

Article

Not peer-reviewed version

Lexicometric and Sentiment Analysis of News in the Spanish Press regarding HIV and Prison

[Enrique J. Vera-Remartínez](#) , [Jose Antonio Zafra-Agea](#) ^{*} , Julio Garcia-Guerrero , [María Pilar Molés-Julio](#)

Posted Date: 14 February 2024

doi: [10.20944/preprints202402.0771.v1](https://doi.org/10.20944/preprints202402.0771.v1)

Keywords: Procesamiento natural del lenguaje; Periódicos como tema; Síndrome de inmunodeficiencia humana; prisiones; España



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Lexicometric and Sentiment Analysis of News in the Spanish Press Regarding HIV and Prison

Enrique J. Vera-Remartínez ¹, José A. Zafra-Agea ^{2,*}, Julio García-Guerrero ³
and María Pilar Molés-Julio ⁴

¹ Penitentiary Center of Castellón I, 12006 Castellón, Spain; enriquevera@gmx.es

² Department of Nursing, Faculty of Health Sciences, UManresa, Fundació Universitària del Bages, Universitat de Vic, Universitat Central de Catalunya, 08242 Manresa, Spain.

³ The Bioethics Committee of the Castellón Health Department, E-12071 Castellón, Spain. garciaj@comcas.es

⁴ Department of Nursing, Faculty of Health Sciences, Universitat Jaume I, 12071 Castellón, Spain. mjulio@uji.es

* Correspondence: jzafra@umanresa.cat

Abstract: Objective To analyze the influence of HIV in penitentiary institutions through an exhaustive lexicometric and sentiment analysis of news in the Spanish written press related to HIV and prison. **Material and method** observational, descriptive, and retrospective design based on news published in Spanish newspapers from 1981 to 2020, linking HIV infection with penitentiary institutions. Natural language processing and lexicometry techniques were used, using the Python programming language and ATLAS Ti v.9 software. **Result:** 318 articles were analyzed, with 10,502 distinct words. The sentiment of news was evaluated in three different periods: the discovery of HIV, and the periods before and after the existence of antiretroviral treatments. N-gram analysis and word cloud creation were carried out for each period. In the initial phase, ignorance predominated, and prominent terms such as "homosexual," "isolate," "test," "trial," or "confirm" were observed. In the period before antiretroviral treatments, news reflected negative feelings and criticism of prison policy, with keywords such as "give birth," "drug," "carrier," or "disease." With the arrival of treatments, news showed an evolution towards a more positive tone, highlighting terms such as "treatment," "release," or "syringes". **Conclusions:** The presence of HIV in Spanish prisons has experienced a significant evolution from an alarming situation to an improvement in the care of inmates, driven by institutional adaptation and advances in antiretroviral treatments.

Keywords: natural language processing; newspapers as topic; human immunodeficiency syndrome; prisons; Spain

1. Introduction

This research focuses on understanding the influence of HIV in penitentiary institutions through lexicometric and sentiment analysis of Spanish written press news related to HIV and prison. Lexicometry, focused on the quantitative analysis of texts, facilitates the identification of linguistic patterns, key term frequencies, and the evaluation of word sentiment in specific contexts, thus optimizing the processing of textual data [1].

The emergence of HIV globally has been one of the most significant pandemics in history. Its severity was fully recognized at the Paris Summit on AIDS in 1994, with the participation of leaders from forty-two nations committed to addressing this health crisis [2].

In the Spanish penitentiary context, the spread of HIV was exacerbated due to the use of drugs by parenteral route. A study indicated that, in 1989-90, 46.2% of new inmates admitted to using this route of consumption, with shared needle use being the main risk factor for HIV infection. The prevalence of HIV-positive prisoners was 28.4% [3], although some other studies showed higher prevalences, 33.6% [4], and more than 60.0% of all HIV-positive individuals were or had been intravenous drug users, with heroin being the main drug of consumption by this route and to a lesser extent cocaine [3,4].



Healthcare in Spanish prisons at that time was precarious, characterized by obsolete facilities, limited resources, and a lack of motivated healthcare personnel, unable to carry out the prevention, promotion, and health education tasks that were at that time almost the only weapons against the pandemic [5].

