Enhancing 'safety' and 'case report' search filters applied in MEDLINE to support PubMed/EMERALD Literature ‘Alerts’ for pharmacovigilance

Candace Norton, NLM Associate Fellow 2016-2017

August 24, 2017

Project Sponsored by Anna Ripple, MLS, Lister Hill National Center for Biomedical Communications

Contents

Abstract

Objective

This project, “Enhancing 'safety' and 'case report' search filters applied in MEDLINE to support PubMed/EMERALD Literature ‘Alerts’ for pharmacovigilance”, was undertaken to augment currently implemented text word-based ‘safety’ and ‘case report’ search filters used in designing MEDLINE search queries for FDA regulatory staff. PubMed literature ‘Alerts’ is a mediated search service that leverages existing functionalities in MyNCBI to provide weekly retrieval emails of the most recently published biomedical citations relevant to drugs and adverse effects to over 100 FDA regulatory reviewers.

Methodology

Three analyses were conducted to identify new candidate text words: (1) compared existing FDA ‘safety filter’ to published adverse effects (AE) search filters, (2) subjected FDA ‘case report’ search filter to successive fractioning/cycling of PubMed citations indexed with ‘case reports’ [publication type] through Boolean ‘NOT’-ing high-frequency candidate text words, and (3) analyzed 68 pharmacovigilance-relevant MEDLINE citations for relevant text words for the current ‘case report’ filter.

Results

Comparison of the FDA 'safety filter' to published AE search filters revealed two candidate text word phrases: "treatment emergent" and "undesirable effect*”. Examination of ‘case report’ [publication type] citations and the 68 pharmacovigilance-relevant citations identified subsets of candidate text words relevant to (1) subjects/patient population and (2) descriptive terms aligned with a ‘case report’ presentation, as well as phrases connoting the rare nature of adverse drug events, such as “previously healthy”, “newly diagnosed”, and “no history”.

Conclusions

The investigation successfully identified several candidate text words and phrases to potentially augment currently employed FDA ‘safety’ and ‘case report’ search filters. Future work will involve validation in a retrieval experiment with relevance feedback.

Background

PubMed/EMERALD Literature 'Alerts' (EMERALD) is a weekly mediated literature alert service targeting the pre-indexing MEDLINE records to support advance detection of emerging adverse drug events in the latest published literature by regulatory review teams at the Food and Drug Administration (FDA). The purpose of this project is to gain experience in pharmacovigilance monitoring and literature search filter development by working in collaboration with an embedded NLM librarian and the Data Mining team at FDA. The Literature 'Alerts' project was identified as an opportunity based on the steady growth of requests from the regulatory review teams of the Center for Drug Evaluation and Research (CDER) at FDA. The project originated as an offshoot of a research collaboration between CgSB/LHC and the Data Mining Team/Office of Translational Sciences (OTS)/CDER that started in 2015. EMERALD is a valuable service to support the regulatory review teams, as these teams require access to the most recent literature possible when monitoring for newly emerging reports of possible adverse drug events. For the portion of MEDLINE in pre-indexed status, MeSH indexing is not yet available, so text word searching to identify relevant citations indicative of emerging adverse drug events is necessary.

The Data Mining Team, along with the embedded NLM librarian, stepped in to offer the mediated search service using PubMed 'MyNCBI' functionalities to the review teams, lending their search expertise to ensure that the necessary results were located in the literature as swiftly as possible. The EMERALD Literature 'Alerts' team devised custom text word search filters tailored to each review team's portfolio of drug agents and saved as weekly alerts in a MyNCBI account. When these alerts pick up new results, the EMERALD team performs a quality control check on the citations before sending the results to the review teams with a link to the FDA's full-text holdings for the publication. EMERALD has been operating for the past two years and has seen steady growth. Now, the EMERALD team supports over 140 regulatory reviewers with close to 50 weekly alerts, meeting with each team for a brief needs assessment before crafting and establishing a weekly alert. Periodic follow-ups are conducted on an ad hoc basis at the request of the review team or the EMERALD team to refine or expand the searches.

