A Natural Language Processing Approach Towards Harmonized Communication of Uncertainties Identified During the European Medicine Authorization Process

Stefan Verweij[1,2,*], Vincent Haverhoek[1,3], Erik Bergman[4], Gabriel Westman[4,5] and Lourens T. Bloem[3]

Within the European Union, the European Medicines Agency's (EMA's) European Public Assessment Report (EPAR) is an important source of information for healthcare professionals and patients that allows them to understand important risks and uncertainties associated with the use of a medicine. However, the EPAR sections describing such important uncertainties can differ substantially in wording, length, and detail, thereby potentially limiting understanding. In this study, we therefore present a natural language processing approach to cluster sentences extracted from the sections on uncertainties in EPARs of centrally authorized medicines, as a steppingstone to harmonization of text describing uncertainties. We used a BERT language model together with dimensionality reduction (Uniform Manifold Approximation and Projection (UMAP)) and clustering (Density-Based Spatial Clustering of Applications with Noise (DBSCAN)) to identify semantic similarities between sentences. Clusters were labeled according to an overarching topic by reviewing the semantically similar sentences. Each cluster was also characterized according to medicine-related characteristics, such as efficacy or side effects. In total, 1,648 medicines were included in this study. For 573 of these medicines (authorized July 27, 2010 to December 31, 2022), we identified an EPAR that described a complete regulatory dossier and contained sections on uncertainties. Of these, 553 EPARs could be attributed to unique active substance-indication combinations. In these 553 EPARs, we identified 13,105 sentences in sections on uncertainties, leading to 26 clusters of which 2 were labeled as noise. The clusters and associated topics provided in this article can be used by regulators and medicine developers as a steppingstone toward a unified way of communicating uncertainties identified during the EMA process to the broader public.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Important uncertainties about the benefits and risks of authorized medicines may be poorly communicated to healthcare professionals and patients by medicine regulators. Although the European Medicines Agency's (EMA) European Public Assessment Report (EPAR) describes such uncertainties, substantial differences in wording, length, and detail may lead to limited understanding of uncertainties.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ We aimed to design a natural language processing (NLP) approach to contribute to understandable and harmonized communication of important remaining uncertainties about authorized medicines to healthcare professionals and patients.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ The clustering approach proposed in our study can be used as a steppingstone to harmonized communication of uncertainties by clustering semantically similar sentences with the use of advanced NLP algorithms. Each cluster contains various uncertainties related to a specific topic, such as age, or pregnancy and fertility, and could thereby serve as candidates for standardization of text describing uncertainties regarding the particular topic.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ Our findings can contribute to more consistent description and communication of uncertainties concerning the benefits and risks of medicines, by understanding the variability in text that is used to describe uncertainties. This is not only relevant for EPARs but may also facilitate communication about uncertainties in other regulatory documents, such as the Summary of Product Characteristics and the package leaflet. Ultimately, this should ensure understandable and harmonized communication of uncertainties, for which we provide recommendations for future steps. In addition, the NLP approach described here illustrates how language models can be used to interpret regulatory and medical documentation.

Received October 6, 2023; accepted January 10, 2024. doi:10.1002/cpt.3195

[1]Dutch Medicines Evaluation Board, Utrecht, The Netherlands; [2]Unit of PharmacoTherapy, Epidemiology and Economics, Groningen Research Institute of Pharmacy, Groningen University, Groningen, The Netherlands; [3]Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands; [4]Swedish Medical Products Agency, Uppsala, Sweden; [5]Department of Medical Sciences, Uppsala University, Uppsala, Sweden. *Correspondence: Stefan Verweij (stefan.verweij@rug.nl)

In the European Union, the European Medicines Agency (EMA) is responsible for formulating an opinion whether new medicines can be authorized. Once a new medicine is authorized by the European Commission, a European Public Assessment Report (EPAR) is published by the EMA on their website to ensure transparency on the specificities underlying that decision. Among others, this EPAR includes extensive details on the evidence concerning quality, safety, and efficacy of the new medicine, which can amount to more than 200 pages for innovative medicines. The final section of the EPAR, the benefit–risk discussion, provides details on how the benefits and risks of the new medicine have been weighed against each other, such that the benefit–risk balance is considered – along with the EMA's opinion – either positive or negative. This includes consideration of important uncertainties that remain at the time of the authorization decision, along with whether and how these will be addressed.[1]

Along with the medicine's label – in the European Union called the Summary of Product Characteristics (SmPC) and package leaflet – the EPAR is an important source of information for healthcare professionals and patients that allows them to understand important risks and uncertainties associated with the use of a medicine. Important uncertainties concern, for example, limited knowledge about effects in certain patient groups, such as those with kidney or liver dysfunction, lacking data on long-term safety and efficacy, or questions whether specific side effects can take place after use of the medicine.[2] However, the EPAR sections describing such important remaining uncertainties can differ substantially in wording, length, and detail. These differences may be due to several reasons. For example, there is no common language or taxonomy to describe uncertainties and every EPAR is written by different authors ("Rapporteurs") from different European countries. Therefore, differences in linguistic backgrounds and preferences may lead to differences in the description of uncertainties between EPARs, and therewith medicines. Moreover, one can imagine that similar uncertainties are described differently over time, due to increased understanding of underlying causes and ways to address them. Consequently, communication of important remaining uncertainties about authorized medicines to healthcare professionals and patients is expected to be hampered.[3]

Natural language processing (NLP) techniques, a popular branch of artificial intelligence (AI) due to the likes of GPT-4 and BERT, are useful for clustering patterns of text and can therefore assist in creating a common consensus of standardized and consistent language.[4,5] Recently, colleagues at the Swedish Medical Products Agency published an example on how to use NLP as an approach to harmonize SmPCs and package leaflets to stimulate the development of the electronic format of the product information.[6] Moreover, the US Food and Drug Administration (FDA) applied NLP techniques to extract data from FDA labeling.[7] However, the EU SmPC and the US label are typically better structured and more standardized than the EPAR, and especially the EPAR sections on important remaining uncertainties. Therefore, in this study, we aimed to design an NLP approach to cluster sentences addressing uncertainties in EPARs of centrally approved medicines in the European Union, ultimately to contribute to understandable and harmonized communication of important remaining uncertainties about authorized medicines to healthcare professionals and patients.