Despite the intensity with which the pandemic affected prisons, there is widespread ignorance about its impact on the prison population. Although research has been conducted on HIV/AIDS in the media in Spain, its influence on penitentiary institutions has not been exhaustively addressed [6,7].

This study aims to analyze how the Spanish written press has addressed the relationship between HIV/AIDS and prison through lexicometric techniques. Relevant news published over forty years (1981-2020) will be collected and analyzed, providing valuable insights for addressing future health crises and preserving a crucial historical record.

2. Material and Method

An observational, descriptive, and retrospective design was adopted to analyze news published in Spanish newspapers from 1981 to 2020.

The newspapers selected for this study were *El País*, *ABC*, *La Vanguardia*, and *El Periódico Mediterráneo*. These media outlets, with printed editions and diverse editorial lines, have a daily circulation of over three hundred thousand readers [8], thus providing a national, regional, and local perspective on the researched topic.

The search was conducted in digitized newspaper archives through subscription to the consulted newspapers. Advanced search techniques were employed, including temporal selection by dates, content format, specific sections, and the combination of terms with Boolean operators and truncations. The descriptors used were: "Human Immunodeficiency Syndrome" or "AIDS," "HIV," "prisoners," "prisons," "jails," and "antiretrovirals."

For *El Periódico Mediterráneo*, access was obtained to the provincial historical archive of Castellón for the years 1981 to 1992 in paper format. Copies between 1993 and 2003, not yet digitized, were consulted at the library of the Jaume I University of Castellón. This search involved the complete review of approximately eight thousand thirty newspapers in paper format. From 2004 to 2020, the newspaper archive was consulted, which already has an electronic version available online.

As inclusion criteria, all news items linking HIV/AIDS with the prison population were collected. As exclusion criteria, news repeated on the same date were eliminated, giving preference to those from national newspapers over regional and local ones. Also excluded were those news items in which the search terms were unrelated to HIV/AIDS infection or the population under study.

The news items were classified into three periods based on scientific literature: the discovery period between 1981 and 1985; the period prior to the appearance of highly active antiretroviral treatments (pre-HAART) between 1986 and 1995; and the period after the appearance of highly active antiretroviral treatments (post-HAART), from 1996 to the present.

Several analyses were performed, including:

Sentiment Analysis: Natural Language Processing (NLP) was applied to determine sentiment, using algorithms that identify and label keywords in the text as positive, negative, or neutral.

N-gram Analysis: Sequences of "n" elements in the text were identified and counted, using tokenization techniques, text normalization, and filtering of meaningless words. Bigrams (grouping of two terms) and trigrams (grouping of three terms) were considered for this analysis.

Word Cloud Analysis: It is a text visualization technique used to display the most frequent words in a text corpus in a visual and easy-to-understand manner.

The analysis was carried out using the Python programming language and ATLAS Ti v.9 software.

Ethical Considerations: Ethical approval was obtained from the Committee on Ethics in Human Research (CEISH) at the Universitat Jaume I de Castellón, under file number "CEISH/64/2023." The project adhered to current regulations on data protection, including the European Regulation 2016/679 and the Organic Law 3/2018 on Personal Data Protection and guarantee of digital rights.

3. Results

A total of 318 newspaper articles addressing the relationship between HIV/AIDS and Spanish prisons were identified. Through lexiconmetric analysis, 10,502 distinct words were extracted from the analyzed corpus.

The news articles were categorized based on their sentiment: positive, neutral, or negative.

Figure 1 shows the average trend of all analyzed press based on the sentiments of the news over time. There is a central axis determining the neutrality of the news sentiment; above that axis and with positive values indicating positive sentiment of the news; and below the axis, with negative values, indicating a negative sentiment of the published news. It is observed that at the beginning and end of the pre-HAART period, a negative sentiment predominated. However, during the post-HAART period, the news generally reflected a positive sentiment, although a notable decline was observed around the year 2017.

