DialogSolutions
Literature · Technology · Services

DOING IT RIGHT: THE RIGOROUS APPROACH TO MEDICAL LITERATURE MONITORING FOR DRUG SAFETY

Finding the One

How to manage the impact of duplicate references in medical literature monitoring

Executive Summary

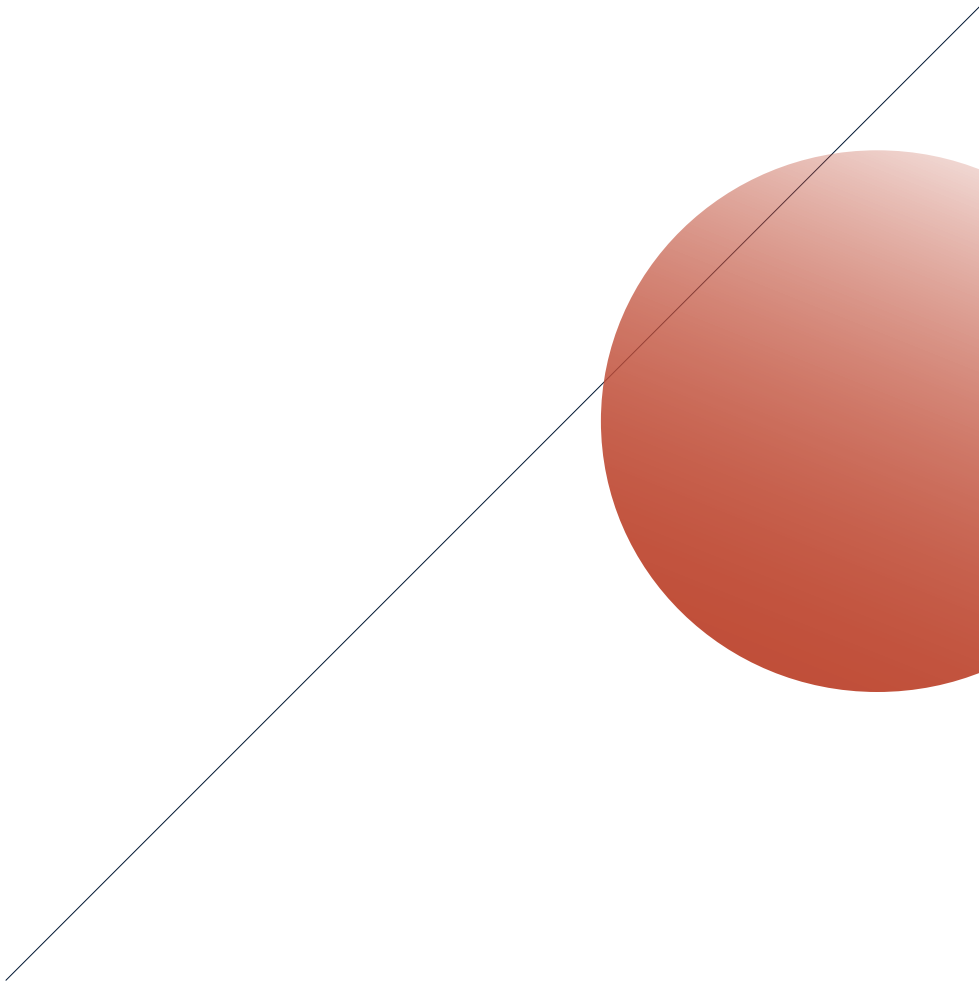
Too much information. In our everyday lives it's an irritation, at worst. But in medical literature monitoring it's a major issue. Because of the huge volume of sources and references in medical literature – and the different ways in which they are indexed – duplicates are a serious headache in the pharmacovigilance (PV) process. Our estimates suggest that on average, one-third of references retrieved in literature monitoring are duplicates; and 40% of references refer to more than one drug.

