



**Lisa
Brooks**

Lisa has worked in drug development for almost 30 years and founded Iris Statistical Computing in 2007. In 2007, Lisa was the biometrics project manager on one of the earliest New Drug Applications to attempt a data submission in CDISC format.

Since then, Lisa and a team of consultants have been extensively involved in many licensing applications.

She has interacted directly with FDA reviewers and staff in the Office of Computational Science and eData regarding sample submissions, data standard strategies, and responses to information requests. Lisa has stayed on top of regulation, technology, and standards through constant industry experience, networking, and volunteerism.

Lisa served 5 years as the Collaborations Director for PHUSE where she collaborated with FDA personnel from the Office of Computational Science, Office of New Drugs, CDER and CBER.



**Piper
Brooks**

Piper is a newcomer to drug development and is currently a Clinical Programmer and Document Specialist for Iris Statistical Computing. They developed a comprehensive training program for new hires. Iris' consulting focuses on supporting CDISC data submissions for licensing applications through their "Be Review Ready" services.



Trends in eSub Evaluation Findings: A Twenty Year Perspective on Data Traceability

Lisa Brooks, Founder and Independent Consultant
Piper Brooks, Clinical Programmer
Ann Shiba, eSubmission Reviewer

Iris Statistical Computing, LLC

23-March-2026

Disclaimer:

Although we are often asked about planning data submissions, we are not Regulatory Professionals. This presentation is based solely on our own experience. Data submission decisions are very complicated and require written FDA agreement.

Agenda

Background

Historical Trends

Methodology

Findings

Summary

Questions

Iris Background

2007

The year Iris evaluated their first CDISC data package

2009

The first year Iris submitted an NDA in Pilot CDISC format

356+

The number of CDISC data packages Iris has evaluated

20

The number of NDAs in CDISC format sent to US FDA with Iris review

US FDA History of Data Standards



Endorsed

2003
FDA ENDORSES CDISC STANDARDS

2000s CDISC Publishes Standards

2008 Pinnacle 21 Launched

2014 FDA Publishes CDISC Validation Rules

2014 FDA Publishes Technical Conformance Guide



Mandated

DECEMBER 17, 2016
FDA REQUIRES CDISC FOR CLINICAL STUDIES

2016 FDA Publishes Technical Reject Criteria

2017 FDA Requires eCTD



Enforced

SEPTEMBER 15, 2021
FDA ENFORCES TECHNICAL REJECTION CRITERIA

2023 FDA Requires SEND for Non-clinical Data

2024 FDA Issues Q&A Guidance on Electronic Data (EHR, RWD, Electronic Signatures)

Trends in CDISC Data Packages in last 20 years

Endorsed



Mandated



Enforced



Data Authenticity
and Reliability

End-to-End Data Traceability

Foundational
Standards

Expanded Standards and
Controlled Terminology

Home Grown
Solutions

Digitalization, Interoperability,
and Automation

Expectation:

- Workflow Improves
- Less Variability in Submission Quality
- Faster Delivery
- Fewer Errors

Has it gotten better?

We think it has
gotten **better**

With information at
our ready we could
know if there are
**fewer errors
overtime**

Iris checks CDISC
packages for
**traceability,
readability,
completeness, and
technical
conformance**

We utilize available
**validation
software**, but do
not replicate their
checks

To Know, We Had to Look...

First, we asked our Sponsors if we had their permission to de-identify the comment logs from our reviews and use them in an analysis



I want to give a special shout out to all those who gave us that approval

A large blue graphic of an envelope with a white 'X' across it. The words 'Thank You' are written in white, bold, sans-serif font across the center of the envelope.

Thank You

A yellow starburst graphic with a dark yellow outline. The number '179' is written in white, bold, sans-serif font inside the starburst.

179

In my previous presentations I introduced a few fictitious data scientists:

Today, you will meet a real one!



Charles helped us to see how we can take personal responsibility with clinical data.



Carrie walked us through how Real World Data could be fit for purpose and become Real World Evidence.

