

MICHAEL STACKHOUSE BIOGRAPHY

Michael Stackhouse holds a bachelor's degree from Arcadia University where he studied Business Administration, Economics, and Statistics. He is currently a MIDS student at UC Berkeley School of Information, studying Data Science. Mr. Stackhouse has extensive CDISC experience, working with both SDTM and ADaM standards as well as serving as a subject matter expert for Define.xml. Mr. Stackhouse's favorite career projects include macro and utility development, as well as molding the minds of young programmer Padawans into the programming Jedis of the future. Mr. Stackhouse is currently a Manager of Statistical Programming at Covance and father to three young, furry, four-legged children.



FRANKLIN E. MENIUS JR.

BIOGRAPHY

Frank Menius is a graduate of the University of North Carolina at Chapel Hill with a BA in International Studies, and is currently studying for his graduate certificate in Applied Statistics and Data Management at NC State University. Mr. Menius is a former police officer, having served on the line for 12 years prior to fulfilling his lifelong dream of putting his uber computer-geek skills to good use. Mr. Menius is a SAS-certified programmer with extensive experience in data standards and define.xml creation. Mr. Menius is currently an Associate Statistical Programmer at Covance, where his primary interests include training and process improvement.



AN AUTOMATED MACRO TO COMPARE DATA TRANSFERS

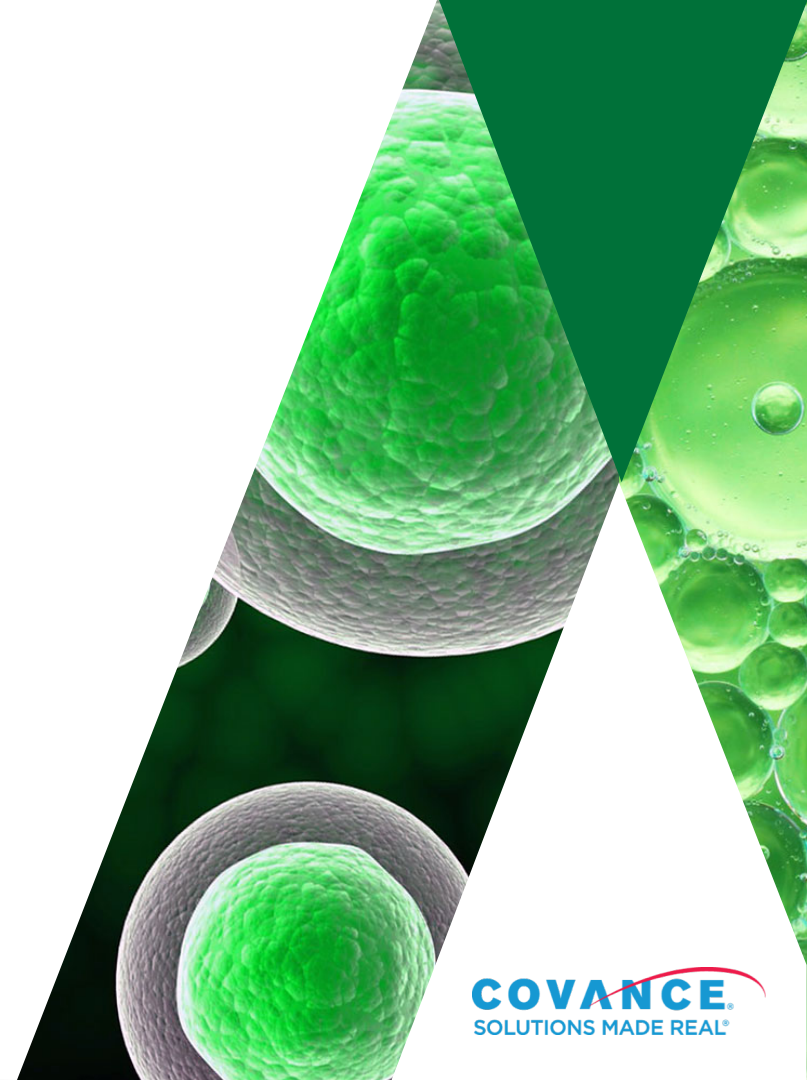
Michael Stackhouse, Covance, Inc.

Franklin E. Menius Jr., Covance, Inc.



