

Extracting and Classifying Drug Discontinuations from Estonian Electronic Health Records: Development and Validation Study

Hendrik Šuvalov, Nikita Umov, Markus Haug, Sven Laur, Marek Oja, Sirli Tamm,
Sulev Reisberg, Jaak Vilo, Raivo Kolde

Submitted to: Journal of Medical Internet Research
on: October 20, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 19

 Figures 20

 Figure 1..... 21

 Figure 2..... 22

 Figure 3..... 23

 Multimedia Appendixes 24

 Multimedia Appendix 1..... 25

Extracting and Classifying Drug Discontinuations from Estonian Electronic Health Records: Development and Validation Study

Hendrik Šuvalov¹ MSc; Nikita Umov² MD; Markus Haug¹ MSc; Sven Laur¹ PhD; Marek Oja¹ PhD; Sirli Tamm¹ MSc; Sulev Reisberg^{1,3} PhD; Jaak Vilo^{1,3} PhD; Raivo Kolde¹ PhD

¹ Institute of Computer Science University of Tartu Tartu EE

² Institute of Family Medicine and Public Health University of Tartu Tartu EE

³ Software Technology and Applications Competence Center Tartu EE

Corresponding Author:

Hendrik Šuvalov MSc

Institute of Computer Science
University of Tartu
Narva mnt 18
Tartu
EE

Abstract

Background: Drug adherence is crucial for chronic disease management, yet treatment discontinuation remains common due to factors such as side effects, inefficacy, or cost. These reasons are often recorded only in free-text clinical notes, making large-scale analysis difficult. While large language models (LLMs) can interpret such unstructured data more effectively than traditional natural language processing methods, few studies have systematically categorized reasons for discontinuation or identified whether the decision was initiated by the patient or the clinician, especially in low-resource languages like Estonian.

Objective: To assess the ability of LLMs to extract and classify reasons for drug discontinuation and identify who initiated it using Estonian electronic health records, and to characterize the observed discontinuation patterns and initiators for statins and antidiabetic medications.

Methods: We combined prescription data with free-text anamneses from a 10% sample of the Estonian population (2012–2019). LLMs (Llama-3.1-70B and GPT-4o) were applied to extract discontinuation phrases and reasons, classify them into a clinician-developed taxonomy, and identify who discontinued the treatment. Performance was evaluated on randomly chosen 100 cases per drug group.

Results: Extraction yielded 625 antidiabetic drug and 233 statin discontinuation cases. Validation confirmed high accuracy, with 93–98% of extracted phrases and 95–96% of extracted reasons judged correct. Classification of discontinuation reasons achieved weighted F1 scores of 0.808–0.836, while classification of who initiated discontinuation achieved weighted F1 scores of 0.645–0.774.

Adverse reactions were the most frequent reason overall, accounting for ~70% of discontinuations for statins and ~44% for antidiabetic drugs. For antidiabetic drugs, treatment inefficacy and contraindications were more common. Patients more often stopped due to adverse reactions or non-medical reasons, while physicians more often initiated discontinuation for contraindications.

Conclusions: LLMs can accurately extract and classify medication discontinuation reasons and initiators from Estonian clinical narratives. Both local and proprietary models performed well, enabling scalable analyses that complement structured health records. This demonstrates the potential of LLMs to unlock information from clinical notes, turning this underutilized EHR component into a valuable resource for monitoring treatment patterns and detecting adverse event signals.

(JMIR Preprints 20/10/2025:86183)

DOI: <https://doi.org/10.2196/preprints.86183>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/>

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://www.jmir.org/>

Original Manuscript

Original Paper

Extracting and Classifying Drug Discontinuations from Estonian Electronic Health Records:

Development and Validation Study

Hendrik Šuvalov, MSc^{1,*}, Nikita Umov, MD², Markus Haug, MSc¹, Sven Laur, PhD¹, Marek Oja, PhD¹, Sirli Tamm, MSc¹, Sulev Reisberg, PhD^{1,3}, Jaak Vilo, PhD^{1,3}, Raivo Kolde, PhD¹

1. Institute of Computer Science, University of Tartu, Tartu, Estonia

2. Institute of Family Medicine and Public Health, University of Tartu, Tartu, Estonia

3. STACC, 51009 Tartu, Estonia

*Corresponding author: Hendrik Šuvalov, MSc, Institute of Computer Science, University of Tartu, Narva mnt 18, 51009 Tartu, Estonia (hendrik.suvalov@ut.ee)

Abstract

Background: Drug adherence is crucial for chronic disease management, yet treatment discontinuation remains common due to factors such as side effects, inefficacy, or cost. These reasons are often recorded only in free-text clinical notes, making large-scale analysis difficult. While large language models (LLMs) can interpret such unstructured data more effectively than traditional natural language processing methods, few studies have systematically categorized reasons for discontinuation or identified whether the decision was initiated by the patient or the clinician, especially in low-resource languages like Estonian.

Objective: To assess the ability of LLMs to extract and classify reasons for drug discontinuation and identify who initiated it using Estonian electronic health records, and to characterize the observed discontinuation patterns and initiators for statins and antidiabetic medications.

Methods: We combined prescription data with free-text anamneses from a 10% sample of the Estonian population (2012–2019). LLMs (Llama-3.1-70B and GPT-4o) were applied to extract discontinuation phrases and reasons, classify them into a clinician-developed taxonomy, and identify who discontinued the treatment. Performance was evaluated on randomly chosen 100 cases per drug group.

Results: Extraction yielded 625 antidiabetic drug and 233 statin discontinuation cases. Validation confirmed high accuracy, with 93–98% of extracted phrases and 95–96% of extracted reasons judged correct. Classification of discontinuation reasons achieved weighted F1 scores of 0.808–0.836, while classification of who initiated discontinuation achieved weighted F1 scores of 0.645–0.774.

