Valderrama-Zurián, J. C., García-Zorita, C., Marugán-Lázaro, S., Sanz-Casado, E. (2021). Comparison of MeSH terms and KeyWords Plus terms for more accurate classification in medical research fields. A case study in cannabis research. *Information Processing & Management*, 58(5), 102658

DOI: https://doi.org/10.1016/j.ipm.2021.102658

Comparison of MeSH terms and KeyWords Plus terms for more accurate classification in medical research fields. A case study in cannabis research

Juan C. Valderrama-Zurián[a,b], Carlos García-Zorita[c,d], Sergio Marugán-Lázaro[c,d], Elías Sanz-Casado[c,d].

[a]. Department of History of Science and Documentation. University of Valencia. Spain.

[b]. Social and Health Research and Information Unit. UISYS Group. University of Valencia. Spain.

[c]. Department of Library Science and Documentation. Carlos III University of Madrid. Spain.

[d]. Research Institute for Higher Education and Science (INAECU). Carlos III University of Madrid, Autonomous University of Madrid. Spain.

Corresponding Author

Elías Sanz-Casado

elias@bib.uc3m.es

INAECU Institute (Research Institute for Higher Education and Science), Carlos III University of Madrid, Calle Madrid 126, Getafe, 28903, Madrid, Spain

Highlights

- MeSH terms and KeyWords Plus terms show different distributions.
- Automatic classification based on MeSH terms and KeyWords Plus terms is uneven.
- Automatic classification of work on humans displays better concordance.
- Caution must be exercised when considering the use of grey boxes with KW Plus.
- There are major restrictions on the use of KeyWords Plus in the biomedical field.

Abstract

KeyWords Plus and Medical Subject Headings (MeSH) are widely used in bibliometric studies for topic mapping. The objective of this study is to compare the two description systems in documents about cannabis research to find the concordance between systems and establish whether there is neutrality in topic mapping. A total of 25,593 articles from

1970 to 2019 were drawn from Web of Science's Core Collection and Medline and analyzed. The tidytext library, Zipf's law, topic modeling tools, the contingency coefficient, Cramer's V, and Cohen's kappa were used. The results included 10,107 MeSH terms and 28,870 KeyWords Plus terms. The Zipf distribution of the terms was different for each system in terms of slope and specificity. The documents were classified into seven topics, and the MeSH system proved better at classification. The kappa coefficient between the two systems was 0.477 (for gamma ≥ 0.2); the topics related with human beings presented higher concordance. The use of KeyWords Plus for topic analyses in biomedical areas is not neutral, and this point needs to be taken into account in interpreting results.

Keywords
MeSH terms; KeyWords Plus; subject mapping; classification systems; cannabis research

1. Introduction

Knowledge mapping can be used in the context of knowledge translation in health sciences to identify areas where research is being done, investigative fronts, emerging trends, and research gaps (Synnestvedt, Chen & Holmes, 2005; Ebener et al., 2006), and therefore knowledge mapping has become a challenge for scientific information professionals (Chen, 2016). Bibliometric studies can be used to obtain mapping knowledge, hotspot maps, trend maps, main lines of research, emerging trends, and research gaps in a topic area (Glänzel, Moed, Schmoch & Thelwall, 2019).

The scientific community is aware of the need for bibliometric studies. The publication of bibliometric studies addressing health science knowledge is on the rise, as corroborated by the facts that on February 25, 2020, PubMed/Medline had 1,553 records published between 1995 and 2019 with the term "bibliometric *" in the title, and the number of records had increased from 45 in the first five years of that period to 994 in the last five years. Of those records, 23 contained the terms "mapping" and "knowledge" in the title, rising from four records in the first five years to 57 in the last five years.

Some bibliometric topic studies are based on the topic categories assigned to journals (Leydesdorff, Comins, Sorensen, Bornmann & Hellsten, 2016). However, such studies may have gaps, for example, when documents published in multidisciplinary journals are misassigned or when a document whose topic is different from the publishing journal's topic is assigned to a knowledge area. For this reason, it was decided to base our article content analysis on author keywords, title words, abstracts, full text, Medical Subject Headings (MeSH), and Web of Science's KeyWords Plus (KW Plus), which are words or phrases that appear frequently in the title or in the bibliographic references (Garfield & Sher, 1993).

Computational semantics methods like Latent Semantic Analysis (LSA) and topic models are used to analyze topics and obtain terms (Natale, Fiore & Hofherr, 2012). For example, the literature on delirium has been studied through analysis of MeSH terms and probabilistic topic modeling of abstracts (McCoy, 2019), and Latent Dirichlet Allocation (LDA) has been used for unsupervised topic identification in the scientific eHealth literature (Drosatos & Kaldoudi, 2020).

Furthermore, topic modeling tools are being supplemented and sometimes replaced in language model analysis nowadays by other tools that incorporate different approaches, such as the inclusion of subjectivity in language, by analyzing personal texts in social media and including for analysis viewpoints such as that of the emotions associated with the words used (Ríssola, Aliannejadi, & Crestani, 2020).

Certain techniques known as word and document embedding are being used to complement analyses based on topic modeling. While embedding techniques use neural network algorithms to create neural embedding models, topic modeling uses LSA, as said before. Topic modeling and neural embedding differ in the context information they use; while topic modeling uses documents, neural embedding uses words as context. This produces different semantic representation models (Sahlgren, 2015). In topic modeling, semantic representation needs to be interpreted by the end user. Furthermore, if terms lack semantic description, they may be incorrectly retrieved due to their ambiguity (Lashkari, Bagheri, & Ghorbani, 2019). However, topics can be labeled automatically by means of neural embedding (Bhatia, Lau, & Baldwin, 2016), or they can help reveal underlying semantic representations of approaches that explore the connections between topic modeling and word embedding, like the neural embedding allocation (Keya, Papanikolaou, & Foulds, 2019).

Our choice is intended to facilitate knowledge mapping by using unsupervised models, where no training data are necessary, inasmuch as the objective is to highlight the problem created by choosing to use one type of indexing (KW Plus) unrestrictedly as opposed to another (MeSH).

The reason for comparing these two indexing systems is that researchers looking for health science documents for analysis commonly use either databases that are specific to the health science area (e.g., PubMed/Medline) or multidisciplinary databases (e.g., Web of Science). Pubmed/Medline is used because it includes Medical Subject Headings, which facilitate searching through a controlled language, providing among other things a powerful method of narrowing results and homing in on what the searcher needs (Shultz, 2007), as well as increasing search effectiveness (Liu & Wacholder, 2017). The commonly used multidisciplinary databases include Science Citation Index Expanded and the Social Science Citation Index from Web of Science's Core Collection, which indexes the world's top-quality journals based on the citations they receive and provides the journal impact factor, the h-index, and other bibliometric indicators (Dettori, Norvell, & Chapman, 2019; Moed, 2009).

The more frequent use of PubMed/Medline and Web of Science in bibliometric studies is reflected by a search in Pubmed Central[1] with the following equation: "bibliometrics" AND (Web of Science OR Medline OR Scopus OR Embase OR Google Scholar). The equation appears in 2,654 of the 4,035 PubMed/Medline records (65.8%), 2,313 of Web of Science's records (57.3%), 1,351 of Scopus's records (33.5%), 908 of Google Scholar's records (22.5%), and 514 of Embase's records (12.7%).

Because these are the databases most often used to evaluate scientific activity in health sciences, the terms these databases use for indexing or describing documents are the words used to represent relationships in knowledge maps. However, since the terminological systems are different, there is a research gap concerning which system of terms affords a better representation of research in the health science area. This paper reports an endeavor to determine accurately which of the two systems of terms, MeSH or KW Plus, is better for representing knowledge maps.

2. Research questions

The objective of this study is to compare two description systems, MeSH and KW Plus, in documents on biomedical subjects to find the level of concordance between the two systems and to establish empirically if there is neutrality in the use of one system or the other to create topic maps.

Moreover, the following four research questions form part of the present study:

---

[1] PubMed Central, a free full-text archive of biomedical and life science journal literature at the U.S. National Institutes of Health's National Library of Medicine.

1. Are term distributions in the MeSH and KW Plus description systems similar?
2. How are documents classified by MeSH and KW Plus terms?
3. What is the degree of concordance between classifications obtained from MeSH terms and classifications obtained from KW Plus terms?
4. Should we use MeSH terms or KW Plus terms for topic mapping?

This study is based on a case study of cannabis research from a multidisciplinary perspective that includes articles on the cannabis plant, its chemical compounds, cannabinoid receptors, drugs, synthetic cannabinoids, laboratory analyses, health effects, use prevention, and legalization.

