AMIA Annu Symp Proc 2016:(in press)

PubMed 'Early Alerts': Towards Better Precision of Literature Searching for Pharmacovigilance Information Based on an Assessment of Relevance Feedback

Anna Ripple, MLS[1]; Alfred Sorbello, DO, MPH[2]; Shahrukh Haider, DDS, MPH[2]; Olivier Bodenreider MD, PhD[1]

[1]Lister Hill National Center for Biomedical Communications, U.S. National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894, USA; [2]US Food and Drug Administration, Center for Drug Evaluation and Research, 10903 New Hampshire Avenue, Silver Spring, MD 20993, USA

Background: The PubMed 'Early Alerts' provide FDA regulatory reviewers with weekly topical searches of the most recently submitted citations to PubMed/MEDLINE to support prospective detection of emerging adverse drug events for specific drugs. We seek to increase the precision of electronic searching based on an assessment of relevance feedback for a subset of retrieved citations for the antidiabetic medications.

Methods: Using a search query optimized for recall and focusing on the titles, abstracts, and keywords, four regulatory evaluators assessed 30 citations each for relevance to drug safety and efficacy drawn from a random sample of 120 of the most recently deposited reports (i.e., not yet indexed). Candidate precision terms were identified by significant word frequency analysis (excluding stopwords) using Word Counter Tool, a free online word counter. We performed a differential frequency analysis of text word occurrences for the relevant compared to non-relevant citations.

Results: Based on reviewers' feedback, half of the reports in the 120 citation random sample were assessed as relevant to safety, efficacy, or both safety and efficacy, whereas the remainder were considered non-relevant for both safety and efficacy. The candidate precision words 'efficacy', 'safety', 'adverse', and 'risk' were identified by expert opinion from the titles or abstracts of the relevant citations among frequently occurring words (see Table 1 below). The text words 'efficacy, 'safety', and 'risk' had no title occurrences in the non-relevant citations. However, 'risk' and 'safety' are not discriminant, because they also occurred in the abstracts of non-relevant citations, leaving 'adverse' as the only candidate precision word derived from the abstract field. Analysis of the keywords field did not yield any candidate precision text words.

Table 1: Relevance Feedback Analysis for 120 PubMed Citations

Relevance Assessment	Citation Count, n /N (%)	Text Word Field and Candidate Precision Words identified from Citations in each subset	
		Relevant citations	Non-relevant citations
Safety only	38/120 (32%)	Title: efficacy; safety; risk Abstract: safety; adverse	Abstract: risk
Efficacy only	49/120 (41%)	Title: efficacy; safety Abstract: safety	Abstract: risk
Safety and efficacy	33/120 (28%)	Title: efficacy; safety Abstract: safety; adverse	Abstract: safety

Conclusions: We identified several candidate text words to potentially increase the precision of our literature search strategy for pharmacovigilance information, but further validation is required. Small citation sample size and restriction to only the title, abstract, and keywords fields may have contributed to the limited capacity to identify such candidate precision words. Future research is planned involving a larger group of citations, a broader set of drugs and the application of more sophisticated analytical techniques such as term frequency-inverse document frequency to detect text words that may enhance the precision of the PubMed 'Early Alerts' as a tool to complement other ongoing pharmacovigilance activities at FDA.

Acknowledgements: Funding support received from the FDA/CDER/OTS and the Intramural Research Program, NIH, National Library of Medicine. **Disclaimer**: The views expressed are those of the authors and do not necessarily represent the views of the US FDA, the NIH, or the US Government.