hence, Victor Garcia, indicated that data were becoming increasingly important in our society and, therefore, they were ever more crucial to journalism. Meanwhile, for Jorge Fernández Maldonado, the establishment of data journalism in sports news was perfectly viable. In fact, he believed this might happen very soon if we bear in mind that "handling big data has become ever more widespread in the world of sport".

There are other journalists who seriously considered this possibility, albeit with important qualifications. For example, Juan Luis Rodríguez (*El País*) stated that while scarcity of resources remains the predominant trend in journalism, just "those media which can afford to allocate human resources to data journalism" can implement it with their editorial teams. Carles Vila was convinced that it would only become a reality if journalists felt the need to develop it. Finally, Javier Sánchez believed that it would be successful but not in sports journalism in particular, since, to his mind, statistics did not contribute much to general-interest media articles, which did not need to enter into highly technical points.

Lastly, several of the professionals participating in the study understood that there was little room for data journalism nowadays in Spain. Economics was the most common reason cited for justifying this argument, but there were also many allusions to the tastes of the target audience. So, according to Eduardo Casado, journalist at *20 Minutos*, before we talk of growth in data journalism "public interest must be aroused, and for the time being, these data or statistics are viewed as being something merely anecdotal". Likewise, he talked about "re-educating" this public, which "was still more interested in more" amateur aspects of sports information.

4. Discussion

As we have seen in the research results, there is no univocal perception about data journalism among the different target groups researched, the typical consumers of sports information. The differences shown in the degree of acceptance and comprehension about the use of advanced techniques for visualization and data prediction, specifically looking at *El País* during its coverage of the 2018 World Cup in Russia, led to a discussion on three key issues: firstly, the usefulness and news value of probability as a concept accepted by journalists; secondly, the editorial commitment that the media is willing to make to nurture the use of data journalism in news reports; and, thirdly, the challenges posed by the development of this new category of journalism for editorial teams for the university training of future professionals.

Firstly, as the results showed, future development of data journalism in the field of sports news in Spain is facing one of its most information challenges which is to become established as a typical modus operandi for specialized professional editorial teams. This way, as indicated by different studies, it will follow in the footsteps of the media in other parts of the world, especially the Anglo-Saxons (Rojas-Torrijos and Rivera, 2016: 173), but also other countries like Germany (Horky and Pelka, 2017) or Scandinavia (Fink and Anderson, 2015), where sports data journalism is already established in the daily cycle of news production as an emerging field of innovation which has brought new ways of telling stories, gathering information and broadcasting news (Segel and Heer, 2010; Borges-Rey, 2016).

However, for the editorial team to implement this, data journalism must also become part of the professional mindset. They must explore the new possibilities different technologies provide to make such journalism possible and also accept new ways of thinking and working as an integral part of the innovation process (Gynnild, 2013). Therefore, data journalism will be transferred and "made much more visible to news consumers" (Bradshaw, 2015: 202).

Obviously, this is a process which will take time and entail training future professionals at universities, who have responded to the questions in this research and also state the need to improve training on this matter in the study plans for Journalism. As stated by different authors, Journalism studies must keep abreast of changes in technology, and requires permanent and critical updating which is in keeping with the new digital panorama and the reality of media companies (López, 2012; Saavedra, Grijalba and Pedrero, 2018).

While this is gradually implemented, professionals will clearly perceive the important potential which lies in data journalism for providing new approaches and for covering sports events, especially football, given the large volume of data that revolve around competitions and also the cyclical and repetitive nature of matches. The statistical character of football competitions in recent years has helped develop automization technologies from the generation of natural language by many sports editorial teams, which produce reports of roboticized data (Graefe, 2016; Túñez, Toural and Cacheiro, 2018; Rojas-Torrijos and Toural, 2019).

As it is added to the creation and distribution of contents processes in the Spanish sports media, this technology also means professionals must accept new methods and work routines based on handling and analysing data with informational purposes, so that, later on, these can be effectively communicated to their respective audiences. Specifically, the statistical model presented by *El País* gave a journalistic value to probability and tried to innovate when seeking how uncertainty could be communicated, but different degrees of acceptance and comprehension among users were found.

The research developed herein, despite covering the phenomenon of sports data journalism from a new perspective, such as the degree to which it is perceived as a specific journalistic initiative by its target audiences, has been limited in scope. Firstly, the sample of the audience that participated in this research could be extended to students at other universities in Spain, so that different perceptions of the same news category for covering competitions may be shared and evaluated.

Moreover, the study was based on just one situation which, although it is innovative and concerns a leading newspaper such as *El País* in important coverage such as the Football World Cup, it is susceptible to comparison with other similar models developed by different communication media in their sports coverage, not just football. In this respect, future lines of research should be set on using data journalism for other important sports events such as the Olympic Games or world cups in other disciplines.

