REDICA Systems

AI for Quality and Compliance Teams

Presented by:
Jerry L. Chapman, Senior GMP Quality Expert at Redica Systems

Another exciting talk on FDA inspections?

A Promise and a Suggestion

A Promise:

You will learn something new

A Suggestion:

Buckle up, there's a lot to cover

Agenda

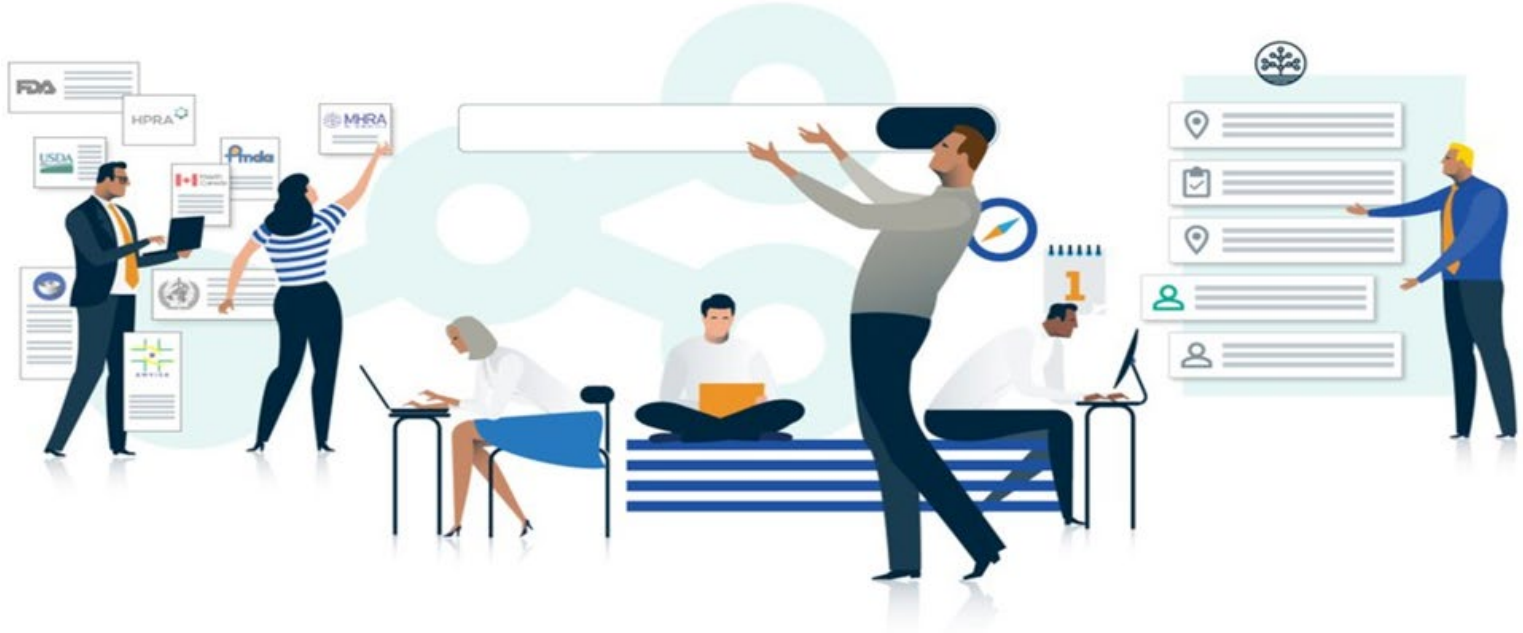
- Intro: Enabling Quality & Regulatory Professionals
- Data Sources, Machine Learning, NLP, AI Tools
- Building and Applying the Models
- Case Study 1: Sterile Manufacturing
- Case Study 2: Data Integrity
- Summary and Conclusions

Agenda

- Intro: Enabling Quality & Regulatory Professionals
- Data Sources, Machine Learning, NLP, AI Tools
- Building and Applying the Models
- Case Study 1: Sterile Manufacturing
- Case Study 2: Data Integrity
- Summary and Conclusions

Introduction

In theory, the mission is simple: Make safe, high quality great products while meeting regulatory requirements and business demands. However, challenges include…



… complex facilities, elaborate supply chains, intricate products, and evolving regulations

Introduction

- Increasing number of regulatory agencies and requirements.
- Health agencies rules and regulations vs. how they are enforced on inspection
- What are their expectations for your facility?
- Where are those changing expectations found? CGMP

Introduction

- Mining enforcement data: 483 observations, warning letter citations.
- Time consuming and difficult, requires expertise.
- Where is enforcement data found?
- Redica Systems has built the largest integrated dataset of inspection and enforcement documents in the world – the data sets used for the case study analyses.

Introduction

- Includes all FDA inspected and registered sites since 2000. 20 times more data than has been released on FDA.gov (targeted FOI).
- Once you get the inspection data, do you have the experts to analyze it?
- And if so, do they still print hard copies and use highlighting markers to analyze them?
- Redica Systems has built better ways to do this: proprietary AI “expert models” using machine learning and honed by industry experts

Agenda

- Intro: Enabling Quality & Regulatory Professionals
- Data Sources, Machine Learning, NLP, AI Tools
- Building and Applying the Models
- Case Study 1: Steril Manufacturing
- Case Study 2: Data Integrity
- Summary and Conclusions

Machine Learning, Natural Language Processing (NLP), AI: Data Sources, Uses

Data Sources

- All data are from PUBLIC sources
- FDA Inspections databases (CLIIL and FACTS) – Dec 8, 2018
- FDA CFR Citations database
- Warning Letters from FDA.gov
- 40,000+ 483s, responses and EIRs obtained via FOIA and FDA.gov
- FDA registration databases (e.g., GDUFA)
- FDA 21 CFR 211 catalogue (Cornell University)
- Experts - Jerry Chapman (GMP), Barbara Unger (GMP), Jane Wastl (GMP), Mark Agostino (med devices), others

Site Tags

- CDER
  - Rx manufacturing (API and FDF)
  - Over-the-Counter (OTC)
  - Other CDER (e.g., unapproved drugs)
  - GCP (Clinical Investigators, IRBs, Sponsors)
  - Compounding Pharmacies
- CBER
- CDRH
- MHRA
- Health Canada
- etc.

How do we use the data?

