PHARMACEUTICAL ONLINE

CLEANING VALIDATION FOR THE 21ST CENTURY

A collection of articles on the application of science, risk and statistics to cleaning

ANDREW WALSH
President, Center for Pharmaceutical Cleaning Innovation

FROM THE AUTHOR

This eBook is a collection of articles written from May of 2017 through December 2019 and are part of the "Cleaning Validation for the 21st Century" Series. The purpose of the series is to provide the reader with deeper insight into using the science-, risk-, and statistics-based approaches to cleaning and cleaning validation that are found in the ASTM E3106-18 Standard Guide for Science-Based and Risk-Based Cleaning Process Development and Validation. The series started in 2011 with the publication of "Cleaning Validation for the 21 Century: Acceptance Limits for Active Pharmaceutical Ingredients (APIs): Part I and Part II," which reviewed the history of setting cleaning validation acceptance limits and introduced the more scientifically justified approach of using the acceptable daily exposure (ADE) limit, now more formally known as the health-based exposure limit (HBEL).

This collection of articles from the series starts with a review of the drivers, both regulatory and industry-based, for the pharmaceutical industry's movement towards science- and risk-based approaches to GMP compliance and how they apply to cleaning. The next four articles discuss the creation of data-derived scales for toxicity, process capability, and method detectability that can be used in the risk assessment of cleaning processes. More specifically, these scales can be substituted for severity, probability, and detectability in FMEAs/FMECAs of cleaning processes. These articles are followed by two that scientifically and statistically explore the potential of visual inspection for use in cleaning validation. The next article describes how the previous articles can be used together to measure and evaluate the risk involved in cleaning. It also explains how the scales in these articles can be used to objectively develop a cleaning control strategy based on actual data — and how they can be combined into a cleaning risk dashboard to visualize the level of cleaning risk. The final article shows how the Toxicity Scale and Process Capability Scale can be combined into a Shirokizawa Matrix that provides a clear visual guide for adjusting the level of effort, formality, and documentation for cleaning validation based on the level of risk as described in ICH Q9.

These articles are the result of the combined efforts of a global team of cleaning validation subject matter experts, pharmaceutical toxicologists, statisticians, and Six Sigma professionals supported by a global peer review team of cleaning validation stakeholders (from Denmark, Germany, Greece, Indonesia, Ireland, Italy, Japan, Malta, Malaysia, Mexico, Spain, Tunisia, United States, and Vietnam), who provided many invaluable comments, insights, and corrections that improved the utility of these articles for readers around the world. I am personally deeply grateful to all of these people for this accomplishment. On behalf of all of us, we sincerely hope that readers find these articles useful in their professional work.

TABLE OF CONTENTS

DEVELOPING A SCIENCE-, RISK-, & STATISTICS-BASED APPROACH TO CLEANING PROCESS DEVELOPMENT & VALIDATION

THOMAS ALTMANN
ALFREDO CANHOTO
MICHEL CREVOISIER
IGOR GORSKY
ROBERT KOWAL
MARIANN NEVEROVITCH
MOHAMMAD OVAIS
OSAMU SHIROKIZAWA
ANDREW WALSH

Cleaning manufacturing equipment to prevent cross contamination of pharmaceutical products is a fundamental aspect of GMPs. Validation of cleaning processes has been required within cGMP industries for a long time and is recognized as an important activity to establish that product cross contamination is controlled to ensure patient safety and product quality.

While cleaning, in and of itself, is a relatively simple process, the pressures of inspection scrutiny and the reactionary programs created by industry to address regulatory concerns have transformed the validation of cleaning into a complex, expensive, and time-consuming activity. From a simple project management analysis, the time that would be required to perform cleaning validation for a facility with multiple products, multiple pieces of equipment, and multiple cleaning procedures can easily run into years.

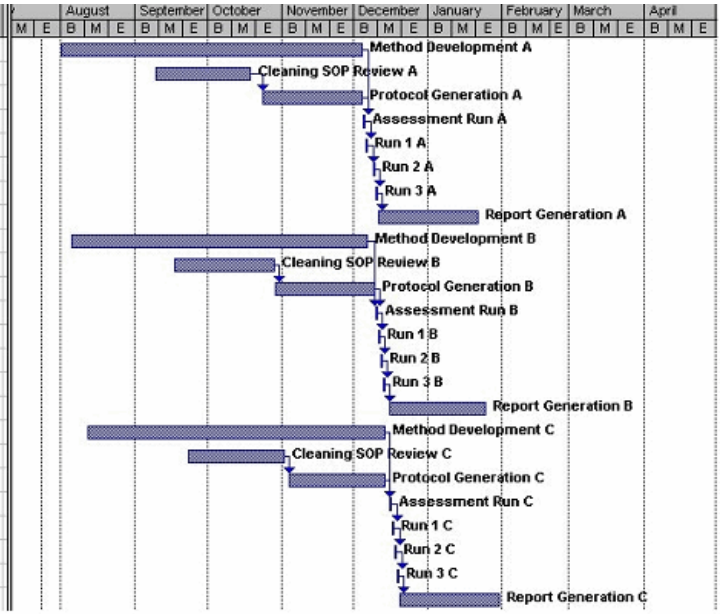
Figure 1 shows a very aggressive, and a clearly hard to accomplish, timeline for concurrently performing three runs for only three cleaning validations allowing only one day between runs and ignoring analysis time and other items. Despite this, the timeline is still six months.

Considering that cleaning validation runs cannot be scheduled and performed every day and the need for method development, protocol development, laboratory analysis,

Figure 1:

Hypothetical timeline for cleaning validation of three products

	Task Name	Duration	Start
1	Method Development A	90 days	Thu 8/1/02
2	Cleaning SOP Review A	30 days	Mon 9/9/02
3	Protocol Generation A	30 days	Thu 10/24/02
4	Assessment Run A	1 day	Thu 12/5/02
5	Run 1 A	1 day	Fri 12/6/02
6	Run 2 A	1 day	Mon 12/9/02
7	Run 3 A	1 day	Tue 12/10/02
8	Report Generation A	30 days	Wed 12/11/02
9	Method Development B	90 days	Mon 8/5/02
10	Cleaning SOP Review B	30 days	Tue 9/17/02
11	Protocol Generation B	30 days	Tue 10/29/02
12	Assessment Run B	1 day	Tue 12/10/02
13	Run 1 B	1 day	Wed 12/11/02
14	Run 2 B	1 day	Thu 12/12/02
15	Run 3 B	1 day	Fri 12/13/02
16	Report Generation B	30 days	Mon 12/16/02
17	Method Development C	90 days	Mon 8/12/02
18	Cleaning SOP Review C	30 days	Mon 9/23/02
19	Protocol Generation C	30 days	Mon 11/4/02
20	Assessment Run C	1 day	Mon 12/16/02
21	Run 1 C	1 day	Tue 12/17/02
22	Run 2 C	1 day	Wed 12/18/02
23	Run 3 C	1 day	Thu 12/19/02
24	Report Generation C	30 days	Fri 12/20/02



and report writing, it is clear that cleaning validation consumes a considerable amount of resources.

Consequently, companies have made various efforts to reduce cleaning activities, such as dedicating equipment or converting to disposable items, but these strategies have their own inefficiencies and costs. Companies have also resorted to strategies such as product grouping, equipment grouping, matrixing, and bracketing to reduce the amount of cleaning they validate, sometimes without acceptable justification. Many companies today validate the cleaning of only one or two "hardest-to-clean" products, selecting them based on the solubility of the API or because the calculated limit is lowest, even though these may not be truly justifiable criteria. Even with such efforts, part of the reality has been that, for all intents and purposes, cleaning validation never seems to be completed. The EU guideline "Good Manufacturing Practice for Medicinal Products for Human and Veterinary Use," Annex 15 outlines in Section 10 that for a worst case product approach a scientific rationale should be provided.[2] It also outlines criteria for determining the worst case, and these criteria

may include solubility, cleanability, toxicity, and potency. This approach makes determining a worst case situation even more complex.

As with many things, the pharmaceutical industry has tended to understand cleaning almost entirely in its relation to regulatory expectations. In particular, cleaning has become closely associated with "process validation." In the late 80s/early 90s, the FDA, as well as other regulatory agencies, began to view cleaning as a process that needed to be validated.[3] At the same time, several legal decisions concerning cleaning made during the resolution of the well-known Barr Labs case solidified this viewpoint.[4] Consequently, companies set about validating existing cleaning procedures without questioning whether the procedures were the most effective or optimal, or even if they were using an appropriate cleaning agent. The cleaning procedures that were subsequently validated may not have been the best choice for their situations.

Cleaning validation incorporates the traditional preapproved protocol, with predetermined acceptance criteria and a three-run process validation approach. Because of the traditional approach, the industry also struggled over how to set the required predetermined acceptance criteria. This process validation approach was adopted without ever asking if three cleaning validation runs and predetermined acceptance criteria were appropriate for the validation of cleaning. Based on that reason, Annex 15 also outlines in Section 10 that the cleaning procedure should be evaluated an appropriate number of times (based on a risk assessment) and meet the acceptance criteria in order to prove that the cleaning procedure is validated.

All these issues underscore the need for effective and efficient cleaning programs that focus efforts and resources where they provide the most value.

Since 2001, there have been many new, and for this highly conservative industry, radical movements from both regulators and within the industry itself. Examples coming from the FDA include "GMPs for the 21st Century,"[5] quality by design (QbD),[6] process analytical technology (PAT),[7] and the agency's 2011 guideline on process validation.[8] Globally, the new International Conference on Harmonisation's guidelines, in particular Q8 and Q9,[9] are major forces driving change in the industry. Movements within pharmaceutical manufacturing itself include lean manufacturing, Six Sigma, and operational excellence (OpEx), which have grown out of the pressures to reduce costs and to better

supply the market. These “planets” have aligned to create a tide drawing the industry toward science-based, risk-based, statistics-based, and cost-effective approaches to ensuring patient safety and product quality during pharmaceutical development and manufacturing. As one of the critical processes in manufacturing, cleaning and its validation can benefit from all these initiatives.

The introduction of the acceptable daily exposure (ADE) in 2010 provided a tool that could be used for setting science-based acceptance criteria for the cleanliness of equipment.[10] Several subsequent publications have revealed how replacing the traditional approaches to setting acceptance criteria with an approach based on the ADE leads to better patient safety and can reduce the validation effort for lower-risk situations,[11-14] and a recent publication discussed how an ADE-derived scale can be used to easily and visually evaluate the risks of cross contamination in manufacturing facilities, including for cleaning.[15]

Cleaning validation programs and master plans could benefit from a risk-based approach to their design and management. Cleaning procedures could benefit through a statistics-based QbD approach resulting in safer and more reliable procedures, and the analytical methods used in cleaning could benefit from PAT, resulting in faster turnaround of equipment. Many of the techniques used in lean manufacturing, Six Sigma, and operational excellence could be used to reduce the time and effort spent, improve the results obtained during cleaning validations, and provide statistics-based means for evaluating and controlling cleaning processes. Perhaps even cleaning, which is certainly a process, should be looked at and evaluated in the manner being suggested in the FDA’s 2011 process validation guidance; indeed, the FDA believes that this guidance is applicable to cleaning.[16]

The authors believe focusing industry efforts where the risks are high will increase patient safety and reducing efforts where the risks are low will ease the regulatory burden on industry and improve operational efficiencies overall.