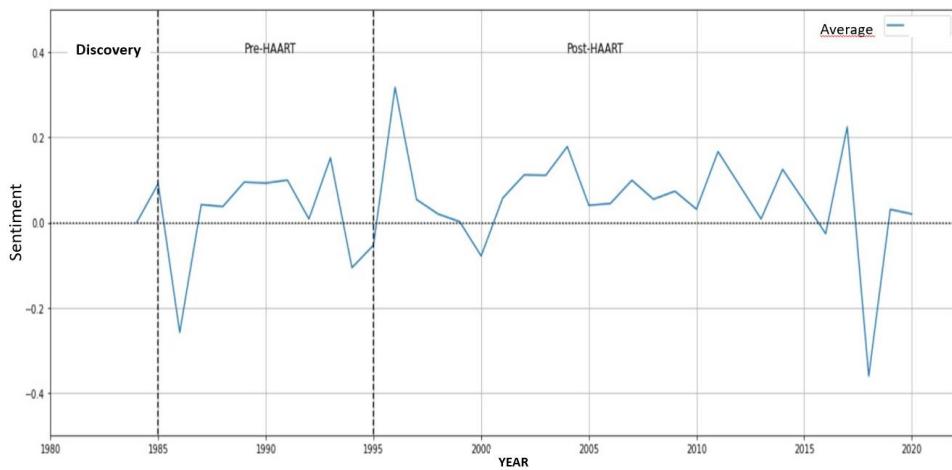


Figure 1. Average Analysis of News Sentiment by Periods.

In the n-gram analysis, the frequency of bigrams and trigrams from all recorded news articles was documented in an attempt to establish relationships between the most frequent terms and their possible interpretation.

The twenty most frequent bigrams were identified with a total of 1,304 occurrences, highlighting terms such as "penitentiary institutions," "inmate population," "penitentiary center," and "AIDS patients," among others, as shown in Figure 2.

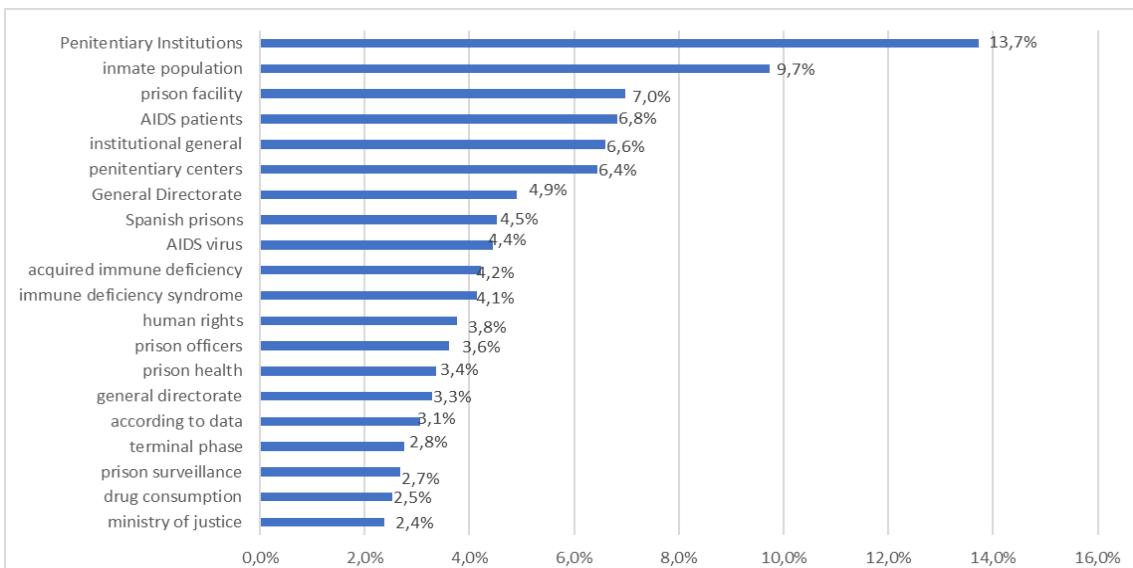


Figure 2. Frequencies of the twenty most repeated bigrams.

Figure 3 presents the most frequent trigrams, highlighting significant combinations of three terms with a total of 455 repetitions.

