Ontology-based Semantic Similarity Measures for Clustering Medical Concepts in Drug Safety

Jeffery L. Painter, MS, JD[1], François Haguinet[2], MS, Gregory E. Powell, PharmD, MBA[1]
Andrew Bate, PhD[3,4]

[1]GSK, Durham, NC, USA; [1]GSK, Wavre, Belgium; [3]GSK, London, UK; [4]London School of Hygiene and Tropical Medicine, London, UK

Abstract

Semantic similarity measures (SSMs) are widely used in biomedical research but remain underutilized in pharmacovigilance. This study evaluates six ontology-based SSMs for clustering MedDRA Preferred Terms (PTs) in drug safety data. Using the Unified Medical Language System (UMLS), we assess each method's ability to group PTs around medically meaningful centroids. A high-throughput framework was developed with a Java API and Python/R interfaces support large-scale similarity computations. Results show that while path-based methods perform moderately with F1 scores of 0.36 for WUPALMER and 0.28 for LCH, intrinsic information content (IC)-based measures, especially INTRINSIC_LIN and SOKAL, consistently yield better clustering accuracy (F1 Score of 0.403). Validated against expert review and standard MedDRA queries (SMQs), our findings highlight the promise of IC-based SSMs in enhancing pharmacovigilance workflows by improving early signal detection and reducing manual review.

Introduction

Semantic similarity measures (SSMs) are essential tools in biomedical informatics, enabling applications in genomics, ontology enrichment, and phenotypic clustering [1]. They support functional clustering of genes and proteins [2,3] and have been used to classify diseases from structured clinical data, including electronic health records (EHRs) [4,5]. Despite their success in genomics and clinical phenotyping, SSMs remain underutilized in pharmacovigilance (PV). This is surprising, as the ability to cluster similar cases effectively is critical for identifying safety issues with medicines as early as possible [6]. Moreover, no single level of specificity is optimal across all use cases in the hierarchical terminologies commonly used in PV [6]. The ability to use semantic distance to cluster safety events, therefore, has considerable potential to improve medicine and vaccine safety.

PV relies heavily on standardized terminologies— WHODrug[1] for medicinal products and MedDRA for adverse events (AEs) [7]. MedDRA plays a central role in aggregating clinically similar outcomes across patients using a rigid four-level hierarchy. However, this structure often fails to reflect more nuanced or mechanistic relationships between terms [8]. For example, *neutropenia* (10029354) and *bone marrow failure* (10065553)—both indicators of myelosuppression—reside in entirely separate High-Level Group Term (HLGT) branches. Widely adopted safety data mining practices typically analyze data at the preferred term (PT) level, which would lead to these two terms being treated independently [6].

In addition, MedDRA imposes a binary grouping model: AEs are either in the same category or not, with no graded scale of similarity and no mechanism to incorporate evolving clinical knowledge or terminology changes. This lack of granularity limits MedDRA's ability to capture the broader semantic landscape that underlies many safety concepts.

Semantic similarity offers a potential solution by quantifying the degree of relatedness between terms on a continuous scale. Rather than relying solely on fixed hierarchies, semantic distance measures—particularly when derived from integrated ontologies—can identify previously unlinked but clinically relevant AEs. Dupuch and Grabar [9] demonstrated this by mapping MedDRA PTs into SNOMED-CT [10] hierarchies via UMLS[2], creating

---

[1] https://who-umc.org/whodrug/whodrug-global/

[2] https://www.nlm.nih.gov/research/umls/index.html

the *ontoEIM* resource. This enriched semantic network introduced intermediate nodes and decomposed complex AEs into semantic primitives (e.g., gastric ulcer → ulcer + stomach), yielding more fine-grained clusters than MedDRA alone. Their study also showed improved alignment with Standardized MedDRA Queries (SMQs) [11] and enhanced signal detection performance.

Recent work has further demonstrated the value of data-driven semantic representations for PV. Gattepaille [12] and Erlanson et al. [13] applied distributional embedding models (vigiVec) to learn vector representations of AEs and drugs from spontaneous reports in WHO's VigiBase. These embeddings captured clinically coherent relationships— often outperforming MedDRA and ATC[3] groupings in tasks such as intruder detection, an evaluation task in which the model must identify the least related term in a group. However, these models derive similarity from co-occurrence patterns in observational data and therefore reflect empirical rather than ontological proximity. These results provide early evidence that semantic approaches—whether ontology-based or corpus-derived—can outperform traditional hierarchical groupings in clustering and relatedness tasks.