Adverse reactions were the most frequent reason overall, accounting for ~70% of discontinuations for statins and ~44% for antidiabetic drugs. For antidiabetic drugs, treatment inefficacy and contraindications were more common. Patients more often stopped due to adverse reactions or non-medical reasons, while physicians more often initiated discontinuation for contraindications.

Conclusions: LLMs can accurately extract and classify medication discontinuation reasons and initiators from Estonian clinical narratives. Both local and proprietary models performed well, enabling scalable analyses that complement structured health records. This demonstrates the potential of LLMs to unlock information from clinical notes, turning this underutilized EHR component into a valuable resource for monitoring treatment patterns and detecting adverse event signals.

Keywords: natural language processing; large language models; LLM; LLMs; NLP; natural language processing; Estonian; healthcare data; drug adherence; drug discontinuations; clinical decision support; data mining

Introduction

Drug adherence and persistence are critical for the management of chronic diseases, as they directly affect both patient outcomes and healthcare costs [1]. Nevertheless, discontinuation of prescribed medications is common [2] driven by factors such as adverse events, lack of perceived efficacy, patient preference, and drug cost [3,4]. While understanding these reasons is essential for improving adherence and informing clinical decision-making, they are rarely standardized or systematically coded in electronic health records (EHR). Instead, they are typically documented in free-text clinical notes, which are difficult to analyze at scale without manual review [5].

Traditional natural language processing (NLP) approaches, including rule-based methods and classical machine learning, can capture common patterns but often fail to detect the nuanced or less frequent circumstances underlying discontinuation decisions. Large language models (LLMs) offer a promising alternative, as they are able to handle semantic complexity and variability in clinical text, making it feasible to classify reasons for discontinuation across thousands of records. Clinical narratives such as patient anamneses, although unstructured and heterogeneous, contain valuable information about patients' experiences and the clinical rationale behind stopping medications, and thus represent an important but underutilized data source.

Prior research on extracting and classifying drug discontinuations has primarily relied on classical NLP methods, such as rule-based approaches or machine learning models based on BERT-type architectures [5,6]. More recently, LLMs have shown strong performance on these tasks. For example, LLMs have been applied to study discontinuations from web-based forums [7], which provide valuable patient-reported insights but often lack clinical context. EHR have been analyzed to identify discontinued medications with LLMs [8], but reasons for discontinuation were not categorized. Research in lower-resource languages is very scarce, and studies addressing whether discontinuation was initiated by the patient or by the clinician are also lacking.

In this study, we use LLMs to examine the discontinuation of antidiabetics and statins among Estonian patients. Specifically, we (1) identify patients who were prescribed a drug from either category and subsequently stopped purchasing it, (2) extract text segments from anamneses that mention discontinuation or reasons for stopping, (3) determine whether the decision was made by the patient or the physician, and (4) categorize the identified reasons using a clinically developed ontology. By linking structured prescription records with unstructured anamneses, we ensure that our cohort truly reflects patients who discontinued therapy and that negative mentions correspond to actual discontinuation of treatment. This approach provides systematic insights into the drivers of medication discontinuation in chronic disease management in Estonia and demonstrates the utility of LLMs for pharmacoepidemiological studies in a low-resource language.

Methods

Dataset Curation

The dataset used for this study was RITA-MAITT, a 10% random sample of the Estonian population ($n = 150\,824$) ranging from 2012 to 2019, consisting of claims, prescription, and EHR data [9]. For this study, only unique and non-empty anamneses were used.

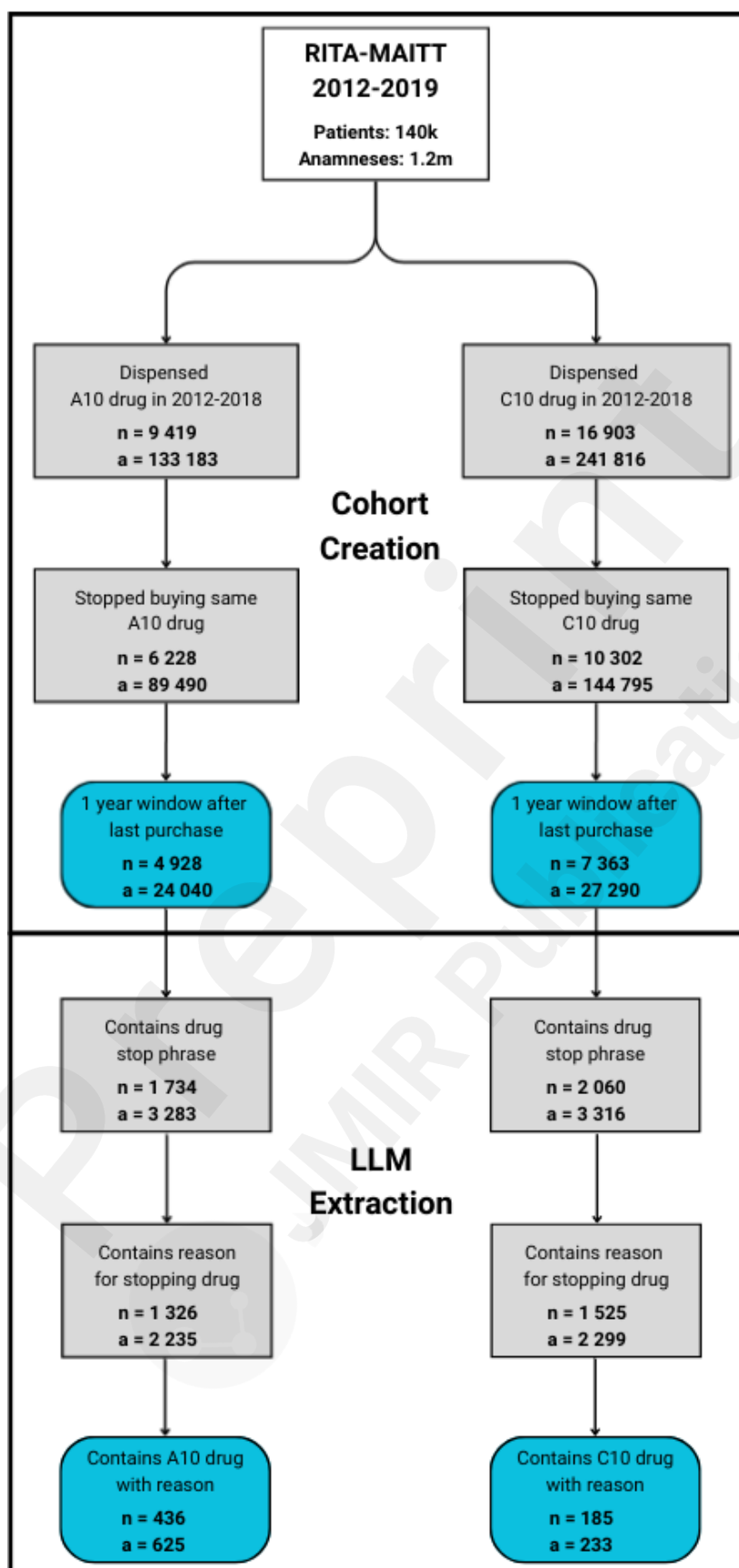
From the prescription records, we identified individuals who had purchased drugs classified under the Anatomical Therapeutic Chemical (ATC) groups A10 (antidiabetics) or C10 (lipid-modifying agents). To detect discontinuation, we focused on purchases made in 2012-2018 and required a subsequent one-year observation window without further purchases of the same drug, ensuring that the absence of refills reflected treatment discontinuation. Importantly, patients were considered to have discontinued a medication if they stopped purchasing one specific ATC code drug, even if they later initiated another drug within the same class (e.g., discontinuation of C10AA05, followed by