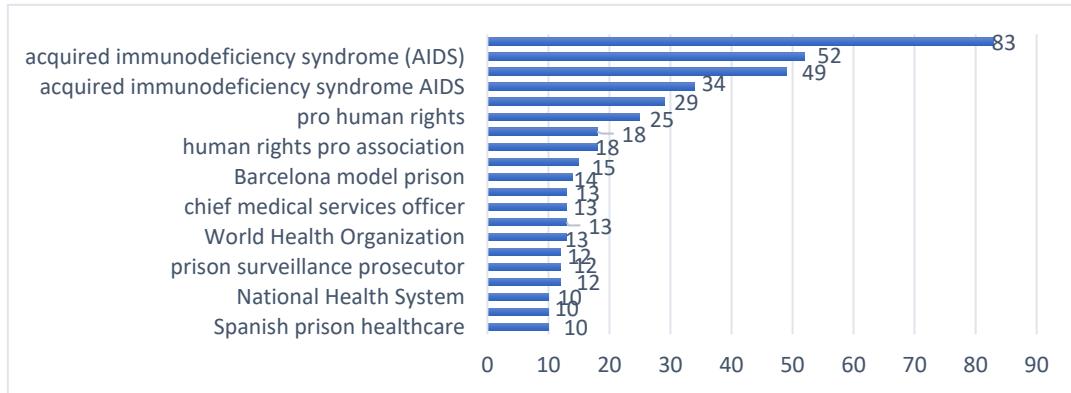


Figure 3. Frequencies of the twenty most repeated trigrams.

Discovery Period (1981-1985): Figure 4 presents the word cloud, highlighting general terms such as "AIDS," "inmate," "prison," and others. Less frequent but relevant terms were also observed, such as "homosexual," "isolate," "declare," "test," "trial," and "confirm."

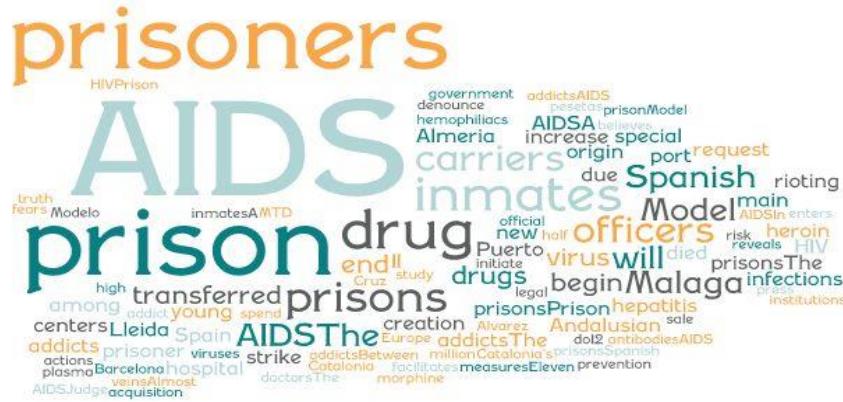


Figure 4. Word cloud in the discovery period (1981-1985).

Pre-HAART Period (1986-1995): Figure 5 shows the word cloud, highlighting terms such as "prisioner", "infect", "drug".



Figure 5. Word cloud in the pre-HAART period (1986-1995).

Post-HAART Period (1996-2020): Figure 6 illustrates the word cloud highlighting terms such as "treatment", "prision", or "aids".

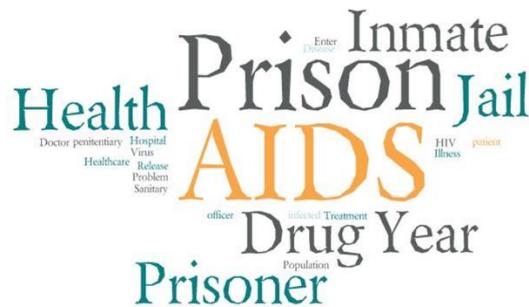


Figure 6. Word cloud in the post-HAART period (1996-2020).

The lexicometric analysis of the headlines reveals significant patterns in both bigrams and trigrams related to health in prisons. The most frequent bigrams include "immune deficiency syndrome," "acquired immune deficiency," and "AIDS virus," while prominent trigrams are "acquired immunodeficiency syndrome (AIDS)" and "Director General of Penitentiary Institutions."

When calculating the Pearson correlation between bigrams and trigrams, a variety of relationships are observed. For bigrams, correlations range from 0.304 to 0.861, whereas for trigrams, they span from 0.263 to 1.639. These findings suggest specific thematic associations or cause-and-effect relationships in headlines related to health in prisons.