METHODS

Identification of EPARs and construction of datasets

For medicines initially authorized in the European Union between January 1, 1995 (the year the EMA was established) and December 31, 2022, EPARs concerning this initial marketing authorization procedure were retrieved from the EMA website (www.ema.europa.eu) as PDF files on April 5, 2023. Using an internal database of the Dutch Medicines Evaluation Board and Utrecht University, these EPARs were linked to attributes of the respective medicines they concerned, such as the medicines' legal basis, active substance, and initially authorized indication(s).

For these EPARs, we first established whether a specific section on uncertainties was present, concerning "beneficial"/"favorable" effects and/or "unfavorable" effects. From the EPARs wherein at least one such section was present, we created two datasets: one that only contained EPARs of medicines for which a complete dossier had been submitted by the pharmaceutical company (i.e., marketing authorization applications referring to Article 8(3) of European Directive 2001/83/EC as the legal basis), thereby excluding EPARs referring to other legal bases, such as generics (Article 10(1)), biosimilars (Article 10(4)), and medicines authorized based on "well-established use" (Article 10a).[8] These latter EPARs were expected to only refer to an already authorized product (the reference product) or contain very few remaining uncertainties. The resulting dataset was named the "innovative medicines dataset." For the second dataset, duplicate EPARs were also removed, that is, EPARs for medicines that contained the same active substance(s) and were authorized for the same indication but as separate medicines with different brand names. For each set of duplicate EPARs, only the EPAR from the earliest authorized medicine was kept in the dataset to ensure that retrieved text described uncertainties for unique medicines. When duplicate medicines had the same authorization date, we randomly chose one of the EPARs. This second dataset was named the "unique innovative medicines dataset" and formed the main dataset for the analyses.

Text acquisition and preprocessing

The EPAR PDF files of both datasets were converted to raw text using the *MuPDF-1.21.1* package in Python 3.10. Consecutively, using regular expressions, the sections describing important remaining uncertainties within the benefit–risk discussion chapter were extracted and categorized as describing uncertainties concerning "favorable" (efficacy) or "unfavorable" (safety) effects, according to the section header in the EPAR. In addition, we assessed the length of the sections on uncertainties for the five countries most often appointed as Rapporteur (i.e., lead assessor, further described as "Rapporteur countries") to gain initial insight into potential country-specific differences in describing uncertainties.

Dimension reduction and clustering

All sentences were encoded into fixed-length semantic vector representations using the pretrained Sentence-BERT (SBERT) all-mpnet-base-v2 model.[9] To allow clustering of nonlinear geometries in embedding space (i.e., being able to identify underlying structure in a

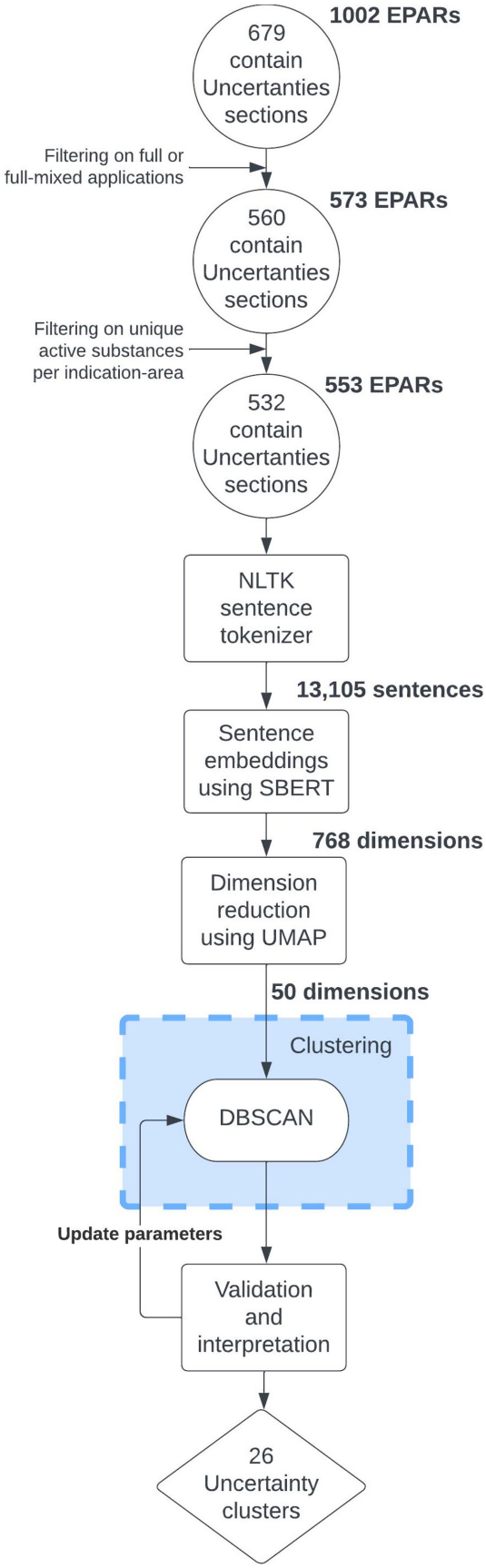

Figure 1 Pipeline of the natural language processing algorithm. EPARs of medicines for which a complete dossier had been submitted were selected ($n=573$). Of these, we included one EPAR for each active substance-indication combination for our main analysis. From the included EPARs ($n=553$), the sections describing uncertainties were extracted and the sentences were tokenized and, consecutively, embeddings were made from these tokenized sentences using SBERT. The dimensions of these embeddings were reduced to improve clustering. The clusters were then manually validated. DBSCAN, Density-Based Spatial Clustering of Applications with Noise; EPARs, European Public Assessment Reports; NLTK, Natural Language Toolkit; SBERT, Sentence-Bidirectional Encoder Representations from Transformers; UMAP, Uniform Manifold Approximation and Projection.