initiation of C10AA07). This ensured that discontinuation was defined at the level of individual drugs rather than drug classes.

After establishing the discontinuation cohort, we extracted free-text clinical notes from the EHRs, specifically anamneses sections documented within one year after the last recorded purchase. Anamneses are particularly informative in this context, as they often capture both patient-reported experiences and clinician assessments. Unlike structured fields, anamneses provide the narrative detail necessary to understand the circumstances surrounding treatment discontinuation, making them a valuable source for identifying and classifying underlying reasons.

We obtained 24 040 unique anamneses from 4 928 patients for A10 and 27 290 unique anamneses for 7 363 patients for C10. The data flow diagram is shown in Figure 1 (Cohort Creation).

Figure 1. Data flow diagram consisting of two parts: 1) Cohort Creation - rule-based filtering of anamneses relevant for discontinuation mentions 2) LLM Extraction - extracting documents containing discontinuations linked with antidiabetics (A10) or statins (C10).



Extracting discontinuation information from clinical narratives

Based on expert review of the data, many anamneses did not explicitly state discontinuation but instead described adverse effects or lack of efficacy. Since structured prescription data confirmed that

drug dispensing had stopped, such negative experiences (e.g., “had no effect,” “caused diarrhea”) were classified as discontinuation reasons even without an explicit mention of stopping.

To systematically identify and structure the information on drug discontinuation from free-text anamneses, we applied an LLM. Specifically, we used Meta-Llama-3.1-70B-Instruct-quantized.w4a16 [10], with a parameter of temperature = 0 to ensure consistent outputs across runs. The model was hosted locally in a secure environment to ensure that no sensitive patient data was transmitted or shared externally. Each anamnesis associated with A10 or C10 drug discontinuation in the previous step was provided to the model together with a structured instruction prompt. In a few-shot fashion, each prompt included ten representative examples curated by a clinician. Although the source data were in Estonian, prompts were written in English, as prior work has shown that few-shot prompting with English instructions yields the best performance on Estonian EHR data [11]. The prompt template is provided in the Supplementary Materials A.1.

To enforce a reproducible and machine-readable format, we constrained the model output using a predefined pydantic [12] schema in Python, which required a valid JSON object with the following fields in Table 1.

Table 1. Extracted fields and their description.

Extracted field	Description
drug_stop_phrase	The full discontinuation-related sentence or clause from the anamnesis (e.g., “patient stopped taking metformin because of headaches”), explicitly including the action of stopping, the drug name, and the reported reason.
reason_for_stopping	A shorter, normalized span capturing only the reason for discontinuation (e.g., “headaches”).
drug_name	The medication name as it appeared in the text.

The extracted drug names were subsequently reviewed by a clinician. Any variants, abbreviations, or non-standard mentions were manually harmonized and mapped to their corresponding ATC codes (A10 or C10). This step ensured alignment between free-text mentions and structured prescription data.

This approach allowed us to capture discontinuation events at the level of individual drugs, with both the narrative context (via the *drug_stop_phrase*) and a shorter representation of discontinuation reasons, suitable for downstream quantitative analysis. This pipeline is illustrated in Figure 1 (LLM Extraction).

Classifying discontinuations

WHO’s five-dimension model attributes poor medication adherence to patient, therapy, condition, system, and socioeconomic factors [13]. However, this framework is not specific to discontinuation and is too broad to be applied as practical labeling categories. Subsequent studies have proposed varying taxonomies for discontinuation reasons, often tailored to the type of medication or clinical context [5,7,14]. In this study, we developed a taxonomy of discontinuation reasons based on the framework proposed by Trevena et al. [7] but adapted it through a qualitative exploration of our data. Two considerations guided these modifications: first, that the source material reflects physician-entered records rather than patient-reported notes; and second, that the drugs of interest fall within the A10 and C10 classes. The full set of classes and definitions is provided in Table 2.

Table 2. Discontinuation reasons and definitions.