3. Related Work

To our knowledge this is the first comparative study of the MeSH and KW Plus systems to ascertain which system is better for classifying topic areas. Previous bibliometric studies have taken different approaches to creating global maps of science, obtaining comparable results at higher levels of granularity (Thijs, 2019); others have employed artificial intelligence to generate new sentences from Title and Abstract Terms and used them to find the most important scientific ideas (Jiang et al., 2019); some have used Author Keywords and KW Plus (Khasseh, Soheili, Moghaddam, & Chelak, 2017; Tran et al., 2018). Still others have added the keywords from Title to determine the main topics of social networks (Maltseva & Batagelj, 2020). Meanwhile, other studies have used Wos' Topic Subject and PubMed's Title and MeSH terms for thematic analysis and to trace the evolution of the use of social media in the field of healthcare research (Chen, Lun, Yan, Hao, & Weng, 2019). Another approach is to analyze an area's MeSH terms using a search in Web of Science (WoS), and afterward to use the documents' DOI or PMID to download the related Medical Subject Headings. Some studies in the neuroscience disciplines and related areas thus utilize the elaborate MeSH indexing system of individual publications in the PubMed database to identify potentially relevant publications (Kocak, García-Zorita, Marugán-Lázaro, Çakır, & Sanz-Casado, 2019).
These studies may present limitations, however, when it comes to classification. It has not yet been determined which of the systems would best classify topic areas. If only KW Plus is used, it may overestimate aspects that the article does not really deal with. For example, in the case of the present study, an article on the analysis of cannabis addiction could be based on studies of other substances, such as opioids or cocaine; then

“opioids” and “cocaine” would appear frequently in the bibliography and therefore in KW Plus. In addition, a percentage of the documents in the Web of Science do not have KW Plus and author keywords, so a large proportion of documents would be underrepresented. In fact, in previous studies only 37% of the documents had both terms (Khasseh et al., 2018), and approximately 50% of the records did not have one or the other (Tripathi, Kumar, Sonker, & Babbar, 2018). There may also be cases where the authors fail to select the best keywords for their paper and so cause the terms to become distorted in the classification process. A recent study has observed that searching by title, abstract and keywords would not be a good enough method for retrieving epidemiological research papers (de Vries, van Smeden, Rosendaal, & Groenwold, 2020).

On the topic of classification comparison, some studies have compared cluster viewing of MeSH term co-occurrence and time-zone viewing of WoS document-term cocitation (Synnestvedt et al., 2005), and others have compared KW Plus and author keywords to find research trends in patient adherence to the treatment, finding that both systems show similar research trends, although KW Plus emphasizes research methods and techniques more than Author Keywords do (Zhang et al., 2016). Some studies have analyzed cited references and Medical Subject Headings as two different knowledge representations (Leydesdorff et al., 2016). Nentidis et al. (2020) have tried to add semantic value to the descriptive capacity of MeSH terms so as to better satisfy the information needs of experts in certain areas of biomedicine, but they use a method with (admittedly weak) supervision. Also, some authors have analyzed the Medline database, the MeSH index tree, and the various options for static mapping from different perspectives and at different levels of aggregation use, in order to investigate the translations and interactions and thus to gain a bibliometric perspective on the dynamics of medical innovations (Leydesdorff, Rotolo, & Rafols, 2012). Other authors have used bicluster high-frequency MeSH terms based on their co-occurrence of distinct semantic types in a MeSH tree to ascertain the structure of a scientific field (Fang, Zhou, & Cui, 2020). Nevertheless, no studies have tackled the question of which method does a better job of classifying by topic areas.

Generally speaking, the work described above focuses on using the terms mentioned to obtain mapping knowledge, main lines of research, emerging trends, and research gaps in a topic area. Little work has compared the various terms that can be used to ascertain which yield a better representation in health sciences. Nor are there studies on the advisability of using MeSH terms or KW Plus terms in the light of the possibility that the resulting studies may be rendered unsuitable or biased by the term system used.

4. Materials and method

4.1. Search strategy

The bibliographic search was run in the Core Collection of Clarivate Analytics' Web of Science, using the connection registered to Carlos III University of Madrid, and in the National Library of Medicine's Medline database, over the PubMed interface. All searches were conducted on the same day (February 21, 2020) to rule out the effect of database updates. The specific search profiles for each of the databases are presented in supplementary material appendix 1. The search was limited to the fifty years between 1970 and 2019. Ethical approval was not needed, since this was a bibliographic study.

4.2. Record sample

Figure 1 shows the record selection and downloading procedure. Only documents considered articles or reviews were selected, and the selection of articles and reviews from each of the databases was made considering the document types shown in supplementary material appendix 2. After selection those documents that bore the same PubMed Identifier (PMID) and three or more KW Plus terms and two or more MeSH terms were included in the study. The definitive sample contained 25,593 records.

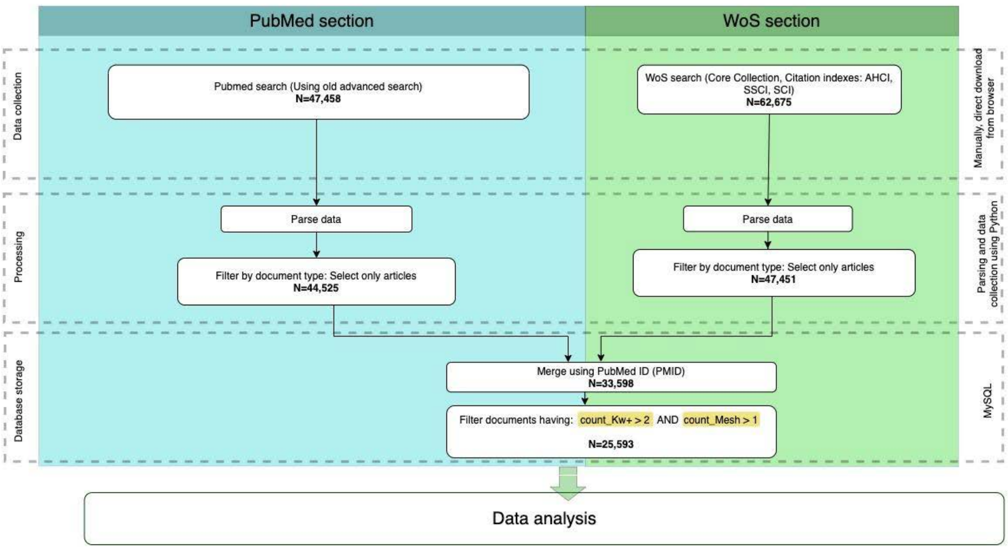

Figure 1. Record selection and downloading procedure.

Term frequency and distribution in each of the two systems were obtained from this sample. The tidytext package (Silge & Robinson, 2016) was used for calculations concerning the terms' frequency distribution. Term frequency distributions were analyzed using the classic version of Zipf's law and the relative frequency (*tf*) of each term in relationship with the total terms of each distribution (Silge & Robinson, 2020).

$$tf \approx \frac{C}{rank^{\alpha}}$$

Distributions can be estimated in logarithm form using simple regression:

$$log10(tf) = log10(C) - \alpha \cdot log10(rank)$$

4.3. Automatic document classification

For automatic document classification, probabilistic topic models (PTMs) were applied separately to both the MeSH terms and the KW Plus terms. We used the Latent Dirichlet Allocation (LDA) algorithm (Blei, Ng, & Jordan, 2003). This kind of algorithm has been used to classify scientific documents on a wide range of occasions, including: analysis of the abstracts of publications indexed in PubMed on "rectal cancer" (Wang et al., 2020) to identify the five areas where the most progress has been made in the last 25 years; combined use with MeSH terms to enhance the performance of biomedical document retrieval tasks (Yu, Bernstam, Cohen, Wallace, & Johnson, 2016); identification of the evolution of research topic popularity over time (Savov, Jatowt, & Nielek, 2020); and development of prediction models that recognize future highly-cited papers (Hu, Tai, Liu, & Cai, 2020). The LDA algorithm has also been used to compare classifications drawn from astrophysics documents in two different sources (Hu et al., 2015).

A number of R statistics packages were used for the calculations, particularly topic models (Grün & Hornik, 2011), for fitting by LDA with Gibbs sampling, and the "tm" package (Feinerer, Hornik, & Meyer, 2008), for building terminological corpora. After some testing, a seven-topic classification was chosen as the best solution.

The MeSH terms and the KW Plus terms were pre-processed to eliminate certain symbols and punctuation marks, such as ";" used as a term separator, single and double quotation marks, and apostrophes. It was not necessary to eliminate empty words. However, some neutral words ("male" and "female") did have to be deleted from the MeSH terms, as did commas in the "inverted" terms (e.g., "receptor, opioid delta"), but ampersands and dashes between words were kept. R-packages were not used to build n-grams, but spaces were replaced by dashes in multiple-word terms in both groups to preserve the terms' semantic meaning. In all cases, the LDA algorithm's own parameters were those shown in supplementary material appendix 3.