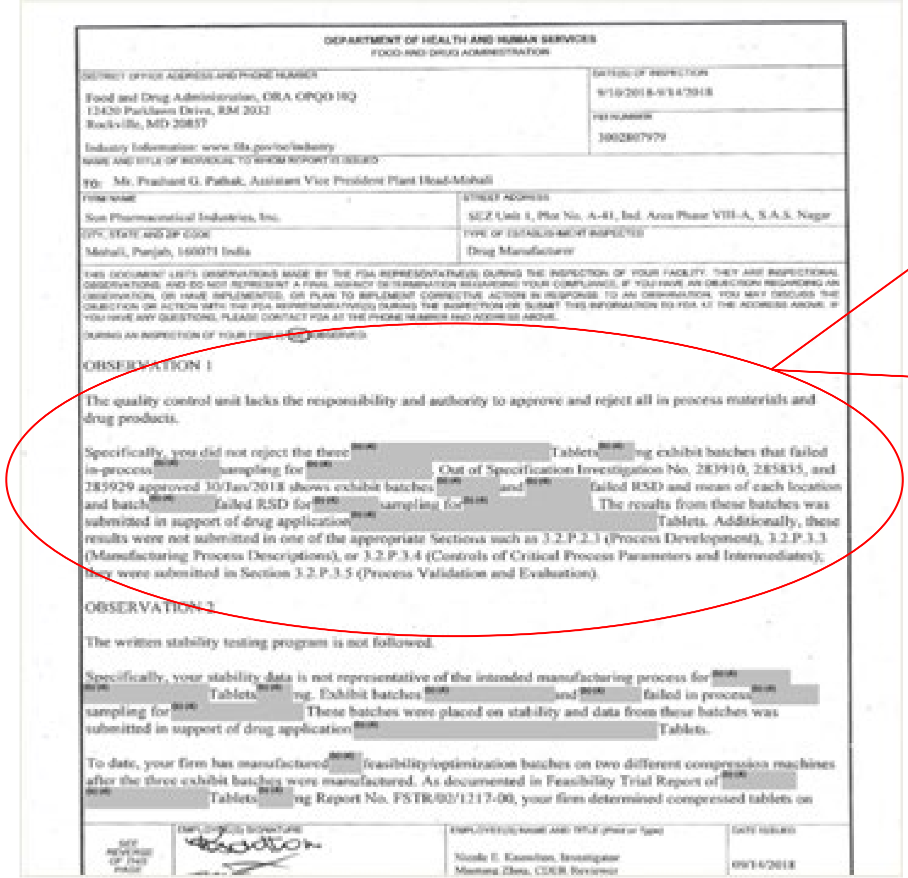
- Investigator profiles
- CMO due diligence
- Vendor selection
- Inspection preparation
- Keeping PQS up to date
- Tracking changes to authoritative sources
- Tracking and trending inspection observations
- Finding observations and citations “hiding in plain sight” (case studies)
- Tip of the iceberg…

Compliance Analysis Algorithm

- Redica Systems has created AI tools – “expert models” – that allow deep and rapid analysis of compliance and other data sets.
- Created using Machine Learning, Natural Language Processing (NLP) and other AI tools, used to analyze FDA warning letters, 483s, and other documents they way an expert would.
- To begin to train the AI algorithm and prepare the documents for examination, there are initial, important steps that must be taken: scanning, parsing, cleaning, tokenizing, POS tagging, stemming, etc.

Making Scanned Documents Searchable and Machine Readable

Processed >40,000 483s, 483R, and EIRs



DEPARTMENT OF HEALTH AND HUMAN SERVICES
FOOD AND DRUG ADMINISTRATION

DISTRICT OFFICE ADDRESS AND PHONE NUMBER
Food and Drug Administration, ORA OPQO HQ
12420 Parklawn Drive, RM 2032
Rockville, MD 20857

Industry Information: www.fda.gov/oc/industry

DATE(S) OF INSPECTION
9/10/2018-9/14/2018

FEI NUMBER
3002807979

NAME AND TITLE OF INDIVIDUAL TO WHOM REPORT IS ISSUED
TO: Mr. Prashant G. Pathak, Assistant Vice President Plant Head-Mohali

FIRM NAME
Sun Pharmaceutical Industries, Inc.

STREET ADDRESS
SEZ Unit 1, Plot No. A-41, Ind. Area Phase VIII-A, S.A.S. Nagar

CITY, STATE AND ZIP CODE
Mohali, Punjab, 160071 India

TYPE OF ESTABLISHMENT INSPECTED
Drug Manufacturer

THIS DOCUMENT LISTS OBSERVATIONS MADE BY THE FDA REPRESENTATIVE(S) DURING THE INSPECTION OF YOUR FACILITY. THEY ARE INSPECTIONAL OBSERVATIONS, AND DO NOT REPRESENT A FINAL AGENCY DETERMINATION REGARDING YOUR COMPLIANCE. IF YOU HAVE AN OBJECTION REGARDING AN OBSERVATION, OR HAVE IMPLEMENTED, OR PLAN TO IMPLEMENT CORRECTIVE ACTION IN RESPONSE TO AN OBSERVATION, YOU MAY DISCUSS THE OBJECTION OR ACTION WITH THE FDA REPRESENTATIVE(S) DURING THE INSPECTION OR SUBMIT THIS INFORMATION TO FDA AT THE ADDRESS ABOVE. IF YOU HAVE ANY QUESTIONS, PLEASE CONTACT FDA AT THE PHONE NUMBER AND ADDRESS ABOVE.

DURING AN INSPECTION OF YOUR FIRM WE OBSERVED:

OBSERVATION 1

The quality control unit lacks the responsibility and authority to approve and reject all in process materials and drug products.

Specifically, you did not reject the three [redacted] Tablets [redacted] mg exhibit batches that failed in-process [redacted] sampling for [redacted]. Out of Specification Investigation No. 283910, 285835, and 285929 approved 30/Jan/2018 shows exhibit batches [redacted] and [redacted] failed RSD and mean of each location and batch [redacted] failed RSD for [redacted] sampling for [redacted]. The results from these batches was submitted in support of drug application [redacted] Tablets. Additionally, these results were not submitted in one of the appropriate Sections such as 3.2.P.2.3 (Process Development), 3.2.P.3.3 (Manufacturing Process Descriptions), or 3.2.P.3.4 (Controls of Critical Process Parameters and Intermediates); they were submitted in Section 3.2.P.3.5 (Process Validation and Evaluation).

OBSERVATION 2

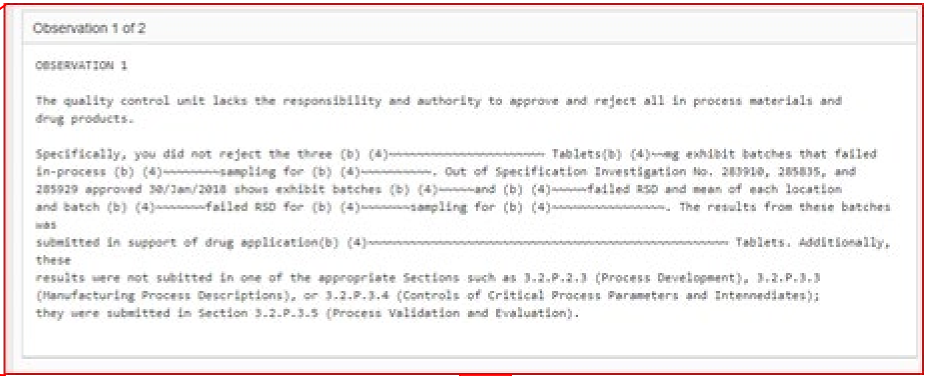
The written stability testing program is not followed.

Specifically, your stability data is not representative of the intended manufacturing process for [redacted] Tablets [redacted] mg. Exhibit batches [redacted] and [redacted] failed in process [redacted] sampling for [redacted]. These batches were placed on stability and data from these batches was submitted in support of drug application [redacted] Tablets.