sentence of a lower intrinsic dimension), manifold approximation and projection (UMAP) dimensionality reduction was conducted prior to clustering, in line with our aim to facilitate future harmonization. UMAP was applied using *umap-learn-0.5.3* with 50 neighbors and 50 dimensions together with the Euclidian distance as metric and the default value (0.50) for the minimal distance.[10] Thereafter, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm was used to aggregate the sentence embeddings into similar clusters using the Python package *scikit-learn-1.2.0*. We used the Silhouette coefficient and Davies-Bouldin and Caliński-Harabasz indices together with the number of clusters and mean cluster size to determine the DBSCAN model parameters, that is, the epsilon (ε) describing the radius of the circle around each data point and the minimum cluster size.[11–13] Stop words have been removed from the naming using *nltk-3.8.1*.

Cluster labeling

Labeling of the clusters was performed by two independent reviewers (authors S.V. and L.T.B.), by manually reviewing the sentences within the clusters to identify the overarching topic(s) covered by most sentences. Any disagreement was discussed between the two reviewers until consensus was reached. Sentences that were considered unrelated to the identified topic(s) were labeled as noise. The extent of noise was then calculated as the proportion of sentences within each cluster that was labeled as such, and expressed as a percentage. Thereafter, each cluster was further characterized as containing sentences relating to one or more of the following medicine-related characteristics: mechanism of action, efficacy, side effects, patient characteristics, interactions, and risk management. An overview of the entire pipeline is provided in **Figure 1**. Finally, the distribution of each cluster over time was visualized in histograms based on the EU authorization dates of the medicines for which sentences had been clustered together.

Sensitivity analyses

We performed four sensitivity analyses to explore the robustness of our approach. First, we applied the clustering algorithm optimized for the main dataset to the innovative medicine's dataset, using the same model parameters (**Figure S1**). Second, we separately optimized the model parameters for the innovative medicine's dataset (**Figure S1**). We then compared the resulting clusters to the clusters identified for the unique innovative medicines dataset and counted overlapping cluster topics. Third, 10,000 bootstrap samples (i.e., random samples with replacement) were drawn from the 13,105 sentence embeddings. Each bootstrap sample was clustered using DBSCAN with the same parameters as used in our main analysis. Per bootstrap sample, similarity between the bootstrap clusters and the clusters identified in our main analysis was measured by calculating the Adjusted Rand Index (ARI), where a value of 1 indicates that clusters were identical – and the approach fully robust – whereas a value of 0 indicates

that overlap between clusters had been identified due to chance. Based on these 10,000 cluster comparisons, we calculated the mean ARI with 95% confidence intervals (CIs).[14,15] Fourth, we repeatedly drew 10,000 random subsamples without replacement of the 13,105 sentence embeddings, using different subsample fractions ranging from 0.00 to 1.00 with a step size of 0.05. Each subsample was clustered using DBSCAN with the same parameters as used in our main analysis. Per subsample, similarity between the subsample clusters and the clusters identified in our main analysis was measured by calculating the ARI. For each fraction size, we calculated and visualized the mean ARI with 95% CIs based on the 10,000 iterations.

RESULTS

Identification of EPARs and sections on uncertainties

A total of 1,648 medicines were centrally authorized in the European Union between 1995 and 2022. Of their EPARs, the first that described uncertainties that were concatenated into a specifically themed section concerned dexamethasone (Ozurdex, EU/1/10/638), which was authorized on July 27, 2010. Therefore, we used this date as the cutoff date for our datasets. From this date, we identified 1,002 EPARs concerning initial authorization of new medicines (**Figure 2**). Of these 1,002 EPARs, only 679 (67.8%) contained at least one section addressing uncertainties. When restricting to the innovative medicines EPAR dataset, this increased to 560 of 573 EPARs (97.7%), whereas for the unique innovative medicines EPAR dataset, it concerned 532 of 553 EPARs (96.2%).

In the dataset used for the main analyses – the unique innovative medicines dataset – 13,105 sentences were identified in the sections on uncertainties, of which 6,899 (52.6%) concerned favorable effects and 6,206 (47.4%) concerned unfavorable effects. The number of words in the sections on uncertainties differed substantially between the top 5 Rapporteur countries, that is, Sweden (78 EPARs), the United Kingdom (71 EPARs), the Netherlands (70 EPARs), Germany (60 EPARs), and Spain (32 EPARs), ranging from 743 words for the Netherlands to 501 for the United Kingdom (**Figure S2**).

Dimension reduction and clustering

SBERT initially yielded 768 dimensions that were reduced to 50 using UMAP. As DBSCAN parameters, an epsilon (ε) of 0.50 and minimum cluster size of 60 samples were chosen based on the 3 previously mentioned clustering validation methods, number of clusters, and mean cluster size (**Figures S3–S7**). Subsequently, 26 clusters were generated from the dataset based on 9,099 sentences (**Table 1**, **Figure 3**). The remaining 4,006 sentences could not be clustered.

Cluster labeling

We labeled the sentences that could not be clustered as outliers. Of the 26 clusters, 2 clusters were fully considered noise (clusters 2 and 24). Of the 24 remaining clusters, on average, 3.9% of the embedded sentences was considered noise. Most clusters were further characterized as addressing efficacy aspects and side effects in relation to the overarching cluster topic (**Table 2**). Examples from 3 clusters are provided in **Table 3**.[16-29] An overview of all clusters can be found in the **Supplementary Information**. Finally, the distribution of each cluster over time is visualized in **Figure S8**.