Together, these findings underscore the complexity of lexical interactions in the realm of prison health and highlight the importance of considering both bigrams and trigrams to better understand emerging themes and dynamics in this field.

4. Discussion:

The exhaustive analysis of 318 newspaper articles addressing the relationship between HIV/AIDS and Spanish prisons has provided a detailed insight into the evolution of media representation over time. The extraction of 10,502 distinct words from the analyzed corpus reveals the complexity and diversity of discourse on this topic in the media.

The sentiment analysis, as depicted in Figure 1, highlights that during the discovery period, there were few collected news items, generally portraying a positive sentiment despite the significant lack of knowledge about this infection. It was when the first cases of infection began to appear in penitentiary centers [9]. By October 1985, a test to determine the presence of antibodies against HIV was available, although confirmation of positive cases was emphasized [10]. The performance of the first voluntary analytical determinations in penitentiary centers, such as in Puerto de Santa María, is reported [11].

During the pre-HAART period, there was substantial writing. The negative sentiment at the beginning of this period was in line with the circumstances of the time. Provincial prisons were old facilities with unhealthy dampness, limited capacity, fairly precarious healthcare, and a significant increase in population [5]. News during this period mainly criticized prison policy and the situation experienced in these centers regarding overcrowding, hygiene, fear of contagion, and protests from both inmates and prison staff. Newspapers like *La Vanguardia* and *ABC* filled pages on these topics with a negative sentiment [12,13]. Also, towards the end of the pre-HAART period between 1994-1995, despite improvements in the creation of new penitentiary centers, reforms in older ones, the establishment of medical services with exclusive dedication, and other measures adopted, there was a new negative sentiment in the news due to the increase in AIDS cases and, simultaneously, the rise in associated mortality, which became the leading cause of death among the prison population [14]. Coinciding with other studies on written press and HIV, the press generally behaved in these years with a somewhat alarmist attitude, mainly facing a disease that was fatal at the time, with no effective treatments available, and acquiring it meant almost a death sentence [7].

During the post-HAART period, the sentiment of the news was generally positive. This period was characterized by significant improvements in healthcare for inmates, the beginning of the

implementation of telemedicine with the first radiological diagnoses [15]; in Catalonia, a regulation was published to humanize prisons dependent on its community, improving healthcare and pharmaceutical services [16]; prison healthcare was awarded by the WHO regional office with European awards for good health practices, in Spanish prisons such as El Dueso, Pamplona, and Alicante [17]; custody judicial hospital units were expanded in public hospital networks [18]; in other words, a ray of hope and therefore, in sentiment analysis, it is a period with predominance of news with positive sentiment. The end of the post-HAART period ends with negative sentiment news, among other reasons, due to limitations in the acquisition of certain drugs by penitentiary centers in their request to the General Directorate of Prison Health, with the latter demanding the requirement of visas for their acquisition [19], without any health criterion explaining it. The increase in deaths among inmates due to drug consumption, becoming the second leading cause of mortality in Spanish prisons, also conditioned this negative sentiment [20]. Other news that contributed to this negative sentiment trend were those regarding the decrease in the number of doctors in penitentiary centers, which began to foreshadow an uncertain future [21].

N-grams analyses, particularly bigrams and trigrams, have allowed the identification of more recurrent terms in the news. The two most repeated joint terms are "penitentiary institutions"; other terms such as "inmate population," "penitentiary centers," or "Spanish prisons" logically point to the scope of the news. In fourth place, terms such as "AIDS patients" appear, alluding to the pathology caused by the virus and referred to inmates. Terms like "AIDS virus" or "acquired immunodeficiency" would also be included. In the early years, there was a lot of writing, with a great lack of knowledge about the differences between the concepts of infection and disease, appearing terms like seropositives, carriers, or AIDS patients used interchangeably. It was even written that several children were born with the disease [22]. Next comes the bigram "human rights," which seems to reflect a denunciation of the situation experienced in penitentiary centers regarding what both the written press and non-governmental organizations estimated as a "thinning" of the right to health protection of inmates. Figures such as "prison officers" appear in the thirteenth position, or the "general director" in the fifteenth. These repetitions are mostly due to all the demands and pressure measures initiated by the collective of prison workers, reflected in writing in the press of the time [23,24], and on the other side is the maximum representative of prison administration in all these conflicts, which is the general director. Entities like "prison healthcare" referred to the healthcare collective working with these patients, in fourteenth place. "Prison surveillance" or "Ministry of Justice" also appear frequently, as it was the ministry that organically depended on all Spanish prisons, except those in Catalonia, until the spring of 1996.