This puts more pressure on already-stretched teams and increases the risk of inconsistent assessments. The same reference may be reviewed several times for ICSRs, aggregate reports and safety signals – or one reference could be sent multiple times for case processing. Either way, duplicates create a growing snowball of unnecessary work. They also create risk: obfuscating the regulatory clock start date, prompting inconsistent review and therefore compliance issues, and skewing aggregate reports.

The ultimate goal is to take note of each duplicate reference while ensuring that only one unique and relevant reference goes through to be reviewed and processed. Achieving this demands a combination of best practice and smart technology. Using multi-database searches can remove many duplicate records from across multiple sources. And ensuring that alerts – automatic, regular search strategies – run against a long memory of their previous results also reduces the risk of duplicates. Finally, applying a robust, proven algorithm can take care of much of the "heavy lifting"

in deduplication, so the reviewer simply has to decide whether or not the reference is a duplicate and needs to be assessed.

In this paper, we explore the issues around deduplication in medical literature monitoring in more depth. At the end, you'll find a checklist to benchmark your own practice. We believe that all PV operations should have a robust process in place for dealing with duplicate references in medical literature: it saves time and money, and reduces risk.



CHAPTER 1

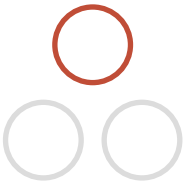
Why Are There So Many Duplicates in Medical Literature Monitoring?

Huge volume of sources and references

In our paper Better Relevance, we discussed the role of precision search in medical literature monitoring, noting that without it, alerts that search for adverse events are likely to produce 30-50% more references. Given the scale of the literature monitoring task in pharmacovigilance, sophisticated searches to identify patient safety issues are essential. While precision search reduces the quantity of irrelevant references, there is still the potential for unwanted duplicate references. This leaves teams with the additional task of deduplication: confirming and identifying duplicates so it is clear which references are relevant for ICSRs, aggregate reports and safety signals.

The problem of duplicates is a direct result of the huge volume of sources and references in medical literature, and the ways in which they are indexed:

- **Multiple databases.** A search strategy may retrieve the same document from multiple databases. *See Figure 1 below.*
- **Multiple publications.** The same material may be published in different journals or presented at different points in time. *See Figure 2 below.*
- **Name and date variations.** The same publication, author or date may be listed in different formats or spellings, creating duplicate references.
- **Duplicates that result from data migration.** If data is migrated from old to new literature monitoring systems, there is a risk of creating further duplicate references.



1/3

of references retrieved in literature monitoring are duplicates, on average

40%

of references refer to more than one drug

Source: Dialog Solutions estimates

- **Publication status changes.** In some medical literature databases, individual articles go through a process of revision/update/publishing status changes, including stages of indexing.