The PTMs classify all the processed documents automatically, and without supervision, into different topics or bags of words. This is one of their main advantages (Chen, 2017).

For this purpose, each document is assigned a probability (gamma parameter, *g*) in each topic. Each of the lists of KW Plus or MeSH terms containing each of the records obtained by our search strategy was considered a document, using the *g* parameter as a threshold for selecting only those documents or lists with a higher estimated likelihood of belonging to one or two topics.

In our analysis, the same group of documents was classified in two different systems, KW Plus and MeSH. At the same time, two groups of documents were selected using different gamma parameter thresholds ($g \geq 0.2$ for one group and $g \geq 0.25$ for the other).

4.4. Association and concordance between the resulting classifications

To evaluate the association and the degree of concordance between the two description systems, based on the results of the LDA analysis, contingency tables were calculated for each of the grouping levels (*g*), with the number of documents classified in each system. The independence between the two classifications was checked with chi-squared tests, and the results were displayed using mosaic graphs. In addition, the degree of association was measured using the contingency coefficient and Cramer's V. For these calculations, the vcd library in R was used (Meyer, Zeileis, & Hornik, 2006; Zeileis, Meyer, & Hornik, 2007).

Cohen's kappa was used (as was the vcd library) to measure agreement. For this purpose, the contingency tables had to be redone using the topics' equivalences. Semantic correspondence between topics was found by an expert in addictions, working with the content of the bags.

5. Results

5.1. Descriptive analysis: Term frequency

The method described above yielded 10,107 MeSH terms and 28,870 KW Plus terms. Seventy-seven MeSH terms and 15 KW Plus terms had more than 1,000 occurrences.

The distributions and their estimates are plotted in figure 2. The statistics for the linear estimates are given in supplementary material appendix 4.

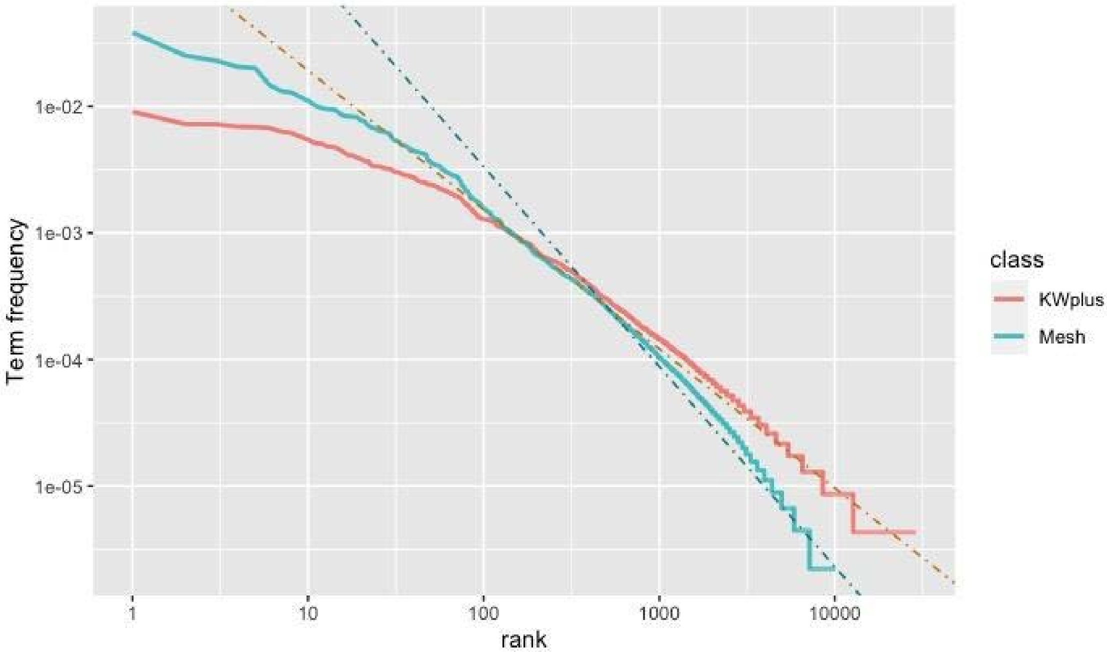

Figure 2. Zipf's Law for MeSH terms and KW Plus terms

The KW Plus terms and the MeSH terms have different distributions (figure 2), with a steeper slope in the case of the MeSH terms (intercept = 0.68, slope = -1.58, R-squared = 0.9669) than in the case of the KW Plus terms (intercept = -0.62, slope = -1.01, R-squared = 0.9631).

As can be observed in figure 2, early on in the relative frequency positions, the MeSH terms show a higher distribution than the KW Plus terms. This is due to the greater presence of common MeSH terms among the most frequent terms. Contrariwise, in the tails the distribution of the KW Plus terms "outdoes" that of the MeSH terms, because the KW Plus terms do not constitute a controlled language, and thus they contain a higher number of used terms that appear infrequently.

Table 1 shows the first 25 MeSH and KW Plus terms appearing in the documents we analyzed, in order of absolute frequency (*freq*). It also gives the relative frequency (*tf*) and inverse frequency (*tf-idf*) of the terms.

Table 1. Frequency of term appearance in the documents.

MeSH terms					**KW Plus terms**			
Word	*freq*	*tf*	*tf-idf*	*rank*	*Word*	*Freq*	*tf*	*tf-idf*
humans	17177	0.0383	0.0000	1	**endocannabinoid-system**	2095	0.0091	0.0063
animals	11295	0.0252	0.0000	2	**alcohol**	1675	0.0073	0.0050
metabolism	10296	0.0229	0.0000	3	**brain**	1668	0.0072	0.0000

drug-effects	9206	0.0205	0.0142	4	**anandamide**	1593	0.0069	0.0048
pharmacology	9006	0.0201	0.0000	5	**drug-use**	1585	0.0069	0.0048
adult	6647	0.0148	0.0000	6	**activation**	1562	0.0068	0.0047
adolescent	5933	0.0132	0.0000	7	**substance-use**	1457	0.0063	0.0044
physiology	5791	0.0129	0.0000	8	**expression**	1425	0.0062	0.0043
epidemiology	5341	0.0119	0.0000	9	**abuse**	1334	0.0058	0.0040
psychology	4979	0.0111	0.0000	10	**cannabinoid-receptor**	1263	0.0055	0.0038
rats	4529	0.0101	0.0000	11	**acid-amide-hydrolase**	1179	0.0051	0.0035
receptor-cannabinoid-cb1	4350	0.0097	0.0067	12	**risk**	1169	0.0051	0.0000
substance-related-disorders	4274	0.0095	0.0066	13	**prevalence**	1108	0.0048	0.0000
endocannabinoids	4249	0.0095	0.0000	14	**rat-brain**	1102	0.0048	0.0033
cannabinoids	3993	0.0089	0.0000	15	**inhibition**	1059	0.0046	0.0032
mice	3771	0.0084	0.0000	16	**united-states**	999	0.0043	0.0000
young-adult	3746	0.0083	0.0000	17	**marijuana-use**	949	0.0041	0.0000
drug-therapy	3723	0.0083	0.0000	18	**delta-9-tetrahydrocannabinol**	944	0.0041	0.0028
antagonists-&-inhibitors	3685	0.0082	0.0057	19	**marijuana**	905	0.0039	0.0027
genetics	3483	0.0078	0.0000	20	**cannabis-use**	889	0.0039	0.0027
chemistry	3412	0.0076	0.0000	21	**dependence**	853	0.0037	0.0026
middle-aged	3160	0.0070	0.0049	22	**mice**	847	0.0037	0.0000
marijuana-abuse	3049	0.0068	0.0047	23	**delta(9)-tetrahydrocannabinol**	782	0.0034	0.0023
methods	3028	0.0067	0.0047	24	**system**	779	0.0034	0.0023
dronabinol	2888	0.0064	0.0000	25	**endogenous-cannabinoids**	773	0.0033	0.0023

The MeSH terms with the highest frequency are terms that appear in many records due to the hierarchical structure of the thesaurus that supports the system (MeSH Tree). The MeSH Tree is divided into 16 branches, which in turn are divided into a varied number of levels and sublevels. For example, branch B refers to the organisms involved in the analyses, such as animals (which includes humans, rats, and mice). Branch M (Named Groups) includes terms about the groups of individuals involved; these include the age groups of persons (such as adult, adolescent, and middle-aged). Other terms called "qualifiers" or "subheadings" are included with the headings. They provide extra information in addition to the descriptors and include terms such as "metabolism," "drug-effects pharmacology," and "physiology."