Copyright © 2018 Covance. All Rights Reserved

Public

A large, stylized graphic on the right side of the slide. It features a large green triangle pointing downwards, which is partially filled with a pattern of green and white bubbles. Below the triangle, there are several large, textured green spheres, some of which are also filled with a bubble pattern. The background of the slide is white, and the overall design is clean and modern.

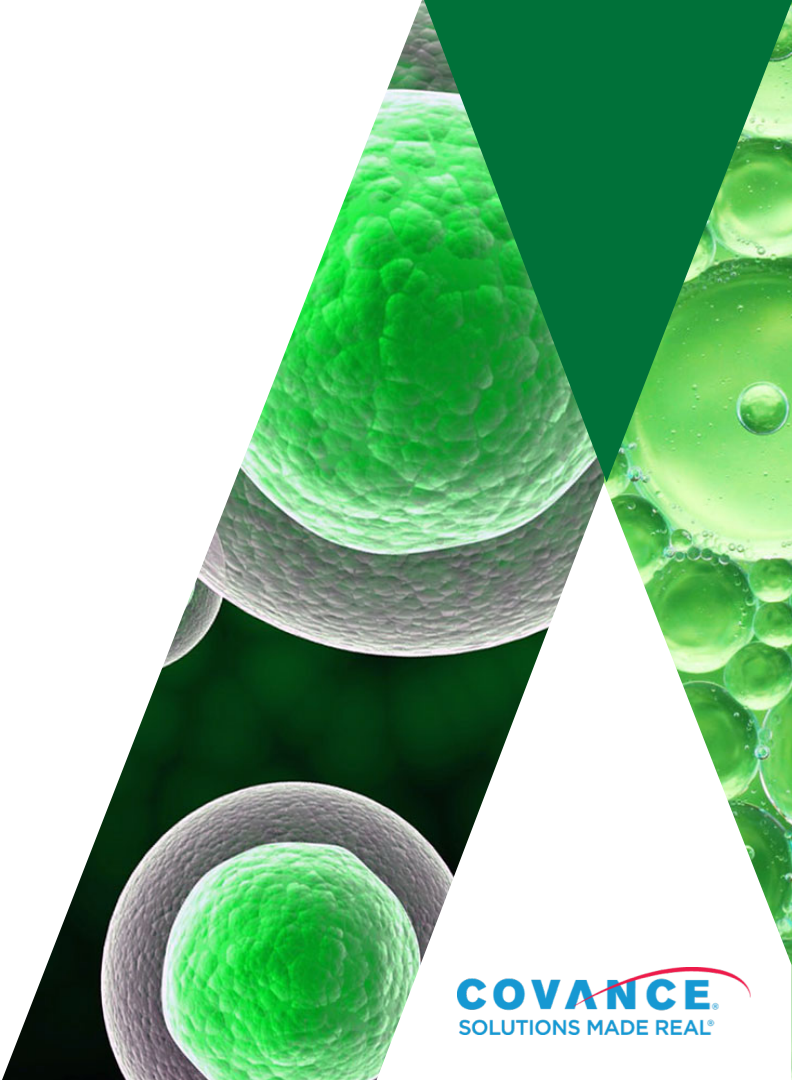
COVANCE
SOLUTIONS MADE REAL®

INTRODUCTION



Copyright © 2018 Covance. All Rights Reserved

Public

A large, stylized graphic on the right side of the slide. It features a large green triangle pointing downwards, which is partially filled with a pattern of green and white spheres, resembling a molecular structure or a cluster of cells. The background of the slide is white, and the graphic is set against a dark green background.

COVANCE
SOLUTIONS MADE REAL®

Data Transfers Happen;

THE NEED FOR PROACTIVE ISSUE IDENTIFICATION

- ▶ Routine data transfers throughout the course of an ongoing study are a necessity for any analysis.
- Issues in tables trace back to SDTM
- Example: Issue found during 4th DMC production where AE counts for subject went **down** compared to last year's delivery!
- What should we do?



What should we do?