In contrast, our approach computes semantic distance from formal biomedical ontologies, leveraging the hierarchical structures and curated relationships encoded in UMLS. We integrated MedDRA, SNOMED-CT, and MeSH [14] to construct comprehensive semantic networks and computed both path-based and information content (IC)-based similarity metrics using an optimized implementation of Apache cTAKES YTEX [15]. Unlike corpus-based models, this ontology-grounded framework assesses concept relatedness based on expert-defined medical knowledge structures rather than usage patterns. To support reproducible, large-scale evaluations, we developed a RESTful API with Python and R interfaces, enabling scalable similarity computations across extensive concept sets.

We benchmarked the performance of multiple SSMs by clustering MedDRA Preferred Terms (PTs) around clinically meaningful centroids. Although we initially explored algorithmic centroid detection (e.g., selecting the PT with the highest average similarity to others in an SMQ), we adopted a more interpretable approach: using narrow SMQs that contain a PT matching the SMQ's name as centroids. This enabled consistent evaluation and straightforward comparison across methods. Cluster validity was assessed against both SMQs and expert medical review. Our evaluation demonstrates that semantic similarity measures (SSMs), when carefully applied to drug safety data, can facilitate the clustering of Preferred Terms (PTs) around medically relevant centroids, offering a scalable, ontology-driven alternative to traditional hierarchical groupings.

Methods

Semantic Similarity Measures Evaluated

To identify effective semantic similarity measures (SSMs) for clustering safety-related concepts, we explored ontological representations of MedDRA, MeSH and SNOMED-CT, integrated via UMLS. From this, we selected six ontology-based SSMs leveraging hierarchical structure and information content (IC): two path-based measures— Leacock and Chodorow (LCH) [16] and Wu and Palmer (WUPALMER) [17]—and four IC-based measures— Resnik (INTRINSIC_RESNIK) [18], Lin (INTRINSIC_LIN) [19], an intrinsic variant of LCH (INTRINSIC_LCH), and Sokal (SOKAL) [20, 21]. Each measure's formal definition follows standard implementations and is outlined below.

Leacock and Chodorow (LCH)

The Leacock and Chodorow measure computes similarity based on the shortest path length between two concepts in the taxonomy, scaled by the maximum depth of the hierarchy. Due to the logarithmic transformation, it is not constrained between 0 and 1:

$$\text{LCH}(c_1, c_2) = -log\left(\frac{\text{length}(c_1, c_2)}{2 \times \text{max\_depth}}\right).$$

---

[3] ATC = Anatomical Therapeutic Chemical classification for drugs and medicines

Wu and Palmer (WUPALMER)

The Wu and Palmer measure quantifies similarity by considering the depths of the two concepts and their lowest common ancestor (LCA) in the hierarchy:

$$\text{WUPALMER}(c_1, c_2) = \frac{2 \times \text{depth}\big(\text{LCA}(c_1, c_2)\big)}{\text{depth}(c_1) + \text{depth}(c_2)}.$$

Resnik (INTRINSIC_RESNIK)

Resnik's measure is based on the IC of the LCA of two concepts. IC is typically derived from concept occurrence frequencies in a corpus or computed intrinsically from ontology structure. Like LCH, this metric is unbounded:

$$\text{RESNIK}(c_1, c_2) = IC\big(\text{LCA}(c_1, c_2)\big)$$

Lin (INTRINSIC_LIN)

Lin's measure extends Resnik's approach by incorporating the IC of the individual concepts, yielding a normalized similarity score:

$$\text{LIN}(c_1, c_2) = \frac{2 \times IC\big(\text{LCA}(c_1, c_2)\big)}{IC(c_1) + IC(c_2)}.$$

Intrinsic Leacock and Chodorow (INTRINSIC_LCH)

This variant modifies the LCH measure by incorporating intrinsic IC, adjusting path-based similarity using IC-weighted scaling. Although not widely standardized, we refer to it as INTRINSIC_LCH for clarity. Like LCH, it is not constrained between 0 and 1.

$$\text{INTRINSIC}_{\text{LCH}(c_1, c_2)} = -\log\left(\frac{IC(c_1) + IC(c_2) - 2 \times IC\big(\text{LCA}(c_1, c_2)\big) + 1}{2 \times max\_IC}\right)$$

Sokal (SOKAL)

The Sokal similarity measure applies IC-based transformations like Lin but incorporates a specific normalization factor. It can be interpreted as an IC-based distance measure reformulated into a similarity metric [20].

$$\text{SOKAL}(c_1, c_2) = \frac{IC\big(\text{LCA}(c_1, c_2)\big)}{2 \times \big(IC(c_1) + IC(c_2)\big) - 3 \times IC\big(\text{LCA}(c_1, c_2)\big)}$$

Evaluation of SSMs

We assessed SSMs by their ability to cluster Preferred Terms (PTs) around centroid PTs. Centroids were drawn from Standardized MedDRA Queries (SMQs) and internal safety-related PT groupings from GSK product global data sheets. Due to limited utility of higher MedDRA hierarchy levels, we focused on narrow SMQs (20–50 PTs) containing a PT identical to the SMQ name, declaring this PT the centroid for clarity. Fifteen SMQs met this criterion, of which 12 overlapped with internal GSK-defined safety concept sets.