As can be seen in table 1, some of the most commonly mentioned MeSH terms have a *tf-idf* value equal to zero. This happens when the *idf* is zero, that is, when the term (e.g., “mice”) is repeated in all the documents to be considered (in our case, the two classifications we are analyzing). The condition of *tf-idf* = 0 identifies the terms as generic terms, as opposed to what we might identify as uncommon or specific terms, whose importance increases inversely to their frequency of appearance, and which therefore have higher *tf-idf* values. Thus, the *tf-idf* value identifies those terms that are more significant to the analysis due to their infrequent appearance in the documents and in the two classifications. A term that is very common in one of the classifications but appears rarely or not at all in the other classification must have a non-zero *tf-idf.* This is what happens in the example of the MeSH terms “drug-effects” and “receptor-cannabinoid-cb1” and the KW Plus terms “endocannabinoid-system” and “alcohol.”

As can be observed in table 2, the KW Plus terms are generally more specific than the MeSH terms (*tf-idf* ≠ 0). This would explain why there are 18,763 terms more in KW Plus than in MeSH and why the MeSH terms display a higher frequency than the KW Plus terms. Because the KW Plus terms are not configured in a controlled language, they open the door to spelling variants and combinations of generic and specific terms. One of the spelling variants is the main psychoactive component of cannabis, “delta-9-tetrahydrocannabinol,” which appears with and without parentheses. However, in MeSH this particular term does not appear at all, as delta-9-tetrahydrocannabinol falls under the term “dronabinol,” a psychoactive compound extracted from the resin of *Cannabis sativa*. Furthermore, generic and specific terms may be combined. For instance, KW Plus contains both the generic term “system” and more specific terms such as “endocannabinoid-system;” this constitutes an additional restriction of the KW Plus description system.

Table 2. Number of records analyzed in each of the systems with $g \geq 0.2$ and $g \geq 0.25$

	Number of documents (%)		
g	KW Plus	MeSH	KW Plus MeSH
$g \geq 0.2$	13,553 (52.2%)	22,962 (89.7%)	10,270 (40.1%)
$g \geq 0.25$	2,302 (9.0%)	8,954 (35.0%)	1,175 (45.9%)

Figure 3 shows the 25 most significant terms in the description of the documents in both systems, listed in order by their *tf-idf* value.

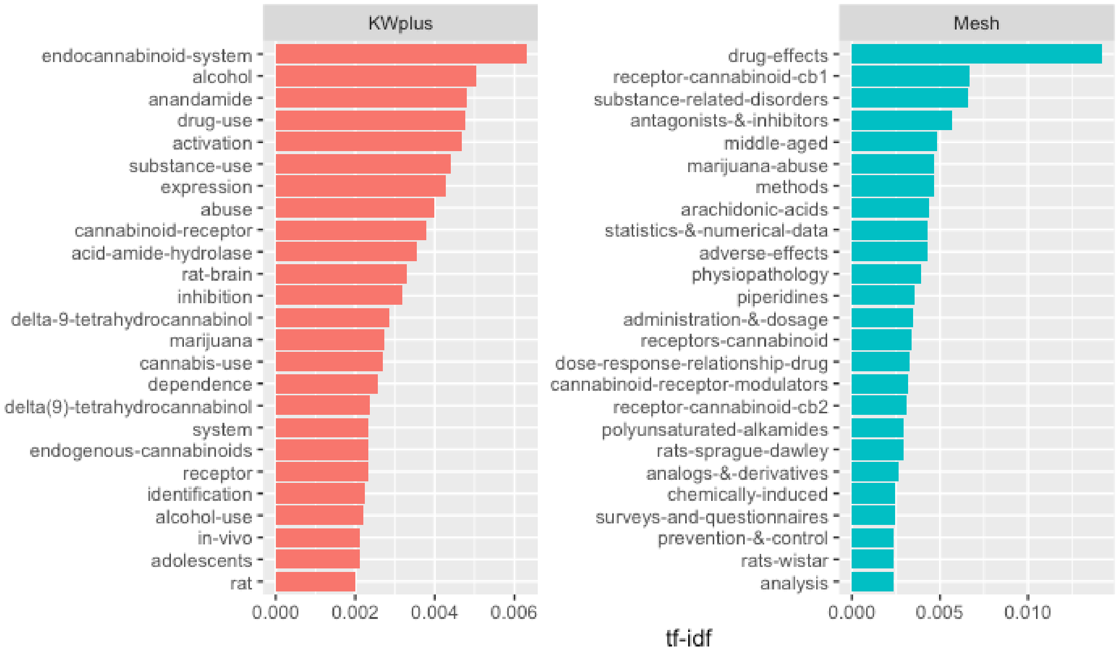

Figure 3. Twenty-five most significant terms for describing the records in each classification system

Comparison of the two lists suggests the following considerations:

- The term “alcohol”/”alcohol-use” is a significant word in KW Plus, but not in MeSH. KW Plus may contain terms related with alcohol due to the use of bibliographic references that include the term “alcohol” for various reasons, including the existence of extensive research into the effects of alcohol on the nervous system and the addiction process that have served as a model for basic research studies into cannabis; or the concomitant consumption of alcohol and cannabis, due to which papers on alcohol would be cited in cannabis-related studies of mental health, toxicology, and public health.
- The chemical compounds related to the endocannabinoid system and cannabinoid receptors are on both lists, but the specific receptors cb1 and cb2 appear only among the MeSH terms. Other chemical compounds appear in MeSH as well, such as: a) piperidine, a component of the hemp plant whose derivatives include rimonabant, an antagonist of the selective cb1 cannabinoid receptor, and b) arachidonic acid, which may be related with the link between the endocannabinoid system and the arachidonic cascade or

the fact that endocannabinoids (like anandamide, which does appear in KW Plus) are synthesized from omega-6 polyunsaturated fatty acid and arachidonic acid.

- Qualifiers or subheadings, such as “drug-effects,” “methods,” “statistics-&-numerical data,” and “analysis” are a feature of the MeSH term list only. They do not appear in the KW Plus terms, because they are not terms that usually appear in papers’ titles, although for MeSH they are descriptive of papers’ contents.
- The use of rats in in-vivo experiments is a source of important descriptors in both lists, although in KW Plus experimental rat use appears generically as “rat” or in connection with “brain,” while the strains used (Wistar and Sprague-Dawley) appear in MeSH.
- Both lists include terms related with substance abuse. MeSH contains “substance-related-disorders” and “marijuana-abuse,” while KW Plus has “cannabis-use,” “marijuana,” and “substance-use.”

5.2. Topic modeling analysis

Applying the gamma parameter (g) as a filter yields two groups, presented in table 2. With $g \geq 0.2$, 10,270 records common to both classifications were found, while, with $g \geq 0.25$, 1,175 records were found.

The classifications obtained with each of the g values show that a higher percentage of documents is classified with MeSH terms than with KW Plus terms. Where $g \geq 0.2$, the MeSH number is 37.5 percentage points higher, and where $g \geq 0.25$, it is 26 percentage points higher. In addition, the documents classified correctly with each of the terms are not the same, because the percentage of documents common to both classifications when $g \geq 0.2$ is 40.1%, and when $g \geq 0.25$ it is 45.9%.

Figures 4 and 5 present, in bags of 20 words apiece, the seven topics into which the documents of each of the terminological models are classified. The name of each topic and the semantic correspondence among the topics obtained through each of the description systems are shown in table 3. To facilitate interpretation of the results, in this table the identifiers of each MeSH topic have been recoded using the same topic number as the KW-Plus equivalent. The beta parameter measures the estimated likelihood of a term’s belonging to a given topic.