Sensitivity analyses

Using the same DBSCAN model parameters for the clustering of sentences of the innovative medicine's dataset led to 30 clusters, of which one was labeled as noise (**Table S1**). Of the 29 remaining clusters, 24 could be matched to the earlier identified clusters for the unique innovative medicine's dataset (**Table S1**). The other way around, all but one of the 24 clusters identified in the main analysis could be linked to the innovative medicine's dataset

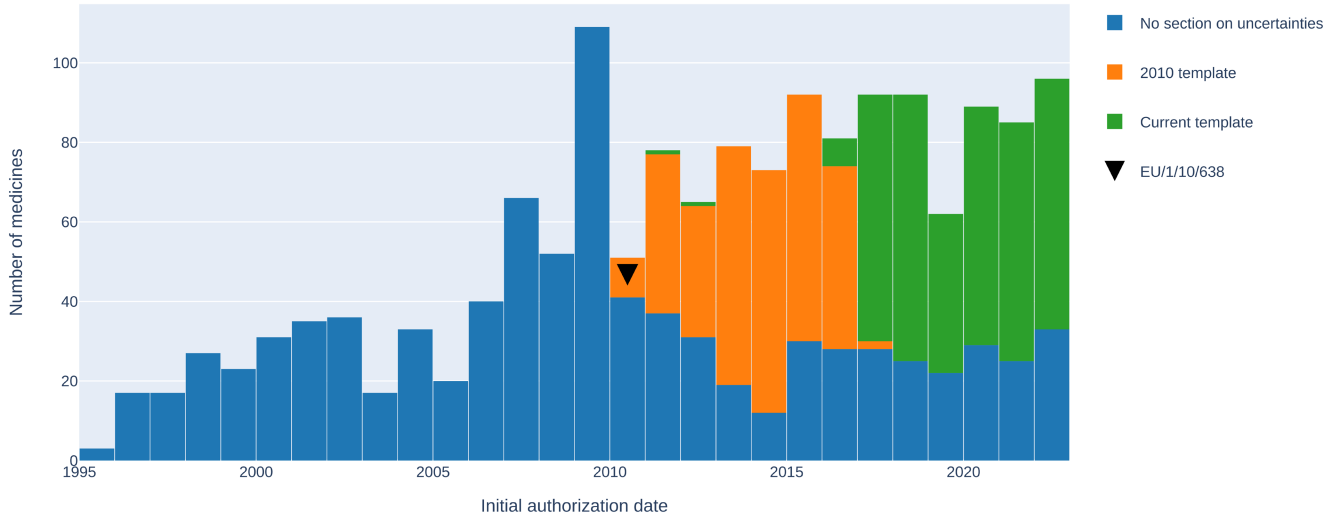

Figure 2 Evolution of the sections on uncertainties in the EPAR over time. Before the authorization of dexamethasone (Ozurdex) on July 27, 2010 (black triangle), there were no occurrences of specific sections on uncertainties (blue). In the 2010 EPAR template, uncertainties were discussed in sections with headers "Uncertainties in the knowledge about unfavorable effects" and "Uncertainties in the knowledge about beneficial effects" (orange). In the current EPAR template, uncertainties are discussed in sections with headers "Uncertainties and limitations about unfavorable effects" and "Uncertainties and limitations about favorable effects" (green). EPARs that were lacking a section on uncertainties after July 27, 2010, could often be attributed to a certain type of application (e.g., generic medicines). EPAR, European Public Assessment Report.

Table 1 Information on the 26 clusters identified in the unique innovative medicine's dataset

Cluster ID	Uncertainties related to	*N* Sentences	*N* Favorable sentences	*N* Unfavorable sentences	% Noise	*N* Unique innovative medicines	*N* Unique active substances
1	Psychiatry and psychopharmacological medicines	127	63	64	2.4%	25	25
2	Noise	3,359	2,196	1,163	100.0%	504	492
3	Immunology and oncology	986	430	556	2.0%	254	249
4	Age	389	196	193	5.9%	185	180
5	Cardiology	207	50	157	14.0%	95	94
6	Vaccines	503	299	204	0.2%	49	47
7	Posology	356	262	94	5.9%	158	155
8	Organ impairment	556	132	424	4.3%	195	193
9	Pain and opioids	88	64	24	4.5%	15	15
10	Antidiabetics	206	81	125	1.5%	25	25
11	Pharmacovigilance and lack of safety data	165	29	136	19.4%	130	129
12	Virology and antivirals	343	251	92	2.9%	69	68
13	Respiratory disease and clotting factors	121	99	22	5.0%	28	28
14	Body weight and GLP1-agonists	289	144	145	10.4%	55	53
15	Ophthalmology	183	45	138	1.6%	38	36
16	Insulin growth factor 1	78	40	38	0.0%	7	7
17	Bone	81	36	45	0.0%	18	18
18	Hematology	224	109	115	0.0%	51	50
19	HIV medicines	93	65	28	1.1%	8	8
20	Pregnancy and fertility	259	39	220	0.8%	97	95
21	Antibiotics and antimycotics	79	65	14	0.0%	15	15
22	Patient group differences in adverse event occurrence	100	2	98	1.0%	65	64
23	Effects on LDL and LDL-lowering medicines	132	54	78	3.8%	14	14
24	Noise	50	45	5	100.0%	12	12
25	Migraine and headache	60	45	15	0.0%	7	7
26	Adverse events and injections	65	8	57	6.2%	57	56
n.a.	Outliers	4,006	2050	1956	100.0%	492	480
Sum (mean for noise)		13,105	6,899	6,206	3.9%		

GLP, glucagon-like peptide; HIV, human immunodeficiency virus; LDL, low-density lipoprotein; n.a., not applicable.
Cluster 2, 24 and the outliers were not taken into account when calculating the mean within-cluster percentage of noise.

(**Table S2**). Using the separately optimized model parameters for the innovative medicines dataset led to the same results.