The bigram "terminal phase" also appears quite frequently, mentioning different stages or phases of the disease, such as stage IV, the most advanced. This bigram appears mainly in the pre-HAART period [25]. It also reflects one of the main causes of contagion in prison, such as "drug use" [26,27]. Among the trigrams, the ones occupying the first four places are the different combinations of the terms "general director of penitentiary institutions" and "human immunodeficiency syndrome," AIDS. Therefore, reference is made to these terms as key themes in many published news. Then come combinations of different penitentiary centers that, in order of repetitions, would be: Puerto de Santa María, Modelo prison in Barcelona, and Castellón penitentiary center. The first two were centers with a long previous trajectory that were protagonists of frequent newsworthy incidents at the beginning of the pandemic, due to protests by inmates demanding improvements in care.

The third due to the influence of a local media, which refers any news on the topic to its closest environment. As positions, "prison surveillance judge" and "medical service chief" appear, characters with great influence in the development of conditional freedoms due to illness, a circumstance that was quite frequent among the group of prisoners with AIDS before the appearance of antiretroviral treatments. They end up in last place with "terminal AIDS phase," a terminology widely used in news publications about conditional freedoms due to terminal illness. And two more terms, "Spanish prison healthcare," corresponding to news from the Spanish Society of Prison Health, which stood out mainly in the post-HAART stage, having a certain

5. Conclusion

The comprehensive analysis conducted on the relationship between HIV/AIDS and Spanish prisons through the examination of newspaper articles spanning several decades has provided invaluable insights into the evolution of media representation and societal perceptions surrounding this complex issue.

The study's findings underscore the profound impact of historical contexts, healthcare advancements, and policy changes on media sentiment and discourse regarding HIV/AIDS within prison settings. From the early stages marked by fear and misinformation to the later periods characterized by improved healthcare and evolving treatment options, the media has played a crucial role in shaping public understanding and responses to the epidemic in carceral environments.

Moreover, the utilization of lexicometric techniques such as bigrams and trigrams has allowed for the identification of recurring themes and key terms across different time periods, shedding light on the predominant narratives and focal points within media coverage. Terms like "penitentiary institutions," "AIDS patients," and "human rights" emerge as central themes, reflecting broader societal concerns and debates surrounding healthcare access, inmate rights, and public health policy.

Overall, this study not only offers valuable historical insights into the intersection of HIV/AIDS and the prison system but also underscores the importance of ongoing research and dialogue in addressing the multifaceted challenges and emerging needs in the prevention, treatment, and management of HIV within carceral contexts. By understanding the past representations and responses to the epidemic, policymakers, healthcare providers, and advocacy groups can better inform future interventions and initiatives aimed at promoting health equity and social justice for incarcerated individuals affected by HIV/AIDS.

Author Contributions: Conceptualization Enrique V., methodology Enrique V. and José Zafra; formal analysis, Enrique V., Pilar M., Julio G., and José Zafra; investigation, all authors; data curation, SIP; writing—original draft preparation and writing—review and editing, all authors. All authors have read and accepted the published version of the manuscript.

Funding: This research did not receive external funding.

Institutional Review Board Statement: The protocol was approved by the Ethics Committee of the Universitat Jaume I, 12071 Castellon, Spain (file number: CEISH/64/2023, 19/05/2023), and each participant signed a written informed consent.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available upon request to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

*Print media is one of the parallel sources of information in this research.