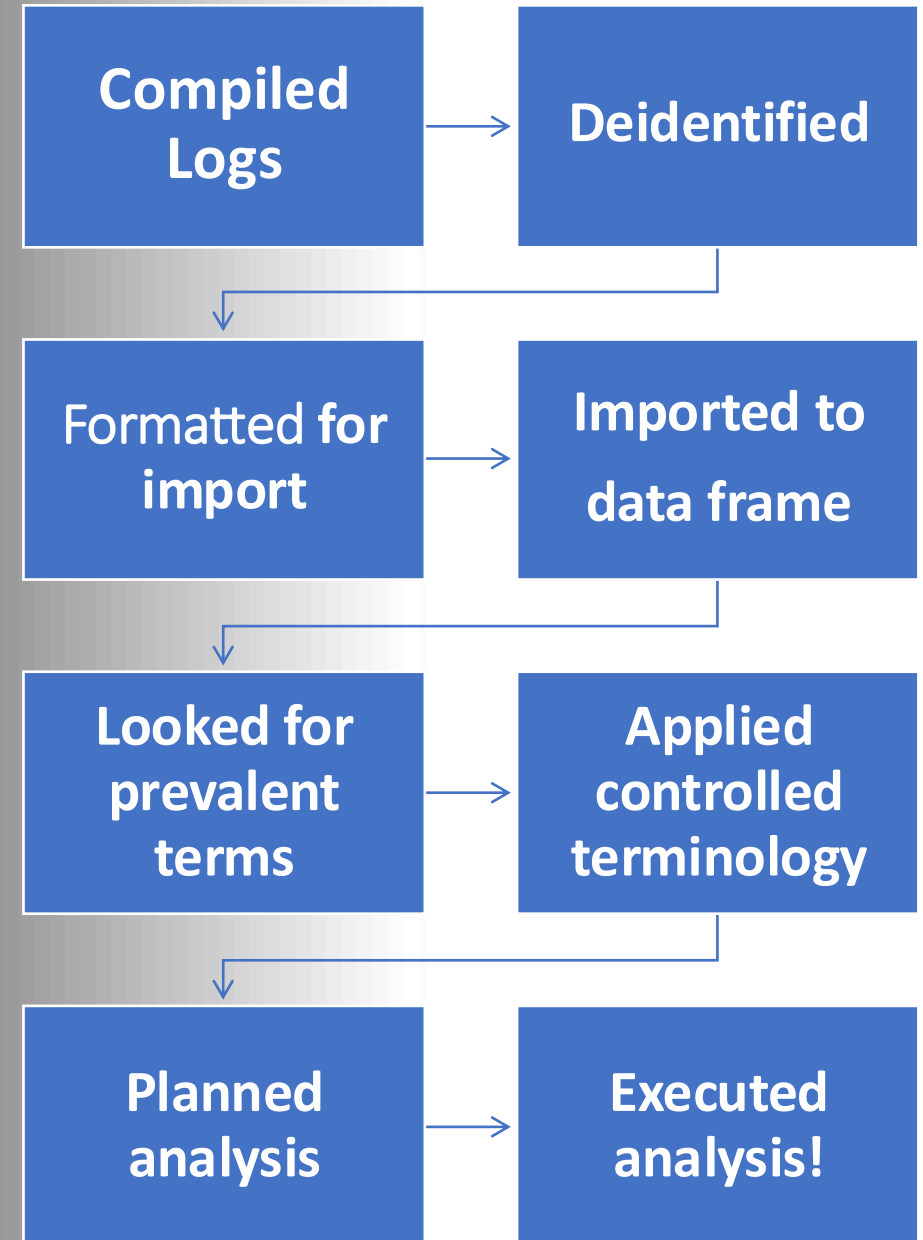
Piper will tell us how they used R and SAS to review tens of thousands of lines of comments to find trends in eSubmission package errors.



Once Piper had a big cup of coffee, they...