Bootstrap sampling with 10,000 iterations resulted in a mean ARI of 0.87 (95% CI: 0.81–0.91), indicating highly similar clusters and thus robustness of our approach. Subsampling with a sample fraction of 0.9 for 10,000 iterations resulted in a mean ARI of 0.88 (95% CI 0.85–0.90), also indicating robustness. The mean ARI for other subsample fractions can be found in **Figure 4**.

DISCUSSION

We illustrated a relatively easy to implement NLP approach to facilitate the harmonization of text describing uncertainties in EPARs by clustering semantically similar sentences. This approach expands on a study by Bergman *et al*. (2022) in which the authors designed a similar approach for medicine labels, typically more structured and standardized medicine regulatory texts than the EPAR [6]. With our NLP approach, we aim to contribute to

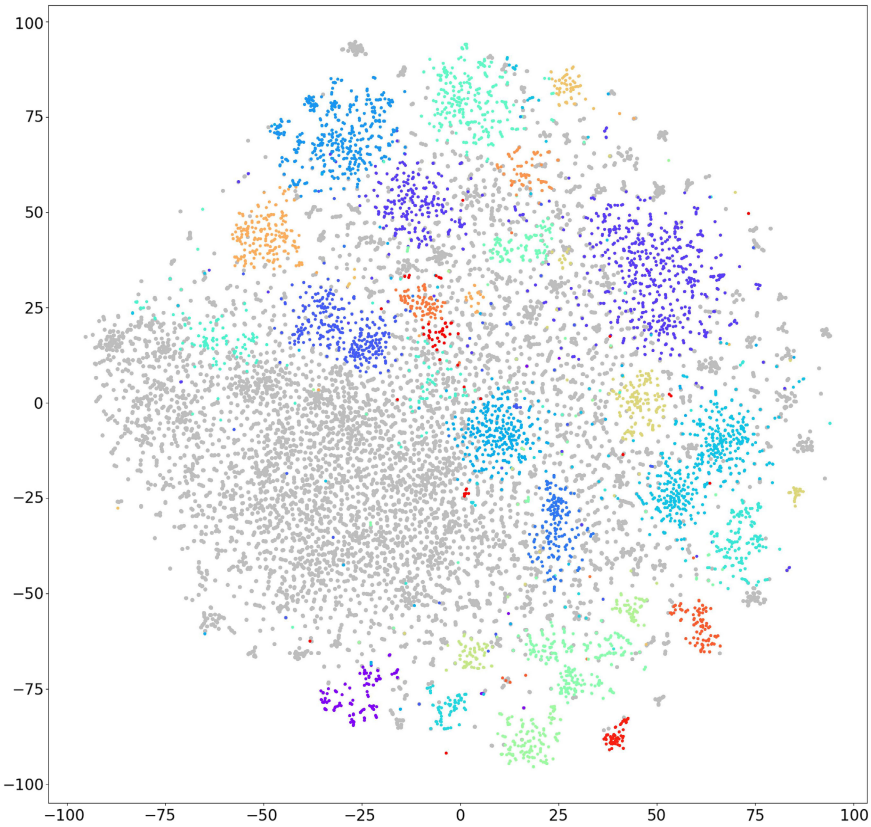

Figure 3 Two-dimensional t-SNE plot of the 13,105 sentences extracted from the EPAR sections on uncertainties. The sentences are colored according to assigned cluster. Each sentence is shown as a circle and each cluster of sentences was assigned a specific color coding. Cluster 2 and 24 are shown in gray, together with the 4,006 outlier sentences. EPAR, European Public Assessment Report.

better understandable and harmonized communication of important remaining uncertainties about the benefits and risks of authorized medicines to healthcare professionals and patients.

We identified 26 clusters of semantically similar sentences in our dataset, of which 24 could be linked to a common topic such as a specific field in medicine (e.g., cardiology) or a specific class of medicines (e.g., antidiabetics). Notably, we observed that some clusters seemed specific to a certain time period, most likely due to the marketing authorization of a class of medicines in that period. For example, the cluster containing uncertainties related to LDL and LDL-lowering medicines contained 101 of 125 sentences linked to medicines containing bempedoic acid as one of their active substances, which were all authorized in 2020. However, for most clusters, there is an increasing trend over time that is in line with the increase in EPARs published over the years. Moreover, we noticed the existence of potential subclusters within certain clusters. For example, we identified a subcluster of sentences related to clotting factors within the cluster related to the respiratory system (cluster 13). We would have expected the sentences concerning clotting factor to be clustered within the hematology cluster (cluster 18) because clotting factors are more closely related to hematology than the respiratory system.[30] The existence of additional layers of subclusters may require further examination.

Within each cluster, we identified multiple types of uncertainties. Most sentences were describing efficacy-related uncertainties or uncertainties about side effects of medicines. Some clusters seemed more efficacy-related, with most sentences coming from the EPAR section describing uncertainties and limitations about favorable effects (e.g., antibiotics and antimycotics; cluster 20), whereas others seemed more related to side effects, with most sentences coming from the section describing uncertainties and limitations about unfavorable effects (e.g., pharmacovigilance and lack of safety data; cluster 11). Future work may focus on exploring the sentiment expressed in these sentences – being more positive or negative – using, for example, Valence Aware Dictionary and sEntiment Reasoner (VADER) and Textblob.[31,32] Such sentiment analysis may elucidate specific semantic orientations, for example, in texts concerning uncertainties about efficacy and side effects, or specific diseases or classes of medicines.