Class	Definition
Adverse reactions	Discontinuation due to adverse side effects, allergic reactions, or negative interactions with other medications
Treatment success	Discontinuation due to successful treatment completion or sufficient improvement in health
Treatment inefficacy	Perceived ineffectiveness of the treatment or loss of belief in its efficacy
Contraindication	Discontinuation due to the emergence or discovery of a medical condition or risk factor that makes the continued use of the treatment unsafe or inappropriate
Non-medical reasons	Discontinuation due to factors unrelated to the patient's health or the treatment's medical effects, such as financial constraints, access issues, personal choice, cultural beliefs, or social circumstances
Other	Other medical reasons not explicitly covered by the other categories
Indeterminate	Unclear or unspecified reasons for discontinuation

To assign each discontinuation instance to one of these categories, we again applied an LLM-based approach. All analyses were first performed locally using Meta-Llama-3.1-70B-Instruct-quantized.w4a16. The model was provided with the extracted *reason_for_stopping* text along with the list of possible categories (Table 2) in the system prompt. Each query was framed in a few-shot format, using ten clinician-curated examples to illustrate the mapping process. To ensure structured outputs, we enforced a predefined pydantic schema, which required a valid JSON response containing the predicted category.

For external validation, the same task was subsequently repeated with GPT-4o [15], as the extraction of narrowly defined phrases excluded potential privacy concerns. All the extracted reasons were manually reviewed beforehand to confirm that no sensitive information was present. This allowed us to compare outputs between a local deployment and a proprietary model while maintaining patient privacy.

In addition, we used the same LLM framework to classify who initiated the discontinuation. The model was prompted (see Supplementary Materials A.1) to assign each case into one of three categories:

- Doctor - discontinuation recommended by the physician
- Patient - the patient stopped independently
- Unspecified - insufficient information to determine initiator

Evaluation Metrics

A random sample of 100 discontinuation cases was evaluated for both statins and antidiabetics. Each case was independently reviewed by a clinician and a non-medical expert, with consensus reached in all instances.

For the extraction task (reasons and phrases), performance was assessed by reporting the proportion of outputs judged valid by the annotators. For the classification tasks (discontinuation reason and initiator of discontinuation), model predictions were compared against the consensus ground truth. We report accuracy (equivalent to micro-averaged F_1), weighted F_1 to account for class imbalance, and Cohen's Kappa, which adjusts for agreement expected by chance.

Ethical considerations

The medical texts were accessed according to the approval of the Estonian Committee on Bioethics and Human Research (1.1-12/1279), which waived the requirement for informed consent for the use of the data. All data were pseudonymized prior to analysis to prevent the identification of individual patients. All data processing and model inference were performed in a secure local environment to ensure patient confidentiality. The only textual information shared with third-party services consisted of extracted discontinuation reasons that were manually verified to contain no sensitive or personally identifiable information. All other analyses were conducted locally, in compliance with applicable data protection regulations and ethical standards governing the use of medical data.

Results

Cohort Overview

We focused on clinical texts that simultaneously mentioned a drug from the target group in the extracted phrase and provided an associated reason for discontinuation. This yielded 625 antidiabetics discontinuation cases with reasons (7.0% of discontinuations) and 233 statin discontinuation cases with reasons (1.8% of discontinuations). The age and sex distributions of patients in the extracted cases were broadly similar to those of all patients who discontinued the respective drugs (see Supplementary Materials A.2), suggesting that the extracted subset is reasonably representative of the overall population. Each identified case was further classified by discontinuation reason and by who initiated the decision (patient or doctor).

Evaluation

Extraction Performance

Validation of 100 sampled cases per drug group showed high accuracy in extracting both discontinuation phrases and reasons. For antidiabetics, 93/100 of extracted phrases and 95/100 of extracted reasons were correct. For statins, the corresponding values were 98/100 for phrases and 96/100 for reasons.

To highlight the variability in how discontinuation reasons are expressed in the free text, Table 3 summarizes the distribution of unique *reason_for_stopping* patterns by frequency and their overall coverage.

Table 3. Distribution of unique discontinuation patterns by frequency and their coverage of the total texts. Coverage represents the percentage of all discontinuation mentions captured by patterns in each frequency range. All the reasons were lowercased and stripped.

Frequency Range	Antidiabetics (A10)		Statins (C10)	
	Patterns	Coverage	Patterns	Coverage
Exactly once	454	72.6%	166	71.2%
2-4 times	58	22.2%	22	22.7%
5-9 times	3	3.2%	2	6.0%
10+ times	1	1.9%	0	0%

Classification performance

For discontinuation reasons and initiator (who stopped), predictions from both the local LLM and GPT-4o were compared against human consensus labels. Table 4 summarizes the performance for both drug groups. Both models achieved strong agreement with human annotations, with kappas consistently in the substantial to excellent range.

Table 4. Performance of local LLM (Llama-3.1-70B) and GPT-4o in classifying discontinuation reasons and initiators based on validation sets (n=100). Accuracy is equivalent to micro- F_1 .

Drug Group	Task	Model	Accuracy	Weighted F_1	Cohen's κ
Antidiabetics (A10)	Reason	Llama-3.1-70B	0.81	0.82	0.70
		GPT-4o	0.80	0.84	0.70
	Initiator	Llama-3.1-70B	0.80	0.78	0.69
		GPT-4o	-	-	-
Statins (C10)	Reason	Llama-3.1-70B	0.81	0.82	0.62
		GPT-4o	0.78	0.81	0.59
	Initiator	Llama-3.1-70B	0.67	0.64	0.46
		GPT-4o	-	-	-

Classifying Discontinuations

Discontinuation Reasons

Both GPT-4o and the local Llama model were used to classify the extracted reasons for treatment discontinuation (Figure 2). Across most categories, the two models produced broadly similar distributions. The main differences were that GPT-4o assigned more cases to *contraindication* and fewer to *other* in the statin group.

Figure 2. Results of classifying extracted reasons with GPT-4o and the local Llama model. Note that 12 cases of GPT-4o labeling for antidiabetics were excluded due to the prompt producing an error, likely due to a profanity filter.