For each of these 12 centroid PTs, we computed pairwise similarity scores using the six selected SSMs across all PTs in MedDRA v26.1.

Metric Selection

The feasibility study initially assessed 10 semantic similarity measures across different ontology configurations (MedDRA, MeSH, SNOMED-CT), yielding 40 SSM-ontology combinations. For each SMQ-SSM pair, we applied agglomerative clustering, identifying the similarity threshold required to include all SMQ PTs, then measuring how

many external PTs (outside the SMQ) exceeded that threshold. As shown in Figure 1, performance was variable across the different measures.
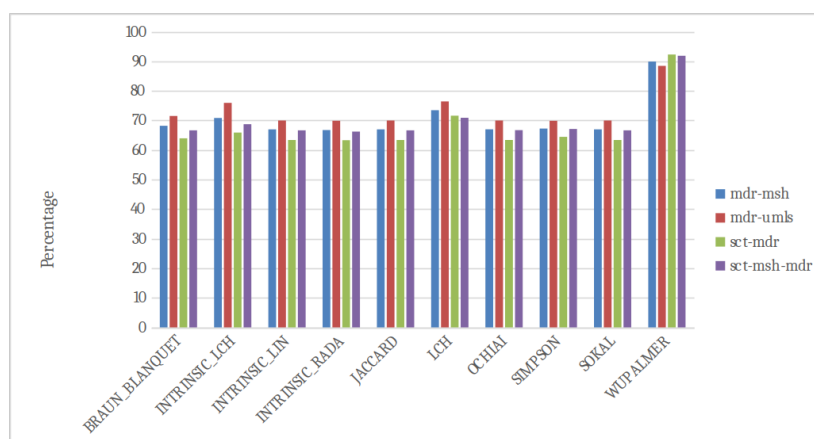

Figure 1: Feasibility analysis of SSMs and ontological views. MDR: MedDRA; MSH: MeSH; SCT: SNOMED-CT; UMLS: Unified Medical Language System

Results indicated that, except for the two path-based metrics (LCH and WU_PALMER), the average proportion of external PTs was consistently lower when using MedDRA+SNOMED-CT integrated ontologies. Among the different SSMs, intrinsic IC-based metrics generally included fewer external PTs compared to path-based metrics. However, no clear distinction was observed between the different intrinsic IC-based metrics. Among path-based methods, LCH and WUPALMER consistently had lower performance, with WUPALMER achieving the weakest performance across all ontology configurations.

Ultimately, the final six selected measures – LCH, WUPALMER, INTRINSIC_RESNIK, INTRINSIC_LIN, INTRINSIC_LCH, and SOKAL – were chosen based on their strong feasibility study performance, established literature support, and methodological complementarity (path-based vs. IC-based approaches). This selection enables a comparative analysis of path-based and IC-based methods, ensuring a rigorous evaluation of SSMs for pharmacovigilance applications.

Based on these findings, we refined our evaluation to focus on MedDRA+SNOMED-CT integrated ontologies and the selected six metrics, aligning with previously reported performance trends [20].

Medical Review and Reference Standards

Six medical experts independently reviewed PT lists generated per centroid PT. Each list was assessed by two reviewers, with discrepancies resolved by a senior expert. Reviewers underwent calibration training and judged whether each PT should be retrieved for the associated safety concept (e.g., Interstitial lung disease).

Twelve SMQs were reviewed (Table 1), each anchored by a centroid PT matching the SMQ name. PTs were drawn from three sources: the SMQ itself, PTs sharing a High-Level Group Term (HLGT) with the centroid, and proprietary "autolist" groupings. Autolists are expert-defined sets of MedDRA PTs typically developed to reflect safety concepts captured in product labeling for which a MedDRA based SMQ may not exist in support of internal monitoring. Unlike SMQs, which are standardized and maintained by CIOMS and ICH[4], autolists offer more targeted groupings, though their proprietary nature may limit external reproducibility.

These PTs formed the non-semantic distance-based inclusion test set. Each SMQ included an average of 88 PTs (median = 82.5), with many additional terms arising from broad HLGT groupings. PTs retrieved solely via semantic similarity served as negative controls.

---

[4] https://cioms.ch/working-groups/meddra/

Table 1: SMQs included in the evaluation.