References

1. Sastriano, C.; Moscoloni, N. Importancia Del Análisis Textual Como Herramienta Para El Análisis Del Discurso. Aplicación En Una Investigación Acerca de Los Abandonos Del Tratamiento En Pacientes Drogodependientes. *Rev. Epistemol. Ciencias Soc.* **2000**, 287–306.
2. WHO Cumbre de Paris Sobre El SIDA; 1994;
3. Martín Sánchez M Programa de Prevención y Control de Enfermedades Transmisibles En Instituciones Penitenciarias (Monográfico de Sanidad Penitenciaria). *Rev. Estud. Penit.* **1990**, Extra, 51–67.
4. Martín, V.; Bayas, J.M.; Laliga, A.; Pumarola, T.; Vidal, J.; Jimenez de Anta, M.T.; Salleras, L. Seroepidemiology of HIV-1 Infection in a Catalonian Penitentiary. *AIDS* **1990**, 4, 1023–1026, doi:10.1097/00002030-199010000-00013.
5. Gámir Meade R *Los Facultativos de Sanidad Penitenciaria*; Editorial Dykinson: Madrid, 1995; ISBN 84-8155-094-9.
6. Martín-Hernández, R. El Sida Ante La Opinión Pública : El Papel de La Prensa y Las Campañas de Prevención Estatales En La Representación Social de SIDA En España. *Rev. humanidades* **2009**, 15, 237–268.

7. Sáez Aramburo, M. del M. Evolución de Los Contenidos Sobre SIDA En La Prensa Escrita Española. *Rev. Española Comun. en Salud* **2014**, 5, 32–55.
8. Orús, A. Medios de Comunicación y Marketing. Mercado Editorial. Número de Lectores de Los Principales Periódicos Españoles En 2021 Available online: <https://es.statista.com/estadisticas/476795/periodicos-diarios-mas-leidos-en-espana/> (accessed on 11 August 2022).
9. Marchena, D. Un Estudio Oficial Revela El Alto Riesgo de Los Presos de La Modelo Ante El SIDA. *La Vanguard*. 1985, 1.
10. Mariño, C. Un Test Positivo Debe Ser Confirmado Con Una Segunda Prueba. *El País* 1985, 1.
11. Agencia EFE Los Reclusos Del Puerto, Sometidos a Pruebas. *ABC* 1985, 1.
12. Bassols, A.M. La Verdad de Nuestras Cárcel. *La Vanguard*. 1988, 1.
13. Adriano SIDA En Las Cárcel. *ABC* 1987, 1.
14. García-Guerrero, J.; Vera-Remartínez, E.; Planell-Ramos, M.V. Causas y Tendencia de La Mortalidad En Una Prisión Española (1994-2009). *Rev. Esp. Salud Pública* **2011**, 3, 245–255.
15. Estalella, A. Los Presos Reciben El Diagnóstico Radiológico Por Red. *El País* 2002, 1.
16. Rios, P. Aprobado Un Reglamento Que Humanizará La Vida En Las Prisiones. *El País* 2006, 1.
17. Girona, C. Premio a La Sanidad Penitenciaria Española. *El País* 2006, 1.
18. Bengoa, A. 41 Hospitales Cuentan Con Unidades de Custodia Hospitalaria Para Presos. *El País* 2016, 1.
19. Rincón, R. Interior Instaura Visados Para Limitar Los Medicamentos de Los Presos. *El País* 2016, 1.
20. Las Muertes Por Consumo de Drogas En Las Cárcel Se Duplican En Un Año. *El País* 2019, 1.
21. De Benito, E. El 37% de Las Plazas de Médicos de Prisiones Están Vacantes. *El País* 2019, 1.
22. Novo, C. Entre El Treinta y El Cuarenta Por Ciento de Los Presos En Cataluña Son Portadores de Anticuerpos. *La Vanguard*. 1987, 1.
23. Continua El Encierro de Funcionarios de Prisiones En Protesta Por El SIDA. *ABC* 1987, 1.
24. Funcionarios de Prisiones Piden Cárcel Especiales Para Los Enfermos de SIDA. *El País* 1987, 1.
25. De la Orden, P. Alta Tensión En Las Cárcel Españolas Por La Propagación Del SIDA Entre Los Reclusos. *ABC* 1987, 1.
26. Europa-Press La Mitad de Los Reclusos Españoles Son Toxicómanos Según Antoni Asunción. *ABC* 1990, 1.
27. Agencias El 55% de Los Presos En España Son Adictos a Drogas. *El Periódico Mediterráneo* 1997, 1.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.