When comparing discontinuations of antidiabetic drugs and statins, *adverse reactions* were by far the most common reason in both groups. For statins, *adverse reactions* accounted for around 70% of all discontinuations, while for antidiabetics they represented about 45%. The lower proportion for antidiabetics can largely be explained by higher frequencies of *contraindication* and *treatment inefficacy* compared with statins.

Among antidiabetic drug discontinuations due to *adverse reactions*, expert review found that gastrointestinal symptoms were most frequent (41%), followed by general malaise (15%) and hypoglycemia (13%). In contrast, for statins, the leading causes were muscle-related complaints

(33%), general malaise (21%), and abnormal liver function (7%).

Discontinuation initiator

The local Llama model was used to classify whether the patient or the doctor initiated discontinuation. For this task, extracted discontinuation phrases (rather than reasons) were analyzed, as they contained the necessary contextual cues. GPT-4o was not applied here, as these longer phrases often included sensitive information that was not appropriate to share with a third-party service. The results are shown in Figure 3.

Figure 3. Results of classifying who initiated the discontinuation based on the extracted phrases using the local Llama model.

Across both drug groups, *adverse reactions* were most often associated with patient-initiated discontinuation. In contrast, *contraindications* were primarily identified as doctor-initiated. For *non-medical reasons*, the decision was again more commonly attributed to patients. An interesting difference emerged in the category of *treatment inefficacy*: for antidiabetics, patient-initiated discontinuations were noticeably more frequent than for statins.

Discussion

Principal Results

In this study, we showed that LLMs can be used to systematically extract and classify reasons for drug discontinuation from Estonian unstructured clinical narratives. By combining structured prescription data with free-text anamneses, we ensured that discontinuations were real events and that narrative mentions reflected meaningful reasons rather than speculation. This step also allowed us to significantly reduce the number of documents to annotate, focusing on only those we knew were relevant (51 330 out of 1.2 million documents, ~4.3%), thereby increasing efficiency and minimizing unnecessary review.

Our results show that the local model (Llama-3.1-70B) achieves high accuracy in extracting the discontinuations. For classifying the reason for discontinuation, both the local model and GPT-4o

achieve strong results. In the case of classifying the initiator of the discontinuation, the local model performs slightly worse than for reason classification. A possible explanation is that, because the documents are written from the perspective of the physician, it can be challenging to identify the subject of the statement from the context.

The most frequent discontinuation reason was *adverse reactions*, which accounted for a much larger share among statin users compared to other drug classes. In contrast, for antidiabetic drugs, other drivers such as *treatment inefficacy* and *contraindications* played a more prominent role, likely reflecting challenges in glycemic control and the need to manage comorbidities. Expert review of reported *adverse reactions* indicated that the most common reactions for statins were muscle-related, whereas for antidiabetic drugs they were primarily gastrointestinal symptoms, consistent with previously published literature [16,17].

Classification of who initiated discontinuation revealed mostly expected patterns, with patients more often stopping due to *adverse reactions* or *non-medical reasons*, and physicians more often initiating discontinuation for *contraindications*. Notably, for antidiabetics, *treatment inefficacy* was much more patient-driven than for statins. Analysis of the texts showed that in these instances, patients measured their blood sugar levels, perceived no effect from the medication and decided to stop taking them.

Limitations

A key limitation for this study is that it remains difficult to estimate recall, that is, how many true discontinuation cases may have been missed. In addition, among the discontinuations that were identified, reasons were explicitly documented in only 7.0% of antidiabetic cases and 1.8% of statin cases. This highlights the need for larger datasets, particularly to capture rare or less commonly documented adverse reactions and to enable analyses at the individual drug level. While the extracted cases do not appear to be biased in terms of age or sex, there may still be other systematic biases in the data that are not captured by these demographic variables.

Future directions

Beyond dataset size, several other directions could further improve and extend this approach. First, improved prompting strategies and experimenting with newer models specifically trained on clinical data may increase accuracy. Second, expanding the taxonomy to include finer-grained classes could capture a broader spectrum of discontinuation drivers. Third, a deeper exploration of adverse reactions, potentially linking clinical text with genetic or patient-level data, could enable prediction of susceptibility to specific adverse effects. Finally, incorporating patient-generated data or other non-EHR sources may reveal discontinuation reasons that physicians do not consistently document, especially in non-adverse contexts such as cost or treatment preferences.

Conclusion

This study demonstrated that LLMs can be effectively applied to Estonian EHR to identify discontinuation phrases, extract underlying reasons, and classify both the causes of discontinuation and the initiator of treatment cessation. By integrating structured prescription data with free-text narratives, we ensured that identified cases reflected real discontinuations rather than speculative mentions. Validation against expert review confirmed high accuracy in phrase and reason extraction, and in most cases good agreement in classification of discontinuation categories and who initiated discontinuation. Our findings showed that *adverse reactions* dominated among statin users, while *contraindications* and *treatment inefficacy* were more frequent for antidiabetic drugs, with patients and physicians contributing to discontinuation in distinct and clinically consistent ways. Together, these results highlight the feasibility of using LLMs for systematic, large-scale analysis of medication discontinuation dynamics, demonstrating their potential to unlock the rich clinical narratives in EHRs and transform this previously underutilized data source into a valuable asset for

pharmacovigilance and routine adverse event signal screening. Achieving this in Estonian, a linguistically unique language spoken by less than a million people, further indicates that such methods could extend to other under-resourced languages, enabling global applications in medication safety surveillance.

Funding

This research was carried out with the support of the Estonian Research Council grants PRG1844 and the Optimal Treatment for Patients With Solid Tumours in Europe Through Artificial Intelligence (OPTIMA), which is funded through the IMI2 (Innovative Medicines Initiative 2) Joint Undertaking and is listed under grant agreement 101034347. IMI2 receives support from the European Union's Horizon 2020 research and innovation program and the European Federation of Pharmaceutical Industries and Associations (EFPIA). This work was also supported by the Estonian Centre of Excellence in AI (EXAI), funded by the Estonian Ministry of Education and Research, by the European Union and Ministry of Education and Research via project TEM-TA72 and by Nordic Innovation (grant 407-7003-P22026). This research was also co-funded by the European Union through the European Regional Development Fund (Project No. 2021-2027.1.01.24-0444, TeamPerMed). This work was also funded by the European Union under the Horizon Europe Action Grant Programme agreement No 101057639. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.