SMQ ID	Centroid PT	SMQ Name	Control PTs	Total PTs
20000024	10002424	Angioedema	101	580
20000004	10007554	Cardiac failure	64	429
20000150	10007636	Cardiomyopathy	136	542
20000073	10012267	Dementia	77	520
20000154	10012305	Demyelination	77	615
20000146	10018304	Glaucoma	63	450
20000042	10022611	Interstitial lung disease	128	569
20000217	10028533	Myelodysplastic syndrome	107	582
20000047	10028596	Myocardial infarction	66	455
20000222	10038695	Respiratory failure	117	584
20000045	10042945	Systemic lupus erythematosus	55	473
20000238	10040477	Sexual dysfunction	88	548
		Total	1,079	6,347

Constructing the Review Datasets

For each centroid PT, similarity scores to all MedDRA PTs (n=26,409) were computed and stored in CSV format. To control reviewer burden, we sampled the top 200 SSM-ranked[5] PTs (excluding non-semantically related PT codes) per centroid. If the 200th ranked PT had more than 50 terms tied at the same score, a maximum of 50 were randomly sampled to maintain diversity, while preventing reviewer overload. On average, 529 PTs were evaluated per SMQ (median = 548). The full selection process for the set of review codes is outlined in Figure 2.
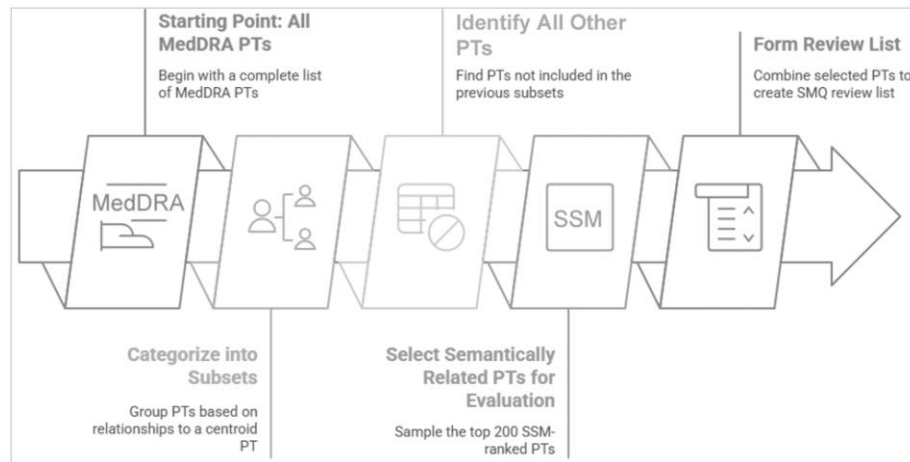

Figure 2: Process for constructing evaluation dataset

[5] SSM measures were computed exclusively under the MedDRA+SNOMED-CT ontological view.

Performance Metrics and Review Process
SSM performance was evaluated against expert judgments using receiver operating characteristic (ROC) curves and precision-recall (P-R) curves. F1-scores were computed across varying similarity thresholds and averaged across all SMQs. To assess how well SSM-based clustering aligned with existing pharmacovigilance (PV) frameworks, we compared clusters against SMQ and HLGT classifications. Cohen's kappa was used to measure agreement with manual review.

Review Process and Adjudication
To reduce reviewer fatigue, PT lists were randomized and split into two equal groups while maintaining ranking integrity within similarity blocks. Across 12 SMQs, a total of 6,347 PTs were manually reviewed, requiring 218 hours of expert evaluation. Discrepancies were adjudicated, and Cohen's kappa was computed per centroid PT to benchmark agreement with SMQ and HLGT groupings, establishing a baseline for SSM-based clustering performance. While the review protocol included adjudication to improve consistency, the process inherently introduces subjectivity. Calibration training helped align reviewers, but some variability inevitably remains, as noted in prior work [22]. This highlights the value of SSMs in supporting—but not replacing—expert curation.
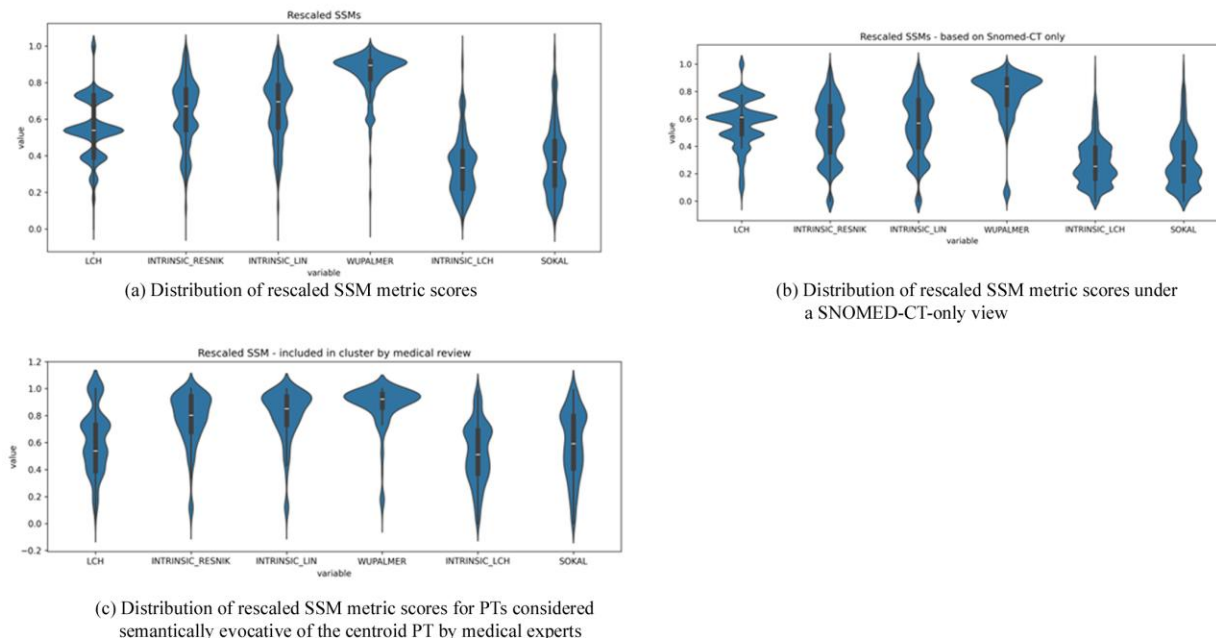
Results