Effective regulatory pharmacovigilance requires timely assessment of information from multiple sources and the citations provided from the literature 'Alerts' weekly updates contributes to the process. The EMERALD team developed two base filters for use to support the review teams at FDA; one for safety monitoring (e.g., adverse drug events, toxicity) and one to identify case reports (e.g., case reports, case studies or series). This project provides an examination of these current base search filters in use in the

EMERALD Literature Alert search service to identify additional candidate terms to enhance the retrieval of relevant results from the pre-indexed portion of MEDLINE for the regulatory review teams.

Objective

This project provides an analysis of two current text-word based search filters ('safety'; 'case report') implemented in the EMERALD Literature Alert search service to identify additional candidate terms for search filter augmentation.

Literature Review

Search filters, also known as optimal search strategies, assist librarians, informationists, and researchers in locating and retrieving different types of evidence from bibliographic databases (Beale, 2014). Search filters are typically designed to locate a particular type of evidence (e.g. randomized controlled trials), consist of controlled vocabulary and text word terms, and are combined with subject terms. Search filters can save time in developing complex search strategies, and can be reused as needed. The InterTASC Information Specialists' Sub-Group (ISSG) is a collaborative group that identifies, tests, and assesses search filters, along with offering critical appraisal guidance for researchers using filters in their work. The Clinical Hedges Study from researchers at McMaster University, led to the development of the PubMed Clinical Queries search filter tool, available for use via PubMed (Wilczynski, 2005), and supports the practice of combing hand searches with online retrieval to identify and assess terms for inclusion in a filter. The practice of using Boolean operator "NOT" to isolate and eliminate results from a reference pool allows for an effective refining process to enhance search filter precision (Wilczynski, 2011).

Pharmacovigilance at the FDA for drugs is traditionally monitored through the analysis of case reports describing possible unintended harmful effects that are submitted by manufacturers and consumers to the FDA Adverse Events Reporting System (FAERS), a database that supports the FDA's post-marketing safety surveillance program (FDA, 2016). Shetty (2011) and Winnenburg (2015) document the use of leveraging MeSH indexing for information mining to improve drug safety, noting the need for developing text word searches to capture the pre-indexed portion of Medline. Golder (2006, 2011, 2012) is an NIHR National Institute of Health Research fellow and staff member at the University of York, specializing in optimizing retrieval of adverse effects (AE) data from bibliographic databases. Golder's work includes validating an AE search filter, as well as a comparison of validated AE search filters from multiple sources (Golder, 2012).

Methodology

In addition to reviewing and analyzing the search filters, regular site visits to FDA and meetings with review teams were conducted to understand reviewer retrieval needs and overall EMERALD Literature Alerts project context and process to better ground the analysis.

Three analyses were conducted to identify new candidate text words:

(1) Compared existing FDA 'safety filter' to published adverse effects (AE) search filters;

(2) Subjected FDA 'case report' search filter to successive fractioning/cycling of PubMed citations indexed with 'case reports' [publication type] through Boolean 'NOT'-ing high-frequency candidate text words;

(3) Analyzed an FDA 'use case' of 68 pharmacovigilant-relevant-identified MEDLINE citations for relevant text words for the current 'case report' filter.

(1) 'Safety' Filter

The FDA safety filter was broken down term by term along with the associated field code in a table format. Golder's most sensitive adverse effects (AE) filters were segmented in the same way in adjacent columns. Each filter occupies a unique column in Table 1, while each term and corresponding field code is in a new row. Filters are created by combining all the individual terms with the Boolean operator "OR". The FDA safety filter was built with input from the regulatory reviewers and relevance feedback using the PubMed interface for MEDLINE (Sorbello, 2015), while Golder's filters were constructed and tested using the Ovid platform (Golder, 2012); subsequently the Golder filters use different field codes. Terms are aligned for comparison based on closest match with definitions provided for the field code abbreviations below the table. It was beyond the scope of this project to compare retrieval in PubMed using the title/abstract field code ([tiab]) versus retrieval in Ovid with either the title/abstract (.ti,ab) or "all fields" (.af) field codes.