Data availability

The medical texts were used according to the Estonian Committee on Bioethics and Human Research approval (1.1-12/1279). These data cannot be shared publicly because they may contain identifiable information about the participants.

Author's Contributions

HŠ conceptualized the study, curated the data, performed formal analysis and validation, conducted the investigation, developed the methodology and software, administered the project, managed resources, created visualizations, prepared the original draft of the manuscript, and contributed to manuscript review and editing. NU contributed to the conceptualization, formal analysis, investigation, and methodology of the study, validated results, and participated in manuscript review and editing. MH performed formal analysis and validation, conducted the investigation, and participated in manuscript review and editing. SL conducted the investigation and supervised the project. MO curated the data, developed the methodology, and supervised the project. ST performed formal analysis, conducted the investigation, and supervised the project. SR curated the data, acquired funding, and administered the project. JV curated the data, acquired funding, administered the project, managed resources, and contributed to manuscript review and editing. RK curated the data, acquired funding, conducted the investigation, administered and supervised the project, and contributed to manuscript review and editing.

All authors reviewed and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix

References:

1. Achterbosch M, Aksoy N, Obeng GD, Ameyaw D, Ágh T, van Boven JFM. Clinical and economic consequences of medication nonadherence: a review of systematic reviews. *Frontiers in Pharmacology*. 2025;Volume 16 - 2025. Available from: <https://www.frontiersin.org/journals/pharmacology/articles/10.3389/fphar.2025.1570359>.
2. Mooses K, Oja M, Malk M, Loorents H, Pajusalu M, Umov N, et al. Systematic evaluation of medication adherence determinants across 137 ingredients on population-level real-world health data. *medRxiv*. 2025. Available from: <https://www.medrxiv.org/content/early/2025/02/12/2025.02.10.25319884>.
3. Kardas P, Lewek P, Matyjaszczyk M. Determinants of patient adherence: a review of systematic reviews. *Frontiers in Pharmacology*. 2013;Volume 4 - 2013. Available from: <https://www.frontiersin.org/journals/pharmacology/articles/10.3389/fphar.2013.00091>.
4. Walsh CA, Cahir C, Tecklenborg S, Byrne C, Culbertson MA, Bennett KE. The association between medication non-adherence and adverse health outcomes in ageing populations: A systematic review and meta-analysis. *British Journal of Clinical Pharmacology*. 2019;85(11):2464-78. Available from: <https://bpspubs.onlinelibrary.wiley.com/doi/abs/10.1111/bcp.14075>.
5. Liu F, Pradhan R, Druhl E, Freund E, Liu W, Sauer BC, et al. Learning to detect and understand drug discontinuation events from clinical narratives. *Journal of the American Medical Informatics Association*. 2019 04;26(10):943-51. Available from: <https://doi.org/10.1093/jamia/ocz048>.
6. Sarraju A, Zammit A, Ngo S, Witting C, Hernandez-Boussard T, Rodriguez F. Identifying Reasons for Statin Nonuse in Patients With Diabetes Using Deep Learning of Electronic Health Records. *Journal of the American Heart Association*. 2023;12(7):e028120. Available from: <https://www.ahajournals.org/doi/abs/10.1161/JAHA.122.028120>.
7. Trevena W, Zhong X, Alvarado M, Semenov A, Oktay A, Devlin D, et al. Using Large Language Models to Detect and Understand Drug Discontinuation Events in Web-Based Forums: Development and Validation Study. *J Med Internet Res*. 2025 Jan;27:e54601. Available from: <https://www.jmir.org/2025/1/e54601>.
8. Shao C, Snyder D, Li C, Gu B, Ngan K, Yang CT, et al.. Scalable Medication Extraction and Discontinuation Identification from Electronic Health Records Using Large Language Models; 2025. Available from: <https://arxiv.org/abs/2506.11137>.
9. Oja M, Tamm S, Mooses K, Pajusalu M, Talvik HA, Ott A, et al. Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model: lessons learned. *JAMIA Open*. 2023 12;6(4):ooad100. Available from: <https://doi.org/10.1093/jamiaopen/ooad100>.
10. Grattafiori A, Dubey A, Jauhri A, Pandey A, Kadian A, Al-Dahle A, et al.. The Llama 3 Herd of Models; 2024. Available from: <https://arxiv.org/abs/2407.21783>.
11. Šuvalov H, Lepson M, Kukk V, Malk M, Ilves N, Kuulmets HA, et al. Using Synthetic Health Care Data to Leverage Large Language Models for Named Entity Recognition: Development and Validation Study. *J Med Internet Res*. 2025 Mar;27:e66279. Available from: <https://www.jmir.org/2025/1/e66279>.
12. Pydantic: Data validation using Python type hints. Version 2.12. Available from: <https://github.com/pydantic/pydantic>
13. Organization WH. Adherence to long-term therapies : evidence for action. World Health Organization; 2003.
14. Golder S, Weissenbacher D, O'Connor K, Hennessy S, Gross R, Hernandez GG. Patient-Reported Reasons for Switching or Discontinuing Statin Therapy: A Mixed Methods Study Using

Social Media. Drug safety. 2022 September;45(9):971—981. Available from: <https://europepmc.org/articles/PMC9402720>.

15. OpenAI, :, Hurst A, Lerer A, Goucher AP, Perelman A, et al.. GPT-4o System Card; 2024. Available from: <https://arxiv.org/abs/2410.21276>.