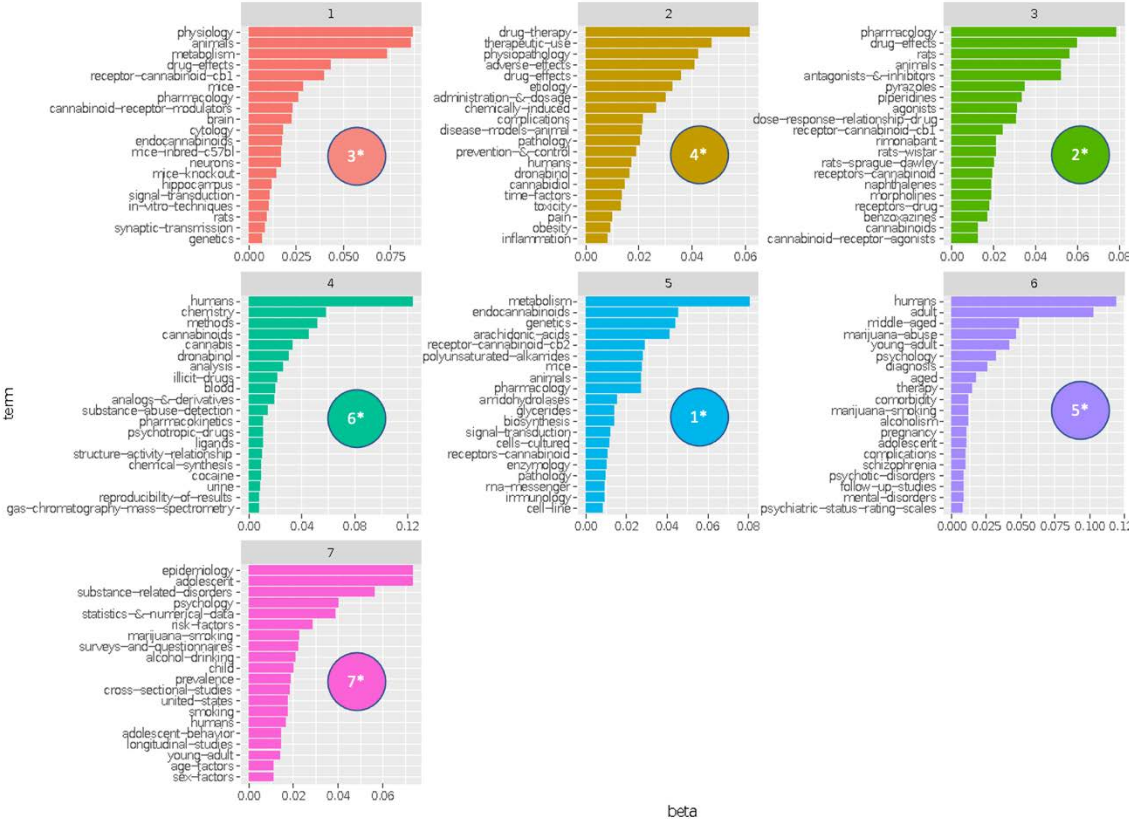

Figure 4. Record classification topics on the basis of MeSH terms with equivalent ID*

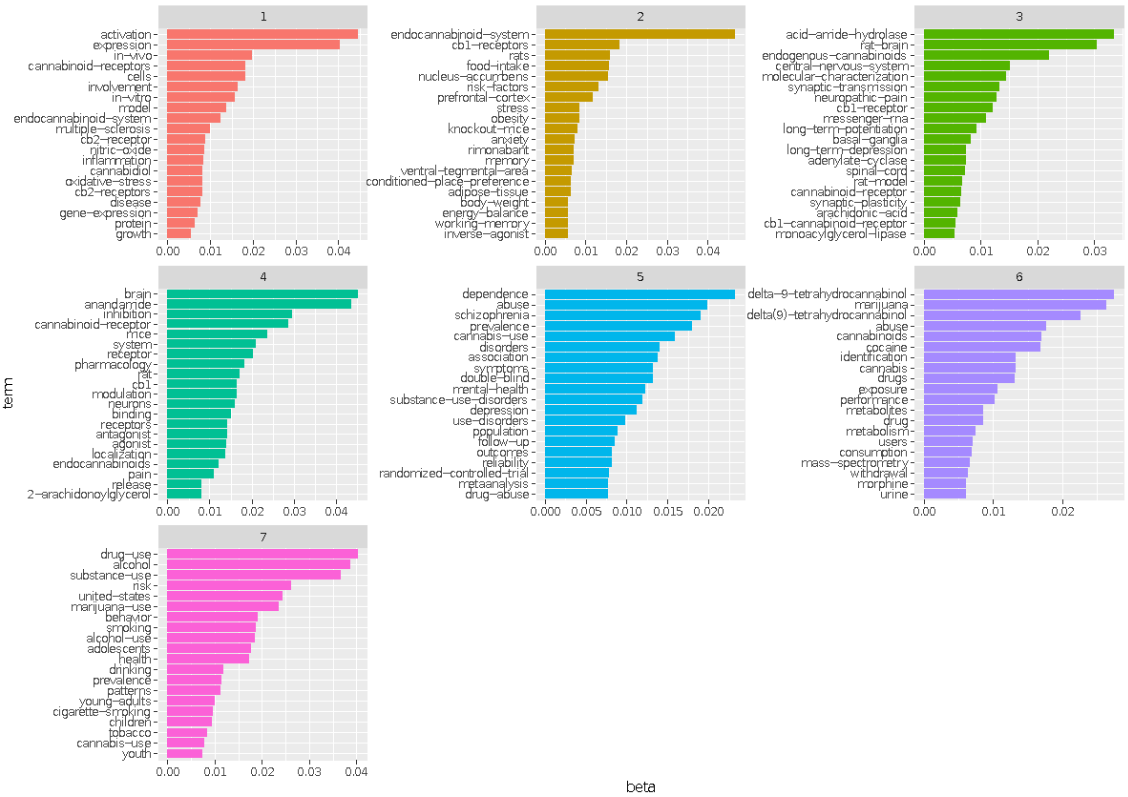

Figure 5. Record classification topics on the basis of KW Plus terms

Table 3. Semantic correspondence between the two classifications

KW Plus Topic subject	KW Plus Topic number	MeSH Topic number	MeSH-equivalent ID*	MeSH Topic subject
Biochemistry, Genetics, & Molecular Biology	1	5	1*	Biochemistry, Genetics, & Molecular Biology
Pharmacology & Pharmacy	2	3	2*	Pharmacology & Pharmacy
Neurosciences	3	1	3*	Neurosciences
Miscellaneous**	4	2	4*	Miscellaneous
Psychiatry & Mental Health	5	6	5*	Psychiatry & Mental Health
Toxicology & Legal Medicine	6	4	6*	Toxicology & Legal Medicine
Public Environmental & Occupational Health	7	7	7*	Public Environmental & Occupational Health

* MeSH ID in equivalent KW-Plus topic number.

**Considered miscellaneous because it covers a large quantity of terms from diverse topics.

The seven topics cover the following subject areas:

1-1*: Biochemistry, Genetics, Molecular Biology, & Immunology: This topic includes terms such as "cells" in KW Plus and "cells-cultured" and "cell-line" in MeSH, and "gene-expression" in KW Plus and "genetics" in MeSH. In KW Plus it includes terms such as "in-vivo" and "in-vitro" studies, "nitric-oxide," "inflammation," and "oxidative-stress," while in the MeSH classification it contains "immunology," "arachidonic-acids," "amidohydrolases," "glycerides," and "signal-transduction."

2-2*: Pharmacology & Pharmacy: This topic includes the terms "pharmacology" and "cannabinoid-receptor-agonists." The term "rimonabant" figures prominently; rimonabant is a selective cb1 cannabinoid receptor antagonist used as an appetite suppressant, and for that reason terms like "obesity" and "food-intake" appear. Another term associated with this topic is "dose-response-relationship-drug."

3-3*: Neurosciences: This topic includes terms such as "physiology," "metabolism," "drug-effects," "cannabinoid-receptor," "endocannabinoids," "brain," "neurons," "hippocampus," and "central-nervous-system." The term "synaptic-transmission" appears in both classifications.

4-4*: Miscellaneous: This is the topic that presents the highest number of terms that fall within other topics as well. The MeSH classification includes terms like "drug-therapy," "therapeutic-use," "physiopathology," "administration-&-dosage," and "complications," while KW Plus includes terms already encompassed by other topics, like "brain," "agonist," "antagonist," "pharmacology," "cb1," and "receptors."

5-5*: Psychiatry & Mental Health: Both classifications include terms related with substance use, "abuse," and "dependence," "psychology," "schizophrenia," "comorbidity," "psychotic-disorders," "depression," "mental health," and "mental disorders," and follow-up studies of pathologies and treatments are plentiful. These studies focus on humans.

6-6*: Toxicology & Legal Medicine: These include active ingredients like "delta-9-tetrahydrocannabinol" and "metabolites," as well as terms related with laboratory analysis techniques or methods, such as "identification," "substance-abuse-detection," and "gas-chromatography-mass-spectrometry," and terms from the field of laboratory analysis, like "chemistry," "analysis," "blood," and "urine."

7-7*: Public Environmental & Occupational Health: Both classifications include terms related with epidemiology, like "prevalence." This topic focuses on young groups and addresses fundamental issues in prevention, like risks and behavior.

Of the seven topics, three are associated more with humans than with animals. These are the topics of Psychiatry & Mental Health, Public Environmental & Occupational Health, and Toxicology & Legal Medicine.

5.3. Statistical analysis

The 10,270 documents assigned in both systems with a threshold of $g \geq 0.2$ and the 1,175 documents assigned with a threshold of $g \geq 0.25$ are presented in contingency tables (tables 4a and 4b).