Table 1. 'Safety' Filter Comparison

FDA Safety Terms [with field codes]	**Golder AE 2012** (Most sensitive search strategy; with fields)	**Golder AE 2012** (Most sensitive search strategy excluding the use of specified named adverse effects; with fields)
safety[tiab]	safety.af	safety.ti,ab
tolerability[tiab]	tolerability.af	tolerability.ti,ab
toxicity[tiab]	toxicity.af	toxicity.ti,ab

“side effects”[tiab]	side effect*.af	side effect*.ti,ab
adverse[tiab]	adverse adj2 effect.af	adverse adj2 effect.ti,ab
“drug-drug interaction”[tiab]	adverse adj2 effects.af	adverse adj2 effects.ti,ab
“drug-drug interactions” [tiab]	adverse adj2 reaction.af	adverse adj2 reaction.ti,ab
“drug interaction”[tiab]	adverse adj2 reactions.af	adverse adj2 reactions.ti,ab
“drug interactions”[tiab]	adverse adj2 event.af	adverse adj2 event.ti,ab
interaction[tiab]	adverse adj2 events.af	adverse adj2 events.ti,ab
interactions[tiab]	adverse adj2 outcome.af	adverse adj2 outcome.ti,ab
mortality[tiab]	adverse adj2 outcomes.af	adverse adj2 outcomes.ti,ab
poisoning[tiab]	undesirable effect*.af	undesirable effect*.ti,ab
toxicology[tiab]	treatment emergent.af	treatment emergent.ti,ab
risk[tiab]	safe.af	safe.ti,ab
risks[tiab]	adrs.af	adrs.ti,ab
“ddi”[tiab]	*(ae OR co OR de).fs*	*(ae OR co OR de).fs*
“ddis”[tiab]	*Specified named adverse effects/ci*	
Field Code Definitions		
[tiab]=title and abstract	.af=all fields; .fs=floating subheading; /ci=subject heading, chemically induced	.ti,ab=title and abstract; .fs=floating subheading
	adj2=proximity searching w/in 2 terms	adj2=proximity searching w/in 2 terms

Table 1. Column 1: Current FDA Safety filter with field codes; Column 2 and Column 3: Adverse effects (AE) search filters developed by Su Golder, published in Golder, S. and Loke, Y. K. (2012), The performance of adverse effects search filters in MEDLINE and EMBASE. Health Information & Libraries Journal, 29: 141–151.

(2) ‘Case Reports’ Filter

To identify candidate terms for recognizing case reports in the pre-indexed records of MEDLINE, I began by searching for all records with the case reports publication type (“case reports” [pt]). Case reports are a designated publication type within MeSH indexing and entries for this field are designated with “case reports” [pt]. From this initial pool of results, I began successively fractioning the results into smaller pools by searching the test term in the title and abstract, followed by eliminating those results with the Boolean operator “NOT”. A sample of the search sequence is available in Table 2, detailing the fractioning process for five test terms.

Table 2. ‘Case Reports’ Successive Fractioning Example

Search #	Query	All Results
1	case reports[pt]	1832985
2	case reports[pt] AND case*[tiab]	860893
3	case reports[pt] NOT case*[tiab]	972092
4	#3 AND report*[tiab]	104965
5	#3 NOT report*[tiab]	867127

6	#5 AND history[tiab]	17922
7	#5 NOT history[tiab]	849205
8	#7 AND histories[tiab]	434
9	#7 NOT histories[tiab]	848771
10	#9 AND commentary[tiab]	175
11	#9 NOT commentary[tiab]	848596

Table 2. Case Reports filter successive fractioning example including search line number, search query, and numerical results from a May 5, 2017 test.