To date, your firm has manufactured [redacted] feasibility/optimization batches on two different compression machines after the three exhibit batches were manufactured. As documented in Feasibility Trial Report of [redacted] Tablets [redacted] mg Report No. FSTR/02/1217-00, your firm determined compressed tablets on

SEE REVERSE OF THIS PAGE | EMPLOYEE(S) SIGNATURE | EMPLOYEE(S) NAME AND TITLE (Print or Type): Nicole E. Knowlton, Investigator; Mancang Zhou, CDER Reviewer | DATE ISSUED: 09/14/2018

Clean Observation Text: OCR, Proofing, Retyping



Observation 1 of 2

OBSERVATION 1

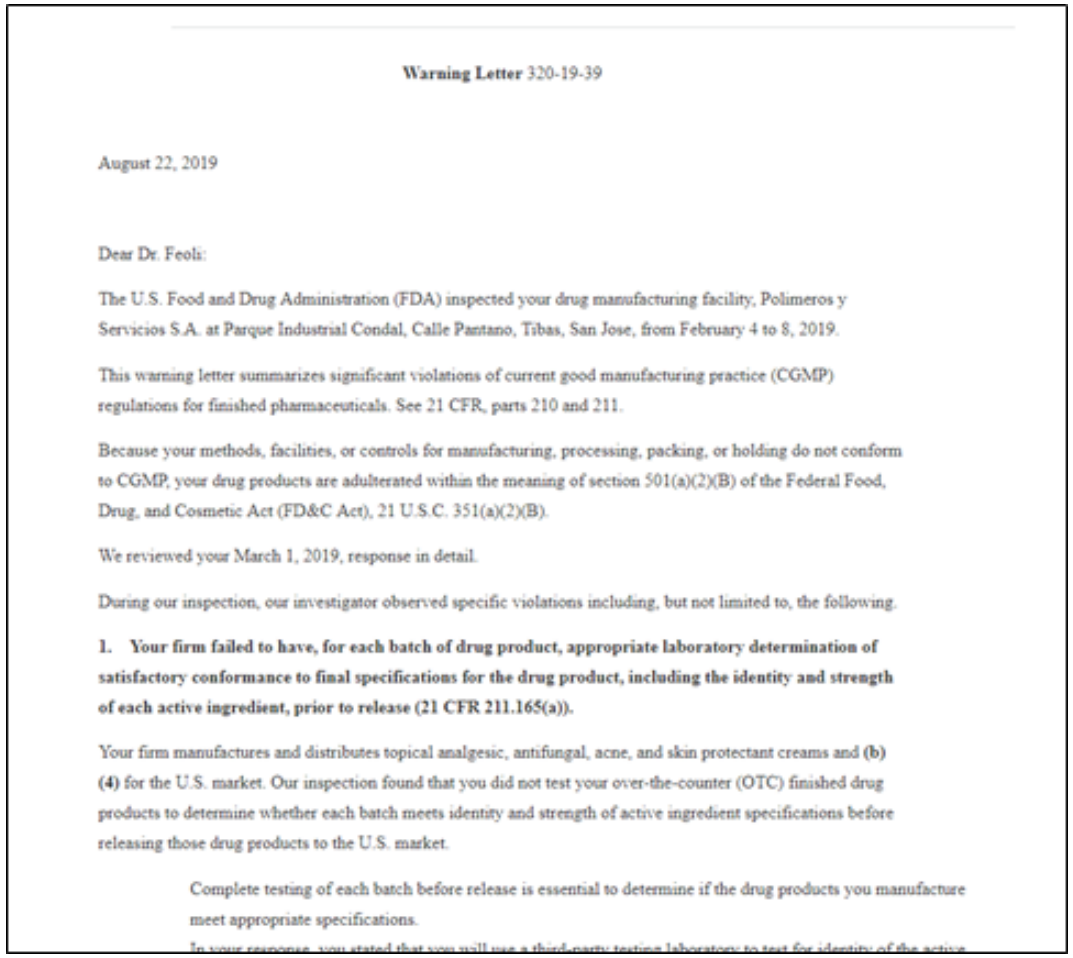
The quality control unit lacks the responsibility and authority to approve and reject all in process materials and drug products.

Specifically, you did not reject the three (b) (4)----------------------- Tablets(b) (4)--mg exhibit batches that failed in-process (b) (4)---------sampling for (b) (4)----------. Out of Specification Investigation No. 283910, 285835, and 285929 approved 30/Jan/2018 shows exhibit batches (b) (4)-----and (b) (4)-----failed RSD and mean of each location and batch (b) (4)-------failed RSD for (b) (4)-------sampling for (b) (4)----------------. The results from these batches was
submitted in support of drug application(b) (4)--- Tablets. Additionally, these
results were not subitted in one of the appropriate Sections such as 3.2.P.2.3 (Process Development), 3.2.P.3.3 (Manufacturing Process Descriptions), or 3.2.P.3.4 (Controls of Critical Process Parameters and Intermediates);
they were submitted in Section 3.2.P.3.5 (Process Validation and Evaluation).

Parse "main topic" of Observation

```
The quality control unit lacks the
responsibility and authority to
approve and reject all in process
materials and drug products.
```

Natural Language Processing: Text Parsing



Warning Letter 320-19-39

August 22, 2019

Dear Dr. Feoli:

The U.S. Food and Drug Administration (FDA) inspected your drug manufacturing facility, Polimeros y Servicios S.A. at Parque Industrial Condal, Calle Pantano, Tibas, San Jose, from February 4 to 8, 2019.

This warning letter summarizes significant violations of current good manufacturing practice (CGMP) regulations for finished pharmaceuticals. See 21 CFR, parts 210 and 211.

Because your methods, facilities, or controls for manufacturing, processing, packing, or holding do not conform to CGMP, your drug products are adulterated within the meaning of section 501(a)(2)(B) of the Federal Food, Drug, and Cosmetic Act (FD&C Act), 21 U.S.C. 351(a)(2)(B).

We reviewed your March 1, 2019, response in detail.

During our inspection, our investigator observed specific violations including, but not limited to, the following.

1. Your firm failed to have, for each batch of drug product, appropriate laboratory determination of satisfactory conformance to final specifications for the drug product, including the identity and strength of each active ingredient, prior to release (21 CFR 211.165(a)).

Your firm manufactures and distributes topical analgesic, antifungal, acne, and skin protectant creams and **(b) (4)** for the U.S. market. Our inspection found that you did not test your over-the-counter (OTC) finished drug products to determine whether each batch meets identity and strength of active ingredient specifications before releasing those drug products to the U.S. market.

Complete testing of each batch before release is essential to determine if the drug products you manufacture meet appropriate specifications.

In your response, you stated that you will use a third-party testing laboratory to test for identity of the active

Drug GMP Warning Letter "parts"

- Name
- Recipient
- Introduction
- **Deficiency title 1**
- Deficiency description 1
- Deficiency action 1
- Deficiency feedback 1
- **Deficiency title 2**
- Deficiency description 2
- Deficiency action 2
- Deficiency feedback 2
- **Deficiency title n**
- Deficiency description n
- Deficiency action n
- Deficiency feedback n
- Cgmp consultant recommended
- Format_type
- Conclusion
- Reply_to
- Identification number
- Footer

Natural Language Processing: Text Parsing

Not all sections are evaluated. **Parsing is key:**

CFR citation: “Your firm failed to clean, maintain, and, as appropriate for the nature of the drug, sanitize and/or sterilize equipment and utensils at appropriate intervals to prevent malfunctions or contamination that would alter the safety, identity, strength, quality, or purity of the drug product beyond the official or other established requirements. (21 CFR 211.67(a))”

Specifically…

It is the “specifically” or “for example” test that contains rich content

Natural Language Processing: Text Cleaning

Text Cleaning is the process of clearing out the “junk” from a sentence.