Figure 3 presents the distribution of agreement (Cohen's Kappa) between medical review and both SMQ and HLGT classifications. Overall, agreement was higher with SMQs, though the variability between centroid PTs was substantial.
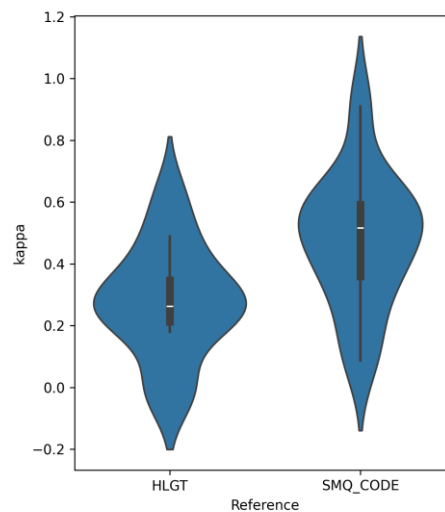
Figure 3: Cohen's Kappa comparing medical review to SMQ and HLGT groupings

Since the evaluated SSMs are not inherently normalized to [0,1], we rescaled them based on their minimum and maximum values across all PTs. The resulting distributions (Figure 4(a)) revealed bimodal patterns for INTRINSIC_RESNIK, INTRINSIC_LIN, INTRINSIC_LCH, and SOKAL. Initially, we hypothesized that the integration of MedDRA and SNOMED-CT contributed to this effect; however, a similar multi-modal distribution was observed when using SNOMED-CT alone (Figure 4(b)), suggesting a more complex underlying structure.