This process was repeated with all the terms currently included in the FDA's case reports filter. After successfully testing out all of the current terms in use in the FDA case reports filter, additional prospective terms were selected from the remaining citations for investigation. The individual terms were selected based on a title/abstract review of three sample batches of results: The first 200, middle 200, and last 200 records were evaluated for potential candidate terms. Supplement 1: "Suppl1_Case Reports Filter Development" contains the line by line search history testing each selected term.

(3) FDA Use Case of 68 pharmacovigilant-relevant citations

To further enhance the identification of candidate test terms, one of the regulatory review teams provided access to a set of 68 case reports delivered via the weekly EMERALD alerts that they had determined to be relevant retrievals. The same successive fractioning process was used for this set of results to identify terms and phrases that are not presently included in the FDA case reports filter. Full search details are included in Supplement 2: "Suppl2_FDA Use Case 68 References".

Results

(1) 'Safety' Filter

Conducting a term by term comparison, the FDA currently uses several terms that were not included in the Golder AE filters, illustrating the development of this custom filter to meet the needs of the regulatory review teams at FDA. The filter comparison revealed that the Golder AE filters include two unique phrases that the FDA team should consider for further investigation as additions to their safety filter: "undesirable effect" and "treatment emergent". Preliminary testing located 103 unique citations not captured by the existing FDA safety filter when the Golder-unique AE terms were added.

Table 3. 'Safety' Filter Unique Terms

FDA Terms [with field codes]	**Golder AE 2012** (Most sensitive search strategy; with fields)	**Golder AE 2012** (Most sensitive search strategy excluding the use of specified named adverse effects; with fields)
safety[tiab]	safety.af	safety.ti,ab
tolerability[tiab]	tolerability.af	tolerability.ti,ab
toxicity[tiab]	toxicity.af	toxicity.ti,ab
"side effects"[tiab]	side effect*.af	side effect*.ti,ab
adverse[tiab]	adverse adj2 effect.af	adverse adj2 effect.ti,ab
"drug-drug interaction"[tiab]	adverse adj2 effects.af	adverse adj2 effects.ti,ab
"drug-drug interactions" [tiab]	adverse adj2 reaction.af	adverse adj2 reaction.ti,ab
"drug interaction"[tiab]	adverse adj2 reactions.af	adverse adj2 reactions.ti,ab
"drug interactions"[tiab]	adverse adj2 event.af	adverse adj2 event.ti,ab
interaction[tiab]	adverse adj2 events.af	adverse adj2 events.ti,ab
interactions[tiab]	adverse adj2 outcome.af	adverse adj2 outcome.ti,ab
mortality[tiab]	adverse adj2 outcomes.af	adverse adj2 outcomes.ti,ab
poisoning[tiab]	**undesirable effect*.af**	**undesirable effect*.ti,ab**
toxicology[tiab]	**treatment emergent.af**	**treatment emergent.ti,ab**
risk[tiab]	safe.af	safe.ti,ab
risks[tiab]	adrs.af	adrs.ti,ab
"ddi"[tiab]	*(ae OR co OR de).fs*	*(ae OR co OR de).fs*
"ddis"[tiab]	*Specified named adverse effects/ci*	
Field Code Definitions		
[tiab]=title and abstract	.af=all fields; .fs=floating subheading; /ci=subject heading, chemically induced	.ti,ab=title and abstract; .fs=floating subheading
	adj2=proximity searching w/in 2 terms	adj2=proximity searching w/in 2 terms

Table 3. Safety filter comparison with the unique phrases from Golder's AE filters identified: "undesirable effect" and "treatment emergent".

(2) 'Case Reports' Filter

The case reports term analysis revealed that candidate terms could be loosely grouped into two categories, those that are descriptive of the document or case being presented, or terms related to the subject of the report (examples in Table 3). This categorical division enables a more targeted investigation into how researchers and clinicians are writing up case reports; even if the explicit publication type is not indicated in the bibliographic record, it may be possible to identify case reports from how the subjects are referenced. Performing a retrieval analysis on a subset of the identified terms in either category would be the next step.