Examples:

- Remove extra spaces
- Make all words lowercase
- Remove typos & misspelling

Natural Language Processing: Tokenization

Splitting up a sentence into tokens. The most basic of which is just to split a sentence into individual words.

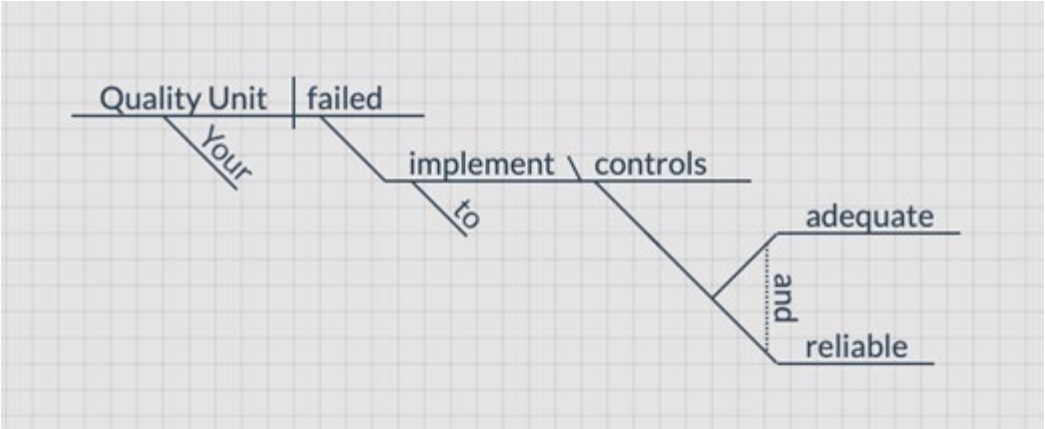
“Your Quality Unit failed to implement adequate and reliable controls for ensuring that distributed drug products always comply with the efficacy and quality they represent to possess.”

['Your', 'Quality', 'Unit', 'failed', 'to', 'implement', 'adequate', 'and', 'reliable', 'controls', 'for', 'ensuring', 'that', 'distributed', 'drug', 'products', 'always', 'comply', 'with', 'the', 'efficacy', 'and', 'quality', 'they', 'represent', 'to', 'possess', '.']

Natural Language Processing: Parts of Speech (POS) tagging

“Part of speech” tagging, a.k.a. sentence diagramming

“Your Quality Unit failed to implement adequate and reliable controls for ensuring that distributed drug products always comply with the efficacy and quality they represent to possess.”

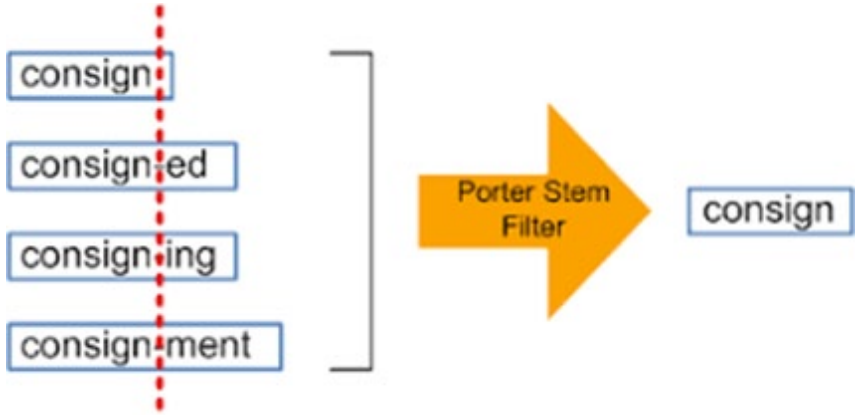


[('Your', 'PRP$'), ('Quality', 'NNP'), ('Unit', 'NNP'), ('failed', 'VBD'), ('to', 'TO'), ('implement', 'VB'), ('adequate', 'JJ'), ('and', 'CC'), ('reliable', 'JJ'), ('controls', 'NNS'), ('for', 'IN'), ('ensuring', 'VBG'), ('that', 'IN'), ('distributed', 'VBN'), ('drug', 'NN'), ('products', 'NNS'), ('always', 'RB'), ('comply', 'VBP'), ('with', 'IN'), ('the', 'DT'), ('efficacy', 'NN'), ('and', 'CC'), ('quality', 'NN'), ('they', 'PRP'), ('represent', 'VBP'), ('to', 'TO'), ('possess', 'VB'), ('.', '.')]

Natural Language Processing: Stemming

Stemming is the process of reducing each word in a written document into its word stem, base or root form. This will not necessarily become a proper word, but all permutations of a word will stem to the same root.

“Your Quality Unit failed to implement adequate and reliable controls for ensuring that distributed drug products always comply with the efficacy and quality they represent to possess.”



your qualiti unit fail to implement adequ and reliabl control for ensur that distribut drug product alway compli with the efficaci and qualiti they repres to possess

Natural Language Processing: N-grams

An n-gram is a contiguous sequence of n items from a given sample of text or speech

“Your Quality Unit failed to implement adequate and reliable controls for ensuring that distributed drug products always comply with the efficacy and quality they represent to possess.”

Bi-gram
('Your', 'Quality')
('Quality', 'Unit')
('Unit', 'failed')
('failed', 'to')

Tri-gram
(**'Your', 'Quality', 'Unit'**)
('Quality', 'Unit', 'failed')
('Unit', 'failed', 'to')
('failed', 'to', 'implement')
('to', 'implement', 'adequate’)…….and longer ones

Subject Matter Experts create n-grams from experience and compliance document language; tested over time in model iterations.

Agenda

- Intro: Enabling Quality & Regulatory Professionals
- Data Sources, Machine Learning, NLP, AI Tools
- Building and Applying the Models
- Case Study 1: Sterile Manufacturing
- Case Study 2: Data Integrity
- Summary and Conclusions

Building and Applying the Models

- Experts create n-grams for each category and subcategory; also, TurboEIR 483 text
- Expert model for human drugs: completed, deployed; it is the most mature model and is used for examples and case studies in this presentation
- Organized by the 6 FDA quality systems + 1 (DI)
- Other expert models, e.g. for **medical devices, GCP, APIs, human cell and tissue/GTP**, have been built and are in the testing phase
- Differentiated by parsing models, n-grams, TurboEIR text, and subject matter experts evaluating them, will be organized differently to match the topic
- *GCP will include at least three different models*

Building and Applying the Models: GMP Classification Categories (Human Drugs)

Quality System

- Agency Notification (4 subs)
- Audit (2 subs)
- CAPA
- Change Control
- Complaint Management
- Records and Reports (13 subs)
- Deviations / Investigations (8 subs)
- Qualified Personnel (3 subs)
- Quality Unit Inadequate (1 sub)
- Risk Mgmt.

