An Automated Literature Analysis Tool to Enhance Literature Searches for Publication Subcommittees (PSCs) and Publication Planning

Michelle Kissner[a], Mark Drinkwater[b], Belinda Stretch[c], Erica Wright[d]

[a] Pfizer Inc, Publications Management, New York, NY, [b] PI2 Solutions, Hertfordshire, UK, [c] Pfizer Inc, Information and Library Services, Walton Oaks, UK, [d] Pfizer Inc, Medical Business Technology, Groton, CT

Scan QR code with smartphone app to view an electronic version of this poster.

Poster #25

If you do not have smart phone access you can access the poster via the internet at: http://congressdownload.pfizer.com/ismpp2013_kissner.html

Objective
In order to provide PSCs with literature searches to assist with meeting business goals, an automated literature analysis tool was piloted to enable easier access to searches that are comprehensive, provided in multiple formats, linked directly to publications/document delivery, and continuously updated.

Research Design and Methods
A lean six-sigma methodology was utilized to identify the tool. We defined the problem, measured and analyzed to find the root cause, made improvements, and put controls in place to ensure searches continue to meet the PSC's needs.

Results
PSCs are provided with targeted search results that: 1) are comprehensive due to searching multiple bibliographic databases that utilize search strategies developed by experienced library science professionals; 2) have greater functionality due to the mining of key concepts and integration with the full text via library subscriptions and document delivery; 3) can be output in multiple formats including Word, Excel, EndNote, RSS feeds and interactive charts; 4) are intuitively organized within the tool to enable rapid assimilation into users workflow; 5) are updated automatically; and 6) saves money.

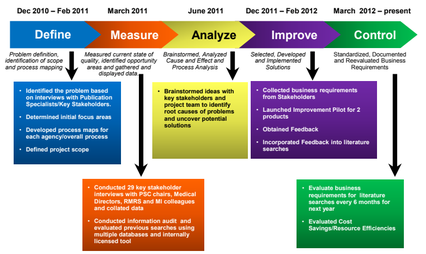
Conclusion
Implementation of this tool has resulted in the ability to automate searches. The comprehensive results are delivered on demand, intuitively organized, and can be outputted in multiple formats in line with the user's need. PSCs can use the tool to keep updated on targeted literature; analyze publication topics aiding in publication planning; save time in reviewing searches due to formatting, timing, and dedicated functionality; and cut costs.

Background
The utilization of literature searches is an important way for publication teams to stay updated on current literature in order to meet business goals. These include awareness of literature related to: 1) products, 2) disease area, 3) scientific landscape, 4) clinical trials, as well as assisting in the development of manuscripts, and identifying unmet medical needs as part of the annual publication planning cycle. With the thousands of research articles published each year, bibliographic databases provide a way to search collections of multiple journals. It is worth emphasizing that the ability to retrieve published literature is dependent on a number of factors. These include: the experience and capabilities of the searcher; quality of the indexing within the database; reliability of the database; coverage and currency of the databases.[1] Published studies have demonstrated that searching one bibliographic database is not enough. Multiple analyses have found that combining databases yields a higher result compared with one or fewer databases.[1-3] Search results from bibliographic databases are also highly dependent on the search strategy used. Therefore, it is important that these strategies are developed in close collaboration with the customer to ensure their needs are met in a targeted and focused fashion. Information requirements should be ascertained through interview and discussion and often go through several iterations to ensure they fully meet customer requirements.

Objective
In order to provide PSCs with literature searches to assist with meeting business goals, an automated literature analysis tool was piloted to enable access to searches that are comprehensive, provided in multiple formats, linked directly to publications/document delivery, and continuously updated.

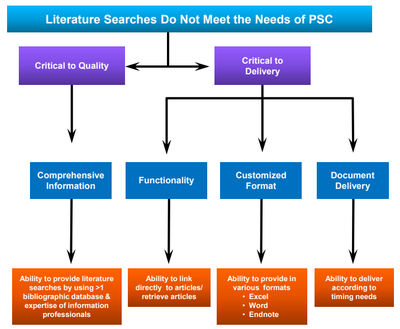
Research Design and Methodology
The process that was used to identify the automated literature analysis tool was through a six-sigma continuous improvement methodology (Figure 1).