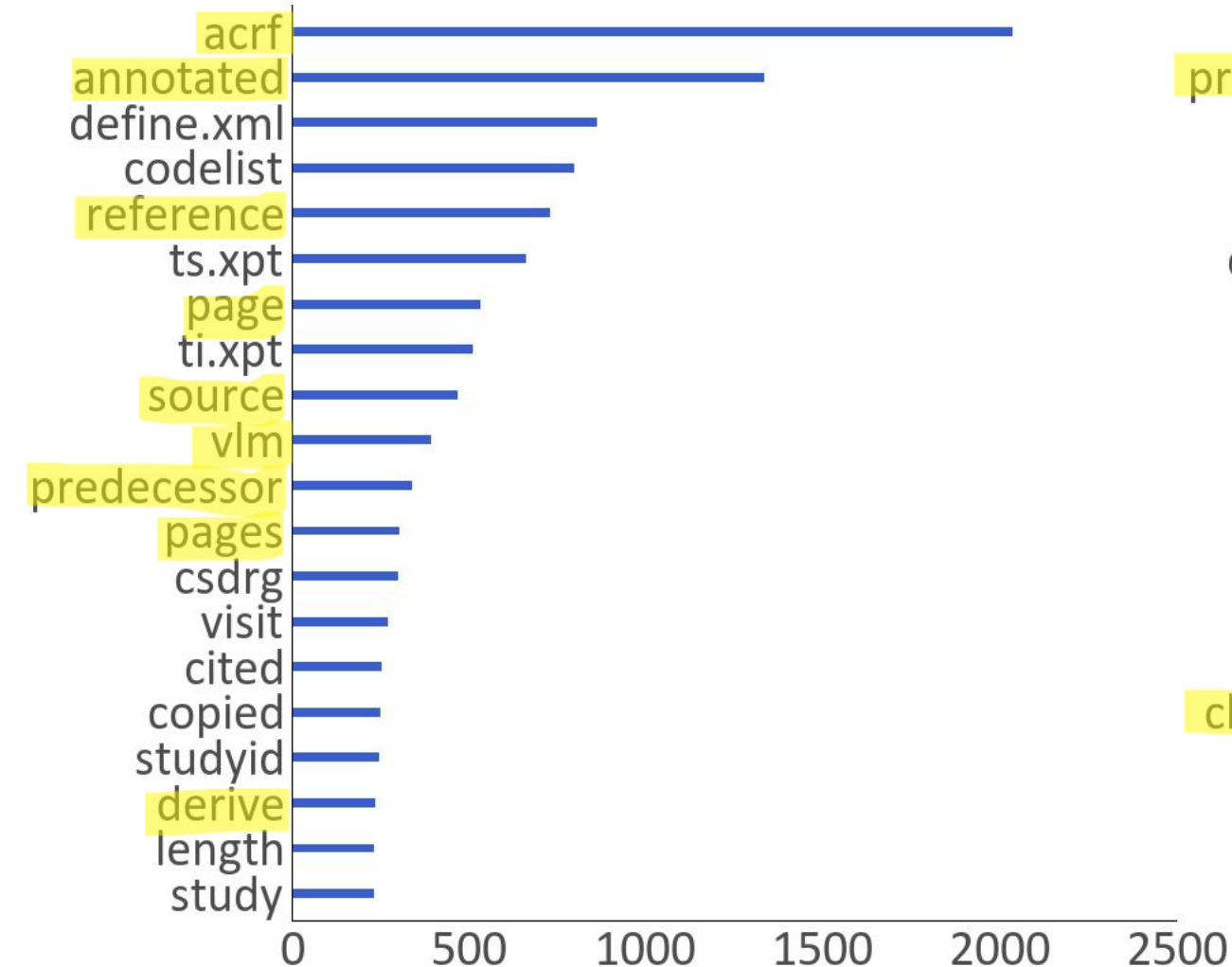


Methodology

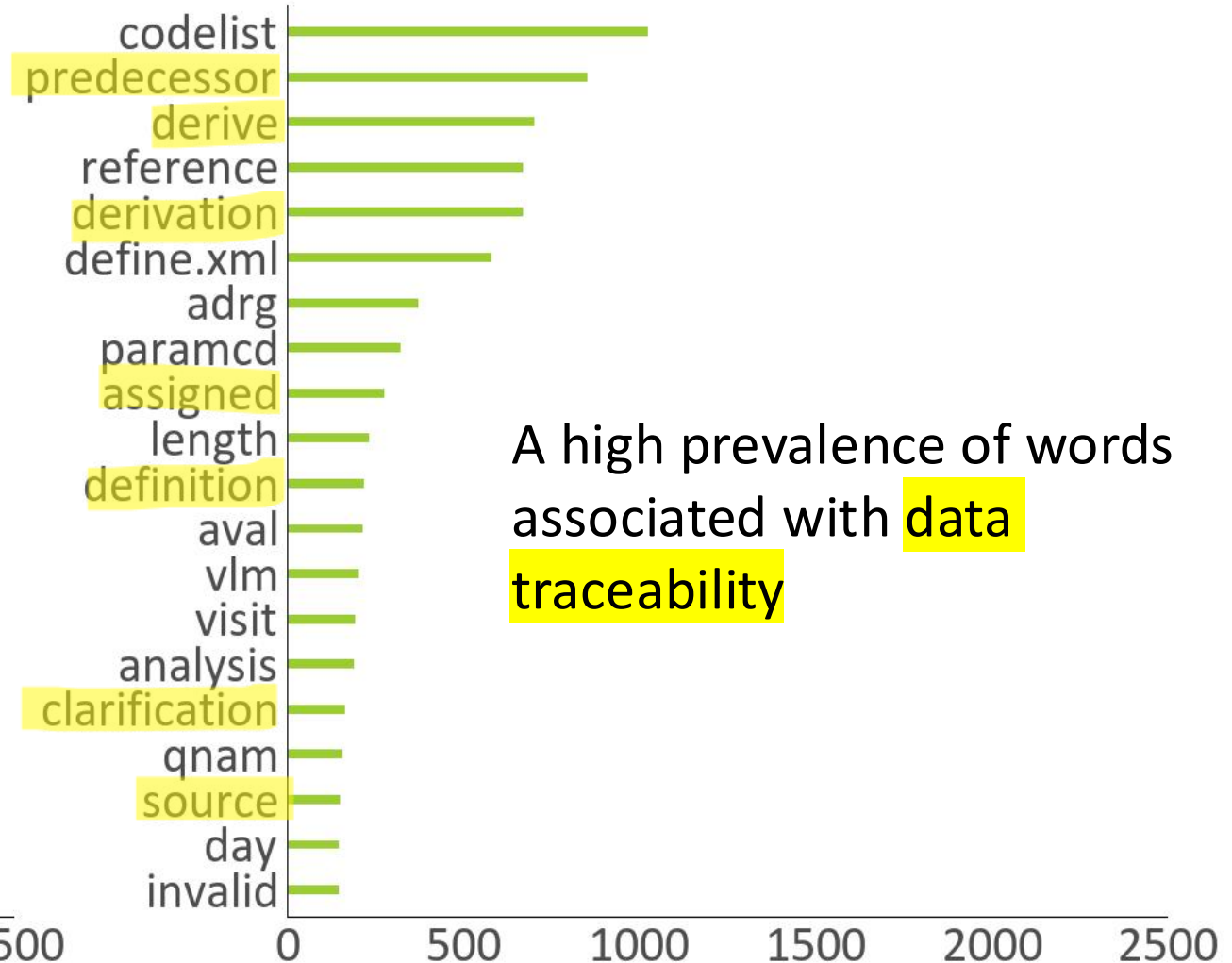


Top 20 Words in SDTM and ADaM Comment Logs

SDTM



ADaM



A high prevalence of words associated with data traceability

An important component of a regulatory review is an understanding of the provenance of the data (i.e., traceability of the sponsor's results back to the CRF data). Traceability permits an understanding of the relationships between the analysis results, analysis datasets, tabulation datasets, and source data. Traceability enables the reviewer to accomplish the following:

Data Traceability



- FDA focused on “**authenticity and reliability**” of electronic records
 - In 1997 with 21 CFR Part 11
 - Later in 2003 for the Scope and Application
- In 2011, FDA emphasized the importance of “**end-to-end Data Traceability**”
 - In *CDER Common Data Standards Issues*
- The importance of “**Data Traceability**” was highlighted in section 8 of the first Study Data Technical Conformance Guide in 2014

Data Traceability Errors In Comment Logs

- Reference to raw or other variables **not in the data submission**
- Reference to variables **without domain**
- **Incorrect use** of 'CRF', 'Derived', 'Assigned', 'Protocol', 'eDT', or 'Predecessor' in origin
- Derivation is **vague, unclear, or incorrect**
- **Multiple origins** or lack of value level metadata (VLM)
- Define reference to CRF page, but **variable not annotated**