FIG. 1 - DUPLICATE ARTICLES WITH DIFFERENT CONVENTIONS FOR AUTHOR AND JOURNAL NAMING

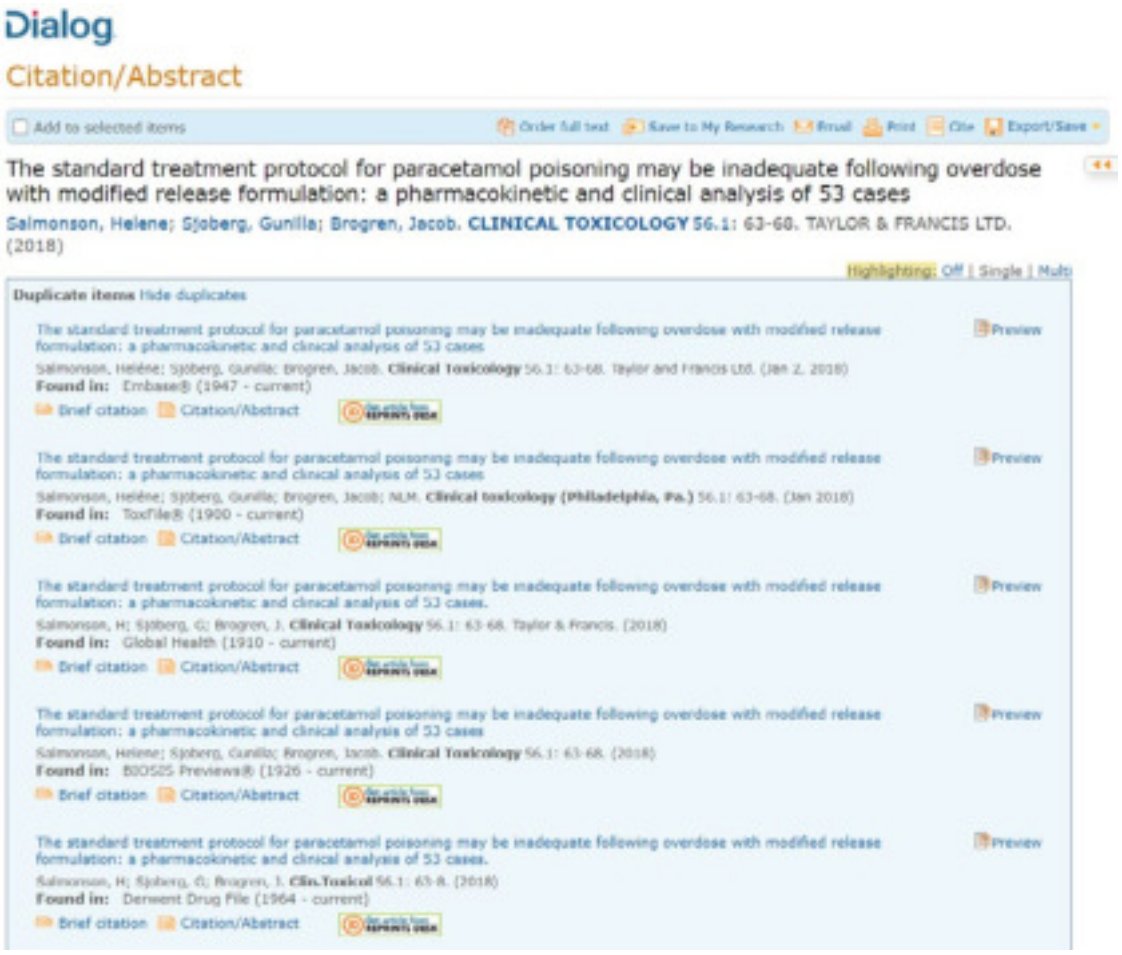
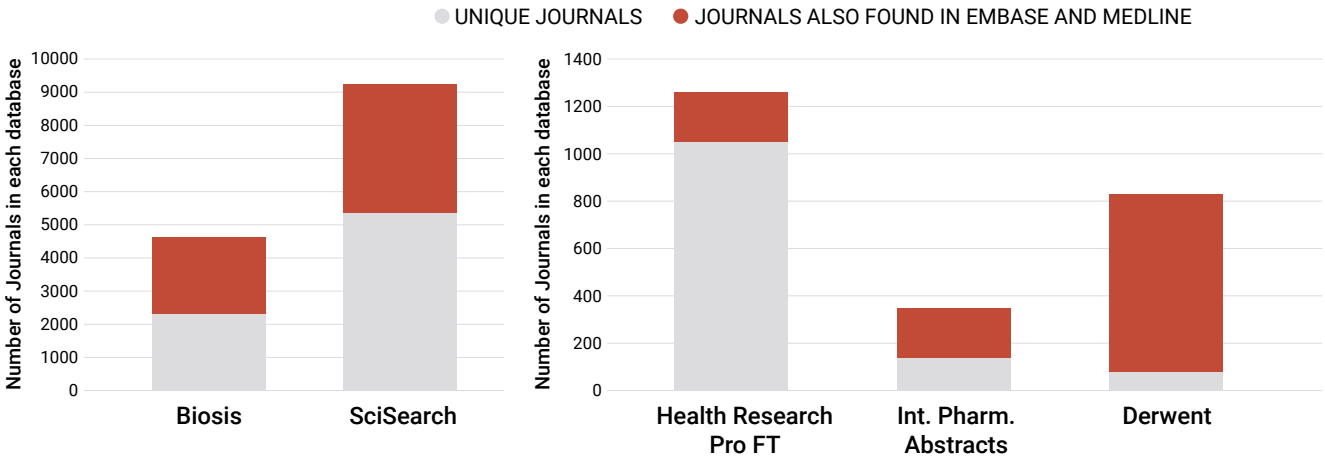


FIG. 2 - GRAPHS SHOWING TOTAL UNIQUE JOURNALS IN EACH DATABASE AND THE NUMBER OF DUPLICATE JOURNALS ALSO FOUND IN EMBASE AND MEDLINE.