Figure 1. Project Overview of Define, Measure, Improve Analyze, Control (DMIAC) Process



Define Phase
During the Define Phase, the problem and project scope was identified; the project scope was based on informal interviews with the Publications Management team members and formal interviews with key PSC team members that included: Medical, Clinical, Medical Information (MI), Outcomes Research (OR) and Regional Medical Research Specialists (RMRSs) (Figure 2). At that point, it was defined what was critical to quality and critical to delivery. The focus of the project was centered on evaluating the quality of the searches and a determination was made to focus on two recently approved oncology products. Process maps were developed that identified the overall process.

Figure 2. Critical to Quality Tree



Measure Phase
During the Measure Phase interviews were conducted with stakeholders (n = 29). The questions asked centered around the needs of the business including specific attributes and other requirements for literature such as format, functionality, delivery preferences, organization of content and usefulness. The interviews revealed that many stakeholders were not utilizing the literature searches and were using other resources.

Moreover, an information audit was conducted by company library science professionals to evaluate the quality of literature searches being provided. The first step was to collate previous search strategies that were conducted and execute strategies that were as closely matched as possible. Secondly, a comparative analysis of results was performed with multiple databases including: Medline, BIOSIS, Embase, CAB Abstracts and Derwent Drug File. During the second step, selected examples were further analyzed from the first step to identify unique references in databases not included by initial searches. Finally, unique references were evaluated to determine the relevance to the product where the product made up the main basis of the article. Results are displayed in Table 1.

Table 1. Information Audit Results

Product	Current Searches Provided	Ovid Databases	Unique references (to Previous Searches)	Qualitative Analysis
Product A	96	219	158	
Product B	40	108	65	64*

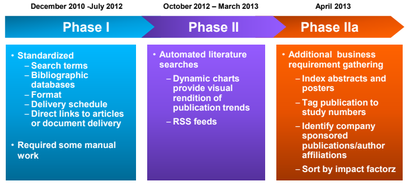
* Inclusive of 49 conference meeting abstracts,3 Letter/opinion/corrections, 3 primary manuscripts, 1 secondary manuscript, 1 review article, 3 short surveys, 1 book section, 2 German articles, 1 news article not relevant

Analyze Phase
To identify possible solutions to the issues raised by the PSCs, the project team (Library Science and Publication Management Professionals) held a brain-storming session with PSC members This session resulted in the recommendation to implement a two-stage pilot: the first stage involved standardizing search terms, databases, formatting, and delivery schedule. In addition, direct links to the articles or populated document delivery ordering forms were provided. Feedback from stakeholders was obtained for the first stage of the pilot that resulted in the team moving to Phase II.

Improve Phase
As part of Phase I, initial enhancements were made which improved the quality of the searches by utilizing multiple bibliographic databases (Medline, Embase, BIOSIS, CAB Abstracts, and Derwent Drug File), leveraging the expertise of company library scientists to gather business requirements and develop search strategies that matched the business needs. These included: deliver searches according to the timing needs, provide multiple formats (Word, Excel, and Endnote), and inclusion of a link to full text embedded in the document or a link to document delivery (article retrieval service).

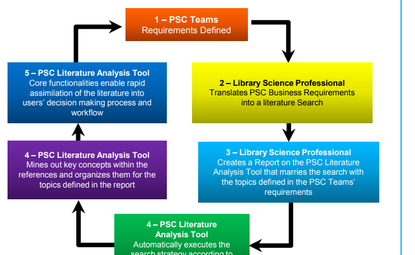
Figure 3 highlights the different phases of the project. Phase II entailed automating the process that was achieved by modifying an existing tool that had capabilities to automatically execute multiline, multibibliographic database searches that were entered and developed by experienced library science professionals. This second stage of the pilot has been implemented and will run for 3 months.