Packaging & Labeling

- Drug product containers and closures (3 subs)
- Label and Packaging Controls
- Line Clearance
- Serialization

Facilities & Equipment

- Cleaning (3 subs)
- Design (9 subs)
- Maintenance (3 subs)
- Alarm Management
- HVAC
- Pest Control
- Records and Reports

Materials

- Distribution
- Material Receipt and Handling (3 subs)
- Material Sampling and Testing (3 subs)
- Material Storage and Control
- Retain Samples

Laboratory

- Method Validation
- OOS/ OOT
- Stability (2 subs)
- Systems Controls
- Testing (4 subs)
- Reagents and Standards
- Records and Reports
- Sample Management

Production

- API
- Batch Records
- Clean Utilities
- Cleaning validation or verification
- Contamination Control
- High Potency/ Allergenic
- Nonsterile products (2 subs)
- Penicillin and Cephalosporin
- Personnel Responsibilities
- Process control (5 subs)
- Process Monitoring / Continued Process Verification
- Process Validation (2 subs)
- Product Contamination
- Records and Reports
- Retain Samples
- Sterile Products (8 subs)

Data Integrity

- Accurate
- Attributable (3 subs)
- Backup and Archival
- Contemporaneous
- Data Destruction
- Data Manipulation
- Legible
- Original Data
- Paper Record Controls
- System Controls
- Testing into Compliance

Building and Applying the Models

- Build the parser and the model
- Process a group FDA warning letters – NLP tools
- Run the algorithm against the training sets
- Experts grade parsing and tagging, “machine learning” – why is it not right? Highly iterative.
- Changes to n-grams, algorithm, fuzzy matching
- Find secondary potential areas of concern in addition to those listed by CFR
- “Hiding in plain sight”

Agenda

- Intro: Enabling Quality & Regulatory Professionals
- Data Sources, Machine Learning, NLP, AI Tools
- Building and Applying the Models
- Case Study 1: Sterile Manufacturing
- Case Study 2: Data Integrity
- Summary and Conclusions

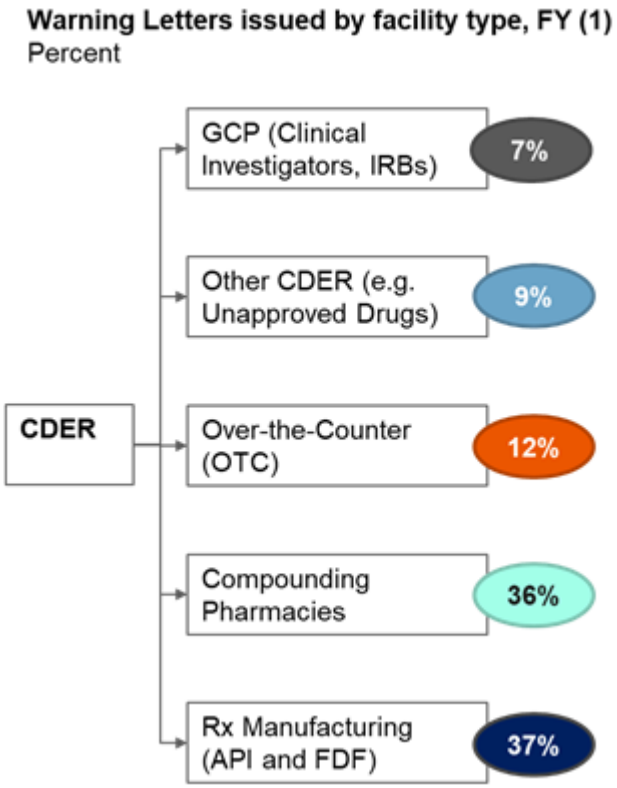
Case Study 1: Pharmacy Compounding and Outsourcing: Defining 503A vs. 503B

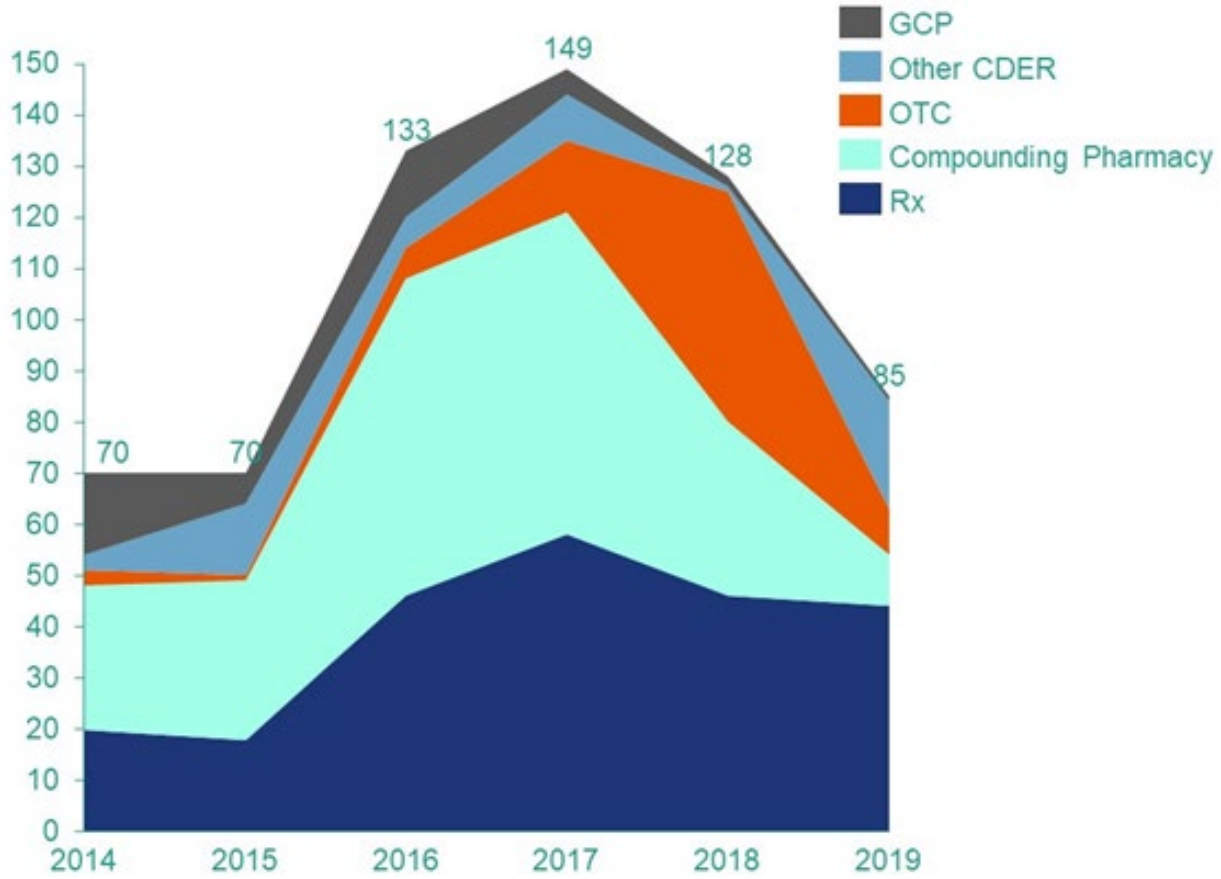
Section 503A (1997)	Section 503B (2013)
Compound according to prescriptions specific to particular patients	May manufacture large batches with or without prescriptions to be sold to healthcare facilities for office use only.
Must Comply with USP <795> and <797> along with state board of pharmacy regulations	Must Comply with USP <795> and <797> along with state board of pharmacy regulations and 21 CFR Part 210 and 211 (CGMP)
Environmental Monitoring must be performed every six months	An Environmental Monitoring program must be developed and performed, at minimum, per production shift in the ISO 5 areas and weekly in ISO 7 and ISO 8

Case Study 1: Challenges Moving from 503A to 503B

- A way to help more patients. But major changes are needed.
- It is not a matter of doing the same work but just more of it.
- Must maintain a quality department as an independent entity with complete autonomy for investigations and releasing product.
- Need to develop systems governed by approved procedures for:
  - Customer complaints
  - Change management
  - Failure investigations
  - Corrective and preventive actions (CAPAs)
  - Aseptic processing
  - Equipment validation, among others.
- Developing / managing these systems requires a different expertise and mindset.