REGULATIONS AND CURRENT GUIDANCE AND THEIR APPLICATION TO CLEANING

This section explores in more detail how the regulations and current guidance mentioned above provide direction on how to implement these approaches to cleaning.

The requirements in 21 CFR 211.67(a) state that "*Equipment and utensils shall be cleaned, maintained, and sanitized 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.*"[17]

Similarly, 21 CFR 111.27(d) states "*You must maintain, clean, and sanitize, as necessary, all equipment, utensils, and any other contact surfaces used to manufacture, package, label, or hold components or dietary supplements.*"[18]

21 CFR 820.70(e) also states "*Contamination control. Each manufacturer shall establish and maintain procedures to prevent contamination of equipment or product by substances that could reasonably be expected to have an adverse effect on product quality.*"[19]

From these statements, several required elements of a cleaning program can be determined: the scope of cleaning, a required schedule for maintenance, and targets to achieve. In order to alter the "identity," "strength," or "purity" of a product, certainly gross contamination would be required. Such high levels should not be found after cleaning. However, in some cases, process residues below the order of gross contamination may still affect patient safety and possibly product quality. One goal of a cleaning program is to verify that no gross contamination remains after cleaning and that any residues that do remain do not jeopardize the safety of the patient or quality of the next product.

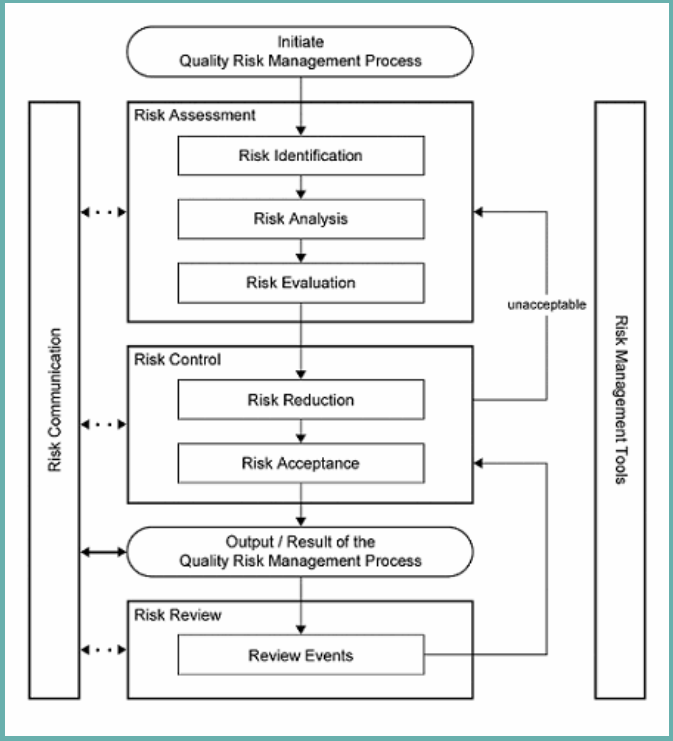
Now let's look at some of the many regulatory guidances that have come out since 2001. While some of them have some degree of application to cleaning, the two guidances that have the most applicability to cleaning are ICH Q9 and the FDA's 2011 process validation guidance.

ICH Q9 GUIDANCE

ICH Q9 outlines basic principles and examples of tools for quality risk management that can be applied to pharmaceutical processes. In ICH Q9 we find two primary principles of quality risk management:[9]

- The evaluation of the risk to quality should be based on scientific knowledge and ultimately link to the protection of the patient; and
- The level of effort, formality, and documentation of the quality risk management process should be commensurate with the level of risk.

Figure 2:
Overview of a typical quality risk management process



If we apply these principles to cleaning, it is apparent that the risks the cleaning processes may present to patient safety and product quality should be ***scientifically assessed***. The extent of any activities, such as cleaning development, cleaning verification, cleaning validation, monitoring, etc., should then be driven by the ***level of risk*** presented. The implementation of these principles offers serious potential for developing useful, effective, and efficient cleaning programs.

In fact, Annex II, "Potential Applications for Quality Risk Management," subsection 6, "Quality Risk Management as Part of Production" under Validation states, "To identify the scope and extent of verification, qualification and validation activities (e.g., analytical methods, processes, equipment and ***cleaning methods***," which clearly encourages the use of ICH Q9 for developing a cleaning validation program. Annex II subsection II.4 "Quality Risk Management for Facilities, Equipment and Utilities" also states that ICH Q9 principles can be applied to setting "acceptable (specified) cleaning validation limits." A precedent for implementing ICH Q9, as it pertains to cleaning, has already been set for this in the International Society for Pharmaceutical Engineering's (ISPE) Risk-Based Manufacturing of Pharmaceutical Products (Risk-MaPP) Baseline® Guide.[10] Figure 2 shows an overview of the ICH Q9 quality risk management process.

As ICH Q9 suggests, the risk management process can also be applied to the cleaning of all manufacturing equipment. Consequently, the risk assessment process should be used to derive criteria that can assist in decision making and control the risks to the patient. For a cleaning process, this should be a systematic and documented process to:

- identify the hazard (e.g., ***cleaning*** process residues)
- assess the severity of the ***cleaning*** process residues
- evaluate means to detect the ***cleaning*** process residues
- determine the levels of ***cleaning*** process residues
- support the implementation and maintenance of appropriate controls.

Risk controls should be commensurate with the level of risk. The ultimate decision on the appropriate controls may rely on both qualitative and quantitative data. The risk assessment should be documented and should include a discussion of all inherent assumptions and limitations. Table 1 shows how cleaning process development and validation maps to the ICH Q9 process.

Table 1:

Map of ICH Q9 Elements to Cleaning

ICH Q9 Steps	Corresponding Cleaning Development and Validation Elements		
Risk Identification	Data collection	Identify possible cleaning process residues - API, cleaning agent, etc. Collect historical cleaning data and other knowledge Methods for detection of process residues	
	Hazard identification	Determine ADEs for cleaning process residues Calculate maximum safe carryover Identify possible failure modes for cleaning process	
Risk Analysis	FMEA (Initial)		
	Severity	Impact of cleaning process failure (e.g., toxicity, product quality)	
	Exposure	Historical cleaning data and other cleaning knowledge Bench scale analysis (process residue characterization, cleanability, cleaning agent selection, critical process parameter determination, design of experiments, "design space" definition) Cleaning process robustness	
	Detectability	Detectability of cleaning process residues	
Risk Evaluation	Collection and evaluation of cleaning data Statistical evaluation of data (Cpk/Ppk) Margin of safety measurement Statistical process control limit determination		
Risk Reduction	Design of experiments Define the cleaning "design space" Cleaning process optimization Training		
Risk Acceptance	FMEA (Final)		
	Severity	Impact of cleaning process failure (e.g., toxicity, product quality)	
	Exposure	Process capability determination (Cpk/Ppk) Margin of safety measurement Statistical process control limit determination	
	Detectability	Statistical process control charting Monitoring program/periodic evaluation Visual inspection PAT applications	
Risk Review	Updates to ADEs based on new clinical/ toxicological data Cleaning failure investigations New product introductions	Cleaning process improvements Statistical process control charting Monitoring program/periodic evaluation	
Risk Communication	Facility cleaning risk assessment Hazard identification report ADE monographs Risk analysis of cleaning procedures Cleaning validation masterplan CV protocols	CV reports Risk evaluation of cleaning procedures Cleaning control strategy Training records New product risk review	

(Note: FMEA is used as an example in this table. Other RA tools may be equally appropriate.)

Table 2:

Map of FDA's Process Validation Guidance Elements to Cleaning

Process Validation Steps		Corresponding Cleaning Dev. and Validation Element
Stage 1 - Process Design	Building and capturing process knowledge and understanding	Historical cleaning data and other cleaning knowledge Bench scale analysis (process residue characterization, cleanability, cleaning agent selection, critical process parameter determination) Design of experiments Define the cleaning "design space" Cleaning process optimization
	Establishing a strategy for process control	Process residue characterization, cleanability determination Cleaning agent selection Critical cleaning process parameters Risk analysis of cleaning procedures Level of cleaning necessary Risk analysis for master planning Product/equipment grouping strategies Cleaning control strategy Analytical method selection (Vis, TOC, etc.)
Stage 2 - Process Qualification	Design and qualification of utilities and equipment	Equipment design for cleanability Cleaning equipment design/qualification
	Process performance qualification	Cleaning process robustness Collection and evaluation of cleaning data Statistical evaluation of data (Cpk)
	PPQ protocol	Sampling strategy Analytical methods (Vis. TOC, HPLC, etc.) Hold Time Studies
	PPQ protocol execution and report	Process capability determination Margin of safety measurement Statistical process control limit determination
Stage 3 - Continued Process Verification		Statistical process control charting Monitoring program/periodic verification Visual inspection PAT applications Net product risk review

PROCESS VALIDATION: GENERAL PRINCIPLES AND PRACTICES

The FDA's process validation guidance[8] aligns with the product life cycle concept and with existing FDA guidance on ICH Q8-Q10 and also describes concepts that are directly applicable to cleaning and cleaning validation. We can simply add ***"cleaning"*** to the elements of the process validation guidance as shown below.

- ***Cleaning*** Process Design — Building and capturing process knowledge and understanding
  - Application of design of experiment to ***cleaning***
  - Multifactorial interactions
  - Using risk analysis tools to screen potential variables
- ***Cleaning*** Process Qualification
  - Use of statistical methods in analyzing all collected ***cleaning*** data
- Continued ***Cleaning*** Process Verification
  - Use of statistical process control techniques
- Continuous Improvement
  - Use of historical data (monitoring, etc.) or technological advances for improvement of ***cleaning*** processes

The elements of the process validation guideline can be easily worked into a framework for a science-, risk-, and statistics-based approach to cleaning. Table 2 shows how cleaning process development and validation maps to the FDA's process validation guidance.

CGMPS FOR THE 21ST CENTURY GUIDANCE

In the FDA guidance "Pharmaceutical cGMPs for the 21st Century — A Risk-Based Approach"[5] we see four principles that have particular relevance to cleaning:

- Encourage the early adoption of new technological advances by the pharmaceutical industry.
- Facilitate industry application of modern quality management techniques, including implementation of quality systems approaches, to all aspects of pharmaceutical production and quality assurance.
- Encourage implementation of risk-based approaches that focus both industry and agency attention on critical areas.

- Ensure that regulatory review, compliance, and inspection policies are based on state-of the-art pharmaceutical science.

Applying these principles to cleaning, it follows that the extent of any activities, such as cleaning development and cleaning validation, should be driven by the level of risk presented and that the use of modern technology is encouraged.