- CRF page has variable annotated, but **no reference in define**



Calculation of Error Rate

In clinical research, the denominator for Error Rate would be the total number of populated fields. Here is the basic calculation:

$$\mathbf{ER} = \frac{\mathbf{Number\ of\ Errors\ Found}}{\mathbf{Total\ Number\ of\ Fields}} \times \mathbf{100}$$

For our purposes, the total number of variables is more meaningful as errors in traceability are on a per-variable basis. Here is our modified calculation:

$$\mathbf{ER} = \frac{\mathbf{Number\ of\ Traceability\ Errors}}{\mathbf{Total\ Number\ of\ Variables\ in\ Define.xml}} \times \mathbf{100}$$

We calculated an ER for each SDTM and ADaM package as well as a combined SDTM and ADaM ER.



Analysis Periods for Summarizing Changes in ER over Time

- Comment Logs available for this analysis were:
 - ✓ Reviewed between 2016 and 2025
 - ✓ For studies that started from the 2000's to 2025
- Periods of interest were:
 - ✓ FDA Endorsed CDISC: 2003 - Nov 2016 ✓
 - ✓ FDA Mandated CDISC: Dec 2016 - Aug 2021 ⚖️
 - ✓ FDA Enforced CDISC: Sep 2021 – Present 📁

Study start date was used to group Error Rates into periods of interest



101 Studies Were Analyzed

Criteria for Inclusion:

- Complete SDTM or ADaM package (not SEND or Legacy data)
- define.xml to count the number of variables

Period	# of Studies	#of SDTM Packages	# of ADaM Packages
Endorsed	14	14	3
Mandated	54	54	53
Enforced	33	28	31