CHAPTER 2

Implications and Costs for Literature Monitoring

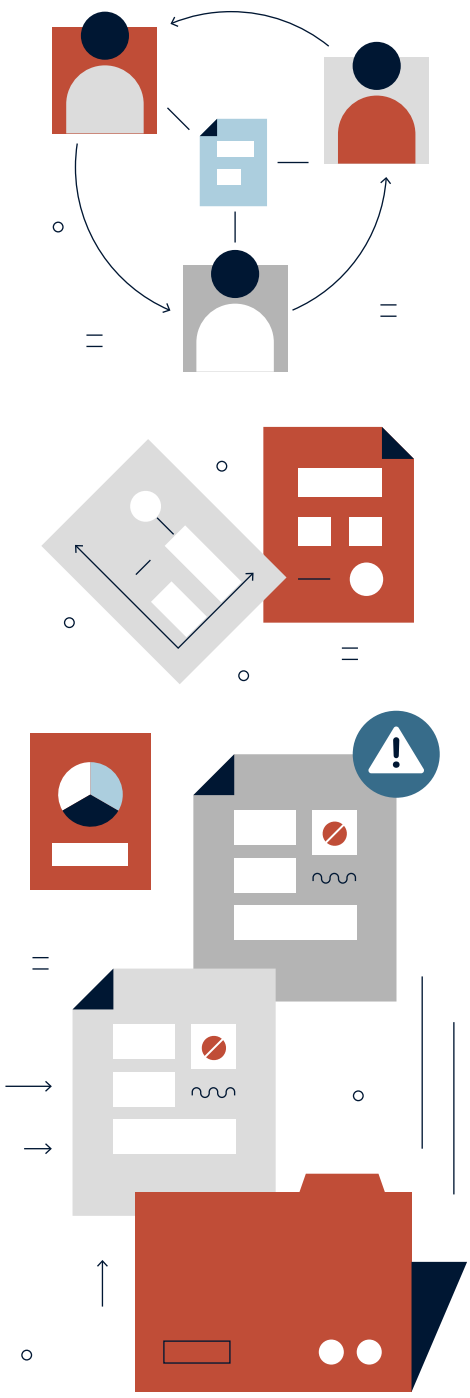
Extra workload, higher costs and the risk of inaccuracy

Medical literature monitoring can tie up huge amounts of time and resource, and the consequences of missing adverse events can be severe. Dealing with duplicates simply puts more pressure on teams that may already be stretched and increases the risk of inconsistent assessments.

THE IMPACTS OF DUPLICATES

- **Increased workload and cost.** At the review stage, for example, the same reference may be reviewed several times for ICSRs, aggregate reports and safety signals. And the same cases may also be sent multiple times to case processing, adding to time and expense downstream. So, dealing with additional duplicate references creates a growing snowball of workload throughout the literature monitoring process.

- **Obfuscation of "day zero".** The regulatory clock start date, or "day zero", is an essential element in the pharmacovigilance timetable and it can be derailed by duplicate references. For example: a search delivers a reference about drug A today. But seven days later, the same piece of literature is indexed in a different database in a different way and is retrieved in relation to drug B. Even though this drug B reference was imported seven days later than when the piece of literature was first found, its "day zero" is the date when it was imported for drug A. So duplicates can lead to late reporting of ICSRs.



