

Detecting Duplicates in Adverse Event Reports for *Cannabis sativa* and *Mitragyna speciosa* (Kratom) Products

Sanya B. Taneja¹, Xiaotong Li¹, Maryann R. Chapin², Sandra L. Kane-Gill¹, Richard D. Boyce¹

¹University of Pittsburgh, Pittsburgh, PA, USA

²Center for Drug Evaluation and Research, FDA, USA



Center of Excellence for
Natural Product-Drug
Interaction Research

INTRODUCTION

BACKGROUND

Spontaneous safety reports submitted to the FDA Adverse Event Reporting System (FAERS)¹ can be used for post-marketing safety surveillance of cannabis and kratom products.

MOTIVATION

Duplicate adverse event reports are a common issue with very few methods for deduplication. Primary report identifiers are not reliable as the reporting of same events is common.

OBJECTIVE

To develop a method that extends probabilistic record matching (VigiMatch²) for deduplication in FAERS and evaluate the methods with a focus on cannabis and kratom reports using FAERS report details and embeddings.

METHODS

Extract & standardize FAERS reports; identify reports with cannabis & kratom

Implement VigiMatch algorithm for deduplication

Extract RxNorm, MedDRA concept codes for drugs and ADRs

Normalize literature references to PubMed IDs with fuzzy string matching

Deduplicate reports based on age, sex matching & drug, ADR similarity

Compare to manually identified duplicate reports

RESULTS

Identification of reports from FAERS database

Identification

Records identified from FAERS:
Cannabis (n = 6,680)
Kratom (n = 738)

Reports removed due to change in vocabulary terms :
Kratom (n = 168)

Automatic Deduplication

VigiMatch algorithm deduplication

Cannabis (n = 207)
Kratom (n = 18)

Age, sex, literature reference, concepts matching

Cannabis (n = 252)
Kratom (n = 13)

ADR string matching
Cannabis (n = 6,167)
Kratom (n = 527)

Cannabis (n = 54)
Kratom (n = 12)

Manual Deduplication

Manually identified duplicates
Cannabis (n = 5,794)
Kratom (n = 520)

Cannabis (n = 886)
Kratom (n = 50)

- Literature references included in 4.7% of the reports; normalized to PubMed IDs (68.4%) and custom IDs (31.6%).

CONCLUSION

- First study to focus on deduplication in FAERS for botanicals combining VigiMatch, approximate matching with literature references, and adverse event reports details.
- Better deduplication with combination of approaches. Embedding similarity could potentially improve deduplication but is more aggressive.
- Manual review showed that extreme duplication remains an issue, specially for reports with missing details.

FAERS: FDA Adverse Event Reporting System
ADR: Adverse Drug Reaction

1. <https://open.fda.gov/data/faers/>
2. Tregunno PM, Fink DB, Fernandez-Fernandez C, Lázaro-Bengoa E, Norén GN. Performance of probabilistic method to detect duplicate individual case safety reports. Drug safety. 2014;37:249–58.