5. Conclusions

The results enabled the state of play for data journalism for the sports editorial teams in Spanish media to be confirmed via the responses gathered to the two questions set at the onset of this research.

Firstly, the apparition of data journalism as a journalistic technique within the field of sports news in Spain is still in its fledgling years and, generally speaking, it seems that typical consumers of this type of contents, as stated in their responses to the questionnaires made, have still not become used to news spots whose main features are data analysis and visualizations of statistical variables. Due to this reality, these contents are often perceived as so inaccessible and complex, they seem to be difficult to understand.

Perhaps the most important point is it is not just amateur consumers in university education who are still in the process of coming to grips with this new category of journalistic work, but also media professionals themselves do not seem to agree on the meaning of data journalism, its potential for the present and future, as well as its value and news purposes.

In this respect, the innovative nature of the results prediction model developed by *El País* during the Football World Cup in Russia came up against an audience that was neither sufficiently accustomed to it nor prepared to fully take in and interpret it, nor was it given the journalistic value the project promoters, especially Kiko Llaneras, had hoped for by setting this innovative model in motion.

However, the evaluation of the model made by *El País* on comparing it with other similar ones which other media and international companies developed for the World Cup was very positive. From the newspaper itself, they argued that their model was finely gauged, since its initial predictions of low probability corresponded with event which almost never happened. Also, it never promised to be more accurate than it actually was. Llaneras himself made it clear that one thing is calculating probabilities which have a news value and another very different thing is making mere conjectures and perhaps this was one of the reasons why the model was not well understood.

Secondly, the study results supported the idea that data journalism is an upcoming category of work which has great potential for development within Spanish sports journalism in forthcoming years. The different audiences consulted agreed, stressing the valuable contributions which may be provided by a data analysis of sports coverage, which not just helps create new stories and ways of building and telling information, but also the graphic resources created, enrich them visually.

In short, a greater future implementation of data by the editorial teams of sports news in Spain would imply tackling coverage of the main events from an original and innovative perspective. This potential does not just mean sports information is moving towards accurate, quality journalism, but also makes its contents more attractive and arouses greater interest among its most typical consumers.

6. Acknowledgment

We thank Toby Wakely for his technical assistance in translation.

7. Bibliographical references

[1] Arias-Robles, F. (2017). Nuevas narrativas digitales en el periodismo deportivo. En J. L. Rojas-Torrijos (Coord.), *Periodismo deportivo de manual* (pp. 203-232). Valencia: Tirant lo Blanch

[2] Arroyo, I.; Bravo, L. C.; Llinás, H. y Muñoz, F. L. (2014). Distribuciones Poisson y Gamma: Una Discreta y Continua Relación. *Prospectiva*, *12*(1), 99-107. http://doi.org/dp72

[3] Borges-Rey, E. (2016). Unravelling Data Journalism. *Journalism Practice, 10*(7), 833-843. http://doi.org/dp73

[4] Bradshaw, P. (07/07/2011). The inverted pyramid of data journalism [Blog]. *Online Journalism Blog*. https://bit.ly/2wum7sy

[5] Bradshaw, P. (2015). Data Journalism. In L. Zion & D. Criag (Eds.), *Ethics for Digital Journalists: Emerging Best Practices* (pp. 202-218). Nueva York/Abingdon: Routledge

[6] Crucianelli, S. (2013). ¿Qué es el periodismo de datos? *Cuadernos de Periodistas, 26*, 106-124. http://bit.ly/32TQCUw

[7] Dader, J. L. (1997). *Periodismo de precisión. Vía socioinformática de descubrir noticias*. Madrid: Síntesis

[8] Dixon, M. J. & Coles, S. G. (1997). Modelling Association Football Scores and Inefficiencies in the Football Betting Market. *Journal of Applied Statistics, 46*(2), 265-280. http://doi.org/bc265q

[9] Ferreras, E. M. (2013). Aproximación teórica al perfil profesional del periodista de datos. *Icono 14*, *11*(2), 115-140. http://doi.org/dp74

[10] Fink, K. & Anderson, C. W. (2015). Data Journalism in the United States. *Journalism Studies, 16*(4), 467-481. http://doi.org/dp75

[11] Glickman, M. E. & Jones, A. C. (1999). Rating the chess rating system. *Chance, 12*(2), 21–28. https://bit.ly/2J1BdZt

[12] Graefe, A. (2016). *Guide to automated journalism*. New York: Tow Center.