Notably, more than half of the 13,105 sentences were labeled as outliers or noise. This seems largely due to how the sections describing uncertainties in the EPAR are written. Some sentences did not address uncertainty but provided medicine-specific context (e.g., *"The efficacy is similar to the established treatment for post-operative pain- IV morphine PCA."*), or referred to other sentences as part of multi-sentence reasoning (e.g, *"This was observed in a higher proportion of patients taking methadone than those not taking methadone."*).[33,34] Alternatively, some sentences could not be clustered due to their uniqueness, such as describing uncertainties specific to characteristics of the medicine (e.g., *"The effect size of the vehicle compared to other common emollients is not known."*).[35] Furthermore, although we did use a sentence tokenizer, we noticed

Table 2 Characterization of the 26 identified clusters according to medicine-related characteristics

Cluster ID	Uncertainties related to	Mechanism of action	Efficacy	Side effects	Patient characteristics	Interactions	Risk management
1	Psychiatry and psychopharmacological medicines		X	X			
2	Noise	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
3	Immunology and oncology		X	X			
4	Age		X	X	X		
5	Cardiology		X	X			
6	Vaccines		X	X		X	
7	Posology		X	X			
8	Organ impairment			X			
9	Pain and opioids		X	X			
10	Antidiabetics			X			
11	Pharmacovigilance and lack of safety data	X		X		X	X
12	Virology and antivirals		X	X			
13	Respiratory disease and clotting factors		X	X			
14	Body weight and GLP1-agonists		X				
15	Ophthalmology		X	X			
16	Insulin growth factor 1		X	X			
17	Bone		X	X			
18	Hematology		X	X			
19	HIV medicines		X	X			
20	Pregnancy and fertility		X	X	X		
21	Antibiotics and antimycotics		X				
22	Patient group differences in adverse event occurrence			X			X
23	Effects on LDL and LDL-lowering medicines		X	X			
24	Noise	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
25	Migraine and headache			X			
26	Adverse events and injections			X			X
n.a.	Outliers	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Sum		1	13	18	2	2	3

GLP, glucagon-like peptide; HIV, human immunodeficiency virus; LD, low-density lipoprotein; n.a., not applicable.

that the context of the period (".") punctuation mark was not always interpreted correctly when extracting sentences from the EPARs and, as a result, sentences were sometimes split on non-sentence-ending periods like decimal points or acronyms (e.g., *"The placebo group (median 70.7 vs.")*).[36] This may have affected the clustering.

The discussion about how regulators should report and communicate uncertainties is not new. In 2014, the US Institute of Medicine organized a workshop about the steps to take in characterization and communication of uncertainty in benefit–risk assessments of medicines.[37] In the same year, the EMA introduced a so-called "effects table" in the EPAR, with an aim to make their decision making about benefits and risks more consistent and transparent.[38] However, this table does not consistently report uncertainties and, if reported, their format may differ from table to table. This can, for example, be observed within the EPAR of tisagenlecleucel (Kymriah) that was initially authorized for two different indications, with the format of the uncertainties reported in the associated effects tables differing between the indications.[39] In relation, Simpkin and Armstrong (2019) reviewed the communication of uncertainties associated with clinical decision making and highlighted challenges and barriers, including communicating uncertainties, in a meaningful way such that it improves decision making.[40] Medicine regulators have an important role in effectively communicating uncertainties, such that it facilitates informed decision making while building trust in the medicine regulatory system.

Our findings can contribute to more consistent description and communication of uncertainties concerning the benefits and risks of medicines in EPARs, by understanding the variability in text that is used to describe uncertainties. However, this is a first

Table 3 Selection of sentences describing uncertainties from three exemplary clusters[16–29]

Cluster ID			
INN	***N* sentences**	***N* unique innovative medicines**	***N* unique active substances**
Uncertainties related to age			
4	389	185	180
Sotrovimab	"Due to the small sample size of participants > **85 years of age**, no meaningful clinical conclusion can be drawn for that population."		
Defatted powder of *Arachis hypogaea* L., semen (peanuts)	"However, the numbers (especially of **12–17-year-old subjects**) are small and a distinction in efficacy between **age groups** cannot be made."		
Ozanimod	"No (controlled) safety data are available for **pediatric** subjects (< **18 years of age**) and **elderly** subjects (> **55 years of age**)."		
Lenvatinib	"There are no data on the use of lenvatinib in **pediatric population**."		
Pandemic influenza vaccine (H5N1) (split virion, inactivated, adjuvanted)	"However, it is not possible to predict whether half the **adult** dose would suffice in **children aged 3 years**."		
Uncertainties related to pharmacovigilance and lack of safety data			
11	165	130	129
Pegcetacoplan	"Full **safety results** for study APL2-302, compiling data across all periods of the study are needed."		
Cerliponase alfa	"**Long-term safety data** for ICV BMN 190 treatment **is limited**."		
Voretigene neparvovec	"**Adverse events not so far reported** may become apparent as more subjects are exposed to the current product."		
Ravulizimab	"**Further safety data will be obtained** from the final Clinical Study Report for Studies ALXN1210-PNH-301 and ALXN1210-PNH-302 a registry study."		
Risankizumab	"Although 7 months additional **safety data** has been submitted by the applicant long-term exposure to risankizumab (> 18 months) **is limited**."		
Uncertainties related to pregnancy and fertility			
20	259	97	95
Vosoritide	"There is no data regarding the use of vosoritide during **pregnancy**."		
Defatted powder of *Arachis hypogaea* L., semen (peanuts)	"In addition, the effect of Palforzia on the immune system of the **mother and fetus** during **pregnancy** is unknown."		
Dupilumab	"However, data from use of dupilumab is too limited to draw any conclusions on potential **embryofetal harms**."		
Brentuximab vedotin	"The precise mechanism of **testicular** toxicity in rats is not known and there is uncertainty on the presence of CD30 in human **spermatogonia/early spermatocytes**."		
Pre-pandemic influenza vaccine (H5N1) (surface antigen, inactivated, adjuvanted)	"Sporadic cases of **pregnancy** were reported in some studies, but the number of cases was very small and no firm conclusion could be drawn."		

