Enhancing Drug Safety through Data Management and Analytics

Ash Baghai

8/25/2024

Overview

Intended Audience

The intended audience for this project includes senior leaders and technical experts within the Analytical Operations and Quality Assurance departments of a biopharma company. These professionals are responsible for ensuring the testing and safety of pharmaceutical products through rigorous analytical methodologies and quality review. The insights from this project will help guide these stakeholders in making informed decisions to enhance drug quality and thus safety.

Existing Literature

The detection of low-level impurities in drug candidates has been a persistent challenge in the pharmaceutical industry. Traditional detection methods often fail to identify impurities present in low concentrations or those that are UV-inactive, posing potential risks to patient safety. Recent incidents of FDA drug recalls due to the presence of carcinogenic, nitrosamine related impurities have highlighted the urgent need for better detection techniques[1,2]. The literature suggests that advanced analytical methods, such as Liquid Chromatography-Mass Spectrometry (LC-MS), could significantly improve impurity detection. However, there is a notable gap in the systematic analysis of LC-MS data to predict and categorize these impurities effectively [3,4].

Anticipated Impact

This project aims to enhance the detection and categorization of toxic impurities in drug candidates, reducing the risk of patient exposure to harmful substances. First, by improving scientific data management to permit the ability to garner insights from integrated evaluations of otherwise siloed but related chemical and experimental data. Second, leveraging data analytics and machine learning to enhance the detection and categorization of toxic impurities. Ultimately, this work is expected to establish a new standard in impurity detection and risk assessment within the company, leading to improved drug safety profiles.

Research Questions

Main Research Question

How can we improve the detection and categorization of toxic impurities in drug candidates using data analytics to cross evaluate analytical data from, LC-MS with compound chemistries?

Sub-Questions:

1. What are the most common toxic impurities that evade detection by traditional methods?
2. How effective is LC-MS in identifying low-level and UV-inactive impurities compared to other techniques, such as GC analyses?
3. Can we develop a predictive model to assess the risk of impurities in future drug candidates based on their chemical structure?
4. Can the proposed data management and modeling systems reduce the incidence of harmful impurities in the pharmaceutical pipeline?
5. What are the cost implications and operational challenges associated with integrating this system into existing drug development workflows?

Definitions

- **Toxic Impurities**

  Substances within drug candidates that are harmful or carcinogenic at low concentrations.

- **GC (Gas Chromatography)**

  An analytical technique used to separate, identify, and quantify mixture components that have first been volatilized into a vapor.

- **LC-MS (Liquid Chromatography-Mass Spectrometry)**

  An analytical technique that combines the physical separation capabilities of liquid chromatography with the mass analysis capabilities of mass spectrometry.

Study Design

Descriptive Phase

We will initiate our study by first configuring the prerequisite data management platform, called Luminata[5]. Data collection and compilation of known toxic impurities will then be carried out, focusing on their chemical structures and corresponding LC-MS results across a diverse range of drug candidates. The goal of this phase is to identify patterns and characteristics of impurities that traditional methods struggle to detect, such as those present in low concentrations or UV-inactive compounds. Data sources will include open-source repositories, historical datasets from previous drug candidates, and newly acquired data from ongoing LC-MS analyses.

Modeling Phase

In the next phase, we will develop a Python-based analytical program designed to cross-compare distinct experimental data sets for each drug candidate accessed from the data management platform. This program will systematically evaluate LC-MS data, identifying low-level impurities by recognizing subtle signal patterns that standard detection techniques might overlook. The program will then categorize these impurities based on their chemical characteristics, allowing us to create a detailed impurity profile for each drug candidate. Following this, we will use machine learning models to predict the risk of similar impurities appearing in future drug candidates, based on their chemical structure and past impurity profiles.