► PROC COMPARE:

- Great tool, but output will be extensive for comparing different cuts of data

► Our Macro:

- Offers a proactive approach to efficiently identify similar issues by reviewing three different extractions on three different levels
 - The datasets available *within* the package
 - The metadata *within* each dataset
 - The count of records per subject *within* each dataset



Ease of Use

- ▶ Goal of development was ease of use. The user only needs to identify three parameters, with a fourth optional parameter.
 - The directory containing the old data
 - The directory containing the new data
 - The directory for the output report
 - Additional subject identifiers in your data (optional)

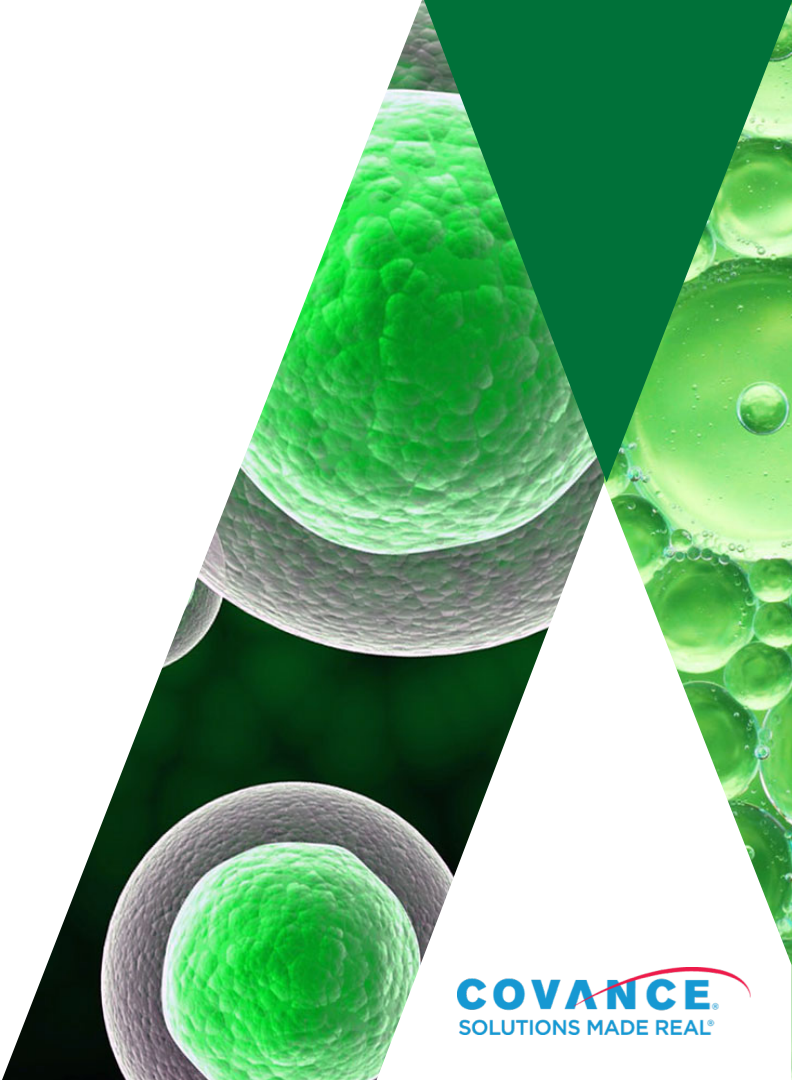


DATASET LEVEL REVIEW



Copyright © 2018 Covance. All Rights Reserved

Public

A large, stylized graphic on the right side of the slide. It features a large green triangle pointing downwards, which is partially filled with a pattern of green and white spheres, resembling a molecular structure or a cluster of cells. The background of the slide is white, and the overall design is clean and professional.

COVANCE
SOLUTIONS MADE REAL®

Dataset Level Review

- ▶ The first check is whether any datasets have been removed or added from one extraction to another.
 - Are there new datasets?
 - Were datasets left out of this transfer?
 - Were the current datasets received under a new name?



Dataset Level Review:

WHAT THE MACRO DOES

- ▶ Finds the directories for each extraction and builds a dataset containing all *sas7bdat* files
- ▶ Completes merge, outputting list of datasets:
 - In old data extraction but not in new
 - In new data extraction but not in old
 - This will also include those datasets that have been renamed



Dataset Level Review:

EXAMPLE REPORT

These datasets are in the OLD cut of data and not in the NEW.

name
ex

These datasets are in the NEW cut of data and not in the OLD.