PAT GUIDANCE

The FDA guidance "PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance"[7] states (again adding ***cleaning***):

- A desired goal of the PAT framework is to design and develop well-understood ***cleaning*** processes that will consistently ensure a predefined quality at the end of the ***cleaning*** process. Such ***cleaning*** procedures would be consistent with the basic tenet of quality by design and could reduce risks to quality and regulatory concerns while improving efficiency.
- Reducing ***cleaning*** cycle times by using on-, in-, and/or at-line measurements and controls

In the PAT guidance we find that, as a process, cleaning should be designed, developed, and well understood, and the use of on-, in-, and at-line measurements and controls is encouraged.

QUALITY BY DESIGN

Although the quality by design initiative as described in ICH Q8-Annex 1 addresses product manufacturing processes, there are principles there that can be applied to cleaning processes as well, such as (once again adding ***cleaning***):[6]

- Selecting an appropriate ***cleaning*** process.
- Identifying a cleaning control strategy (CS).
- A systematic evaluation, understanding and refining of the ***cleaning*** process, including:
    - Identifying, through e.g., prior knowledge, experimentation, and risk assessment, the material attributes and ***cleaning*** process parameters that can have an effect on ***cleaning*** critical quality attributes (CQAs);

- Determining the functional relationships that link material attributes and ***cleaning*** process parameters to ***cleaning*** CQAs.

- Using the enhanced ***cleaning*** understanding in combination with quality risk management to establish an appropriate control strategy which can, for example, include a proposal for design space(s) and/or real-time release.

Using a systematic approach such as those described in the Q8-Annex 1 could enable continual improvement and innovation of cleaning processes without being locked into previously validated parameters and restricted by onerous change control procedures.

ICH Q7 GUIDANCE

ICH Q7 Section 5.2.5 states that "*Acceptance criteria for residues and the choice of cleaning procedures and cleaning agents should be defined and justified*".[20]

Through the use of the word "justified," this simple sentence implies that science-based approaches should be employed in setting acceptance criteria for cleaning development and cleaning validation. The risk-based approach to cleaning validation is further recommended in point 12.70:

> "*In general, cleaning validation should be directed to situations or process steps where contamination or carryover of materials poses the greatest risk to API quality.*"

Although ICH Q7 applies specifically to APIs, the concept that science-based and risk-based approaches should be employed in cleaning development and cleaning validation can be extended to all pharmaceuticals.

ANNEX 15 OF THE EU GMP GUIDE

Annex 15 states that:[2]

> "*Limits for the carryover of product residues should be based on a toxicological evaluation to determine the product specific Permitted Daily Exposure (PDE) value[1]. The justification for the selected PDE value should be documented in a risk assessment which includes all the supporting references*"

1 See EMA Guideline on setting health based exposure limits for use in risk identification in the manufacture of different medicinal products in shared facilities

As in ICH Q7, the use of the word "justification" implies that science-based and risk-based approaches should be employed in setting acceptance criteria for cleaning development and cleaning validation. **Note**: Annex 15 is applicable to pharmaceutical products.

OPERATIONAL EXCELLENCE AND SIX SIGMA

Operational excellence can be defined as conducting business in a manner that satisfies customer demand, improves quality, and generates higher yields, faster throughput, and less waste. Six Sigma can be defined as a disciplined, data-driven approach and methodology for eliminating defects in any process.

These two approaches provide statistical tools to improve processes and increase quality. Since cleaning is a process that can be measured, these techniques can be effectively used to improve the cleaning process and enhance the safety and quality of pharmaceutical products.

SUMMARY

The guidance discussed above can be applied to create a new approach to cleaning and cleaning validation that is based on science, risk, and statistics. It offers clear ways of making sensible changes in cleaning that would reduce the complexity, lower the costs, and shorten the process while providing an even higher probability that cleaning of pharmaceutical manufacturing equipment has been effective. By implementing a truly science-based approach, such as the use of the ADE for risk analysis, with appropriate risk assessments, and with cleaning process development in place, a streamlined cleaning program may be readily developed that ensures patient safety and product quality while lightening the regulatory burden on industry.

PEER REVIEWERS

James Bergum, Ph.D.; Pernille L. Damkjær; Mallory DeGennaro; Michael Hrytsack, Ph.D.; Michael Schousboe; and Joel Young

REFERENCES:

1. This article is an excerpt and adaption from “Science, Risk & Statistics-based Guide to Cleaning Process Development and Validation” (United States Copyright Office Number TXu001950765) that was submitted in August 2013 to the International Society for Pharmaceutical Engineering (ISPE) for potential publication as a Baseline Guide. Portions of this article were previously published in Pharmaceutical Engineering November/December 2011 Vol. 31 No. 6.
2. EudraLex - Volume 4 - Good Manufacturing Practice (GMP) guidelines - Annex 15 Qualification and validation.
3. FDA Guide to Inspections Validation of Cleaning Processes, Section IV. Evaluation of Cleaning Validation, July 1993, U.S. Food and Drug Administration (FDA), www.fda.gov.
4. United States vs. Barr Laboratories, Inc. Civil Action No. 92-1744, U.S. District Court for the District of New Jersey: 812 F. Supp. 458. 1993 US Dist. Lexis 1932; 4 February 1993, as amended 30 March 1993.
5. Pharmaceutical CGMPs for the 21st Century Industry – A Risk-Based Approach: Final Report, September 2004, U.S. Food and Drug Administration (FDA), www.fda.gov.
6. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH Harmonised Tripartite Guideline, Pharmaceutical Development – Q8(R2), August 2009, www.ich.org.
7. FDA Guidance for Industry: PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, September 2004, U.S. Food and Drug Administration (FDA), www.fda.gov.
8. FDA Guidance for Industry: Process Validation - General Principles and Practices January 2011, U.S. Food and Drug Administration (FDA), www.fda.gov.
9. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH Harmonised Tripartite Guideline, Quality Risk Management – Q9, Step 4, 9 November 2005, www.ich.org.
10. ISPE Baseline® Guide: Risk-Based Manufacture of Pharmaceutical Products (Risk-MaPP), International Society for Pharmaceutical Engineering (ISPE), First Edition, September 2010.
11. Walsh, A., “Cleaning Validation for the 21st Century: Acceptance Limits for Active Pharmaceutical Ingredients (APIs): Part I,” Pharmaceutical Engineering, July/August 2011, Vol. 31, No. 4, pp. 74- 83.

ABOUT THE AUTHORS

THOMAS ALTMANN is RD&E global technical manager – life sciences at Ecolab. He has 18 years' experience in consulting in the pharmaceutical and biotechnology industry, as well as API production related to cleaning, validation, and disinfecting of product contact surfaces.

ALFREDO CANHOTO, PH.D., is a senior manager of quality assurance validation for Alnylam Pharmaceuticals and has expertise in cleaning development and validation.

MICHEL CREVOISIER retired from Novartis Pharma in 2015. In his last position, he was a senior QA expert for qualification and validation.

IGOR GORSKY is a Pharmaceutical, Biopharmaceutical and Medical Device industry expert with over 35 years of experience in multiple disciplines including Quality Assurance. His started his career at Alpharma and continued it at Wyeth and Shire where he led global Pharmaceutical Technology Validation. For the last six years he has been consulting globally for Valsource, LLC.

ROBERT KOWAL is a retired cleaning validation subject matter expert (SME) for Johnson & Johnson.

MARIANN NEVEROVITCH is MS Pharmacy Research Scientist at Bristol-Myers Squibb. Cleaning Validation Subject Matter Expert with 15+ years of experience in cleaning verification analytical method development and support. Since 2010 leading cleaning verification program in Support of Clinical Supply Operations.

MOHAMMAD OVAIS, MPharm is a Cleaning Validation Consultant. He has more than thirteen years of hands-on experience in validating cleaning processes. He provides training and consulting services in process and cleaning validation.

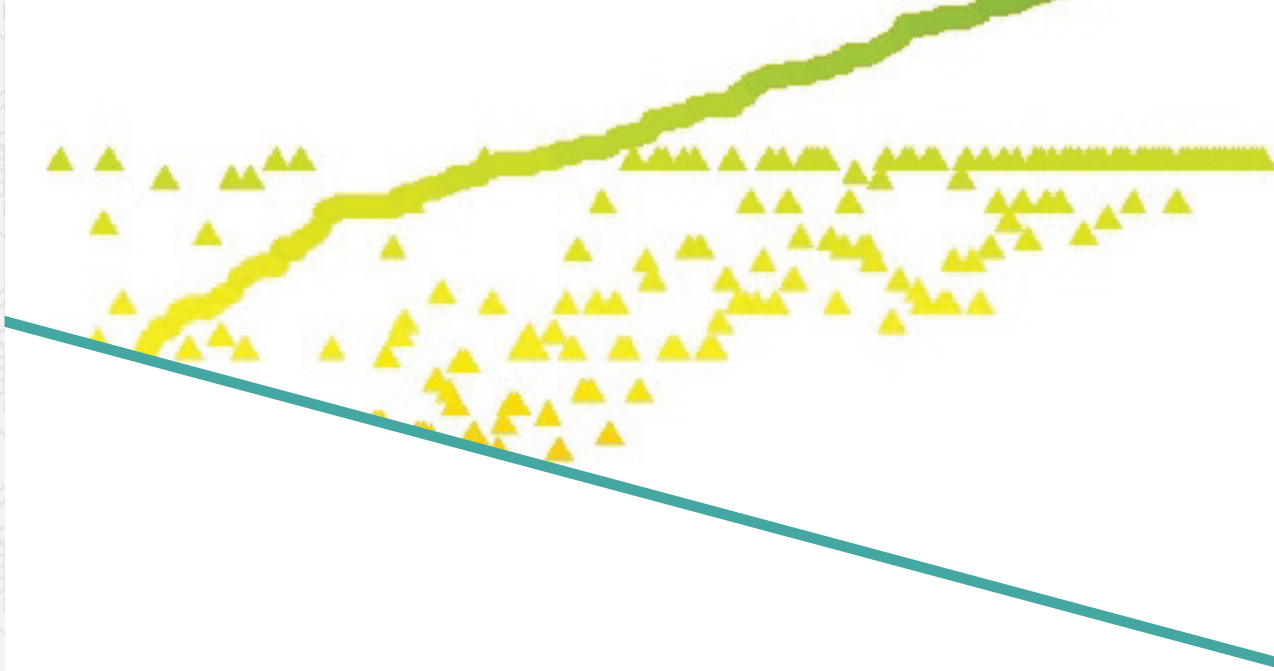
OSAMU SHIROKIZAWA is President of Life Scientia Limited, a pharmaceutical engineering and consultancy firm.

ANDREW WALSH is president of the Center for Pharmaceutical Cleaning Innovation, a non-profit research and educational organization.