- **Inconsistent review, and compliance issues.** Duplicate references can potentially be reviewed and assessed differently by different individuals. This can create compliance issues if, for example, one reviewer assesses the reference as an ICSR and the other as a non-ICSR.
- **Skewed reporting.** Data analysis of reports that include duplicates can lead to skewed conclusions. Duplicates also create extra effort for aggregate reporting review and can lead to skewed aggregate reports.
- **Risk when analysing signals.** Duplicate references can contribute to incorrect counts and therefore influence decisions related to signals.

CHAPTER 3

Solutions and Best Practice

Aggregate sources, focus on alerts and choose proven algorithms

How do you solve the problem of duplicates? Ultimately, the goal is to take note of each duplicate reference while ensuring that only one unique and relevant reference goes through to be reviewed and processed.

Deduplication algorithms work by assessing the various fields attached to a typical literature reference. Most use a combination of them to assess duplicates. However, given the many permutations of each field, these algorithms need to be carefully built to ensure they are both comprehensive and accurate enough for the task.

SINGLE OR MULTIPLE DATABASES?

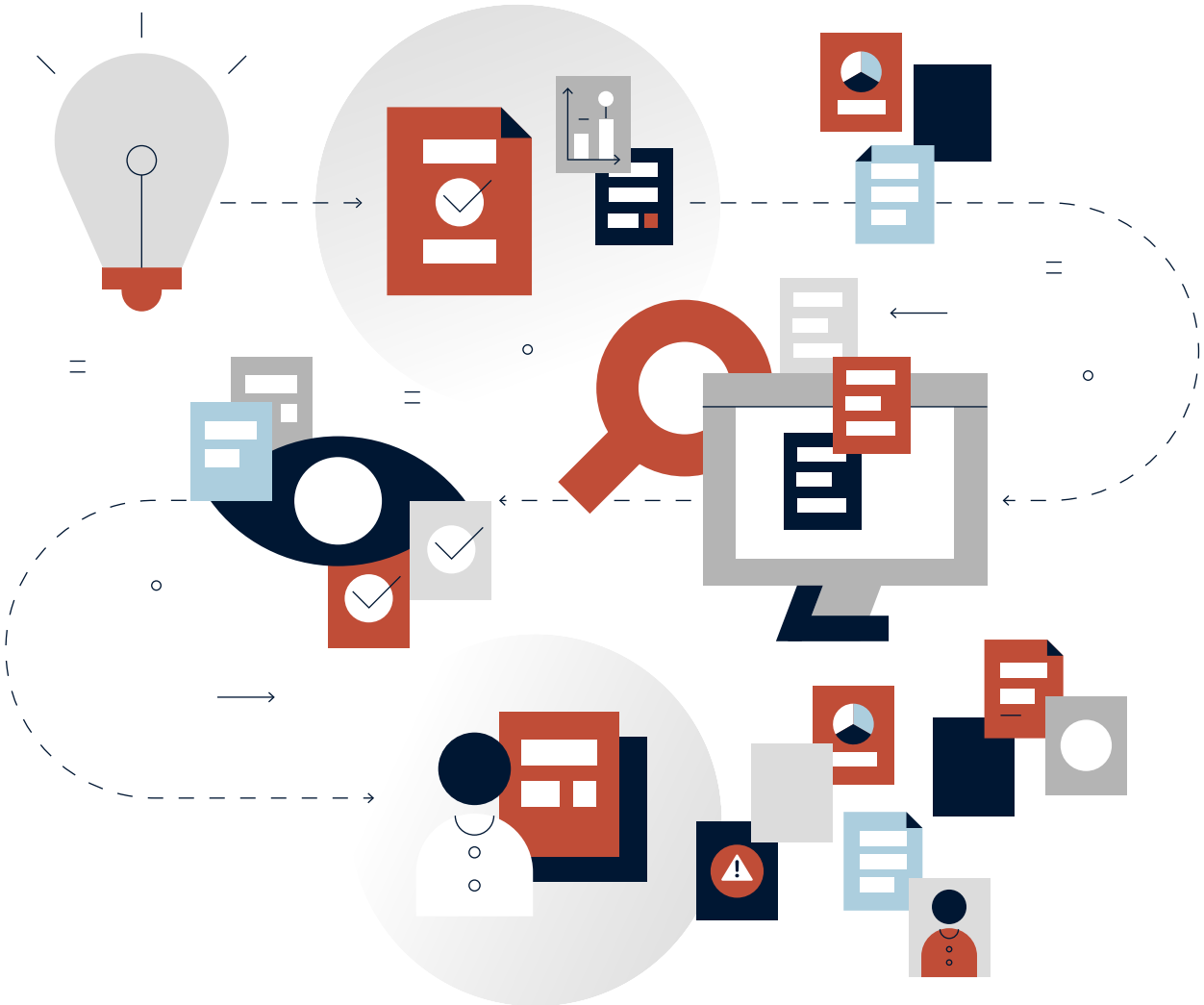
The first step in deduplication is to look at where your references are coming from. Searches derived from a multi-database search are efficient because they remove many duplicate records from across multiple sources. And they are most effective when they are accompanied by a normalisation of the fields that contribute to the duplication.