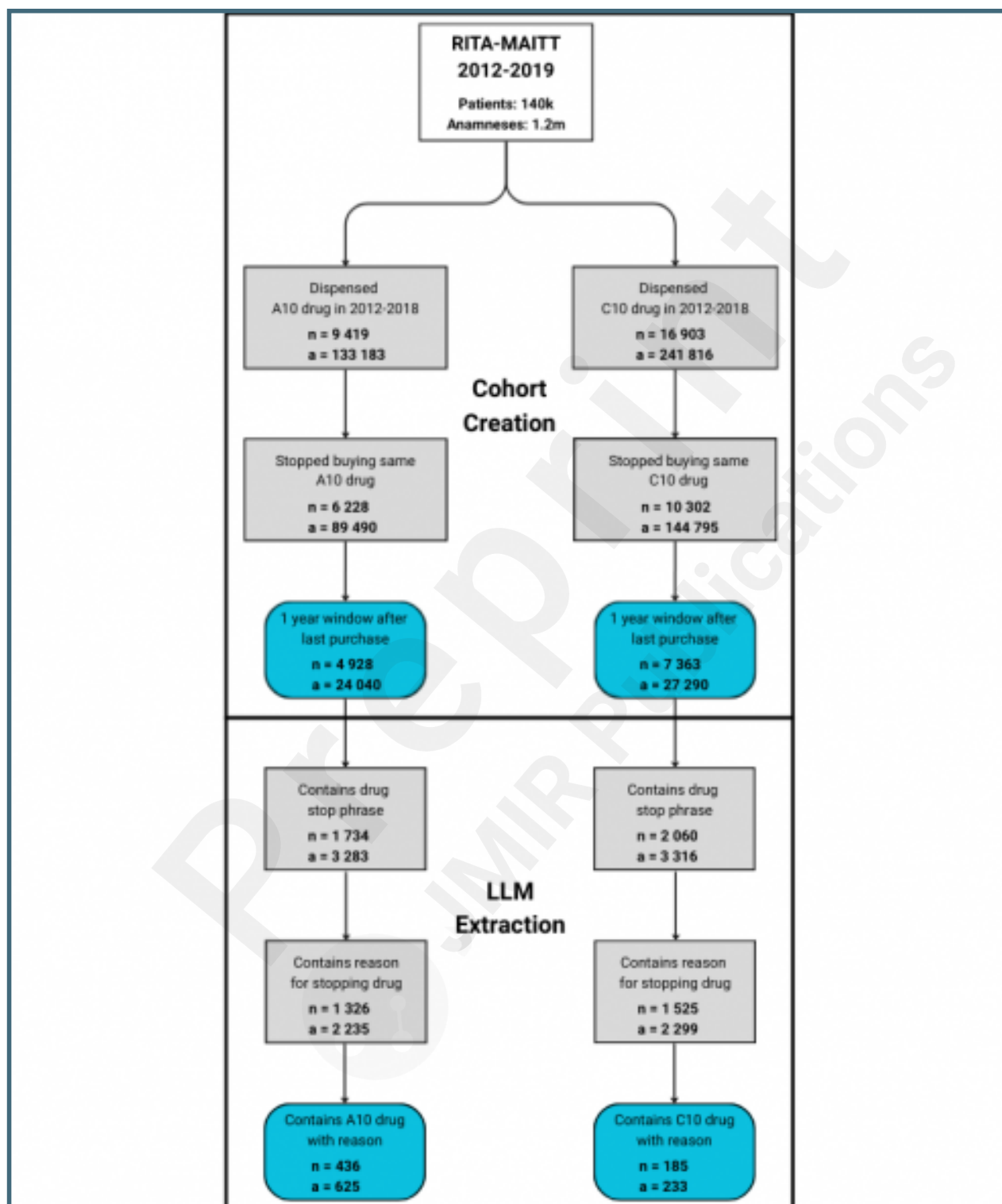
16. Ruscica M, Ferri N, Banach M, Sirtori CR, Corsini A. Side effects of statins: from pathophysiology and epidemiology to diagnostic and therapeutic implications. *Cardiovascular Research*. 2022 03;118(17):3288-304. Available from: <https://doi.org/10.1093/cvr/cvac020>.

17. Shurrab NT, Arafa ESA. Metformin: A review of its therapeutic efficacy and adverse effects. *Obesity Medicine*. 2020;17:100186. Available from: <https://www.sciencedirect.com/science/article/pii/S2451847620300063>.