12. Walsh, A., "Cleaning Validation for the 21st Century: Acceptance Limits for Active Pharmaceutical Ingredients (APIs): Part II," Pharmaceutical Engineering, September/October 2011, Vol. 31, No. 5, pp 44-49.
13. Walsh, A., Ovais, M., Altmann, T., and Sargent, E.V., "Cleaning Validation for the 21st Century: Acceptance Limits for Cleaning Agents," Pharmaceutical Engineering, November/December 2013, Vol. 33, No. 6, pp 12-24.
14. Walsh, Andrew, Michel Crevoisier, Ester Lovsin Barle, Andreas Flueckiger, David G. Dolan, Mohammad Ovais, "Cleaning Limits—Why the 10-ppm and 0.001-Dose Criteria Should be Abandoned, Part II," Pharmaceutical Technology 40 (8) 2016.
15. Walsh, Andrew, Ester Lovsin Barle, Michel Crevoisier, David G. Dolan, Andreas Flueckiger, Mohammad Ovais, Osamu Shirokizawa, and Kelly Waldron, "An ADE-Derived Scale For Assessing Product Cross-Contamination Risk In Shared Facilities," Pharmaceutical Online, May 2017.
16. "Clarifying Questions Upfront is Key in Process Validation, US and EU PV Principles in Alignment —CDER's McNally" in http://www.ipqpubs.com/uncategorized/clarifying-questions-upfront-is-key-in-process-validation-us-and-eu-pv-principles-in-alignment-cders-mcnally-stresses/.
17. 21 CFR Part 211.67 – Current Good Manufacturing Practice for Finished Pharmaceuticals, Equipment Cleaning and Maintenance, U.S. Code of Federal Regulations, U.S. Food and Drug Administration (FDA), www.fda.gov.
18. 21 CFR Part 111.27 – Current Good Manufacturing Practice in Manufacturing, Packaging, Labeling, or Holding Operations for Dietary Supplements, Equipment and Utensils; What Requirements Apply to the Equipment and Utensils that You Use? U.S. Code of Federal Regulations, U.S. Food and Drug Administration (FDA), www.fda.gov.
19. 21 CFR Part 820.70 – Quality System Regulation, Production and Process Controls, U.S. Code of Federal Regulations, U.S. Food and Drug Administration (FDA), www.fda.gov.
20. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH Harmonised Tripartite Guideline, Quality Risk Management – Q9, Step 4, 9 November 2005, www.ich.org.

AN ADE-DERIVED SCALE FOR ASSESSING PRODUCT CROSS-CONTAMINATION RISK IN SHARED FACILITIES

ANDREW WALSH
ESTER LOVSIN BARLE, PH.D.
MICHEL CREVOISIER, PH.D.
DAVID G. DOLAN, PH.D.
ANDREAS FLUECKIGER, MD
MOHAMMAD OVAIS
OSAMU SHIROKIZAWA
KELLY WALDRON



Beginning in 2004, a global team of pharmaceutical toxicologists, industrial hygienists, quality assurance professionals, a cleaning validation professional, and a representative from the US FDA participated in the development of the International Society for Pharmaceutical Engineering's (ISPE) Risk-Based Manufacturing of Pharmaceutical Products (Risk-MaPP) Baseline Guide.[1] The purpose of the Risk-MaPP Guide was to provide a "scientific, risk-based approach to manage the risk of cross-contamination."[1] The Risk-MaPP Guide was based on the International Conference on Harmonisation's Quality Risk Management Guideline (ICH Q9)[2] and encouraged the "selection of the appropriate risk control strategies to be implemented on a case-by-case basis to maintain patient safety and assure product quality."[1]

GUIDANCE ON MANAGING RISK IN SHARED FACILITIES

From a cGMP (quality) perspective, Risk-MaPP provided guidance on how to manage the risk of product cross-contamination in shared facilities, from the four main sources the Risk-MaPP guide identified:

1. Mix-up (wrong material being used)
2. Retention (residues left in equipment after cleaning)

3. Mechanical transfer (transfer of material on contaminated non-product contact surfaces)
4. Airborne transfer (material movement through air)

During the planning meetings for the Risk-MaPP Guide, the FDA specifically requested that this guide:[1]

- provide an approach to quantify the toxicity of drugs
- provide a risk management/assessment model that gives a clear view on how to address the controls to comply with 21 CFR 211.42(c) — for separation/dedication in facilities, etc.
- discuss how the approach fits into cleaning validation.

ADE AS A METRIC FOR ASSESSING RISK

In 2010, the Risk-MaPP Guide was published and introduced the acceptable daily exposure (ADE) as an appropriate metric for assessing pharmaceutical manufacturing risks for worker exposure, and for patient safety (e.g., in cleaning validation), as it is a value based on all the available pre-clinical and clinical data for a compound. In accordance with the principles of ICH Q9, the level of controls required — and the level of effort, formality, and documentation needed (including validation) — are commensurate with the level of risk. Consequently, it is critical to have a means of measuring the level of risk. It should be understood from this article that simply dividing compounds into two classes (highly hazardous and non-hazardous) is fallacious, and that the hazards that drugs present to patients should be viewed on a continuum, as envisioned by Risk-MaPP. This article presents a new hazard rating scale approach derived from the ADE concept that can be used as a visual tool to quickly evaluate and prioritize the relative hazards posed by drugs manufactured in a shared facility or equipment train.

ICH Q9 AND APPLYING RISK IN THE MANUFACTURING ENVIRONMENT

The principles behind applying risk in pharmaceutical and biopharmaceutical manufacturing were introduced in ICH Q9, which was formally adopted by US FDA in 2006. ICH Q9 mentioned its applicability to cleaning (including acceptance limits) in Annex II.4 and to validation in Annex II.6. According to ICH Q9, risk is defined as the combi-

nation of the probability of occurrence of harm and the severity of that harm. This can be expressed as:

$$Risk = f\,(Severity\ of\ Harm,\ Probability\ of\ Occurrence\ of\ That\ Harm)$$

Risk, in terms of a hazard (i.e., the potential source of harm), can also be expressed as:

$$Risk = f\,(Severity\ of\ a\ Hazard,\ Level\ of\ Exposure\ to\ That\ Hazard)$$

If the hazard is **intrinsic to an active pharmaceutical ingredient (API)**, then this general equation can be further refined to:

$$Risk = f\,(Toxicity_{API},\ Level\ of\ Exposure_{API})$$

CALCULATING CLEANING RISK

Patient risk for adverse effects increases as exposure to a given API or other compound rises above the ADE (which is synonymous with the permissible daily exposure [PDE]; and health-based exposure limits [HBEL], as used in the EU[3]). This risk is a function of the unique dose-response-duration relationship for each compound and the level of exposure to the compound from, for example, residual API carryover after cleaning. Note that this methodology can be applied to any agent or compound, and is therefore not exclusive to APIs. Consequently, the ADE provides a value that can be used as a surrogate for severity to calculate the potential cleaning risk, as shown in the following equation:

$$Cleaning\ Risk = f\,(ADE_{API},\ Level\ of\ Exposure_{API})$$

This equation tells us that the risk to a patient from cleaning is a function of the toxicity of the drug and the level of exposure (residues) found after cleaning.

INTRODUCING THE TOXICITY SCALE AND TOXICITY SCORES

Since ADE values vary over several orders of magnitude, they are hard to compare directly. However, to facilitate such a comparison, the values can be converted into a logarithmic scale in a manner similar to that used to create the pH scale. By converting the ADE values into units of grams per day and taking their negative logarithms, a continuous toxicity scale can be generated, as shown in Table 1. The resulting toxicity

Table 1:
ADE-Based Toxicity Scores

ADE Value	In grams/day	Toxicity Score -log(ADEgrams/day)
100 pg/day	0.0000000001	10
1 ng/day	0.000000001	9
10 ng/day	0.00000001	8
100 ng/day	0.0000001	7
1 µg/day	0.000001	6
10 µg/day	0.00001	5
100 µg/day	0.0001	4
1 mg/day	0.001	3
10 mg/day	0.01	2
100 mg/day	0.1	1
1 gm/day	1	0

scores are shown ranging from 0 to 10. However, it should be understood that toxicity score values above 10 and below 0 can still be legitimately obtained.

Similar to how the pH scale provides a convenient means of quickly assessing and comparing the acidity or alkalinity of solutions, this toxicity scale can be used as a means of comparing the toxicities or hazard levels of pharmaceutical compounds.

PRACTICAL EXAMPLES AND INSIGHTS

For example, in Figure 1, the ADE values for four well-known compounds were converted to this toxicity scale. The first compound is dioxin (i.e., 2,3,7,8-Tetrachlorodibenzo-p-dioxin), a well-known and very hazardous compound with an ADE of 35 picograms/day (toxicity score = 10.5). The second compound is arsenic trioxide, a relatively hazardous compound used to treat acute promyelocytic leukemia, with an ADE of 13 µg/day (toxicity score = 4.9). The APIs also include two low-hazard compounds: aspirin, with an ADE of 5 mg/day (toxicity score = 2.3), and sodium chloride, with an ADE of 26 mg/day (toxicity score = 1.6).

While it is not unexpected to see aspirin and sodium chloride occupying the low end of the toxicity scale and dioxin at the high end, readers may find it surprising to see arsenic trioxide, a compound most people would consider to be quite hazardous, occupying a place at the midpoint of the scale. This illustrates the ADE concept at work: The toxicity of a compound is dose-dependent, so while higher doses of a compound may generate extreme adverse effects, somewhat lower doses may be harmless. The toxicity scale helps to reveal this in relation to other compounds/APIs. It should also be understood that exceeding the ADE by a small amount does not necessarily put patients at risk. For example, a swab result exceeding the ADE-derived limit by a small amount during a cleaning validation study is not an immediate cause for alarm.