Table 4a. Number of documents classified in both models ($g \geq 0.2$)

	MeSH-ID* topic number							
(a)	1*	2*	3*	4*	5*	6*	7*	Total
ID-KW Plus topic number	$g \geq 0.20$							
1	915	142	122	202	6	78	0	1,465
2	122	601	323	334	121	16	6	1,523
3	359	397	821	163	8	24	1	1,773
4	408	518	222	93	3	186	0	1,430
5	4	7	4	188	1,222	17	720	2,162
6	11	66	4	69	139	934	78	1,301
7	1	0	0	10	472	21	2,508	3,012
Total	1,820	1,731	1,496	1,059	1,971	1,276	3,313	12,666**

** A document may be assigned to more than one topic.

Table 4b. Number of documents classified in both models ($g \geq 0.25$)

	MeSH-ID* topic number							
(b)	1*	2*	3*	4*	5*	6*	7*	Total
KW Plus topic number	$g \geq 0.25$							
1	62	1	0	2	0	1	0	66
2	4	36	6	20	3	0	0	69
3	26	22	112	3	0	0	0	163
4	10	16	3	3	0	4	0	36
5	0	0	0	12	107	0	18	137
6	0	1	0	0	0	227	0	228
7	0	0	0	0	5	0	471	476

Total	102	76	121	40	115	232	489	1,175

It should be borne in mind that, where g is lower, some of the documents are ascribed to more than one topic in one of the two classifications. As a result, the total in the corresponding contingency table is higher than the total number of documents considered. However, in the case of $g \geq 0.25$, each document is assigned to a single topic in each of the description systems.

Tables 4a and 4b present an initial comparison of the equivalences between the two classifications shown in table 3 (the recoded MeSH topic numbers have been used). While in table 3 equivalence was found through expert review of the bags of words to invest the words with semantic content, now the expert's opinion is supported by statistical analysis.

5.4. Measurements of association

The first step is independence testing and calculation of the usual measurements of association linked to the chi-squared test in simple contingency tables. The results of these measurements are shown in table 5.

Table 5. Statistics measuring independence and association between the two models ($g \geq 0.2$, $g \geq 0.25$)

	$g \geq 0.2$	$g \geq 0.25$
Pearson's chi-squared test	23392 (df = 36, p-value < 0.05)	4119.9 (df = 36, p-value < 0.05)
Contingency coefficient	0.805	0.882
Cramer's V	0.555	0.764

In both cases ($g \geq 0.2$ and $g \geq 0.25$), it must be accepted (p-value < 0.05) that the two classifications are not independent and do have some association between them. The contingency coefficient furthermore reveals that overall this relationship is intense and grows greater in the case of the more demanding criterion. The values of Cramer's V corroborate the higher degree of association in the case of the higher threshold.

Mosaic plots are used to visualize the associations among the seven topics of each of the classifications. The seven KW Plus topics are ranged along the vertical axis (*topic.x*), and the corresponding MeSH-ID* topics are ranged along the horizontal axis (*topic.y.*).

The areas of the rectangles represent the joint co-occurrence frequencies (length is MeSH and width is KW Plus). Those cells where the frequencies are higher than expected are colored blue, while the cells where the reverse is the case are colored red.

For example, in figure 6, which graphs the data in table 4a, many documents classified into topic 1 in KW Plus may be expected to be classified into topic 1* (topic number 5 in fig. 4) by MeSH, and few of them may be expected to be classified into MeSH topics 2* or 6* (3 and 4 in fig. 4).

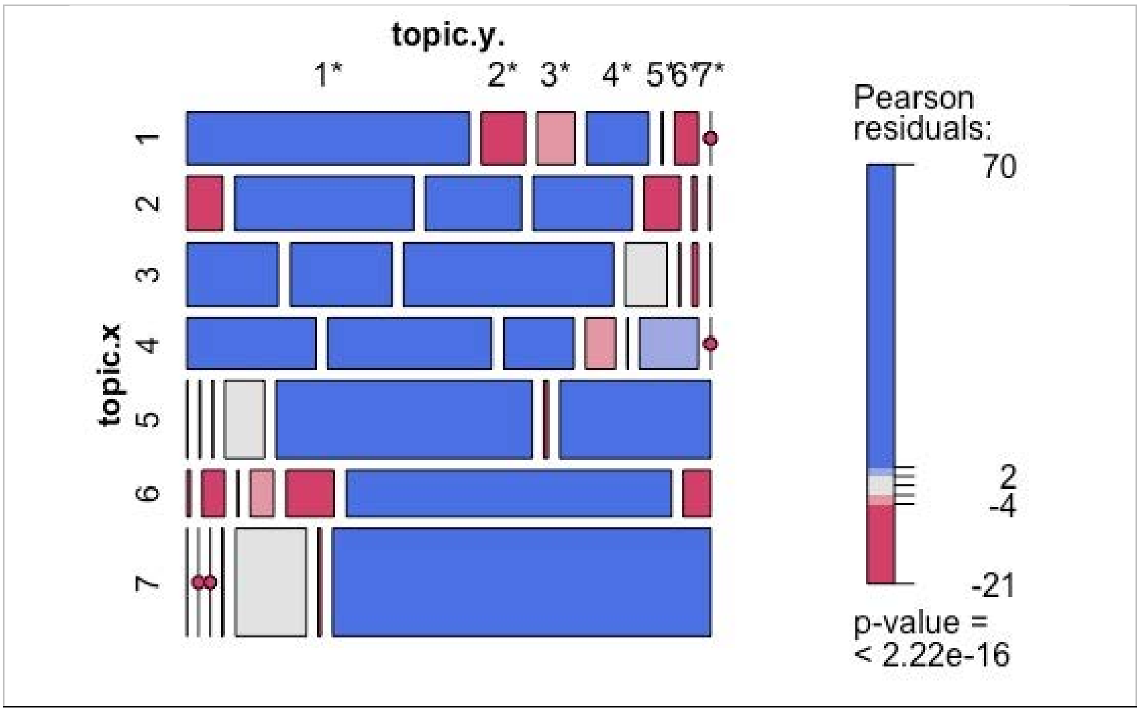

Figure 6. Mosaic plots of the contingency tables of KW Plus terms and MeSH terms, with $g \geq 0.2$

Let us observe the graph in figure 7, plotted from the data in table 4b, where the correspondence between classifications is established on the basis of the documents most likely to be properly assigned to a given topic (highest gamma). The results match the expert's observations fairly well. For example, the documents classified by KW Plus into topic number 7 may be expected to be classified by MeSH into topic 7 (7*) as well. The same applies to the combinations KW Plus-1/MeSH-5 (1*), KW Plus-3/MeSH-1 (3*), KW Plus-5/MeSH-6 (5*), and KW Plus-6/MeSH-4 (6*).

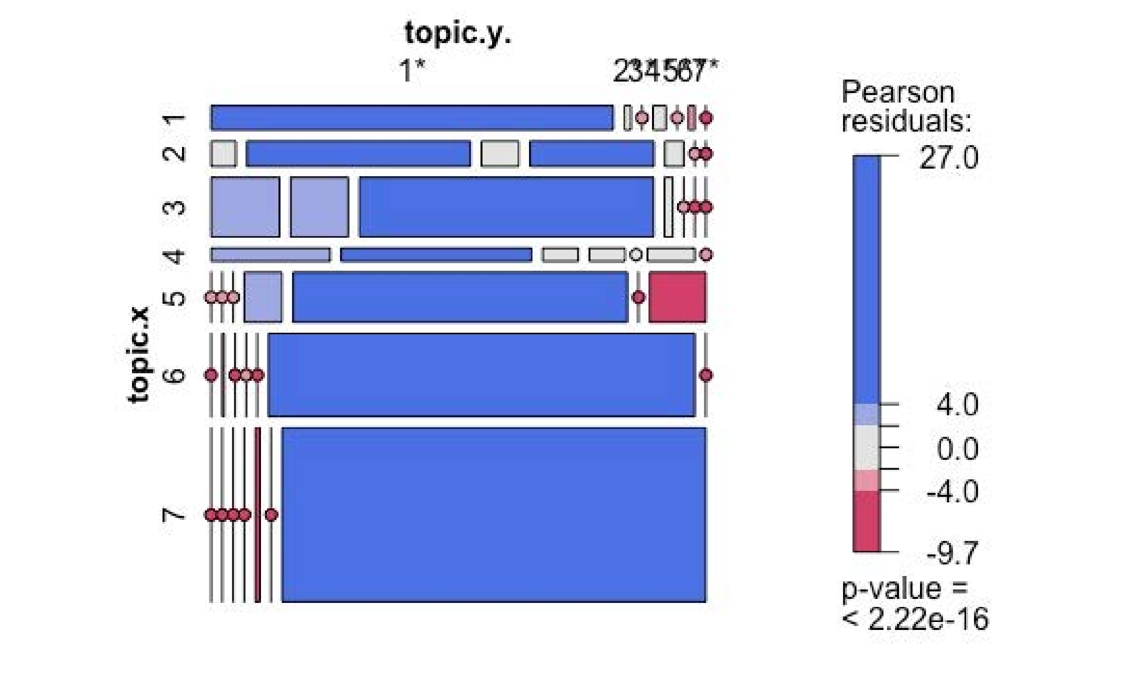

Figure 7. Mosaic plots of the contingency tables of KW Plus terms and MeSH-ID* terms, with $g \geq 0.25$

There are two major discrepancies between the statistical results and the semantic assignment. The first is the assignment of KW Plus-2 as being compatible with not only MeSH-3 (2*), but also MeSH-2 (4*). The second discrepancy is in KW Plus-4, which we have identified semantically as “Miscellaneous,” and which the statistical analysis assigns, in varying degrees, to the first five MeSH topics, with a larger number of documents classified in MeSH-3 (2*) than in MeSH-2 (4*). These discrepancies grow greater when we consider a lower gamma value, which indicates greater ambiguity in document classification in both models.