Topic-related keywords within the sentences are shown in bold.
INN, international nonproprietary name.

step. We recommend examining how the variability identified in the text describing uncertainties affects interpretation and understanding of these uncertainties by healthcare professionals and patients, and what wording is most effective. Based on these insights, text describing uncertainties in EPARs should be standardized where possible. In addition, this can facilitate the development of understandable and consistent communication about uncertainties in other regulatory documents important for clinical decision making, such as the SmPC and the package leaflet, where such communication is currently missing.[41] Moreover, our findings can fuel activities and discussions related to the recently published multi-annual AI workplan of the Heads of Medicines Agencies and EMA's combined Big Data Steering Group, which aims to "harness the capabilities of AI for personal productivity, process automation and systems efficiency, increased insights into data and strengthened decision-support for the benefit of public and animal health".[42] For example, our NLP approach, together with other regulatory NLP approaches such as those developed by Bergman *et al.* (2023), can contribute to the development of knowledge mining and communication support roadmaps.[42,43]

Our study has several limitations. First, whereas our sensitivity analyses indicated substantial robustness of our clustering method, we would like to highlight the role of the dataset used in identifying certain specific clusters. Using the innovative medicine's dataset, we identified four clusters specifically related to risk management, such as references to the SmPC, and containing sentences describing the necessity of post-authorization data generation. Although we considered these four clusters relevant for harmonization of communicating uncertainties and how they are addressed, we did

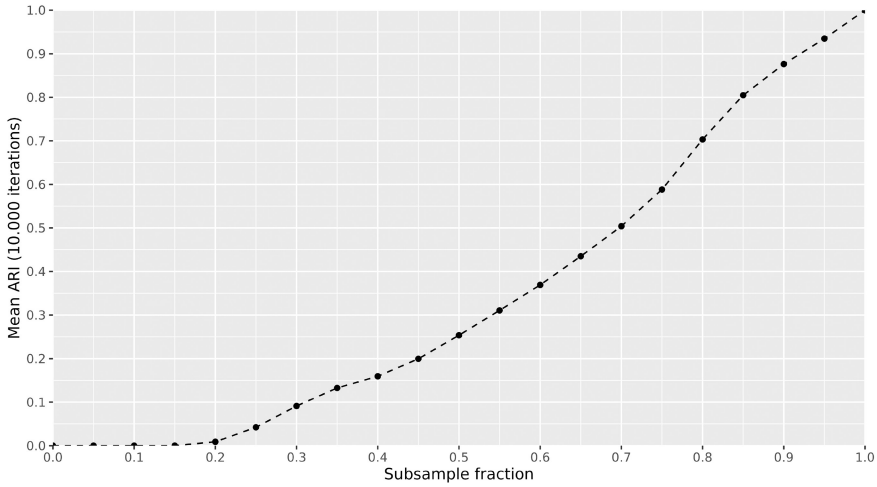

Figure 4 Mean ARI of 10,000 iterations per subsample fraction. We drew 10,000 random subsamples for a set of subsample fractions ranging from 0 to 1 (using a step size of 0.05). Each subsample was clustered using DBSCAN and the ARI was calculated as a measure of similarity between the subsample clusters and the clusters identified in the main analysis. For each subsample fraction, the mean ARI was then calculated based on these 10,000 subsample comparisons. ARI, Adjusted Rand Index; DBSCAN, Density-Based Spatial Clustering of Applications with Noise.

not identify them in our main analysis. Second, we were not able to cluster sentences according to certain medicine-specific characteristics, such as marketing authorization pathways (e.g., conditional or exceptional authorization), product types (e.g., advanced therapy medicinal products) or clinical development support through the PRIME scheme, because there were relatively few medicines with these characteristics in our study period. Third, because our dataset of 13,105 sentences contained relatively few data points per cluster, we were not able to perform formal analyses of differences in cluster representation over time. However, we did visualize distributions over time through histograms.

In conclusion, we designed a straightforward NLP approach to cluster similar sentences extracted from sections on uncertainties in the EPAR that can be used by regulators and medicine developers as a steppingstone toward harmonized communication of uncertainties concerning benefits and risks of medicines to the broader European public.

SUPPORTING INFORMATION

Supplementary information accompanies this paper on the *Clinical Pharmacology & Therapeutics* website (www.cpt-journal.com).

ACKNOWLEDGMENTS

The authors thank Anne C. Taams, MSc (Dutch Medicines Evaluation Board and Utrecht University) for sharing her expertise in categorizing uncertainties in European benefit–risk discussions and contributing to the labeling of uncertainties.

FUNDING

No funding was received for this work.

CONFLICT OF INTEREST

The authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

S.V., V.H., E.B., G.W., and L.T.B. wrote the manuscript. S.V. and L.T.B. designed the research. V.H. performed the research. S.V., V.H., and L.T.B. analyzed the data. E.B. and G.W. contributed new analytical tools.

DISCLAIMER

The views expressed in this article are the personal views of the authors and may not be understood or quoted as being made on behalf of or reflecting the position of the Dutch Medicines Evaluation Board, the Swedish Medical Products Agency, the European Medicines Agency, or one of their committees or working parties.

1. *European Medicines Agency, Day 150 and Day 195 Joint Response Assessment Report (clinical template rev 05.21)* (European Medicines Agency, Amsterdam, 2021) .
2. Bloem, L.T. *et al*. European conditional marketing authorization in a rapidly evolving treatment landscape: a comprehensive study of anticancer medicinal products in 2006–2020. *Clin. Pharmacol. Ther.* **114**, 148–160 (2023).
3. EUPATI. Risk communication in medicines <https://toolbox.eupati.eu/resources/risk-communication-in-medicines/> (2015). Accessed December 20, 2023.
4. Devlin, J., Chang, M., Lee, K. & Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv.org* (2018) https://doi.org/10.48550/arXiv.1810.04805.
5. Ecoffet, A. GPT-4 technical report. OpenAI arXiv.org (2023) https://doi.org/10.48550/arXiv.2303.08774.
6. Bergman, E., Sherwood, K., Forslund, M., Arlett, P. & Westman, G. A natural language processing approach towards harmonisation of European medicinal product information. *PLoS One* **17**, e0275386 (2022).