name
exrename
vs

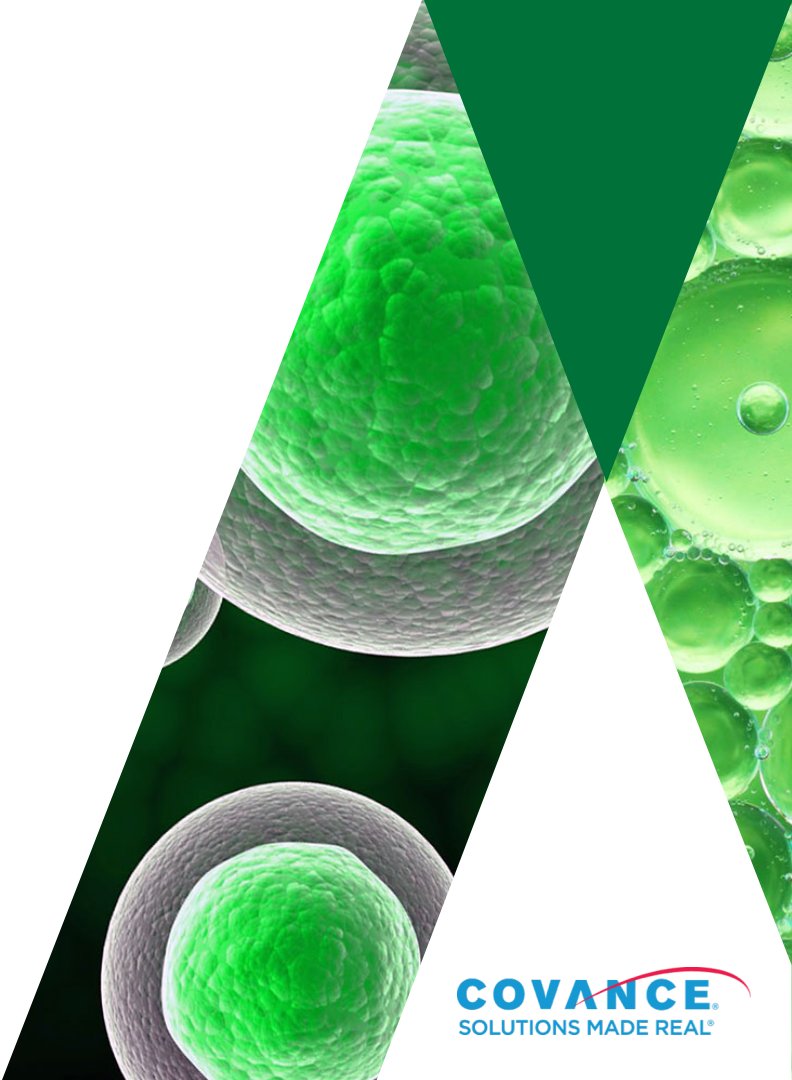


METADATA LEVEL REVIEW



Copyright © 2018 Covance. All Rights Reserved

Public

A large, stylized graphic on the right side of the slide. It features a large green triangle pointing downwards, which is partially filled with a pattern of green and white bubbles. Below the triangle, there are several large, textured green spheres, some of which are also filled with a bubble pattern. The background of the graphic is a dark green gradient.

COVANCE
SOLUTIONS MADE REAL®

Metadata Level Review

- ▶ Now the macro compares matched datasets for metadata changes of interest
 - Were variables added?
 - Were variables dropped?
 - Did the variable type change?
 - Did the variable length change?
 - ▶ (other attributes, including formats, can be added to the comparison but these have the largest implications for programmers)



Metadata Level Review:

WHAT THE MACRO DOES

- ▶ Assembles the subset of metadata needed from the SASHELP.VCOLUMN dataset pulling in data where LIBNAME in("NEW", "OLD").
- ▶ Iterates over every dataset that was in both OLD and NEW and compares the metadata.
- ▶ Creates two checks
 - Variables added and dropped (NAME was in OLD and not in NEW and vice versa)
 - TYPE or LENGTH did not match



Metadata Level Review:

EXAMPLE REPORT

*Differences in metadata between OLD.DE and NEW.DE.
Variables in NEW not in OLD.*

Column Name
DEORRES

*Differences in metadata between OLD.DS and NEW.DS.
Differences in variables.*

Column Name	Column Type	Column Length	Column Type	Column Length
DSSTDTC	char	8	num	8