[13] Groll, A.; Kneib, T.; Mayr, A. & Schauberger, G. (2016). *Who's the Favourite? A Bivariate Poisson Model for the UEFA European Football Championship 2016*. Technical Report Number, 195. Munich: Institut für Statistik, Universität München. http://doi.org/dp76

[14] Gynnild, A. (2013). Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets. *Journalism, 15*(6), 713-730. http://doi.org/dp77

[15] Horky, T. & Pelka, P. (2017). Data Visualisation in Sports Journalism. *Digital Journalism, 5*(5), 587-606. http://doi.org/gf3mwn

[16] Hvattum, L. M. & Arntzen, H. (2010). Using ELO ratings for match result prediction in association football. *International Journal of Forecasting, 26*(3), 460-470. http://doi.org/fgs4rg

[17] Igartua, J. J. y Humanes, M. L. (2009). *Teoría e investigación en comunicación social*. Madrid: Síntesis

[18] Kang, M. (15/06/2015). Exploring the 7 Different Types of Data Stories. *Mediashift*. https://bit.ly/2vErEMR

[19] Jackson, M. (2015). Data Journalism in the UK: a preliminary analysis of form and content. *Journal of Media Practice, 16*(1), 55-72. http://doi.org/dp78

[20] Llaneras, K. (10/09/2018). Evaluando nuestras predicciones durante el Mundial de Rusia. *Medium*. http://bit.ly/2QETRuy

[21] López, X. (2012). La formación de los periodistas para los entornos digitales actuales. *Revista de Comunicación, 11*, 178-195. https://bit.ly/33BfctO

[22] Marrero-Rivera, O. (2011). *Fundamentos del periodismo deportivo*. San Juan: Terranova

[23] Mayer-Schönberger, V. y Cukier, K. (2013). *Big data. La revolución de los datos masivos*. Madrid: Turner Publicaciones

[24] Rodríguez-Gómez, G.; Gil-Flores, J. y García-Jiménez, E. (1999). *Metodología de la investigación cualitativa*. Málaga: Aljibe

[25] Rogers, S. (2013). *Facts are sacred. The power of data*. Londres: Guardian Books

[26] Rogers, S. (2014). Data journalism is the new punk. *British Journalism Review, 25*(2), 31-34. http://doi.org/dp79

[27] Rojas-Torrijos, J. L. y Rivera, A. (2016). El Español y El Confidencial, exponentes del periodismo deportivo de datos en los medios nativos digitales españoles. *Revista Doxa Comunicación, 23*, 171-193. http://doi.org/dp8b

[28] Rojas-Torrijos, J. L. y Toural, C. (2019). Periodismo deportivo automatizado. Estudio de caso de AnaFut, el bot desarrollado por El Confidencial para la escritura de crónicas de fútbol. *Revista Doxa Comunicación,* 29, 235-254. http://doi.org/dp8c

[29] Saavedra, M.; Grijalba, N. y Pedrero, L. M. (2018). Hacia una redefinición de las competencias y perfiles profesionales del comunicador audiovisual en el ecosistema digital. *Doxa Comunicación, 27*, 369-385. http://doi.org/dp8d

[30] Segel, E. & Heer, J. (2010). Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics, 16*(6), 1139-1148. http://doi.org/dnqx5m

[31] Silver, N. (17/08/2015). Nate Silver Comes Full Circle With Sports Data Journalism Mecca. *Sporttechie.com*. https://bit.ly/2WwFtbb

[32] Túñez, J. M.; Toural, C. y Cacheiro, S. (2018). Uso de bots y algoritmos para automatizar la redacción de noticias: percepción y actitudes de los periodistas en España. *El Profesional de la Información, 27*(4), 750-758. http://doi.org/dp8f

[33] Veglis, A. & Bratsas, C. (2017a). Reporters in the age of data journalism. *Journal of Applied Journalism & Media Studies,* 6(2), 225-244. http://doi.org/dp8g

[34] Veglis, A. & Bratsas, C. (2017b). Towards a taxonomy of data journalism. *Journal of Media Critiques,* 3(11), 109-121. http://doi.org/dp8h

Notes

1. The first explanation of the *El País* prediction model can be read by clicking on the following link: http://bit.ly/2QLQEJU

2. Throughout the World Cup *El País* made up to four updates more of its model. These were as follows:

    a. "Las opciones de cada selección para estar en octavos del Mundial" /"The options for each national team to be in the knock-out phase of the World Cup" (2018, 24 th of June): http://bit.ly/37TtFCq

    b. "De los equipos en octavos, ¿cuál es el favorito para ganar el Mundial?"/"Out of the teams in the knock-out phase, which is the favourite to win the World Cup?" (2018, 29th of June): http://bit.ly/2tQm495

    c. "El "big data" es el pulpo Paul del Mundial de Rusia de 2018 y decía que España será subcampeona" /"Big data" is the Paul the Octopus of the 2018 Russia World and it said Spain would be sub champion" (2018, 1st July): http://bit.ly/36Puo7A

    d. "Los favoritos para ganar su cruce de cuartos y el Mundial"/"The favourites to win their cross of quarter finals and the World Cup" (2018, 4th of July): http://bit.ly/37Z8LBR

3. The *El Español* control chart which is part of the questionnaire can be consulted at the following link: http://bit.ly/2RcWZwT