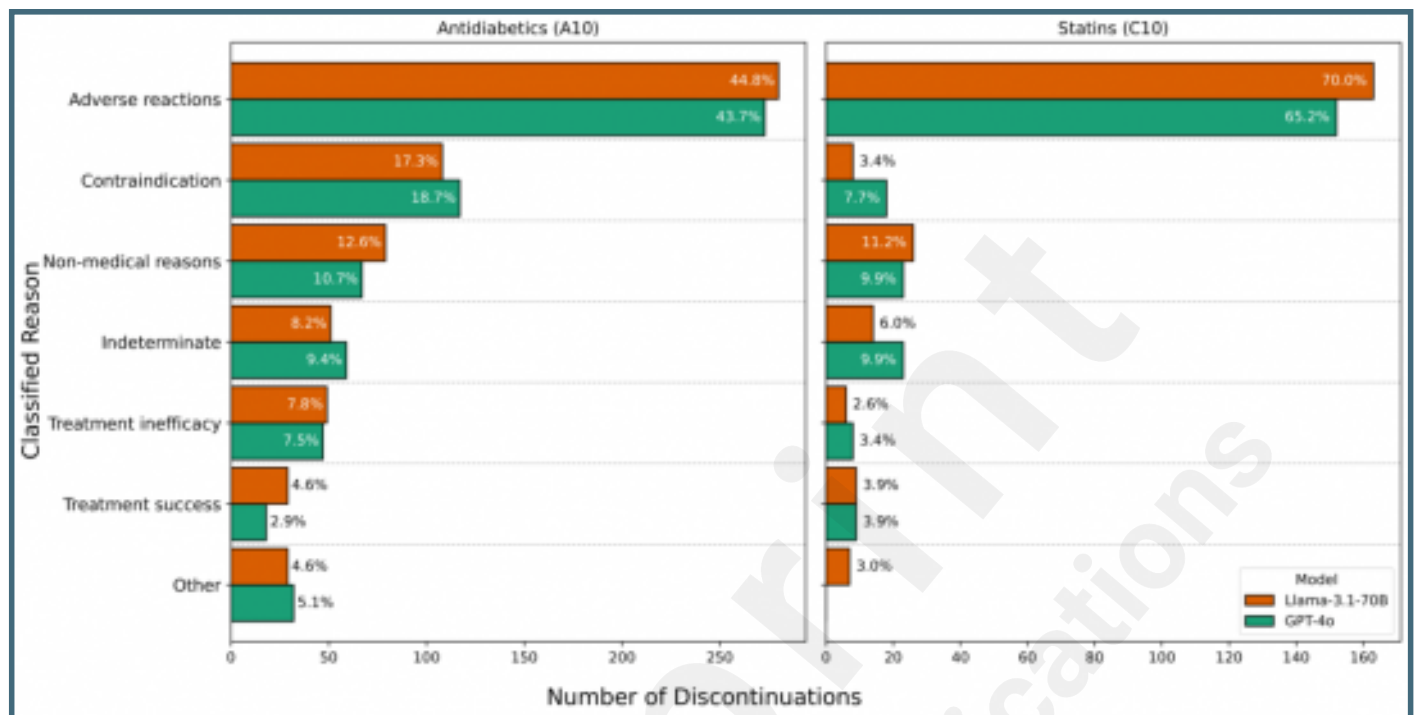
Supplementary Files

Figures

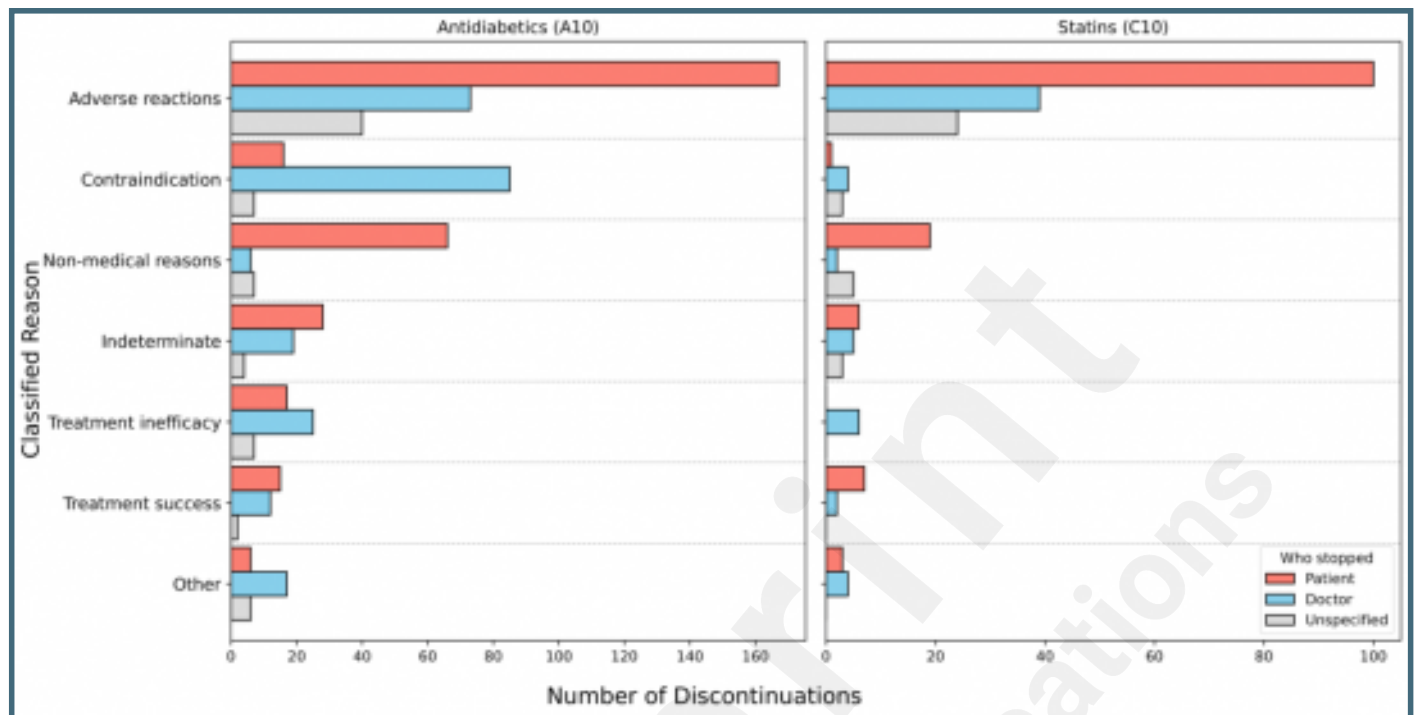
Data flow diagram consisting of two parts: 1) Cohort Creation - rule-based filtering of anamneses relevant for discontinuation mentions 2) LLM Extraction - extracting documents containing discontinuations linked with antidiabetics (A10) or statins (C10).



Results of classifying extracted reasons with GPT-4o and the local Llama model. Note that 12 cases of GPT-4o labeling for antidiabetics were excluded due to the prompt producing an error, likely due to a profanity filter.



Results of classifying who initiated the discontinuation based on the extracted phrases using the local Llama model.



Multimedia Appendixes

Prompt templates and age and sex comparison of extracted discontinuation cohorts vs base cohorts.

URL: <http://asset.jmir.pub/assets/99b4f8ebcdbe94666c02c2a06dcb01b0.docx>