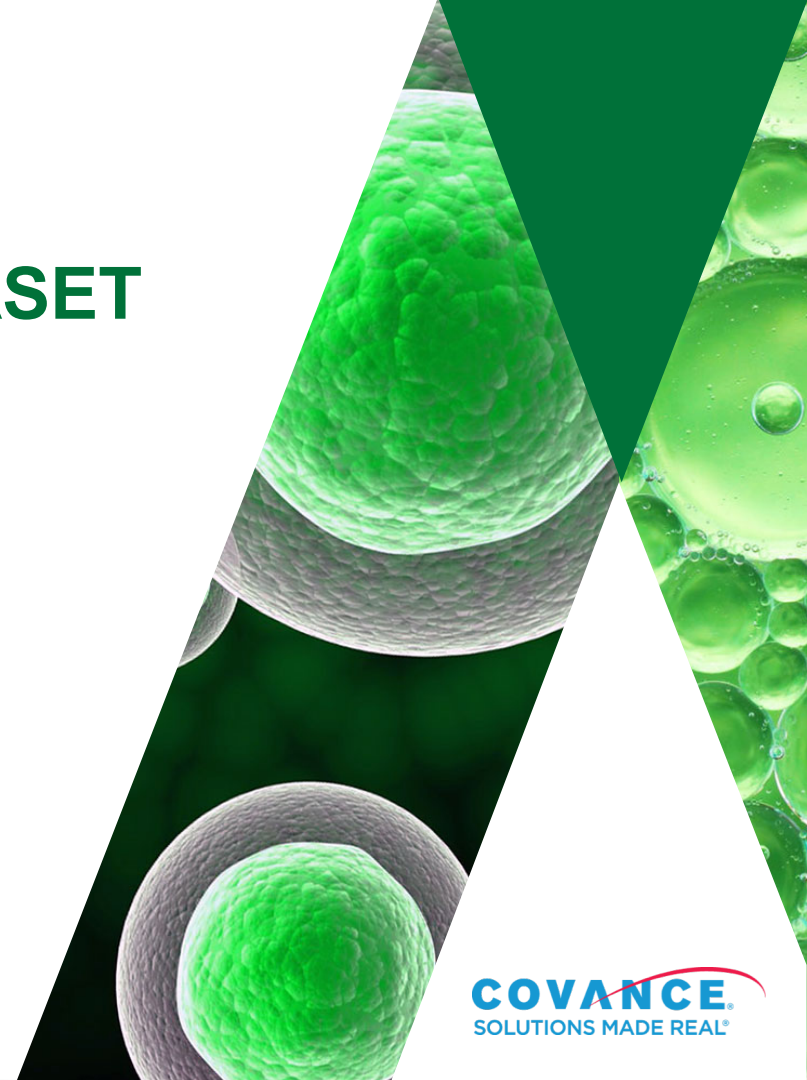


SUBJECT COUNTS PER DATASET



Copyright © 2018 Covance. All Rights Reserved

Public

A large, abstract graphic on the right side of the slide. It features a dark green triangle pointing downwards, which is partially filled with a lighter green, textured, spherical shape. To the right of this triangle is a vertical strip showing a close-up of green, bubbly liquid, possibly representing a cell or a chemical reaction.

COVANCE
SOLUTIONS MADE REAL®

Subject Counts Per Dataset

- ▶ Next the macro compares the number of subject counts per dataset to identify data loss.
 - Examples where this could occur:
 - A data extraction where an issue resulted in one subject inadvertently being removed from all CRF datasets.
 - A data extraction between Q1 and Q3 DMC where 14 AEs were removed for one subject without obvious explanation



Subject Counts Per Dataset:

WHAT THE MACRO DOES

- ▶ If the program can establish a subject identifier, it runs frequencies using that identifier and merges the datasets.
- If the **old** count (previous extraction) is GREATER THAN the **new** count (new extraction), then the record is flagged and output.



Subject Counts Per Dataset:

WHAT THE MACRO DOES (CONT.)

- ▶ Determines the subject identifier using a set of potential candidates:
 - PT, SUBJECT_NUMBER, SUBNUM, SUBJECT, USUBJID, SUBJID.
- ▶ If a subject identifier needs to be added to this list, the user can update the %let addsubj=; parameter at the beginning of the code.
 - %let addsubj= CHECK1 CHECK2 CHECK3;
- ▶ If the program does not find a matching identifier it will generate a warning in the log:
 - WARNING: SUBJECT IDENTIFIER NOT COMPENSATED FOR - COULD NOT FIND IDENTIFIER WITHIN THE LIST *<list of variables>* IN THE DATASET *<dataset being reviewed>*.