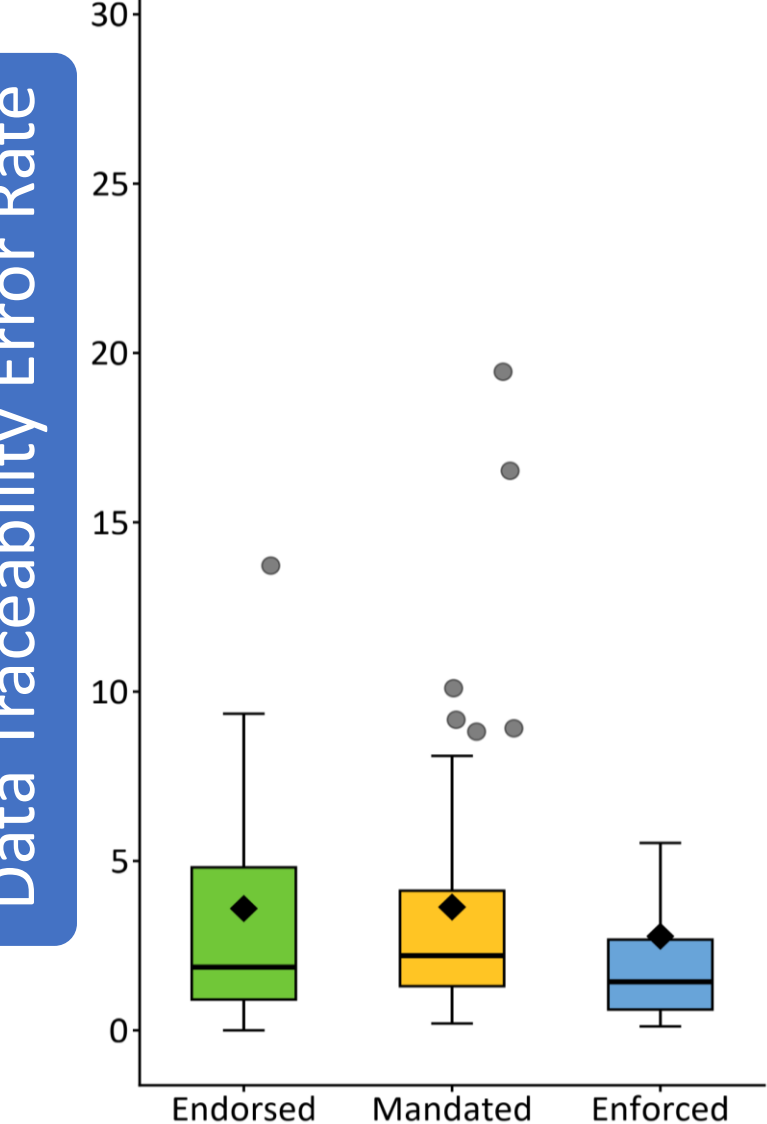


Data Traceability Error Rate

SDTM

ADaM

SDTM and ADaM



N	14	54	28
Mean	3.6	3.6	2.8
95% CI	(1.3, 5.9)	(2.6, 4.7)	(0.5, 5.1)

Endorsed Mandated Enforced

N	3	53	31
Mean	6.2	4.2	2.2
95% CI	(-5.3, 17.8)	(2.8, 5.6)	(1.5, 2.9)