Case Study 1: FDA Warning Letters by Facility Type





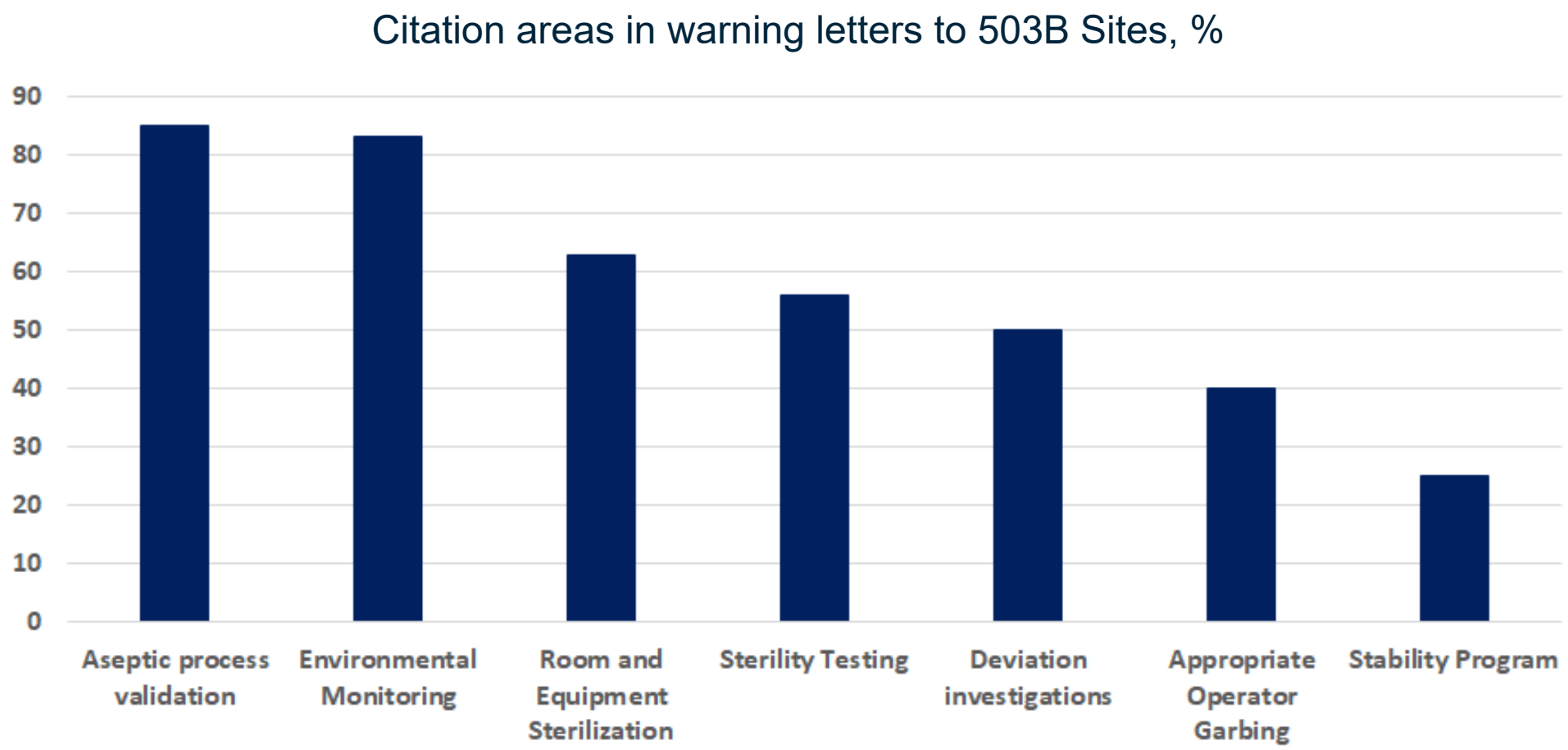
Dates: FDA Fiscal Year used (October – September), data through Aug 18,2019

Case Study 1: Warning Letter Population for Analysis

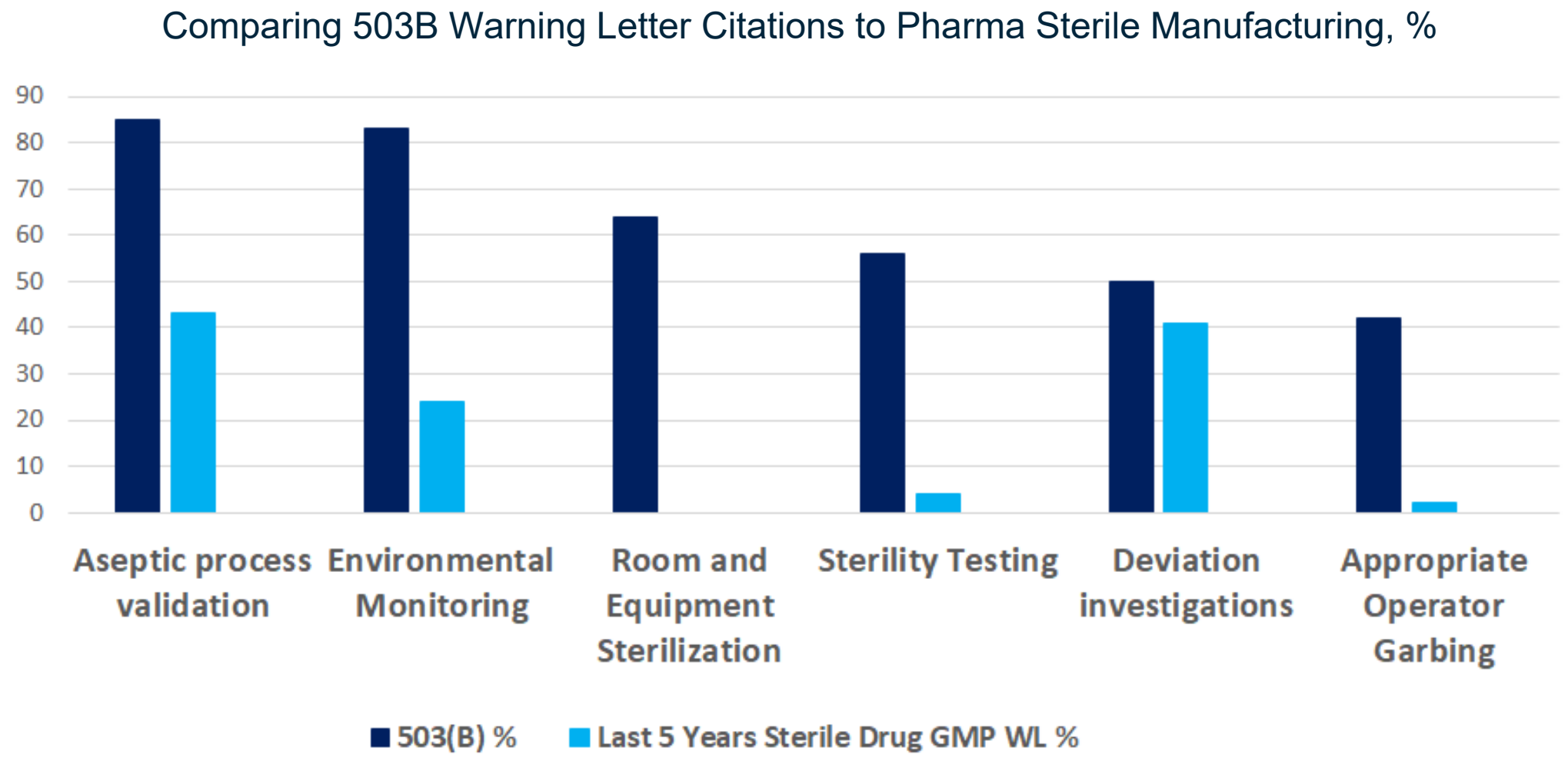
503B FDA warning letters to date since 2014*	59
Number with CFR citations (mostly after 2014)	52
Last 5 years of ALL FDA Drug GMP Warning Letters* (N=205: 76 API, 80 Non-Sterile, *49 Sterile*)	49
Searched and analyzed by CFR citation	101