Lastly, we can use our analysis to find the concordance between the two classification models using Cohen’s kappa coefficient. This coefficient measures how closely the two systems agree when assigning documents to topics. It is calculated from the values of the diagonal of a matrix where the rows and columns represent the same groups. Because this is not the case in our analysis, since the groups are assigned a number randomly, the topics have been recoded according to the list established in table 3 and to a certain extent corroborated by table 4b and figure 7.

The values found for the matches between the two models, by gamma level, were kappa = 0.477 (for $g \geq 0.2$) and kappa = 0.8236 (for $g \geq 0.25$). As can be seen, while the concordance between the two description systems is moderate for the first level, when documents with a higher probability of being unequivocally classified (higher gamma) are compared, however, the concordance between the two systems is much greater.

As may be observed in tables 4a and 4b and figures 6 and 7, the topics associated more with humans, like "Psychiatry & Mental Health" (KW Plus-5, MeSH-6/5*), "Public Environmental & Occupational Health" (KW Plus-7, MeSH-7/7*), and "Toxicology & Legal Medicine" (KW Plus-6, MeSH-4/6*), are the topics presenting the greatest concordance between the two systems. However, in the data obtained when $g \geq 0.25$, some of the documents classified by KW Plus as "Psychiatry & Mental Health" (KW Plus-5) are unexpectedly classified by MeSH as "Public Environmental & Occupational Health" (MeSH-7/7*).

6. Conclusions and discussion.

An analysis was performed comparing the use of the MeSH and KW Plus classification systems for automatic document classification and the use of each system to map knowledge in the health sciences area.

The terms in each system were studied to determine whether they followed the same model. The distributions of MeSH and KW Plus terms were found to behave differently. Behavior is clearly linked to the documentary nature of each of the systems. The MeSH system uses hierarchical, controlled language and therefore tends to use fewer terms to describe documents (10,107 MeSH per 28,870 KW Plus). For this same reason, among the most frequent terms (i.e., terms ranking lower than 100), there is a larger number of documents that use the MeSH terms. However, the opposite occurs for highly ranked terms (i.e., ranking above 1,000); then the uncontrolled nature of the KW Plus descriptors makes a larger number of terms necessary to describe documents.

One significant point of this work is that automatic classification of documents on the seven defined topics on the basis of MeSH and KW Plus terms, using $g \geq 0.2$, shows that nearly 50% of the documents included in Web of Science are not included in the seven topics. Therefore, they are not well classified. In addition, the concordance found between the two systems' classification of documents for the less demanding gamma ($g \geq 0.2$) is kappa=0.477. This means that, if gamma is made less demanding, a very weak degree of concordance will be found. One possible reason for the low concordance in topic categories is the wide dispersion of KW Plus terms. Another point to note is that the three topics concerning studies with human beings ("Psychiatry and Mental Health," "Public Environmental & Occupational Health," and "Toxicology and Legal Medicine") are the topics that displayed the highest degree of concordance in document classification. This fact suggests that differentiating between studies in human beings and other studies (primarily laboratory studies and basic research) could prove an advantageous strategy when topic mapping in the biomedicine area.

Furthermore, grey-box tools like VOSviewer and CiteSpace, or more recently Bibliometrix, frequently offer KW Plus-based topic maps as outputs. As the use of grey-box tools spreads and becomes increasingly accepted (due to researchers' relative ability to intervene in the calculation systems), this paper's findings become more germane to the interpretation of their outputs. This is especially so since topic analyses based on the categories into which WoS (or Scopus) classifies journals are being questioned more and more, yet the need to run article-level topic analyses is becoming clearer and clearer. Therefore, the main practical application of this study is that it shows that the use of WoS subject delimiters for topic analyses in biomedical documents is not neutral. Our findings determine that this is particularly significant when WoS's KW Plus terms are used, because, compared to the medical descriptive standard (MeSH), the KW Plus description system is effective only in a small portion of highly specialized, unequivocally classified documents— in our case just 4.6% of the documents processed. Therefore, MeSH terms should be used instead of KW Plus terms in the study of topic areas in biomedicine.

One limitation of the work reported here is that we did not differentiate MeSH terms by their hierarchical levels of classification. Therefore, we included generic and specific terms from the same branch in the corpus we analyzed. This could produce a certain amount of classification redundancy. Furthermore, due to the lack of standardization in KW Plus terms, the same term could be included with different terminological variations in different topics. Nevertheless, this posed no obstacle to the automatic assignment of the terms to the document topics.

One possible line of further research in this area is comparative study of automatic document classification between MeSH terms and abstracts, for more accurate classification in medical research fields.

Funding: Juan Carlos Valderrama-Zurián was the beneficiary of a grant from the Valencian Regional Ministry of Innovation, Universities, Science, and Digital Society to carry out postdoctoral research at Carlos III University of Madrid (BEST/2020/121).
This research has also been funded by Program H2019/HUM-5744 CM-INCLUSIVA-CM, under the Community of Madrid's 2019 call for grants for R&D programs in social sciences and humanities.

References

1. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022. https://jmlr.org/papers/volume3/blei03a/blei03a.pdf
2. Bhatia, S., Lau, J. H., & Baldwin, T. (2016). Automatic labelling of topics with neural embeddings. *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 953–963. https://www.aclweb.org/anthology/C16-1091.pdf
3. Chen, C. (2016). Grand challenges in measuring and characterizing scholarly impact. *Frontiers in Research Metrics and Analytics,* 1, 4. doi: 10.3389/frma.2016.00004
4. Chen, L.-C. (2017). An effective LDA-based time topic model to improve blog search performance. *Information Processing & Management*, *53*(6), 1299–1319. doi: 10.1016/j.ipm.2017.08.001
5. Chen, X., Lun, Y., Yan, J., Hao, T., & Weng, H. (2019). Discovering thematic change and evolution of utilizing social media for healthcare research. *BMC Medical Informatics and Decision Making*, 19(S2), 50. doi: 10.1186/s12911-019-0757-4
6. de Vries, B. B. P., van Smeden, M., Rosendaal, F. R., & Groenwold, R. H. (2020). Title, abstract, and keyword searching resulted in poor recovery of articles in systematic reviews of epidemiologic practice. *Journal of Clinical Epidemiology*, 121, 55-61. doi: 10.1016/j.jclinepi.2020.01.009
7. Dettori, J. R., Norvell, D. C., & Chapman, J. R. (2019). Measuring academic success: The art and science of publication metrics. *Global Spine Journal*, 9(2), 243-246. doi: 10.1177/2192568219831003