A LARGE-SCALE STUDY DEMONSTRATING THE TOXICITY SCALE CONCEPT

In our previous article[3], the ADEs of 304 APIs were compared to their 0.001 dose-based limits as a demonstration of how inaccurate and overly conservative the 0.001 dose-based approach is in estimating safe levels for exposure in patients, which could lead to impractical and unachievable limits. The distribution of these ADE values fits

Figure 1:
Comparison of pH and toxicity scales

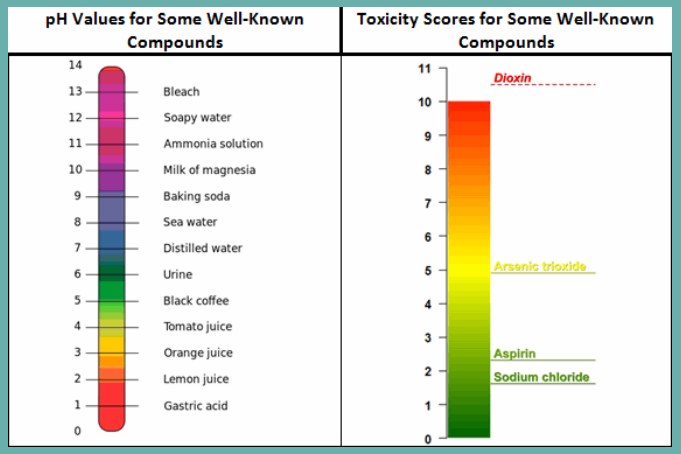


Figure 2:

Histogram of ADE-derived toxicity scores for 304 AOIs

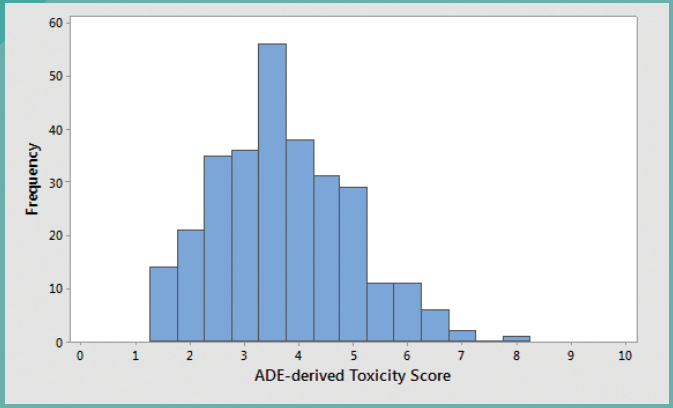
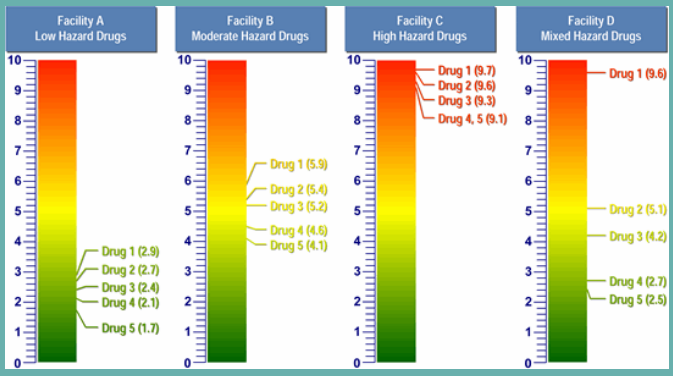


Figure 3:

Comparison of four facilities with five products each, using the toxicity scale



well within the presented scale, with an average toxicity score of 3.7; a median of 3.6; a mode (16 compounds) of 3.3; and a slightly positive skewness of 0.42, indicating a fairly normal distribution for these values (Figure 2). While the analysis did not include all known pharmaceutical compounds, a significant sample of these compounds was represented, with the results indicating that there was not a bias toward highly toxic or relatively nontoxic compounds in the dataset. The analysis also demonstrates that the toxicity scale encompasses the typical range of ADEs quite well.

PRACTICAL USES OF THE TOXICITY SCALE

The toxicity scale can be used to visualize and, quite quickly, understand the relative hazard of different compounds that are manufactured in a common facility. The scale can also satisfy the US FDA's desire for a tool that both identify the hazard of a drug from a maximum safe carryover (MSC) perspective and provides a linkage to cleaning validation. Figure 3 provides examples of how the toxicity scale can be used to visualize the hazards of drug products (as expressed by their toxicity scores and, thus, their ADEAPI values) manufactured in facilities where the hazards are low, moderate, and high; and one where the hazards are mixed.

In the above examples we can clearly see that, in Facility A, the products are of very low hazard. To manage the risks of operator and patient exposure, this facility would not need to employ the same level of controls as a facility with high-hazard products. Conversely, Facility C handles highly hazardous products and would need to have significant controls in place. It should be understood that, all other factors being relatively equal, lower-hazard products mean lower manufacturing risk and higher-hazard products mean higher manufacturing risk. Using this model, a company can quickly assess whether one facility requires a greater degree of controls than another facility, or, as suggested by the Risk-MaPP Guide, whether a given product is appropriate for introduction and manufacture in an existing facility with its equipment and cleaning practices, or whether adaptation of one or the other is necessary.

APPLICATIONS TO CLEANING VALIDATION

From a cleaning perspective, it should also be clear that Facility A has very low hazards associated with its products, and therefore should not have to put the same cleaning

validation program in place that Facility C may require. Facility A may actually be able to use "visually clean" as its sole acceptance criterion for cleaning validation if the maximum safe surface residues (MSSRs) calculated from these ADEs are well above the level that is visible and all surfaces are capable of being inspected. There may also be limited need, or no need, for any continued monitoring based on risk management criteria of an individual facility. On the other hand, the MSSRs for Facility C would most likely be well below the level that is visible for these APIs, so swab/rinse sampling would be required and specific analytical methods may even be needed. Continued monitoring may be necessary. Facility D presents a unique mixture of low, medium, and high hazards. In a case such as this, comprehensive manufacturing controls, swab/rinse, sampling, and continued monitoring may be necessary after Drug 1 (toxicity score = 9.6). Decontamination and other cross-contamination mitigation steps may even be necessary. But Drug 4 (toxicity score = 2.7) and Drug 5 (toxicity score = 2.5) may only require "visually clean" as their sole acceptance criterion for cleaning validation, especially if they are followed by products with low maximum daily doses.

BENEFITS TO REGULATORS

The toxicity scale can benefit regulators, as well, since an inspector seeing an API toxicity scaling graphic within a facility risk assessment may choose to focus less inspection effort on the cross-contamination controls for Facility A, and move on to Facility C, which presents higher cross-contamination risk. Based on the relatively hazardous products manufactured in Facility C, it would be appropriate for an inspector to spend more energy exploring the cross-contamination controls for such a facility. Facility D has a mix of low-, moderate-, and high-hazard products, and may have different levels of controls in place for handling them. As such, inspectors might elect to investigate how this facility handles these products and may want to focus their inspection on the risk reduction and cross-contamination controls due to Drug 1 and how these products interact with the other products (e.g., mix-up, retention, mechanical transfer, and airborne transfer).

CONCLUSION

As stated earlier, the US FDA, EU, and other health authorities have, during the development of the Risk-MaPP Guide and through other published guidelines, expressed a

strong interest in an "approach for identifying highly hazardous drugs." We believe that dividing pharmaceutical compounds into two classes, highly hazardous and non-hazardous, cannot be scientifically justified, and that the hazards that drugs present to patients should be viewed on a continuum. The new toxicity scale described in this article provides such an approach, and potentially could be used:

1. to visually compare the hazard levels or toxicities of pharmaceutical compounds processed within the same or across different facilities
2. as a measure of severity in assessing cross-contamination risks in facilities and identifying or determining appropriate control strategies
3. to communicate relative severity levels of hazards to internal (e.g., QA, RA) and external (e.g., customers, regulators) stakeholders
4. by manufacturers as a tool to quickly assess if a new product can be introduced into an existing facility
5. to assist inspectors in identifying facilities and areas with the greatest risk to focus on prior to or during their inspections, potentially accelerating inspections — a benefit to both regulator and industry
6. as a health-based severity scale for use in cleaning process FMEAs/FMECAs (failure mode and effects analysis/failure modes, effects, and criticality analysis).

FUTURE PUBLICATIONS

The toxicity scale presented in this article provides a way to measure the relative toxicity of compounds with regard to cross-contamination. A subsequent article will discuss a new scale for probability based on process capability values. Another article will discuss a new scale for detectability based on visual residue limits (VRLs) and maximum safe surface limits. A final article will discuss how these three scales in combination can be used in FMEAs/FMECAs for risk assessment of cleaning or other processes.

PEER REVIEWERS

Thomas Altmann; James Bergum, Ph.D.; Alfredo Canhoto, Ph.D.; Parth Desai; Mallory DeGennaro; Igor Gorsky; Michael Hyrtzak, Ph.D.; Robert Kowal, Mariann Neverovitch; and Joel Young

ABOUT THE AUTHORS

MICHEL CREVOISIER retired from Novartis Pharma in 2015. In his last position, he was a senior QA expert for qualification and validation.

DAVID DOLAN, PH.D., DABT is the toxicologist specializing in occupational, environmental, and product quality toxicology at Amgen Inc. over the past 12 years, after previously specializing in occupational toxicology and occupational epidemiology at Merck & Co., Inc.

ANDREAS FLUECKIGER, MD, was in charge of occupational health including occupational toxicology in the Roche Group for 32 years before his retirement in August 2018. His department also established Roche's health-based exposure limits values used in GMP cleaning.

ESTER LOVSIN BARLE, DVM, MSC, PHD, MScTox is the Head of Corporate Toxicology at Lonza and responsible for Health topics at Corporate Lonza EHS. Previously she has been the Head of Health Hazard Assessment in Novartis Global HSE & BCM.

MOHAMMAD OVAIS MPharm is a Cleaning Validation Consultant. He has more than thirteen years of hands-on experience in validating cleaning processes. He provides training and consulting services in process and cleaning validation.

OSAMU SHIROKIZAWA is President of Life Scientia Limited, a pharmaceutical engineering and consultancy firm.

KELLY WALDRON is a senior consultant with Valsource and a regulatory science researcher with the Pharmaceutical Regulatory Science Team at the Dublin Institute Technology.

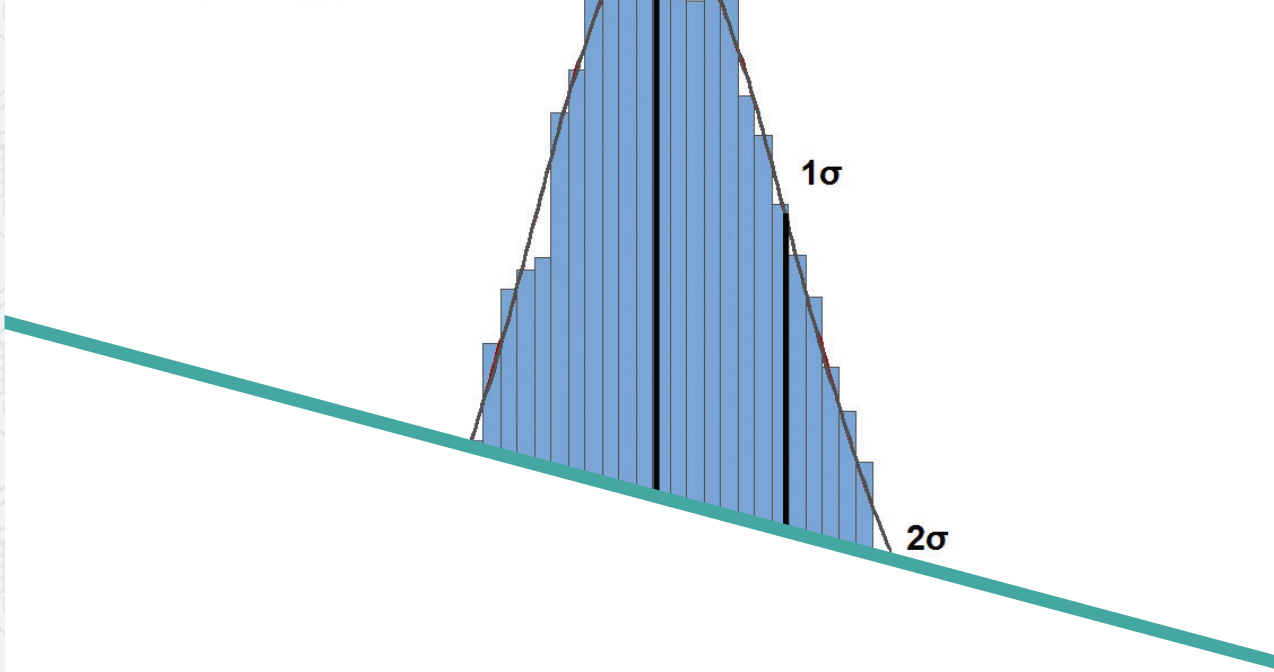
ANDREW WALSH is president of the Center for Pharmaceutical Cleaning Innovation, a non-profit research and educational organization.