(a) Distribution of rescaled SSM metric scores

(b) Distribution of rescaled SSM metric scores under
a SNOMED-CT-only view

(c) Distribution of rescaled SSM metric scores for PTs considered
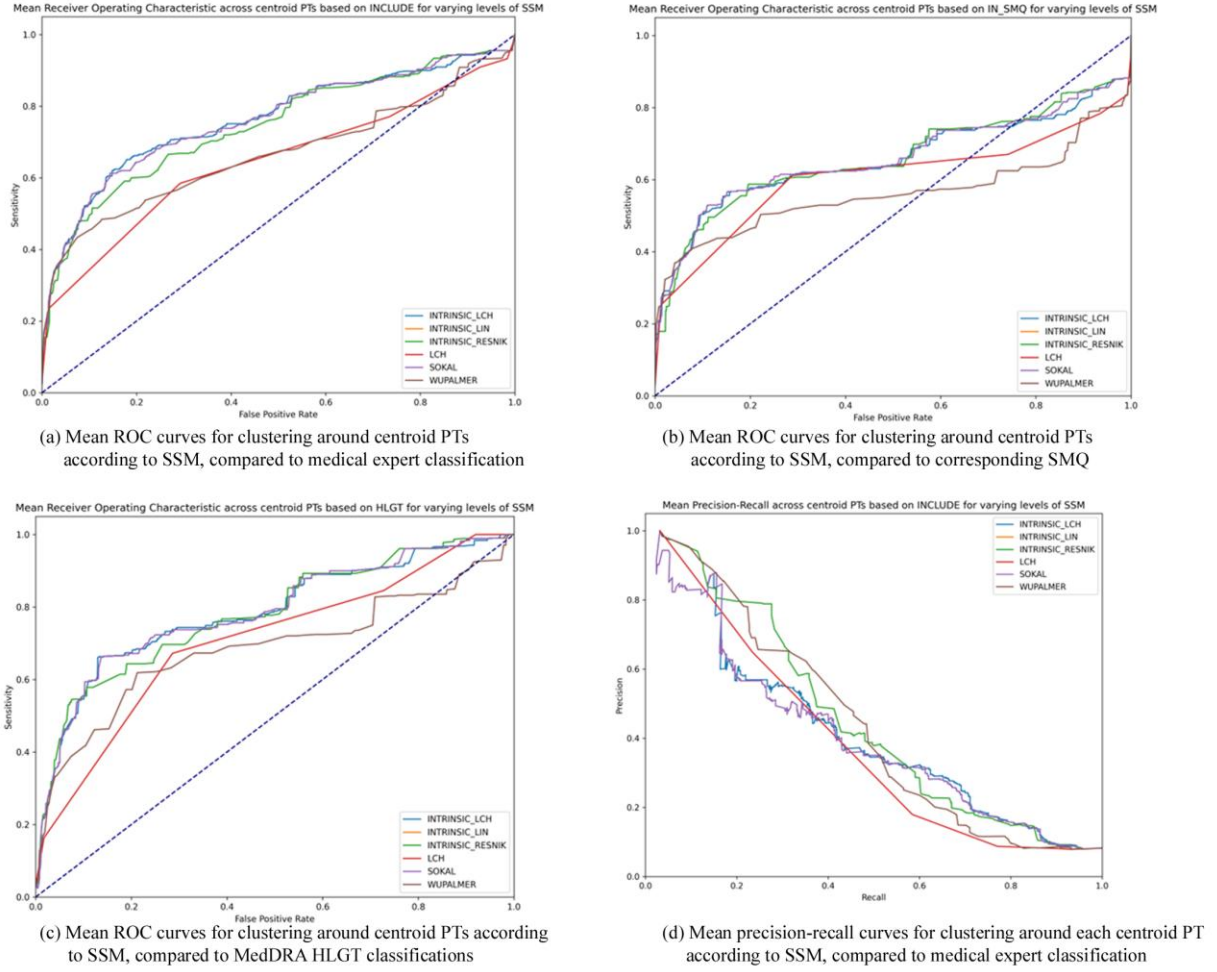semantically evocative of the centroid PT by medical experts

Figure 4: SSM Distributions

PTs included in the expert-determined clusters had higher SSM values than excluded PTs (Figure 4(c)). The difference in distributions was more pronounced for INTRINSIC_LCH and SOKAL, likely due to their narrower and lower-value distributions.

Figure 5(a) displays the mean ROC curves across centroid PTs based on medical review. Path-based SSMs (LCH, WUPALMER) underperformed compared to IC-based SSMs. Notably, LCH exhibited a step-like ROC curve due to its limited set of distinct values. Among IC-based SSMs, INTRINSIC_LIN and SOKAL produced identical curves due to their strictly monotonic relationship. INTRINSIC_RESNIK performed less favorably than the other IC-based measures.

When SMQs were used as the reference standard ROC curves (Figure 5(b)) exhibited similar performance for all IC-based SSMs. However, past a false positive rate of 0.75, these curves dipped below the identity line. Under this criterion, LCH showed closer alignment with IC-based curves, whereas WUPALMER's performance deteriorated more sharply. HLGT-based results (Figure 5(c)) closely resembled those obtained using medical review.

The mean precision-recall (P-R) curves derived from medical review revealed that the positive predictive value (PPV) of INTRINSIC_LIN, INTRINSIC_LCH, and SOKAL was lower at sensitivity levels below 0.5 but subsequently increased (Figure 5(d)). In contrast, WUPALMER exhibited the opposite trend. This pattern became clearer when examining the F1-score as a function of SSM values (Figure 6). The highest F1-score was achieved by INTRINSIC_LCH (0.404), followed closely by INTRINSIC_LIN and SOKAL, both scoring 0.403. The path-based SSMs achieved lower performance with F1 scores of 0.36 for WUPALMER and 0.28 for LCH. Notably, the F1-score curve of INTRINSIC_LIN mirrored that of SOKAL but was shifted toward higher SSM thresholds. When SSM thresholds increased, the increase in F1 score seemed to follow mainly a two steps shape, with a first sharp increase between from 0.2 to 0.3 for INTRINSIC_LCH and SOKAL and between 0.55 and 0.7 for INTRINSIC_RESNIK and INTRINSIC_LIN. The second sharp increase was between 0.4 and 0.45 for INTRINSIC_LCH, between 0.45 and 0.5 for SOKAL, and between 0.7 and 0.8 for INTRINSIC_RESNIK and INTRINSIC_LIN.