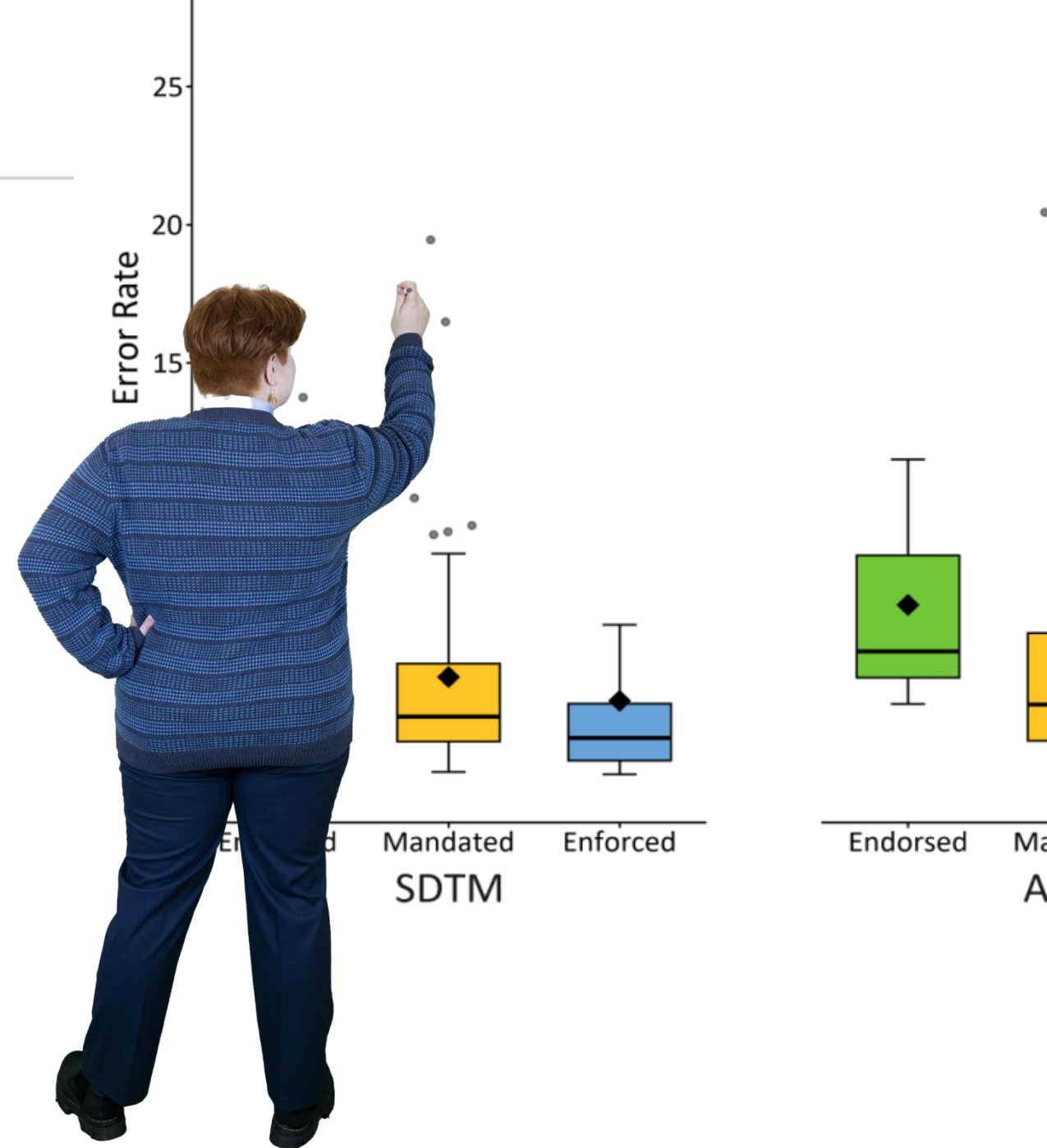
Endorsed Mandated Enforced

N	14	54	33
Mean	3.8	3.8	2.6
95% CI	(1.6, 5.9)	(2.8, 4.8)	(1.5, 3.7)