REFERENCES:

1. ISPE Baseline Guide: Risk-Based Manufacture of Pharmaceutical Products: A Guide to Managing Risks Associated with Cross-Contamination. (ISPE, Tampa, FL, First Edition, 2010), Vol. 7, p. 186.
2. ICH Q9: Quality Risk Management (Step 4) (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use [ICH], London, 2005).
3. EMA Guideline on Setting Health Based Exposure Limits for Use in Risk Identification in the Manufacture of Different Medicinal Products in Shared Facilities, www.ec.europa.eu/health/files/gmp/2013_gmp_pc_en.pdf.

A PROCESS CAPABILITY-DERIVED SCALE FOR ASSESSING THE RISK OF COMPOUND CARRYOVER IN SHARED FACILITIES

ANDREW WALSH
DAVID G. DOLAN, PH.D.
ANDREAS FLUECKIGER, M.D.
IGOR GORSKY
ROBERT KOWAL
ESTER LOVSIN BARLE, PH.D.
MOHAMMAD OVAIS
OSAMU SHIROKIZAWA
KELLY WALDRON



A previous article discussed the development of a scale for measuring the risk of compound carryover in shared facilities in terms of the toxicity of the follow-on compound based on the acceptable daily exposure (ADE).[1] This article will present another new scale based on the process capability of a cleaning process that can be used to evaluate the probability of cross-contamination by compounds manufactured in a shared facility or equipment train. This approach can be used to evaluate sample results from cleaning validations or monitoring studies for all types of cleaning, including manual, semi-automated, or automated cleaning.

As stated in the previous article, the core principles behind evaluating "risk" in pharmaceutical manufacturing were introduced in the International Conference on Harmonisation (ICH) Q9 guideline (formally adopted by the U.S. FDA in 2006),[2] which mentions its applicability to cleaning (including acceptance limits) in its Annex II.4, and to validation in Annex II.6. According to ICH Q9, risk is defined as the combination of the probability of occurrence of harm and the severity of that harm. This can be expressed as:

$$Risk = f\,(severity\ of\ harm,\ probability\ of\ occurrence\ of\ that\ harm) \qquad \textit{(Equation 1)}$$

Risk, in terms of hazard (i.e., the potential source of harm), can also be expressed as:

$$Risk = f\,(severity\ of\ a\ hazard,\ probability\ of\ exposure\ to\ that\ hazard) \qquad \textit{(Equation 2)}$$

If the hazard is **intrinsic to an active pharmaceutical ingredient (API)**, this general equation can be further refined to:

$$Risk = f\,(toxicity_{API},\ probability\ of\ exposure_{API}) \qquad \textit{(Equation 3)}$$

As discussed in the first article in this series, the ADE (which is a dose that is considered to be protective of health for all patient populations, by all routes of exposure, every day for a lifetime) provides a value that can be converted into a toxicity score and utilized in the calculation of a potential "cleaning risk" (as shown in Equation 4).

$$Cleaning\ Risk = f\,(toxicity\ score_{API},\ probability\ of\ exposure_{API}) \qquad \textit{(Equation 4)}$$

What was missing from this equation in the first article was a comparable value for the probability of exposure$_{API}$. Since the probability of exposure is always 100 percent, and only the degree of exposure varies, this term needs to be refined. What we are more specifically interested in is the probability of residues remaining after cleaning that would exceed the ADE and put patients at risk. Consequently, what we are looking to measure is the probability of cleaning validation samples failing the limit calculated from the ADE, which can be simplified as the probability of cleaning failure. This article will explore the use of the process capability of the cleaning process as a means to measure the probability of cleaning failure$_{API}$ as shown in Equation 5.

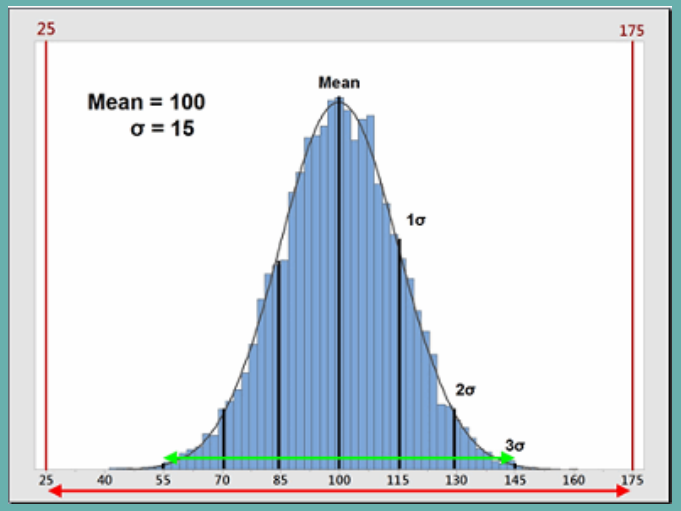
$$Cleaning\ Risk = f\,(toxicity\ score_{API},\ cleaning\ process\ capability_{API}) \qquad \textit{(Equation 5)}$$

(Note: This equation can be used with any compound that has a calculated ADE/PDE [permitted daily exposure], including cleaning agents).

BASICS OF PROCESS CAPABILITY

Process capability (Cp) is a simple, straightforward comparison of the spread of the process data (its variability) to the spread of the specification limit for that process. Basically, it is a measure of how well the data fits within the specifications. Figure 1 shows a plot of a hypothetical dataset that has a mean of 100 and a standard deviation of 15. In this example, suppose the specification limits for this process are 25 to 175. As we can see, the data fits well within these specifications, and they are centered within the specification limits. The process capability for this data is calculated using the following equation:

Figure 1:
Example of Process Capability (Cp)



$$Cp = \frac{Upper\ Specification\ Limit - Lower\ Specification\ Limit}{6\sigma} \quad \textit{(Equation 6)}$$

or

$$Cp = \frac{175 - 25}{6 \times 15} = \frac{150}{90} = 1.67 \quad \textit{(Equation 7)}$$

Sometimes, data is not centered within the specification range and is significantly closer to one specification limit than the other. In these cases, a modification of the Cp calculation is used that only looks at the distance of the mean to whichever specification limit is closest to the mean. This is called the process capability index (Cpk). Also, for data that has no upper or lower specification limits (such as cleaning data), a variation of the Cpk can be used instead that calculates a process capability based only on one specification (one-tail calculation). These are the Cpu (upper) and Cpl (lower) and are also simple comparisons of the spread of the data (its variability), specifically *the distance from the data mean to the upper specification limit (USL) or to the lower specification limit (LSL)*. Since cleaning validation data does not have lower specification limits, the Cpu equation should be used as a technique to quantify the probability of being exposed to an API at or above its ADE. The calculation for the Cpu can be seen in Equation 8.

$$Cpu_{(upper\ limit)} = \frac{Upper\ Specification\ Limit - Mean}{3\sigma} \quad \textit{(Equation 8)}$$

The terms in this equation can be obtained using data derived from cleaning swab or rinse studies as shown in Equation 9.

$$Cpu_{(ADE\ limit)} = \frac{ADE - derived\ Limit - Mean\ of\ Swab\ Data}{3\sigma\ of\ Swab\ Data} \quad \textit{(Equation 9)}$$

$$Cpu = \frac{175 - 100}{3 \times 15} = \frac{75}{45} = 1.67 \quad \textit{(Equation 10)}$$

While this approach can obviously be applied to data that is above the quantitation limits, readers may assume that this cannot be applied to data where some, or even all, of the data are below the ***quantitation*** limit (censored data), and calculating a mean and standard deviation would seem impossible. This situation often occurs with cleaning data obtained using high performance liquid chromatography (HPLC). However, there are valid ways of dealing with censored data to obtain acceptable estimates of the mean

Figure 2:
Example of Process Capability Index for Upper Specification (Cpu)

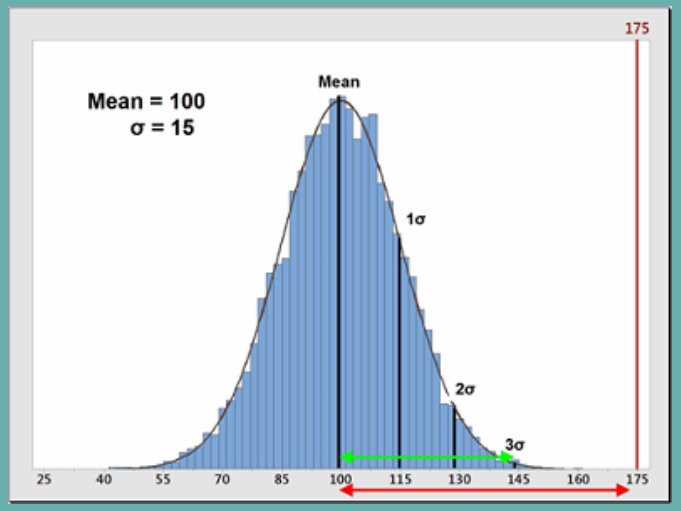


Figure 3:

Process Capability of 2.0

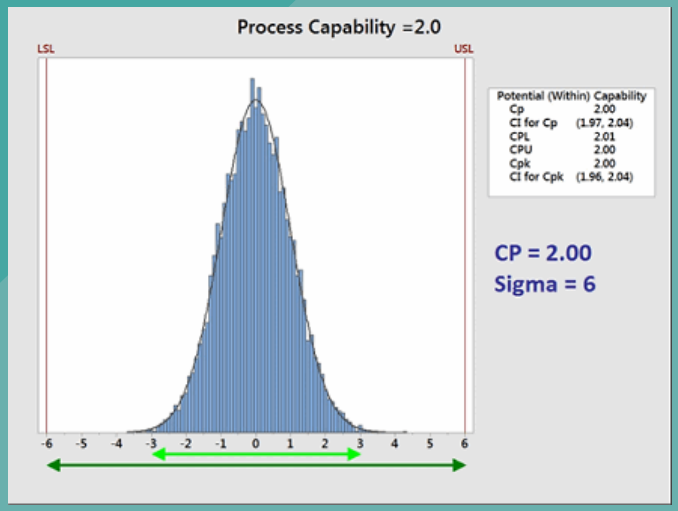
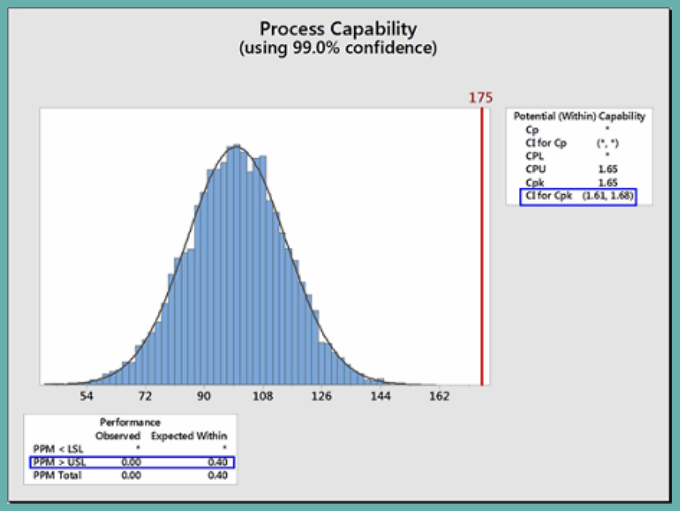


Figure 4:

Example of Confidence Limit and PPM Calculations for Cpu



and standard deviation that have been used in other disciplines that deal frequently with censored data.[3] Therefore, these calculations can still be performed with cleaning data that have points below the quantitation limit.