Outcome

The final deliverable of this study will be a comprehensive report that includes the following:

1. A detailed list of identified impurities and their chemical classifications.
2. A risk assessment for future drug candidates, highlighting compounds that are more likely to harbor undetected toxic impurities.
3. Cost itemization for projected expenditures to scale the system.
4. Recommendations for integrating this predictive impurity detection system into existing quality assurance workflows to enhance drug safety and efficacy. This report will serve as a critical decision-making tool for senior leaders and technical experts within the Analytical Operations and Quality Assurance departments.

Data

We will utilize a combination of open-source databases and proprietary data from the biopharma company's LC-MS and GC testing results. Key data sources will include:

- **Compound Structures:** Detailed chemical structures of drug candidates from both existing databases and proprietary sources.
- **GC and LC-MS Spectra:** GC and High-resolution LC-MS spectra, capturing both typical and atypical signal patterns for each drug candidate.
- **Known Impurities:** Historical data on known toxic impurities, particularly those that have previously evaded detection by traditional methods.

If gaps are identified during the descriptive phase, we will propose the collection of additional data, such as new LC-MS or GC tests or acquiring supplementary datasets from external sources.

Data Systems requirement

A scientific data management platform, called Luminata, will be used that consolidates both compound libraries and diverse drug experiment data types in a searchable and quantifiable format. A prerequisite to this project will be the pricing of and configuration of this platform. The justification and setup of this type of data management system already has precedent[5] .

Hypotheses

Null Hypothesis (H0): There is no significant improvement in the detection of toxic impurities using the proposed system compared to traditional methods.

Alternative Hypothesis (H1): The proposed system significantly improves the detection and categorization of toxic impurities compared to traditional methods.

Potential Risks

1. **Scientific Validity**

   There is a risk that the program may fail to detect certain impurities, leading to incomplete or misleading results. To mitigate this, we will conduct rigorous validation tests using both historical data and blind sample tests.

2. **Stakeholder Expectations**

   Balancing the innovative approach of this system with stakeholders' comfort levels with traditional methods is critical. There may be resistance to adopting a new system, particularly if it challenges long-established practices. We will address this by clearly demonstrating the system’s advantages through case studies and pilot implementations.

3. **Ethics and Data Security**

   Given the proprietary nature of the data involved, safeguarding sensitive information is essential. We will implement stringent data security measures such as access controls.

4. **Model Generalizability**

   There is a risk that the predictive model may perform well on certain drug classes but poorly on others. To address this, we will test the model across a broad range of compounds and refine it to improve generalizability.

5. **Cost and Resource Allocation**

   The development and maintenance of the predictive model may incur significant costs, particularly in terms of computational resources (such as data management platform) and ongoing model updates. A cost-benefit analysis will be conducted to ensure that the long-term benefits justify these expenses.

References

1. https://www.fda.gov/drugs/drug-safety-and-availability/drug-recalls
2. Bharate, S. S. (2021). Critical Analysis of Drug Product Recalls due to Nitrosamine Impurities. Journal of Medicinal Chemistry, 64 (6), 2923-2936. https://doi.org/10.1021/acs.jmedchem.0c02120
3. Shaik, K. M., Sarmah, B., Wadekar, G. S., & Kumar, P. (2022). Regulatory Updates and Analytical Methodologies for Nitrosamine Impurities Detection in Sartans, Ranitidine, Nizatidine, and Metformin along with Sample Preparation Techniques. *Critical Reviews in Analytical Chemistry, 52*(1), 53-71. https://doi.org/10.1080/10408347.2020.1788375
4. Marlés-Torres, A., López-García, R., Bessa-Jambrina, S., & Galán-Rodríguez, C. (2023). Ultra-high-speed liquid chromatography combined with mass spectrometry detection analytical methods for the determination of nitrosamine drug substance-related impurities. *Journal of Separation Science, 46*(13), e2300125. https://doi.org/10.1002/jssc.202300125
5. https://www.acdlabs.com/resource/how-pfizer-is-using-luminata-to-support-pharmaceutical-development/