*Through February 2020

Case Study 1: 503B Warning Letter Citations



Case Study 1: 503B Warning Letter Citations

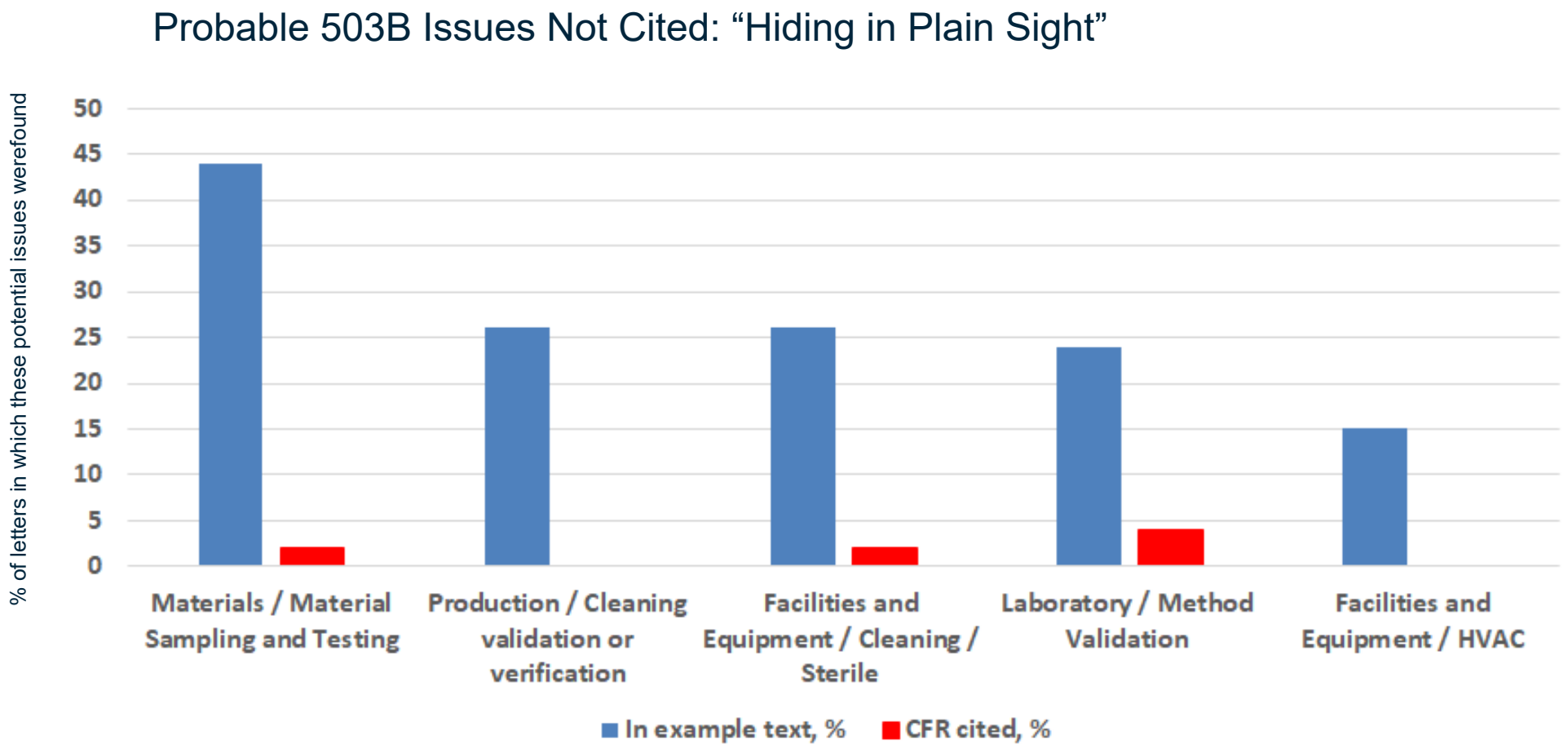


Case Study 1: 503B Warning Letters

Question

When our algorithm is applied to the last five years of drug 503B warning letters, what does it reveal, “hiding in plain sight?”

Case Study 1: Probable 503B Issues Hiding in Plain Sight



Agenda

- Intro: Enabling Quality & Regulatory Professionals
- Data Sources, Machine Learning, NLP, AI Tools
- Building and Applying the Models
- Case Study 1: Sterile Manufacturing
- Case Study 2: Data Integrity
- Summary and Conclusions

Case Study 2: Data Integrity

Case Study 2: Data Integrity

How are data integrity citations found in warning letters?

Hint: Not by searching for “data integrity”

Case Study 2: Data Integrity

DI issues are cited in the area in which they are found – for example:

- **§ 211.100(b)** Written procedures; deviations
- **§ 211.22(d)** Responsibilities of quality control unit
- **§ 211.194(a)** Laboratory records
- **§ 211.68(b)** Automatic, mechanical, and electronic equipment
- **§ 211.180(a)** General requirements (records and reports)
- **§ 211.192** Production record review
- **§ 211.186(a)** Master production and control records
- **§ 211.188** Batch production and control records

Case Study 2: Data Integrity

So, can we just look for these CFR citations?

No. For example:

§ 211.100 Written procedures; deviations.
(b) Written production and process control procedures shall be followed in the execution of the various production and process control functions and shall be documented at the time of performance…

Pertains to not following procedures, which is not a data integrity issue, but this is the one generally cited when documentation is not contemporaneous (“at the time of performance”).

§ 211.22 Responsibilities of quality control unit.
(d) The responsibilities and procedures applicable to the quality control unit shall be in writing; such written procedures shall be followed…

Refers to quality unit procedures not being in writing or not being followed, which is not a data integrity issue, but is the one generally cited if the quality unit is allowing data integrity issues to happen.

Case Study 2: Data Integrity

So, how do you keep track of data integrity issues if they can be cited in so many different ways?

Yesterday (?): An expert who reads through each letter and analyzes it and keeps a running total. Hard copy, highlighter,…

Today: An expert AI system.