Table 4. 'Case Reports' Sample Terms by Category

Descriptive	Case*[tiab] OR report*[tiab] OR presentation*[tiab] OR present*[tiab] OR management[tiab] OR rare*[tiab] OR following[tiab] OR isolated[tiab] OR associated[tiab] OR history[tiab] OR study[tiab] OR studie*[tiab] OR diagnos*[tiab] OR unusual*[tiab] OR fail*[tiab] OR new*[tiab] OR treat*[tiab] OR therapy[tiab] OR complicat*[tiab]
Subject	Human[tiab] OR individual[tiab] OR patient[tiab] OR patients[tiab] OR patient's[tiab] OR child[tiab] OR children[tiab] OR girl[tiab] OR boy[tiab] OR woman[tiab] OR man[tiab] OR adult[tiab] OR female[tiab] OR male[tiab] OR infant[tiab]

Table 4. 'Case Reports' Sample filter additions, identified by category: Descriptive terms and subject terms.

(3) FDA Use Case of 68 pharmacovigilant-relevant citations

The analysis of the 68 citation pharmacovigilant-relevant citations use case identified an additional four search phrases that had not been included in the existing case reports filter nor previously tested in the first case report experiment: "previously healthy", "previously reported", "newly diagnosed", and "no history". Preliminary testing located a small number of unique results when these new phrases were added to the existing FDA case report filter.

Discussion

The safety filter comparison provides timely confirmation to the EMERALD team that the terms identified in their current FDA filter are in line with the pharmacovigilance research community who also monitor the literature for emerging adverse drug events. Without the benefit of MeSH indexing, the EMERALD team is reliant upon text-word searching to locate the most current, not yet indexed PubMed citations for the regulatory review teams, emphasizing the need for current best term selection. The comparison to Golder is a tidy peer comparison as her work in the field has been validated and peer reviewed on multiple occasions and is regarded as the gold standard of AE filter development. By confirming the significant overlap of terms between the Golder filters and the FDA filter, this places the FDA filter in good stead. The comparison revealed two phrases from Golder that are not presently included in the FDA filter and it is my recommendation that these additions be considered for further inclusion and testing in a retrieval experiment.

Case reports are an extremely valuable source of early signal detection when identifying adverse drug events. One of the key challenges for the EMERALD team at FDA is constructing a text-word search that

is robust enough to detect the multitudinous phrases that authors may use to describe a case report without retrieving a surplus of irrelevant results that the regulatory review teams must eliminate. The case report search filter investigation is intended as an exploration in identifying candidate terms that could be added to the existing case reports filter and later assessed for relevance by a full retrieval analysis experiment.

These explorations successfully identified candidate terms for both the safety filter and the case reports filter. Future work would include designing and implementing a full retrieval analysis experiment to evaluate how well each term performs in providing relevant results to the regulatory review teams. Avenues of future investigation could also include the addition of the FDA-internal Designated Medical Events (DME) terms list to the weekly EMERALD literature alerts. DMEs are adverse events that are considered rare, serious, and associated with a high drug-attributable risk. DMEs can constitute an alert for the FDA Safety Reviewers to investigate with as few as one to three case reports for a pharmacovigilant adverse drug event signal.

One of the key learning experiences for me has been the growing awareness of how the needs of the regulatory review teams can change and expand rapidly when performing pharmacovigilance monitoring. As part of this project, I joined the EMERALD Team at FDA in their meetings with several of the regulatory review teams they support. Attending these meetings allowed me to observe how other federal agencies work with NLM resources to further health information sharing. It was a remarkable experience to witness the impact of an interagency collaboration in real time.

Through my site visits to FDA, I learned that even informal needs assessments and reference interviews can have a profound impact on the level of trust and communication between research teams and the information professionals that serve in an embedded capacity alongside them. Often the team meetings were brief, a half hour or less, but in that time, the EMERALD team could confirm the current performance satisfaction of the weekly alerts, inquire about changing needs, soliciting feedback directly from the reviewers and allowing an opportunity for both groups to reach a consensus on the best 'Alerts' strategies for their pharmacovigilance needs.