Figure 3. Design Phases of Project Improvements



The literature analysis tool utilizes an importer that automatically executes multiline, multidatabase — Medline, Embase, BIOSIS, CAB Abstracts, and Derwent Drug Files via the OVID platform — developed and added to the tool by experienced library science professionals. The library science professionals work in close collaboration with the PSC to ensure that their topics of interest are understood and then translated into search strategies. Once the search strategy is developed, it is inputted into the tool and a literature search is generated. These auto searches are performed at preset frequencies determined by the business and the results are downloaded and imported into the tool for instant processing. This approach ensures that results are comprehensive while duplicate references are automatically removed. An overview of the process is highlighted in Figure 4.

Figure 4. Literature Analysis Tool Process Overview

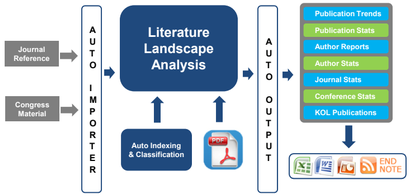


Control Phase
The Control Phase of the pilot focused on standardizing the parameters of the implementation. Therefore, business requirements for literature searches are evaluated and revisions are made as necessary.

Results
Results from stakeholder interviews and the information audit from Phase I found that searches were not comprehensive due to limited bibliographic databases, access to incomplete lists of synonyms, and nontargeted search strategies. Additionally, literature searches were not provided on demand, formatted according to the team's needs, and the ability to link directly to the article or to a document-delivery form when a library subscription was not available.

Phase II resulted in the implementation of the literature analysis tool developed in conjunction with the company PSC teams. This tool provides a dynamic and graphical rendering of predigested published literature that is automatically imported and developed according to the business requirements of the PSC teams. Figure 5 highlights a schematic of how the literature analysis tool works.

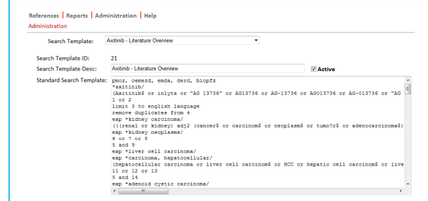
Figure 5. Automated Literature Analysis Tool – Schematic



The automated searches, are delivered on demand, comprehensive, of high quality, intuitively organized, and outputted into multiple formats. The expertise of company colleagues (Library Science Professional) was leveraged and cost savings were realized.

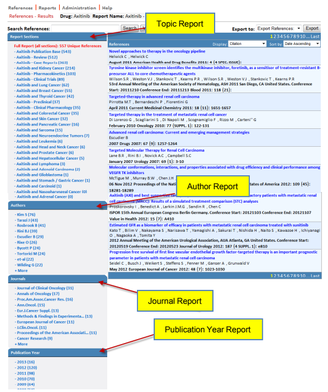
The library science professional creates search strategies that are entered into the administration section of the tool resulting in the generation of a report within the tool that links the search with the topics defined in the PSCs business requirements (Figure 6).

Figure 6. Search Strategies Entered into Administration Section of Literature Analysis Tool



The search strategy is automatically executed according to the frequency required and key concepts are mined out within the references and organizes them for the topics defined in the report (Figure 7).

Figure 7. Hit List Report with Key Concepts Mined



The core functionalities enable rapid assimilation of the literature into users' decision-making process and workflow. Sorting and searching the references become intuitive (Figure 8).

Figure 8. Hit List Report Demonstrating Ability to Search and Sort

Once the user has analyzed publication topics in the tool, information can be exported in a variety of formats to enable further publication planning tasks to be completed. Figure 9 shows the functionality of exporting into multiple formats including Word, Excel, and EndNote.

Figure 9. Exporting Functionality

Figures 10a and 10b are examples of Word and Excel output from the export.

Figure 10a. Word File Literature Search Results with Table of Contents

Figure 10b. Excel File Literature Search Results

Additionally, XML feeds can populate SharePoint sites, spreadsheets containing each reference on its own row allows the data to be analyzed in MS Excel and Endnote, or other reference bibliography tools can be created and updated directly from the literature analysis tool. References can then be output directly into MS Word to facilitate manuscript development.

Through the tool, enhanced functionality is realized due to direct links to the full text or document delivery once the button below is selected in (Figure 11).