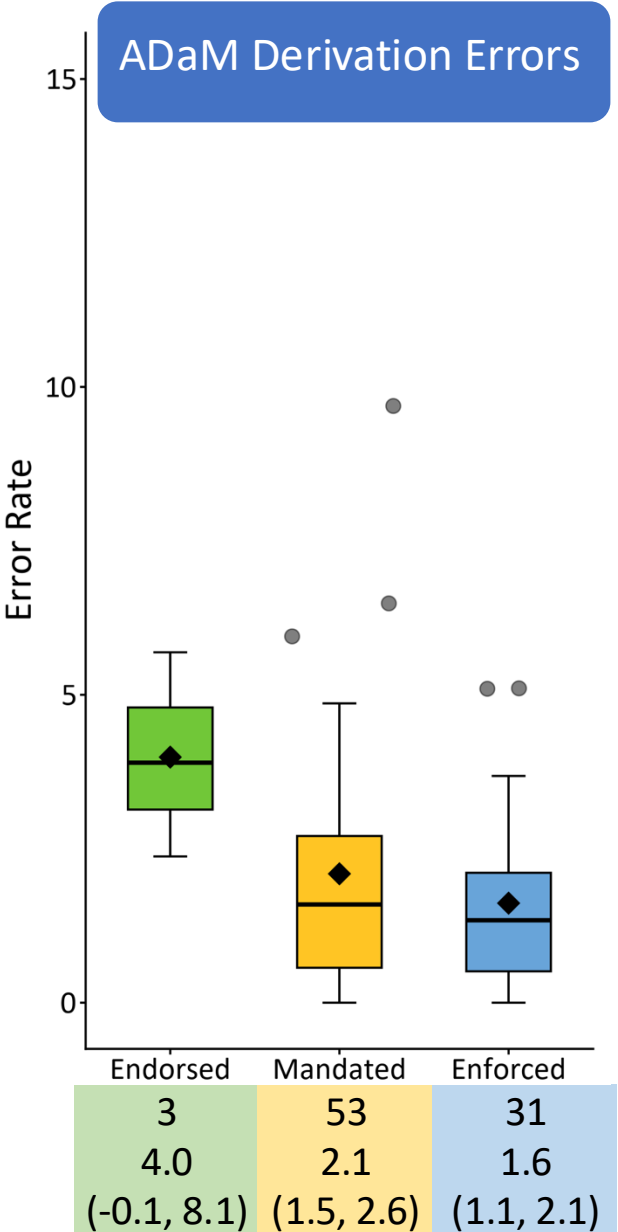
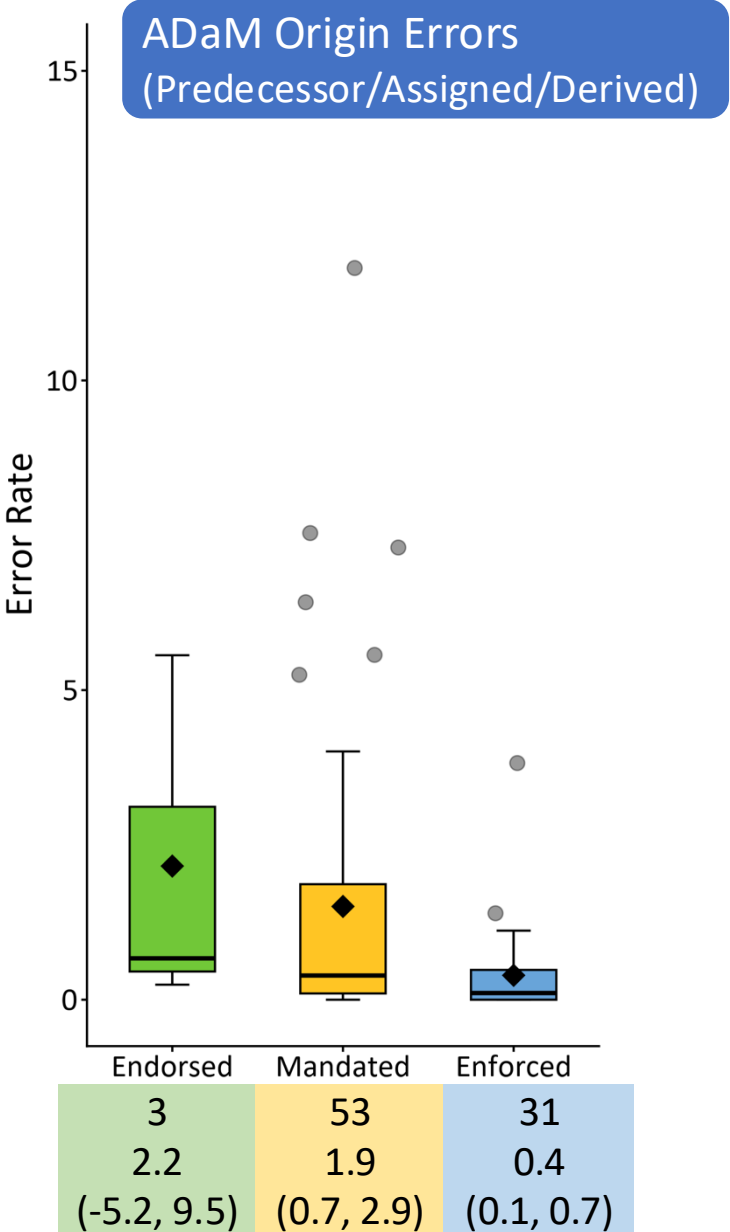
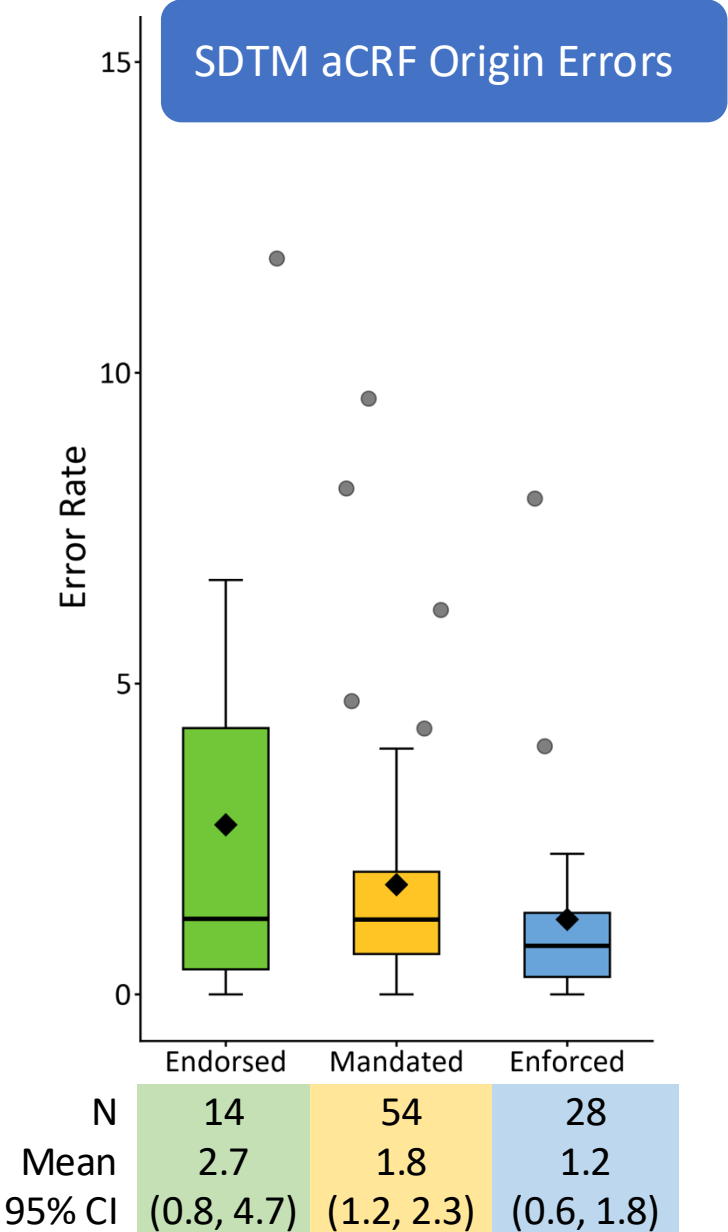
Outliers