In Six Sigma or Operational Excellence programs, the values generated by these process capability calculations are considered to have significance in interpreting how acceptable a process is. The guidelines that are widely used for these values are shown in Table 1.

Table 1:

Six Sigma Definitions of Process Capability Values

Cp Value	Six Sigma Definitions
Cp < 1.0	Unacceptable or "not capable"
1.0 ≤ Cp < 1.33	Fair
1.33 ≤ Cp < 1.66	Acceptable
1.66 ≤ Cp < 2.0	Exceptional
Cp > 2.0	Goal of Six Sigma

The goal of these so-called "Six Sigma" programs is to develop or improve manufacturing processes such that they have an additional 3 standard deviations (sigma) of room on both sides of their process data, which mathematically calculates to a Cp of 2.0 (Figure 3). It should be noted that, in practice, many companies have been satisfied just to reach Five Sigma (1.66) and feel that striving for Six Sigma (2.0) is not worth the extra cost and effort. Therefore, achieving a process capability of 2.0 should be considered very good.

Like other statistical parameters that are estimated from sample data, the calculated process capability values are only estimates of true process capability and, due to sampling error, are subject to uncertainty. Hence, to account for these uncertainties, it is recommended to report and use the lower confidence limit of the Cpu from these calculations instead of just the Cpu itself. Almost all statistical software in use today can provide confidence intervals for process capability values. Figure 4 shows an example using Minitab 17®. In Figure 4 below, the text box on the right from the output from Minitab reports the Cpu as 1.65 and that the 99 percent confidence intervals

(CIs) for the Cpu range from 1.61 to 1.68 (Note: Minitab reports the CI for the Cpk, which is either the Cpu or the Cpl; in this case it is the Cpu).

Minitab can also report the expected number of possible failures out of a million based on the process capability analysis. In this example, the lower text box reports, based on this data, that there are 0.4 possible failures out of 1 million (i.e., exceeding the upper specification limit). Although possible, there is a very low probability of a failure in this example.

Table 2:

Cpu-based Scale for Probability of Exposure

(Probability of) Exposure Scale			
Cpu (cleaning process)	**(1/Cpu x 10)**	**Failures (PPM)***	**Rating**
1	10.0	1350	Unacceptable
1.11	9.0	434	Poor
1.25	8.0	88	Fair
1.42	7.0	10	Acceptable
1.66	6.0	0.3	Good
2	5.0	0.001	Very Good
2.5	4.0	3.19E-08	Excellent
3.3	3.0	2.08E-17	
5	2.0	3.67E-45	
10	1.0	4.91E-192	Exceptional
100	0.1	<2.23E-308	

* Potential failures (in parts per million) were calculated using Minitab 17 and without the 1.5 Sigma shift

These values could then be used with a severity scale, such as the toxicity scale, to calculate a risk priority number (RPN) that could be used to rank risks identified in an FMEA for the handling or cleaning of drug compounds (Table 3).

However, in order to arrive at a valid RPN value, the values of the factors used to calculate it must be from a ratio scale. As the values in this Cpu-based scale are derived directly from data and satisfy the criteria for a ratio scale, these values may be multiplied

Table 3:

RPN Scores based on Toxicity Scores and Process Capability Scores

Toxicity Score	Process Capability Score	Traditional RPN Score
10	10	100
9	9	81
8	8	64
7	7	49
6	6	36
5	5	25
4	4	16
3	3	9
2	2	4
1	1	1

Table 4:

Risk Scores based on Toxicity Scores and Process Capability Scores

Compound	Toxicity Score	Process Capability Score	Traditional RPN Score
Drug 1	10	1.0	10
Drug 2	1	10.0	10

Table 5:

Level of Risk Scores based on Toxicity Scores and Process Capability Scores

Toxicity Score	Process Capability Score	SO (Risk) Scores	
10	10.0	10	10
9	9.0	9	9
8	8.0	8	8
7	7.0	7	7
6	6.0	6	6
5	5.0	5	5
4	4.0	4	4
3	3.0	3	3
2	2.0	2	2
1	1.0	1	1

with a scale for severity to arrive at a measure of the risk associated with cleaning of a product or compound. However, the previously described toxicity scale is a logarithmic scale, and values from it cannot be multiplied with values from the process capability scale to derive any meaningful or useful result. Unfortunately, one of the issues with the scales typically used for FMEAs is that they are ordinal scales and the values in these scales cannot be meaningfully multiplied as is normally done in FMEAs. Several authors have already pointed out that the scales typically used in FMEAs to calculate RPNs may not yield useful or valid results.[4-7]

For example, in Table 4 we see two compounds that have very different toxicity scores and process capability scores and would clearly present different levels of risk. Drug 1 has a toxicity score of 10 (severe hazard at low ADE values) and has a corresponding process capability score of 1.0 (excellent cleanability), which is a low-risk situation. On the other hand, Drug 2 has a toxicity score of 1 (very low hazard due to low ADE value) but has a corresponding process capability score of 10.0 (poor cleanability), which represents a potentially high-risk situation. Yet when these two drugs are scored, their resulting RPN scores are identical, and thus this metric fails to appropriately discriminate among risks. Clearly, the result of simply multiplying these numbers could be seriously misleading. But more importantly the ***specific value and information provided by each score is lost***.

One suggested improvement to the FMEA has been to substitute an SO (severity and occurrence) score, which is simply a listing of the raw scores side-by-side in place of the RPN/risk score.[5] Table 5 shows SO scores for pairs of toxicity scores and process capability scores.

This approach is somewhat similar to the scoring of gymnasts in the Olympics. Gymnasts receive a score based on the difficulty of the performances and another based on their execution. Such an approach improves the transparency of the evaluation process. For instance, the "10/10" in Table 5 would mean it's a very toxic compound and its cleanability (ease of removal) is very poor, which translates into a very high-risk situation. A "10/1" would mean it's a very toxic compound but cleaning is extremely effective, which would result in a very low-risk situation. A "5/5" would mean it's a moderately toxic compound and cleaning is good, so it is a low risk. A "5/10" would mean it's a moderately toxic compound and cleaning is poor; it is a high risk. A "5/1" would mean it's also a moderately toxic compound, but its cleaning is extremely good, so it is a very low risk.

Table 6:
Ranking by Level of Risk

Compound	SO Score		Traditional RPN Score	Risk Rank? (Rank 1-10)
Drug 1	2	9	18	?
Drug 2	9	2	18	?
Drug 3	7	3	21	?
Drug 4	5	5	25	?
Drug 5	10	4	40	?
Drug 6	1	10	10	?
Drug 7	4	6	24	?
Drug 8	8	1	8	?
Drug 9	6	8	48	?
Drug 10	3	7	21	?

Now consider the SO scores shown in Table 6. The cleaning validation expert's judgment should include questions such as: Which drug poses the greatest risk (10=highest), and what is the capability of the cleaning processes (higher = worse)? What is the level of confidence to classify these compounds knowing these scores? How would this ranking compare to other rankings, and would the results be consistent?

Just looking at these scores, it should be quickly obvious that Drug 1 is a low-hazard compound and the cleaning procedure is not very effective, while Drug 2 is a high-hazard compound but the cleaning procedure is very effective and robust. As a consequence, the cleaning procedure for Drug 1 needs considerable improvement to assure that any residues after cleaning are at safe levels, while Drug 2's cleaning procedure does not. Drug 4 is a moderate hazard compound and the cleaning procedure is good, yet the traditional RPN approach scores Drug 4 as a higher risk than Drug 1. The traditional RPN Scores provide us with little information to evaluate the risk involved with these drugs. Obviously, multiplying these scores can obscure important information.

As stated in the first article, manufacturers could use the toxicity scale to evaluate new products for possible introduction into their facility or a manufacturing area. However, a new product that simply has a high toxicity score should not be rejected as a candidate for manufacture in the facility based solely on that score. A new product introduction should depend on how well the manufacturer can manage the cross-contamination risks presented by the introduction of this product. For instance, if a facility's cleaning program is capable of effectively removing residues of this new product to safe levels, then introducing this product into the facility actually presents a low risk. This new process capability scale can be used to guide the decision of whether or not a facility's cleaning program may be acceptable for the introduction of the new product.

For example, a facility's existing cleaning validation data (e.g., swab data) can be used to calculate a predicted process capability as a measure of how well the facility's current cleaning processes may be able to clean the new product. This is as simple as substituting the ADE-derived limit of the new product into Equation 9 along with the mean and standard deviation of the existing data and calculating what the process capability would be for this new product based on the existing data. This analysis could be used to guide the decision as to whether the product could be successfully cleaned by the existing cleaning process, or whether cleaning process development would be required.

If the predicted process capability analysis indicates that the current cleaning process is capable of effectively removing residues of the new product to safe levels, the next step would be to run a lab-scale "cleanability" test to quickly confirm whether the new compound is significantly different in cleaning requirements than the other products manufactured. If this is not the case, the product could be quickly moved to launch, with a single verification study performed on cleaning after the first batch.

These new scales can also be used for performing cleaning FMEAs/FMECAs, which are risk assessment tools specifically for identifying the potential failures of a cleaning process that could put a patient at risk. Cleaning FMEAs/FMECAs are equivalent to worker exposure FMEAs[8] that have been in use for some time, except that the focus of cleaning FMEAs/FMECAs is on patient exposure rather than on worker exposure. If the failure affects the process capability, that could result in higher process residues remaining in, or on, manufacturing equipment. Data from a cleaning process with a defined cleaning design space can be analyzed to determine whether the failure has a detrimental effect on the process capability. For example, if the process time of the cleaning is off by a few minutes, analysis of the cleaning design space may reveal a robust cleaning process and that this would not have a significant effect on the process capability or on residues of the API. Conversely, if the process time is off and the existing process capability data shows that this cleaning process is not very robust, this could have a significant effect, resulting in unsafe residues of the API remaining on manufacturing equipment. A cleaning FMEA/FMECA could also assess whether such a failure could be easily detected or might go undetected. It should be understood that cleaning FMEAs/FMECAs specifically target cleaning process failures that can result in process residues remaining and consider the severity (toxicity) of the compound that might remain and the likelihood that these residues may be present at unsafe levels. It's important to note that cleaning FMEAs/FMECAs are essentially different from equipment FMEAs, which are more focused on failures of the equipment or instruments and the likelihood of a piece of equipment or an instrument failing. Such equipment or instrument failures may or may not have an impact on patient safety.