Apply what we discussed earlier regarding machine learning,

natural language processing to data integrity.

Let’s look at a few examples our algorithm found…

Case Study 2: Data Integrity

Actual warning letter citations and what the algorithm found in the “specifically” or “for example” text regarding data integrity:

- Your firm failed to maintain adequate written records of major equipment maintenance **(21 CFR 211.182).**
  - Algorithm found: “maintenance work orders in trash bags” “document retrieved from the waste receptacle” “handwritten notes”

- Your firm failed to ensure that each person engaged in the manufacture, processing, packing, or holding of a drug product has the education, training, and experience… **(21 CFR 211.25(a)).**
  - Algorithm found: “form in the trash” “falsification” “manipulation of employee training records”

- Your firm failed to prepare batch production and control records for each batch of drug product that include complete documentation… **(21 CFR 211.188(b)(7)).**
  - Algorithm found: “falsified batch records”

Case Study 2: Data Integrity

Tool applied to the last five years of drug GMP warning letters:

Drug GMP Warning Letters, 2015 -2019 (76 API, 80 Non-Sterile, 49 Sterile)	205
Deficiency keywords / n-grams	4108
Data Integrity keywords / n-grams	911
Warning letters with one or more Data Integrity n-gram	140 (68%)

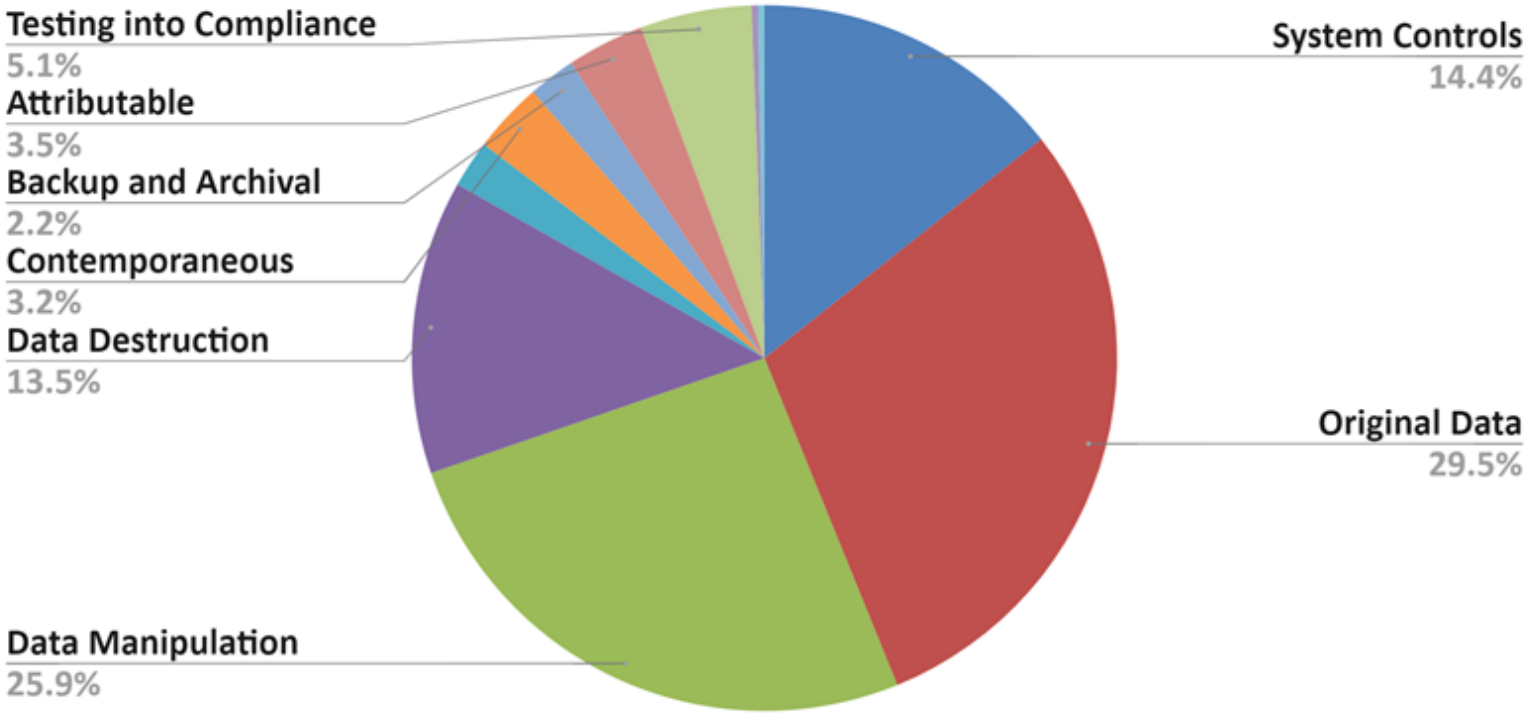
Case Study 2: Data Integrity

Question

What types of data integrity issues did the model find? Remember that Data Integrity is one category area with 11 sub-categories.

Case Study 2: Data Integrity

Data integrity Finding Categorized into the 11 Sub-Topics:



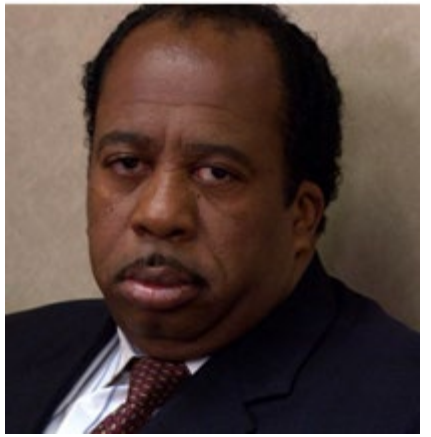
Case Study 2: Data Integrity

Question

Are data integrity issues just found in companies in India and China?

Case Study 2: Data Integrity

No. Data integrity is a global issue.



The ability to run large data sets can produce surprises.

Case Study 2: Data Integrity

Case Study 2: Data Integrity

Question

What happens when we compare the total number of warning letters with the number that contain data integrity issues?

Case Study 2: Data Integrity

Data Set: Countries with 7 or More Drug GMP Warning Letters

Country	Number WLs with DI n-grams	Total Number of WLs	Percent with DI Issues
Canada	10	11	97
India	40	51	78
Japan	5	7	71
USA	9	13	69
China	40	62	65
South Korea	7	14	50

Agenda

- Intro: Enabling Quality & Regulatory Professionals
- Data Sources, Machine Learning, NLP, AI Tools
- Building and Applying the Models
- Case Study 1: Steril Manufacturing
- Case Study 2: Data Integrity
- Summary and Conclusions

Summary and Conclusions

- Presentation title and part of our mission: *Enabling Quality and Regulatory Professionals with Structured Data and Actionable Insights Using AI*
- How do we do that? With **unmatched data sets and proprietary analysis tools**, freeing up experts to interpret what the AI finds and make action plans to fill the gaps
- Discover what agencies are looking for, **even if it is not directly cited**
- These data sets, tools, and more, are **available to you**
- Expect **more surprises/insights** as we examine various data sets with increasingly sophisticated tools, or as you examine your areas of interest

Empowering the champions of quality and safety with actionable data intelligence

www.redica.com

jerry.chapman@redica.com