(a) Mean ROC curves for clustering around centroid PTs according to SSM, compared to medical expert classification

(b) Mean ROC curves for clustering around centroid PTs according to SSM, compared to corresponding SMQ

(c) Mean ROC curves for clustering around centroid PTs according to SSM, compared to MedDRA HLGT classifications

(d) Mean precision-recall curves for clustering around each centroid PT according to SSM, compared to medical expert classification

Figure 5: ROC Curve Analyses

F1-scores computed using SMQ-based evaluation were generally lower than those based on medical review, with a maximum score of 0.367 for the three best-performing SSMs. Similarly, HLGT-based evaluation yielded slightly lower F1-scores, with 0.398 for INTRINSIC_LCH and 0.387 for both INTRINSIC_LIN and SOKAL.
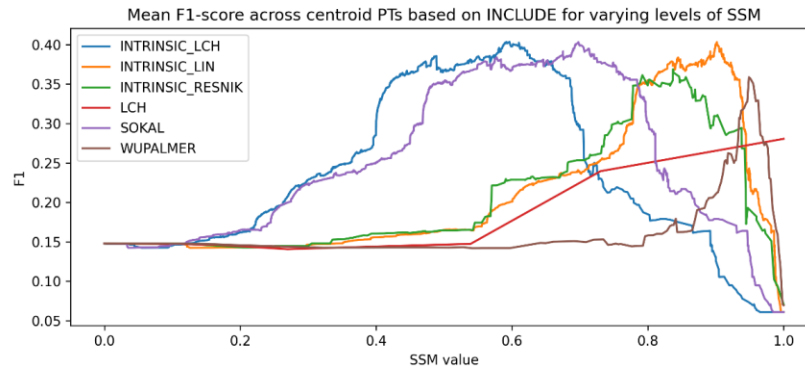
Figure 6: F1-score as a function of SSM thresholds for clustering around centroid PTs according to each SSM, compared to medical expert classification

Discussion

Our evaluation demonstrates that semantic similarity measures (SSMs), when carefully applied to drug safety data, can effectively cluster Preferred Terms (PTs) around clinically relevant centroids. Among the six tested SSMs, intrinsic information content (IC)-based measures—particularly INTRINSIC_LCH, INTRINSIC_LIN, and SOKAL—consistently outperformed path-based approaches. These measures likely benefit from incorporating both IC and conceptual overlap, enabling more nuanced differentiation between related terms. In contrast, the relatively lower performance of INTRINSIC_RESNIK suggests that using only the IC of the lowest common ancestor (LCA) may be insufficient for fine-grained clustering.

Interestingly, several SSMs exhibited bimodal or multimodal similarity score distributions, even within SNOMED-CT. These patterns likely reflect the influence of ontology-specific properties, such as term granularity and concept density. Future work may explore ontology preprocessing techniques, including branch pruning or IC recalibration, to address such distributional artifacts. Although the IC-based SSMs showed improved performance over path-based methods, the overall gains remain incremental. This underscores the importance of future research into hybrid or ensemble models that may yield stronger clustering effects.

While SSM-based clustering approximated expert judgment, it did not fully replicate it. Comparisons with established references (SMQs, HLGT) revealed that HLGT-based groupings unexpectedly outperformed SMQs in some cases, potentially due to the composition of the selected SMQs. These findings reinforce the value of SSMs as decision-support tools that complement—but do not replace—clinical expertise.

Our evaluation framework—anchored by expert-reviewed centroid PTs—ensures practical relevance but introduces subjectivity, echoing challenges identified by Pedersen et al. [22] regarding alignment between expert raters. However, the focus on 12 SMQs and internal autolist concepts limits generalizability. These terms were selected for their relevance to known safety concerns, but additional work is needed to assess the method's performance on a broader array of pharmacovigilance scenarios, including newer or rare adverse events. Moreover, the number of centroids evaluated was constrained by the availability of medical reviewers, a common limitation in semantic similarity studies. Furthermore, the evaluation did not include external benchmarks such as word embeddings, string-based similarity, or clustering via alternate ontological models. Future comparisons with these methods may help situate our approach within the broader landscape of semantic clustering techniques.

An important limitation of this study—and of most PT-level clustering analysis is the assumption that MedDRA PTs consistently and accurately reflect the underlying clinical events. In practice, the selection of PTs in safety reports can be influenced by reporter expertise, site conventions, or data entry constraints, introducing variability that may affect both clustering outcomes and evaluation. As a result, even perfectly grouped PTs may not always correspond to distinct or consistent clinical phenomena. Addressing this limitation may require incorporating contextual data from narratives or leveraging probabilistic models to account for PT variability in future work.