Acknowledgements

This project would not have been possible without the dedication and support of my project sponsor Anna Ripple. She championed me and my involvement in this project from the beginning and I have learned so much. Additional thanks to Olivier Bodenreider for supporting my involvement with the FDA

through NLM's IAA; this was an excellent opportunity to participate as an embedded librarian with another agency. Thanks as well to Alfred Sorbello and the FDA Data Mining team for welcoming me into their space and allowing me to explore their operations.

My program mentor Liz Amos offered much insight and wisdom over the course of the fellowship year and was instrumental in my introduction to Anna and my subsequent involvement with this project. Kathel Dunn and the 2016-2017 Associate cohort (Megan Fratta, Kendra Godwin, and Tyler Moses) demonstrated persistent patience and curiosity throughout the course of this work, and helped keep me on track as the scope of the work shifted. Final thanks to the NLM leadership and Library Operations in particular for making the Associate Fellowship program possible.

References

Golder, S., & Loke, Y. (2009). Search strategies to identify information on adverse effects: a systematic review. *J Med Libr Assoc, 97*(2), 84-92. doi:10.3163/1536-5050.97.2.004

Golder, S., & Loke, Y. K. (2010). Sources of information on adverse effects: a systematic review. *Health Info Libr J, 27*(3), 176-190. doi:10.1111/j.1471-1842.2010.00901.x

Golder, S., & Loke, Y. K. (2012). The performance of adverse effects search filters in MEDLINE and EMBASE. *Health Info Libr J, 29*(2), 141-151. doi:10.1111/j.1471-1842.2012.00980.x

Golder, S., & Loke, Y. K. (2012). Sensitivity and precision of adverse effects search filters in MEDLINE and EMBASE: a case study of fractures with thiazolidinediones. *Health Info Libr J, 29*(1), 28-38. doi:10.1111/j.1471-1842.2011.00972.x

Golder, S., McIntosh, H. M., Duffy, S., & Glanville, J. (2006). Developing efficient search strategies to identify reports of adverse effects in MEDLINE and EMBASE. *Health Info Libr J, 23*(1), 3-12. doi:10.1111/j.1471-1842.2006.00634.x

InterTASC Information Specialists' Sub-Group. (2017). The InterTASC Information Specialists' Sub-Group search filter resource. Retrieved on August 22, 2017 from https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home

Shetty, K. D., & Dalal, S. R. (2011). Using information mining of the medical literature to improve drug safety. *J Am Med Inform Assoc, 18*(5), 668-674. doi:10.1136/amiajnl-2011-000096

Sorbello A, Ripple A, Bodenreider O. PubMed 'Early Alerts': A pilot study to support prospective detection of emerging adverse drug events. [poster] AMIA Annu Symp Proc 2015:1699. ; NIH Intramural Research Festival 2015; MLA/CHLA/ABSC/ICLC (Medical Library Association (MLA), the Canadian Health Libraries Association/Association des bibliothèques de la santé du Canada (CHLA/ABSC), and the International Clinical Librarian Conference (ICLC)) 2016.

U.S. Food and Drug Administration. (2016). Questions and answers on FDA's Adverse Event Reporting System (FAERS). Retrieved on August 22, 2017 from http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htm

Wilczynski, N. L., McKibbon, K. A., & Haynes, R. B. (2011). Search filter precision can be improved by NOTing out irrelevant content. *AMIA Annu Symp Proc, 2011*, 1506-1513.

Wilczynski, N. L., Morgan, D., & Haynes, R. B. (2005). An overview of the design and methods for retrieving high-quality studies for clinical care. *BMC Med Inform Decis Mak, 5*, 20. doi:10.1186/1472-6947-5-20

Winnenburg, R., Sorbello, A., Ripple, A., Harpaz, R., Tonning, J., Szarfman, A., Bodenreider, O. (2015). Leveraging MEDLINE indexing for pharmacovigilance - Inherent limitations and mitigation strategies. *J Biomed Inform, 57*, 425-435. doi:10.1016/j.jbi.2015.08.022