8. Drosatos, G., & Kaldoudi, E. (2020). A probabilistic semantic analysis of eHealth scientific literature. *Journal of Telemedicine and Telecare*, 26(7-8), 414-432. doi: 10.1177/1357633X19864006
9. Ebener, S., Khan, A., Shademani, R., Compernolle, L., Beltran, M., Lansang, M. A., & Lippman, M. (2006). Knowledge mapping as a technique to support knowledge translation. *Bulletin of the World Health Organization*, *84*, 636-642. doi: 10.2471/BLT.06.029736
10. Fang, L., Zhou, X., & Cui, L. (2020). Biclustering high-frequency MeSH terms based on the co-occurrence of distinct semantic types in a MeSH tree. *Scientometrics*, *124*(2), 1179-1190. doi: 10.1007/s11192-020-03496-4
11. Feinerer, I., Hornik, K., & Meyer, D. (2008). Text Mining Infrastructure in R. *Journal of Statistical Software*, *25*(5), 1–54. doi: 10.18637/jss.v025.i05
12. Garfield, E., & Sher, I. H. (1993). KeyWords Plus™ algorithmic derivative indexing. *Journal of the American Society for Information Scie*nce, 44(5), 298-299. http://www.garfield.library.upenn.edu/papers/jasis44(5)p298y1993.html.
13. Glänzel, W., Moed, H. F., Schmoch, U., & Thelwall, M. (Eds.). (2019). Springer Handbook of Science and Technology indicators. Springer, Cham. doi: 10.1007/978-3-030-02511-3
14. Grün, B., & Hornik, K. (2011). Topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, *40*(13), 1–30. doi: 10.18637/jss.v040.i13
15. Hu, B., Dong, X., Zhang, C., Bowman, T. D., Ding, Y., Milojević, S., Ni, C., Yan, E., & Larivière, V. (2015). A lead-lag analysis of the topic evolution patterns for preprints and publications. *Journal of the Association for Information Science and Technology*, *66*(12), 2643–2656. doi: 10.1002/asi.23347
16. Hu, Y.-H., Tai, C.-T., Liu, K. E., & Cai, C.-F. (2020). Identification of highly-cited papers using topic-model-based and bibliometric features: the consideration of keyword popularity. *Journal of Informetrics*, *14*(1), 101004. doi: 10.1016/j.joi.2019.101004
17. Jiang, G., Huang, C.-K., Zhang, X., Lv, X., Wang, Y., Yu, T., & Cai, X. (2019). Wnt signaling in liver disease: Emerging trends from a bibliometric perspective. *PeerJ*, 7, e7073. doi: 10.7717/peerj.7073

18. Keya, K. N., Papanikolaou, Y., & Foulds, J. R. (2019). Neural embedding allocation: Distributed representations of topic models. *CoRR*, *abs/1909.04702*. http://arxiv.org/abs/1909.04702
19. Khasseh, A. A., Soheili, F., Moghaddam, H. S., & Chelak, A. M. (2017). Intellectual structure of knowledge in iMetrics: A co-word analysis. *Information Processing & Management*, *53*(3), 705-720. doi: 10.1016/j.ipm.2017.02.001
20. Kocak, M., García-Zorita, C., Marugán-Lázaro, S., Çakır, M. P., & Sanz-Casado, E. (2019). Mapping and clustering analysis on neuroscience literature in Turkey: a bibliometric analysis from 2000 to 2017. *Scientometrics*, *121*(3), 1339-1366. DOI: 10.1007/s11192-019-03259-w.
21. Lashkari, F., Bagheri, E., & Ghorbani, A. A. (2019). Neural embedding-based indices for semantic search. *Information Processing & Management*, *56*(3), 733–755. doi: 10.1016/j.ipm.2018.10.015
22. Leydesdorff, L., Comins, J. A., Sorensen, A. A., Bornmann, L., & Hellsten, I. (2016). Cited references and Medical Subject Headings (MeSH) as two different knowledge representations: Clustering and mappings at the paper level. *Scientometrics*, *109*(3), 2077-2091. doi: 10.1007/s11192-016-2119-7
23. Leydesdorff, L., Rotolo, D., Rafols, I. (2012). Bibliometric perspectives on medical innovation using the Medical Subject Headings (MeSH) of PubMed. *Journal of the American Society for Information Science and Technology*, 63(11), 2239-2253. doi: 10.1002/asi.22715
24. Liu, Y.-H., & Wacholder, N. (2017). Evaluating the impact of MeSH (Medical Subject Headings) terms on different types of searchers. *Information Processing & Management*, *53*(4), 851-870. doi: 10.1016/j.ipm.2017.03.004
25. McCoy Jr, T. H. (2019). Mapping the delirium literature through probabilistic topic modeling and network analysis: A computational scoping review. *Psychosomatics*, 60(2), 105-120. doi: 10.1016/j.psym.2018.12.003
26. Maltseva, D., & Batagelj, V. (2020). Towards a systematic description of the field using keywords analysis: Main topics in social networks. *Scientometrics*, *123*(1), 357-382. doi: 10.1007/s11192-020-03365-0

27. Meyer, D., Zeileis, A., & Hornik, K. (2006). The strucplot framework: Visualizing multi-way contingency tables with vcd. *Journal of Statistical Software*, *17*(3), 1–48. https://epub.wu.ac.at/3984/1/strucplot.pdf
28. Moed, H. F. (2009). New developments in the use of citation analysis in research evaluation. *Archivum Immunologiae et Therapiae Experimentalis*, 57(1), 13-18. doi: 10.1007/s00005-009-0001-5
29. Natale, F., Fiore, G., & Hofherr, J. (2012). Mapping the research on aquaculture. A bibliometric analysis of aquaculture literature. *Scientometrics*, 90(3), 983-999. doi: 10.1007/s11192-011-0562-z
30. Nentidis, A., Krithara, A., Tsoumakas, G., & Paliouras, G. (2020). Beyond MeSH: Fine-grained semantic indexing of biomedical literature based on weak supervision. *Information Processing & Management*, 57(5), 102282. doi: 10.1016/j.ipm.2020.102282
31.Ríssola, E. A., Aliannejadi, M., & Crestani, F. (2020). Beyond modelling: Understanding mental disorders in online social media. In J. M. Jose, E. Yilmaz, J. Magalhães, P. Castells, N. Ferro, M. J. Silva, & F. Martins (Eds.), *Advances in Information Retrieval*. Vol. 12035, pp. 296–310. doi:10.1007/978-3-030-45439-5_20.
32. Sahlgren, M. (2015). A brief history of word embeddings (and some clarifications). [Online; accessed 30-September-2015] https://www.linkedin.com/pulse/brief-history-word-embeddings-some-clarifications-magnus-sahlgren
33. Savov, P., Jatowt, A., & Nielek, R. (2020). Identifying breakthrough scientific papers. *Information Processing & Management*, *57*(2), 102168. doi: 10.1016/j.ipm.2019.102168
34. Shultz, M. (2007). Comparing test searches in PubMed and Google Scholar. *Journal of the Medical Library Association: JMLA*, 95(4), 442-445. doi: 10.3163/1536-5050.95.4.442
35. Silge, J., & Robinson, D. (2016). tidytext: Text mining and analysis using tidy data principles in R. *JOSS*, *1*(3), 37. doi: 10.21105/joss.00037
36. Silge, J., & Robinson, D. (2020). Analyzing word and document frequency: Tf-idf. In J. Silge, & D. Robinson. *Text Mining with R*. Boston: O'Reilly Media, pp. 31-44. https://www.tidytextmining.com/tfidf.html

37. Synnestvedt, M.B., Chen, C., & Holmes, J. (2005). CiteSpace II: Visualization and knowledge discovery in bibliographic databases. *AMIA 2005 Symposium Proceedings*, 2005, 724-728. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1560567/pdf/amia2005_0724.pdf
38. Thijs, B. (2019). Science mapping and the identification of topics: Theoretical and methodological considerations. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall. *Springer Handbook of Science and Technology indicators*. Springer, Cham. pp. 213-233. doi: 10.1007/978-3-030-02511-3
39. Tran, B., Pham, T., Ha, G., Ngo, A., Nguyen, L., Vu, T., Ho, R. (2018). A bibliometric analysis of the global research trend in child maltreatment. *International Journal of Environmental Research and Public Health*, 15(7), 1456. doi: 10.3390/ijerph15071456
40. Tripathi, M., Kumar, S., Sonker, S. K., & Babbar, P. (2018). Occurrence of author keywords and keywords plus in social sciences and humanities research: A preliminary study. *COLLNET Journal of Scientometrics and Information Management*, 12(2), 215-232. doi: 10.1080/09737766.2018.1436951
41. Wang, K., Feng, C., Li, M., Pei, Q., Li, Y., Zhu, H., Song, X., Pei, H., & Tan, F. (2020). A bibliometric analysis of 23,492 publications on rectal cancer by machine learning: Basic medical research is needed. *Therapeutic Advances in Gastroenterology*, *13*, 175628482093459. doi: 10.1177/1756284820934594
42. Yu, Z., Bernstam, E., Cohen, T., Wallace, B. C., & Johnson, T. R. (2016). Improving the utility of MeSH® terms using the TopicalMeSH representation. *Journal of Biomedical Informatics*, *61*, 77–86. doi: 10.1016/j.jbi.2016.03.013
43. Zeileis, A., Meyer, D., & Hornik, K. (2007). Residual-based shadings for visualizing (conditional) independence. *Journal of Computational and Graphical Statistics*, *16*(3), 507–525. doi: 10.1198/106186007X237856
44. Zhang, J., Yu, Q., Zheng, F., Long, C., Lu, Z., & Duan, Z. (2016). Comparing keywords plus of WOS and author keywords: A case study of patient adherence research. *Journal of the Association for Information Science and Technology*, 67(4), 967-972. doi: 10.1002/asi.23437