Conclusion

This study presents a systematic evaluation of six ontology-based semantic similarity measures (SSMs) for clustering MedDRA PTs in a PV setting. IC-based metrics, especially INTRINSIC_LIN, INTRINSIC_LCH, and SOKAL, demonstrated strong potential for improving automated clustering of MedDRA PTs. These methods offer a practical path toward reducing manual curation workloads and enhancing early signal detection in large-scale drug safety analyses.

Future research may expand the ontology base beyond MedDRA and SNOMED-CT, explore hybrid models that integrate lexical or contextual similarity, or combine SSMs with statistical signal detection metrics to improve overall predictive accuracy.

Acknowledgements

Declarations

GSK covered all costs associated with the conduct of the study and the development of the manuscript and the decision to publish the manuscript. J.P., F.H, G.P, and A.B. are employed by GSK and hold financial equities. This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Lord PW, Stevens RD, Brass A, Goble CA. Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation. Bioinformatics. 2003;19(10):1275–83.
2. Pesquita C, Faria D, Falcao AO, Lord P, Couto FM. Semantic similarity in biomedical ontologies. PLoS Comput Biol. 2009;5(7):e1000443.
3. Guzzi PH, Mina M, Guerra C, Cannataro M. Semantic similarity analysis of protein data: assessment with biological features and issues. Brief Bioinform. 2012;13(5):569–85.
4. Martínez S, Sánchez D, Valls A. A semantic framework to protect the privacy of electronic health records with non-numerical attributes. J Biomed Inform. 2013;46(2):294–303.
5. Hauben M, Aronson JK. Defining 'signal' and its subtypes in pharmacovigilance based on a systematic review of previous definitions. Drug Saf. 2009;32:99–110.
6. Bate A, Brown EG, Goldman SA, Hauben M. Terminological challenges in safety surveillance. Drug safety. 2012 Jan;35:79-84
7. Fescharek R, Kübler J, Elsasser U, Frank M, Güthlein P. Medical Dictionary for Regulatory Activities (MedDRA) data retrieval and presentation. Int J Pharm Med. 2004;18:259–69.
8. Brown EG. Methods and pitfalls in searching drug safety databases utilising the Medical Dictionary for Regulatory Activities (MedDRA). Drug Saf. 2003;26:145–58.
9. Dupuch M, Grabar N. Semantic distance-based creation of clusters of pharmacovigilance terms and their evaluation. J Biomed Inform. 2015;54:174–85.
10. SNOMED International. SNOMED CT. 2025. Available from: https://www.snomed.org/
11. Mozzicato P. Standardised MedDRA queries: their role in signal detection. Drug Saf. 2007;30:617–19.
12. Gattepaille LM. Using the WHO database of spontaneous reports to build joint vector representations of drugs and adverse drug reactions, a promising avenue for pharmacovigilance. In2019 IEEE international conference on healthcare informatics (ICHI) 2019 Jun 10 (pp. 1-6). IEEE.
13. Erlanson N, China JF, Taavola H, Norén GN. Clinical Relatedness and Stability of vigiVec Semantic Vector Representations of Adverse Events and Drugs in Pharmacovigilance. Drug Safety. 2025 Jan 20:1-3.
14. Dhammi IK, Kumar S. Medical subject headings (MeSH) terms. Indian J Orthop. 2014;48(5):443.
15. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc. 2010;17(5):507–13.
16. Leacock C, Chodorow M. Combining local context and WordNet similarity for word sense identification. In: Fellbaum C, editor. WordNet: An Electronic Lexical Database. Cambridge (MA): MIT Press; 1998.
17. Wu Z, Palmer M. Verb semantics and lexical selection. arXiv [cmp-lg]. 1994. Available from: https://arxiv.org/abs/cmp-lg/9406033
18. Resnik P. Using information content to evaluate semantic similarity in a taxonomy. arXiv [cmp-lg]. 1995. Available from: https://arxiv.org/abs/cmp-lg/9511007
19. Lin D. An information-theoretic definition of similarity. In: Proceedings of ICML. 1998;98:296–304.
20. Sokal RR, Rohlf FJ. The comparison of dendrograms by objective methods. Taxon. 1962;11(2):33–40.
21. Sánchez D, Batet M. Semantic similarity estimation in the biomedical domain: an ontology-based information-theoretic perspective. J Biomed Inform. 2011;44(5):749–59.
22. Pedersen T, Pakhomov SVS, Patwardhan S, Chute CG. Measures of semantic similarity and relatedness in the biomedical domain. J Biomed Inform. 2007;40(3):288–99.