USING ALERTS

Unlike searches, alerts automatically run search strategies on a regular basis. The advantage of alerts is that they can have a “memory”: in some cases for 180 days and in others for the lifetime of the alert. This means that every alert runs against a memory of all the articles it has ever retrieved, reducing the risk of retrieving duplicate references.

APPLYING A PROVEN DEDUPLICATION ALGORITHM

Robust deduplication algorithms can do the heavy lifting of automatic deduplication. Depending on the content provider, data can be "dirty" and not normalized, so an additional deduplication algorithm can be applied within the literature monitoring software. This way, potential duplicates are brought to the attention of the reviewer, who then decides whether or not the reference is a duplicate and needs to be assessed. Dialog Solutions provides robust auto-deduplication and our **Drug Safety Triager** literature monitoring software provides an additional manual deduplication option.



CHAPTER 4

Checklist to Benchmark Your Own Process

Assess how you could reduce the impact of duplicates on literature monitoring

All pharmacovigilance operations need to have a robust process in place for dealing with duplicate references in medical literature. This checklist helps you to assess how your organisation measure up against best practice, so you can identify areas for review.

- **Do you have a clear goal for suppressing duplicates?** Is your process documented?
- **Does your content provider have a transparent deduplication policy?**
- **Does your literature monitoring software have a deduplication feature and does it track what was labelled as duplicate?** It's important that the algorithm used by vendors is clear and well documented, and can give you the confidence that you are neither missing potential valid references nor unnecessarily processing duplicates.
- **Do you track your rate of duplicates and if so, what percentage of duplicate references do you typically get?** Tracking the rate of duplicates in literature monitoring helps you to assess the efficiency of your process.
- **Is your process optimised for downstream applications?** Best practice ensures that case processing does not have to deal with duplicate references or conflicting assessments of the same reference.

DialogSolutions
Literature · Technology · Services

Our Other Whitepapers

Better Relevance

How precision search reduces cost, effort and risk in medical literature monitoring

DOWNLOAD

When Best is Not Enough

Ensuring quality management and validation in medical literature monitoring

DOWNLOAD

Dialog Solutions
Literature · Technology · Services

Get in touch to discuss how we can make your medical literature monitoring more effective:
go.dialog.com/Dialog_Solutions_MLM/

Find out more about our end-to-end approach to medical literature monitoring:
dialog.com/what-we-do/medical-literature-monitoring/

About Dialog Solutions

We are Dialog Solutions. Our goal is to simplify the research process for anyone, in any organisation. We do this through our technology and services, combined with the access we provide to the world's best peer-reviewed content.

Our origins are back before the Internet, and even before the dawn of personal computing. Dialog, our core search product, was launched in 1966 and is now part of our suite of research tools that includes Drug Safety Triager, Dialog Alerts Manager and PinPoint.

But we provide our customers with more than just software; we help them improve the way they do their research. Combined with our innovative approach to software development, we are a secure and stable partner for any organisation that uses research to make better business decisions.

Dialog Solutions is proud to be part of ProQuest LLC.