Figure 11. Link Functionality

Figures 12-15 demonstrate how dynamic charts are generated, which outputs a subset analysis and provides interactive visual renditions of the data that helps with the identification of publications trends and publication planning. Components of the charts in the tool are live and link back to the hit list literature search report. These charts can be copied and pasted into PowerPoint for management reports, presentations, and/or publication planning slides.

Figure 12. Hit List View to Generate Charts and Figures

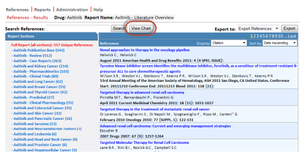


Figure 13. Bar Chart by Topic with XY-Axis Swapped

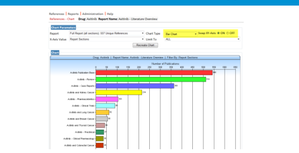


Figure 14. Pie Chart by Journal (Top 10)

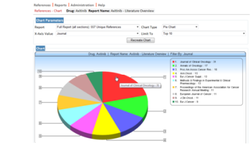
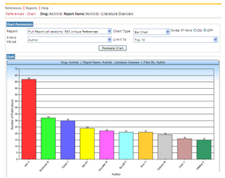


Figure 15. Bar Chart by Author (Top 10)



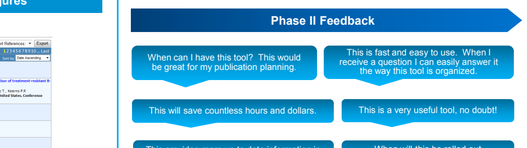
Users are notified by email after an alert is set when new references have been retrieved for the topics in which they are interested, so they are immediately updated on current literature. Standard functionalities such as searching, sorting, and filtering enable the references to be sliced and diced to provide users with quick and easy subsets of data.

During Phase I of the project, a hard savings of $12,829, as well as a released capacity of $115,500 was realized, resulting in a total savings of $128,328 during this five-month pilot. Released Capacity is the monetary value of the total time saved by all colleagues who were involved in the task or process. In this case, it involved reducing the number of literature searches colleagues performed on their own and/or the time saved due to having a link available to request articles. Phase II cost savings will be analyzed in the future, however, significant savings may be realized by not having to pay for agency literature analysis reports or publication planning assessments in addition to the aforementioned savings realized in Phase I.

Feedback received for both phases of the project are acknowledged in Figures 16a and 16b.

Figure 16a. Feedback During Phase I of the Project

Figure 16b. Feedback During Phase II of the Project



Conclusions
The literature analysis tool has resulted in the delivery of high quality, comprehensive searches, due to customized and targeted search strategies, and multiple databases utilized by leveraging the expertise of company library science professionals. These searches are delivered and formatted to meet the needs of stakeholders and link directly to full text or document delivery resulting in time saved in reviewing or performing searches, retrieving articles, and formatting search results. Additionally, the dynamic graphing capabilities can be useful during the publication planning cycle. As a result, PSCs can use the tool to keep updated on current literature, analyze publication topics aiding in publication planning, which can save time and money.

Disclosures
Authors of this presentation disclose the following concerning possible financial or personal relationships with commercial entities that may have a direct or indirect interest in the subject matter of this presentation.

Michelle Kissner: Employee, Pfizer Inc. Ownership of Pfizer stock.

Mark Drinkwater: Employee, PI^2 Solutions Ltd. Vendor for Literature Analysis Tool.

Belinda Stretch: Employee, Pfizer Inc. Ownership of Pfizer stock.

Erica Wright: Employee, Pfizer Inc. Ownership of Pfizer stock.

References
1. McDonald S, Taylor L, Adams C. Searching the right database: a comparison of four databases for psychiatry journals. *Health Libraries Review.* 1999;16:151-156.
2. Suarez-Almazor ME, Belseck E, Homik J, et al. Identifying clinical trials in the medical literature with electronic databases: MEDLINE alone is not enough. *Controlled Clinical Trials.* 2000;21:476-487.
3. Woods D, Trewheellar K. Medline and Embase complement each other in literature searches. *BMJ.* 1998;316:1166-1167.