Predominant issues contributing to the outliers:

- Gross errors in listing aCRF page numbers in the origin.
- A fundamental misunderstanding of the definition of origin in SDTM and ADaM.
- ADaM computational methods not reviewed by a statistician.



Origin and Derivation Sub-Categories

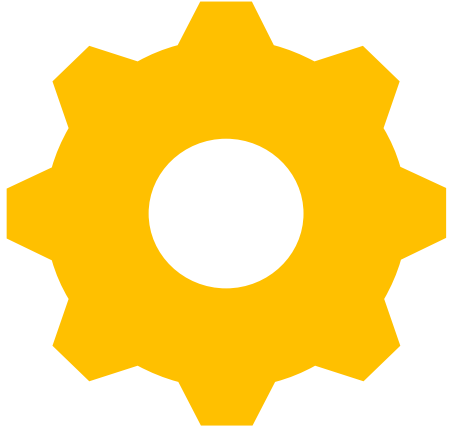


Has the quality of electronic data submissions improved since FDA endorsed CDISC standards?

- The answer is yes. We observed a downward trend in error rates related to data traceability in ADaM packages over time.
- Over the periods of Endorsed, Mandated, and Enforced, our analysis shows that the quality of clinical trial data submissions have improved.
- Following the enforcement of standardized data submission requirements in 2021, the trend is more pronounced.
- Regulatory enforcement of Technical Rejection Criteria, maturing standards, and increased understanding of standards over time created a measurable positive impact on submission quality.



Traceability Issues Persist



- Technical conformance alone is not sufficient
- Validation software detects structural issues
- It cannot fully assess data origins, derivations and relationships that foster reviewer confidence



- Invest in education, process rigor, and multidisciplinary collaboration
- Evaluation of traceability must be built into the quality control process
- Do not leave traceability as a documentation afterthought

What should we do come Monday?



Implement traceability review earlier in our processes



Incorporate more robust quality control around origin definitions



Add a statistical review of computational methods

```
mirror_mod = modifier_ob.  
Set mirror object to mirror.  
mirror_mod.mirror_object  
operation == "MIRROR_X":  
mirror_mod.use_x = True  
mirror_mod.use_y = False  
mirror_mod.use_z = False  
operation == "MIRROR_Y":  
mirror_mod.use_x = False  
mirror_mod.use_y = True  
mirror_mod.use_z = False  
operation == "MIRROR_Z":  
mirror_mod.use_x = False  
mirror_mod.use_y = False  
mirror_mod.use_z = True  
selection at the end -add  
mirror_ob.select= 1  
modifier_ob.select=1  
context.scene.objects.active  
("Selected" + str(modifier_ob.  
mirror_ob.select = 0  
= bpy.context.selected_object  
data.objects[one.name].select  
print("please select exactly  
-- OPERATOR CLASSES --  
types.Operator):  
X mirror to the selected  
object.mirror_mirror_x"  
mirror X"  
context):  
context.active_object is not
```

Due to data privacy issues, comment logs and comment log data are not available as supplemental material.



Questions



Lisa@IrisStatComp.com
Piper@IrisStatComp.com

23-March-2026