Subject Counts Per Dataset:

EXAMPLE REPORT

These subjects in NEW.AE have less records than in OLD.AE.

Subject Identifier for the Study	Frequency Count	Frequency Count
7802	4	.

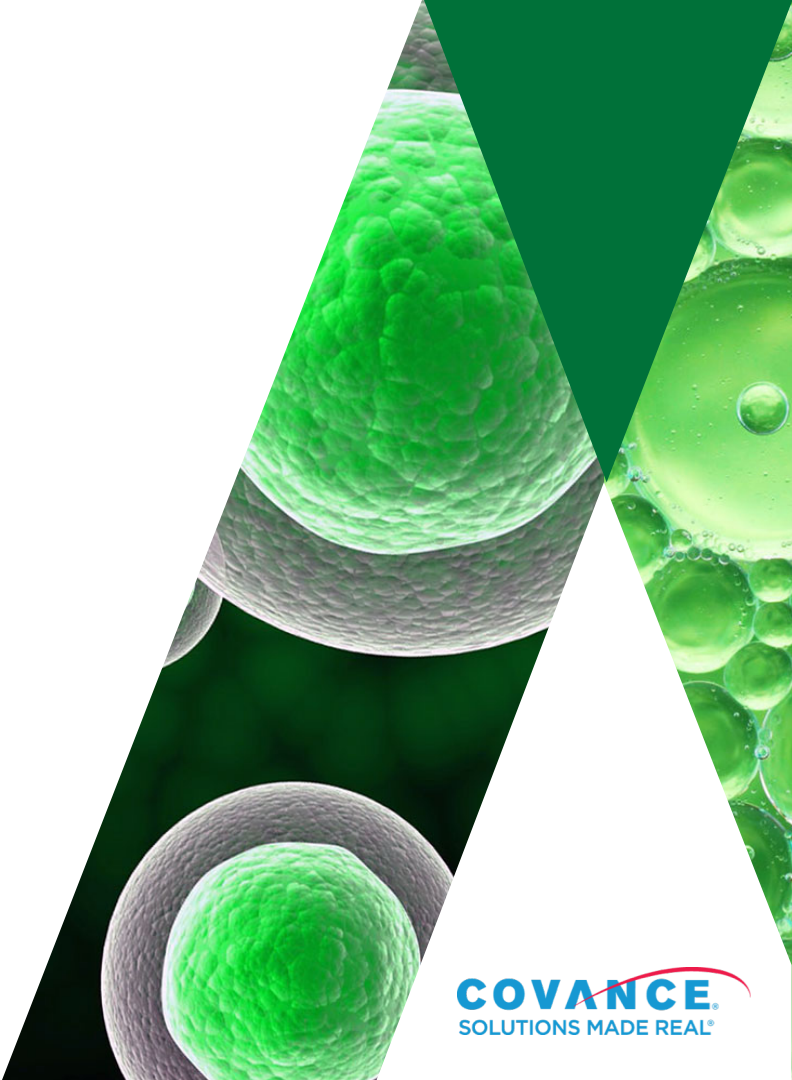


POTENTIAL LIMITATIONS



Copyright © 2018 Covance. All Rights Reserved

Public

A large, stylized graphic on the right side of the slide. It features a large green triangle pointing downwards, which is partially filled with a pattern of green and white bubbles. Below the triangle, there are several large, textured green spheres, some of which are also filled with a bubble pattern. The background of the graphic is a dark green gradient.

COVANCE
SOLUTIONS MADE REAL®

Potential Limitations

► Potential of false positives

- As data is cleaned some records may be determined appropriate to delete, this macro will continue to flag them
- Advised to confirm with data management that these records should indeed have been removed.



Potential Limitations (cont.)

- ▶ Flexibility and portability were prioritized, so not all issues can be caught.
 - Old transfer: 10 records
 - Subject added: 5 records
 - Subject dropped: 2 records
 - New transfer: 13 records
- ▶ To Limit this situation it is advised to run the program at regular intervals through out the life of the study.