7. Shi, Y., Ren, P., Zhang, Y., Gong, X., Hu, M. & Liang, H. Information extraction from FDA drug labeling to enhance product-specific guidance assessment using natural language processing. *Front. Res. Metr. Anal.* **6**, 670006 (2021).
8. European Parliament and Council of the European Union. Directive 2001/83/EC of the European Parliament and of the council of 6 November 2001 on the community code relating to medicinal products for human use. *Official Journal L* **311**, 0067–0128 (2001).
9. Reimers, N. all-mpnet-base-v2. *HuggingFace* <https://huggingface.co/sentence-transformers/all-mpnet-base-v2> (2021).
10. Allaoui, M., Kherfi, M.L. & Cheriet, A. Considerably improving clustering algorithms using UMAP dimensionality reduction technique: a comparative study. In *International Conference on Image and Signal Processing* (Springer, Cham, Switzerland, 2020). https://doi.org/10.1007/978-3-030-51935-3_34.
11. Rousseeuw, P.J. & Kaufman, L. *Finding Groups in Data: An Introduction to Cluster Analysis* (John Wiley & Sons, Hoboken, New Jersey, 2009).
12. Davies, D.L. & Bouldin, D.W. A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **2**, 224–227 (1979).
13. Caliński, T. & Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat. – Theory Meth.* **3**, 1–27 (1974).
14. Hubert, L. & Arabie, P. Comparing partitions. *J. Classif.* **2**, 193–218 (1985).
15. Efron, B. Bootstrap methods: another look at the jackknife. In *Breakthroughs in Statistics: Methodology and Distribution* 569–593 (Springer, New York, 1992).
16. Committee for Medicinal Products for Human Use. *European Public Assessment Report of Xevudy* (European Medicines Agency, Amsterdam, 2021).
17. Committee for Medicinal Products for Human Use. *European Public Assessment Report of Palforzia* (European Medicines Agency, Amsterdam, 2020).
18. Committee for Medicinal Products for Human Use. *European Public Assessment Report of Zeposia* (European Medicines Agency, Amsterdam, 2020).
19. Committee for Medicinal Products for Human Use. *European Public Assessment Report of Lenvima* (European Medicines Agency, London, 2015).
20. Committee for Medicinal Products for Human Use. *European Public Assessment Report of Pumarix* (European Medicines Agency, London, 2011).
21. Committee for Medicinal Products for Human Use. *European Public Assessment Report of Aspaveli* (European Medicines Agency, Amsterdam, 2021).
22. Committee for Medicinal Products for Human Use. *European Public Assessment Report of Brineura* (European Medicines Agency, London, 2017).
23. Committee for Medicinal Products for Human Use. *European Public Assessment Report of Luxturna* (European Medicines Agency, London, 2018).
24. Committee for Medicinal Products for Human Use. *European Public Assessment Report of Ultomiris* (European Medicines Agency, Amsterdam, 2019).
25. Committee for Medicinal Products for Human Use. *European Public Assessment Report of Skyrizi* (European Medicines Agency, London, 2019).
26. Committee for Medicinal Products for Human Use. *European Public Assessment Report of Voxzogo* (European Medicines Agency, Amsterdam, 2021).
27. Committee for Medicinal Products for Human Use. *European Public Assessment Report of Dupixent* (European Medicines Agency, London, 2017).
28. Committee for Medicinal Products for Human Use. *European Public Assessment Report of Adcetris* (European Medicines Agency, London, 2012).
29. Committee for Medicinal Products for Human Use. *European Public Assessment Report of Aflunov* (European Medicines Agency, London, 2010).
30. Pipe, S.W. Recombinant clotting factors. *Thromb. Haemost.* **99**, 840–850 (2008).
31. Hutto, C.J. & Gilbert, E.E. *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*, Vol. **8**. 2 (Eighth International Conference on Weblogs and Social Media (ICWSM-14), Washington D.C., 2014).
32. Loria, S. *Textblob* (Github, San Francisco, 2018) <https://github.com/sloria/textblob>.
33. Committee for Medicinal Products for Human Use. *European Public Assessment Report of IONSYS* (European Medicines Agency, London, 2015).
34. Committee for Medicinal Products for Human Use. *European Public Assessment Report of Moventig* (European Medicines Agency, London, 2014).
35. Committee for Medicinal Products for Human Use. *European Public Assessment Report of Staquis* (European Medicines Agency, Amsterdam, 2020).
36. Committee for Medicinal Products for Human Use. *European Public Assessment Report of Uptravi* (European Medicines Agency, London, 2016).
37. Claiborne, A.B., English, R.A. & Caruso, D. *Characterizing and Communicating Uncertainty in the Assessment of Benefits and Risks of Pharmaceutical Products: Workshop Summary* (National Academies Press, Washington D.C., 2014).
38. Committee for Medicinal Products for human use working group. In *Update on work package 5: effects table pilot (phase I). In Benefit-Risk Methodology Project* (European Medicines Agency, London, 2014).
39. Committee for Medicinal Products for Human Use. *European Public Assessment Report of Kymriah* (European Medicines Agency, London, 2018).
40. Simpkin, A.L. & Armstrong, K.A. Communicating uncertainty: a narrative review and framework for future research. *J. Gen. Intern. Med.* **34**, 2586–2591 (2019).
41. Davis, C. *et al*. Communication of anticancer drug benefits and related uncertainties to patients and clinicians: document analysis of regulated information on prescription drugs in Europe. *BMJ* **380**, e073711 (2023).
42. HMA-EMA Big Data Steering Group. *Multi-Annual AI Workplan 2023–2028* (European Medicines Agency, Amsterdam, 2023).
43. Bergman, E., Pasmooij, A.M.G., Mol, P.G.M. & Westman, G. A full-document analysis of the semantic relation between European public assessment reports and EMA guidelines using a BERT language model. *PLoS One* **18**, e0294560 (2023).