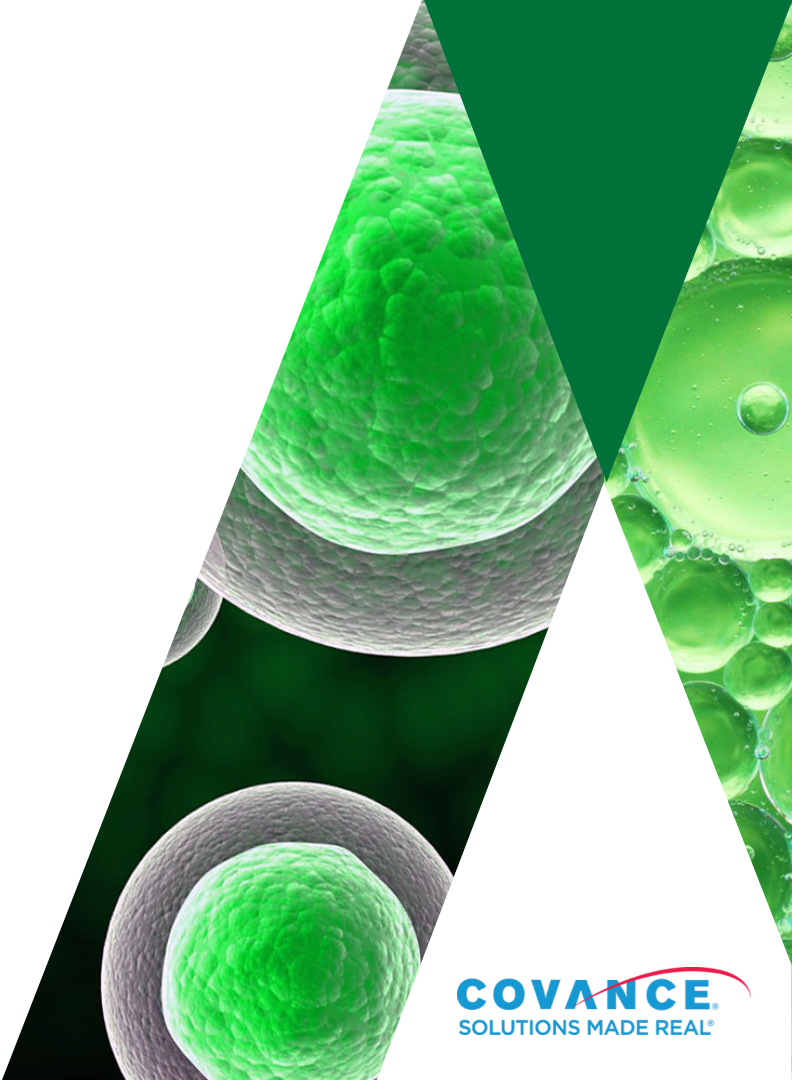


CONCLUSION AND Q&A



Copyright © 2018 Covance. All Rights Reserved

Public

A large, stylized graphic on the right side of the slide. It features a large green triangle pointing downwards, which is partially filled with a close-up image of a green, textured sphere (resembling a virus or cell). To the right of this triangle is a vertical strip showing a close-up of green bubbles. Below the triangle is a smaller green sphere, also with a textured surface.

COVANCE
SOLUTIONS MADE REAL®

Conclusion and Q&A

- ▶ Data extractions are common and can cause changes to the existence, structure, or attributes of source data.
- ▶ It is important to proactively and efficiently detect those changes as soon as possible
- ▶ This macro has the potential to save programmers time and effort on rework when ran at regular intervals
- ▶ Though designed to be portable and dynamic small changes can be made to address specific needs of the programmer or study



Conclusion and Q&A

► Questions?



COVANCE[®]

SOLUTIONS MADE REAL[®]

Covance Inc., headquartered in Princeton, NJ, USA, is the drug development business of Laboratory Corporation of America Holdings (LabCorp). COVANCE is a registered trademark and the marketing name for Covance Inc. and its subsidiaries around the world.



Copyright © 2018 Covance. All Rights Reserved

Public

Name: Michael Stackhouse
Organization: Covance, Inc.
Address: 4000 Centergreen Way # 300
City, State ZIP: Cary, NC 27513
E-mail: Michael.Stackhouse@Chiltern.com
Web: <https://www.linkedin.com/in/michael-s-stackhouse/>

Name: Frank Menius
Organization: Covance, Inc.
Address: 4000 Centergreen Way # 300
City, State ZIP: Cary, NC 27513
E-mail: Frank.Menius@Chiltern.com
Web: <https://www.linkedin.com/in/frank-menius/>