CONCLUSION

During the development of Risk-MaPP[9] (the International Society for Pharmaceutical Engineering's Risk-Based Manufacturing of Pharmaceutical Products), the U.S. FDA had

expressed a strong interest in a tool to identify the degree of hazard of drugs. Such a tool would allow inspectors to quickly identify those specific manufacturers, facilities, and manufacturing areas that are handling the riskiest APIs. As a result, the focus of their inspections could shift to ensure the appropriate level of control exists at the locations that present the greatest risk to patient safety. The ADE-derived toxicity scale provided a means to quickly and visually identify facilities, manufacturing lines, and equipment that handle more hazardous compounds.[1] But this only indicates whether a hazardous compound is present or not and its degree of hazard. The toxicity scale by itself does not indicate how well the facility is handling that compound or removing residues of that compound. The use of the newly proposed process capability scale can provide a realistically defined measure of cleaning performance, and, when combined with the toxicity scale, can indicate the level of risk for compound carryover in shared facilities.

According to ICH Q9, the two primary principles of quality risk management are:

> "The evaluation of the risk to quality should be based on scientific knowledge and ultimately link to the protection of the patient; and
>
> The level of effort, formality, and documentation of the quality risk management process should be commensurate with the level of risk."

Working from these principles, these scales could also be used as a guide for a facility to decide how much cleaning validation is necessary to demonstrate that a cleaning process is effective and consistent.

The authors believe that cleaning is a deserving candidate for adoption of a science- and risk-based approach. We believe that the ADE and the cleaning process capability provide the scientific justification, and the analysis using the toxicity scale and the process capability scale provide the measure of risk as formulated in ICH Q9. To summarize, the risk involved in cleaning should be evaluated through the ADE-derived toxicity scale, which informs us which products are more hazardous than others. The process capability scale informs us of the difficulty in cleaning the products to safe levels. These two scales combined inform us where our cleaning process development efforts should be focused, and they can even help us to assess the reliability of our cleaning processes, ultimately resulting in the improvement of patient safety.

ABOUT THE AUTHORS

DAVID DOLAN, PH.D., is an occupational, environmental, and quality toxicologist at Amgen Inc.

ANDREAS FLUECKIGER has been in charge of occupational health including occupational toxicology in the Roche Group for 30+ years. His department also establishes Roche's health-based limits values used in GMP cleaning.

IGOR GORSKY is a senior consultant at ConcordiaValsource and a frequent speaker/writer on topics such as cleaning validation, critical utilities, process scale-up and validation, and knowledge management.

ROBERT KOWAL is a retired cleaning validation subject matter expert (SME) for Johnson & Johnson.

ESTER LOVSIN BARLE is head of health hazard assessment in Novartis Global HSE & BCM, where she is responsible for development of health-based exposure limits (HBELs), occupational toxicology, and patient safety-related process in support of manufacturing in Novartis globally.

MOHAMMAD OVAIS is scientific affairs manager for Xepa-Soul Pattinson (M) Sdn Bhd, Malaysia and a long-time student of cleaning validation.

OSAMU SHIROKIZAWA is a director and senior consultant of Life Scientia Limited, a pharmaceutical engineering and consultancy firm.

KELLY WALDRON is a senior consultant with Valsource and a regulatory science researcher with the Pharmaceutical Regulatory Science Team at the Dublin Institute Technology.

ANDREW WALSH is president of the Center for Pharmaceutical Cleaning Innovation, a non-profit research and educational organization.

The next article will discuss a new scale for detectability based on visual residue limits and maximum safe surface limits. A final article will discuss how these scales might be used together for performing cleaning FMEAs/FMECAs.

PEER REVIEWERS:

Sarra Boujelben, Parth Desai, and Mallory DeGennaro

REFERENCES:

1. Walsh, A., et.al., "An ADE-Derived Scale For Assessing Product Cross-Contamination Risk In Shared Facilities," May 2017, Pharmaceutical Online, www.researchgate.net/publication/317176627_An_ADE-Derived_Scale_For_Assessing_Product_Cross-Contamination_Risk_In_Shared_Facilities
2. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH Harmonised Tripartite Guideline, Quality Risk Management – Q9, Step 4, 9 November 2005, www.ich.org.
3. Helsel, Dennis R., "Statistics for Censored Environmental Data Using Minitab® and R" 2012, 2nd Edition, John Wiley and Sons, ISBN 978-0-470-47988-9
4. Wheeler, D., "Problems with Risk Priority Numbers – Avoiding More Numerical Jabberwocky," Quality Digest, June 2011, www.qualitydigest.com/inside/quality-insider-article/problems-risk-priority-numbers.html
5. Hubbard, D. and D. Evans, "Problems with scoring methods and ordinal scales in risk assessment." IBM Journal of Research and Development. Volume 54, Number 3. May/June 2010
6. Gilchrist, W., Modeling Failure Modes and Effects Analysis. International Journal of Quality & Reliability Management, 1993, 10 (5), 16-23.
7. Waldron, Kelly, "Risk Analysis and Ordinal Risk Rating Scales - A Closer Look" Journal of Validation Technology Dec. 2015, Vol. 21 Issue 4, p11
8. Cadwallader, L. C., "The Worker Exposure Failure Modes and Effects Analysis" Fusion Science and Technology Vol. 47, Issue 4, 2005
9. ISPE Baseline® Guide: Risk-Based Manufacture of Pharmaceutical Products (Risk-MaPP), International Society for Pharmaceutical Engineering (ISPE), First Edition, September 2010.

A SWAB LIMIT-DERIVED SCALE FOR ASSESSING THE DETECTABILITY OF TOTAL ORGANIC CARBON ANALYSIS

ANDREW WALSH
THOMAS ALTMANN
ALFREDO CANHOTO, PH.D.
ESTER LOVSIN BARLE, PH.D.
DAVID G. DOLAN, PH.D.
ANDREAS FLUECKIGER, M.D.
IGOR GORSKY
ROBERT KOWAL
MARIANN NEVEROVITCH
MOHAMMAD OVAIS
OSAMU SHIROKIZAWA
KELLY WALDRON

This article will discuss how the detection limit for analytical methods can be combined with cleaning validation swab limits to create a detectability scale similar to that described in our article on visual inspection.[1] This new detectability scale can assist in determining whether an analytical method is acceptable for use in a cleaning validation or verification. Combined with the HBEL-derived toxicity scale[2] and Cpu (process capability)-derived probability scale,[3] it can also provide for a total measure of risk in cleaning.

Note: This article uses the term health-based exposure limit (HBEL), which is synonymous with the terms acceptable daily exposure (ADE) and permitted daily exposure (PDE).

SELECTION OF ANALYTICAL METHODS IN CLEANING

Analytical methods typically used in cleaning validation fall into the two broad categories of specific methods and nonspecific methods, and the decision for using one or the other should be science-based and risk-based.[1] Figure 1 presents a hierarchy for selecting analytical methods in reference to the HBEL-derived toxicity scale.[2] For low-risk situations, visual inspection may be the only method needed, supported by nonspecific methods or by specific methods as necessitated by the increasing hazard level. As the level of the hazard increases, the rigor required of the analytical method should also increase. However, as indicated by the question marks, the transitions from using only

a visual inspection to needing a nonspecific analysis such as total organic carbon (TOC) and then to needing a specific analysis are not obvious. The use of the scale discussed in this article may provide a tool to help resolve these questions for analytical methods as was shown for visual inspection.[1] This article is focused on TOC as an example for nonspecific methods, as compounds containing organic carbon are the most common; however, this scale could be applied to other analytical methods.

DETERMINATION OF ANALYTICAL DETECTION LIMITS

Detection limits (DLs) and how they are determined are fundamental to this discussion. It is fairly well known that DLs for HPLC are normally determined by evaluating the signal-to-noise ratio. As stated in ICH Q2(R1):[4]

> "Determination of the signal-to-noise ratio is performed by comparing measured signals from samples with known low concentrations of analyte with those of blank samples and establishing the minimum concentration at which the analyte can be reliably detected. A signal-to-noise ratio between 3 or 2:1 is generally considered acceptable for estimating the detection limit."

For methods where there is no specific background noise to measure, such as TOC, other techniques may be employed, such as the standard deviation of the blank. ICH Q2 states:

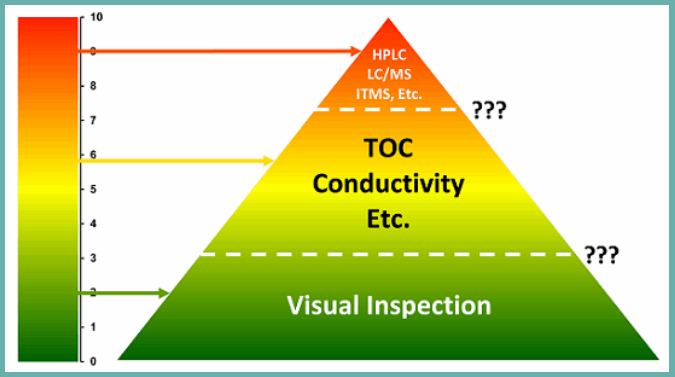
> Measurement of the magnitude of analytical background response is performed by analyzing an appropriate number of blank samples and calculating the standard deviation of these responses.

Similarly, a multiple of 3 is applied to the standard deviation of the blank and set as the DL. For example, for a blank with a mean of 100 and a standard deviation of 30, the DL would be set at 190 (100 + 30 x 3 = 190). This type of approach has been used for TOC.

DETECTION LIMITS FOR TOC

Acceptance of TOC for use in cleaning validation has grown over the past 20 years, with a number of articles being published on its application to APIs and cleaning agents.[5-19] Of the 15 articles cited, only eight addressed the DL for TOC in one way or another.

Figure 1:
Risk hierarchy of analytical methods (Note: Toxicity scale is based on −log(**HBEL**) where **HBEL** is the acceptable daily exposure in grams)



Since the methods for calculating DLs are a matter of debate among analytical chemists,[20] this may account for the lack of information on DLs for TOC in past articles. However, for the purposes of this article, the DL of TOC is very important.

An early article on using TOC for cleaning validation by R. Baffi, et. al.[5] examined its use for biologic compounds. The authors mention in their abstract that "...a limit of detection of approximately 0.1 ppm," but the text offered no details on how this DL was derived.

Gavlik, et. al. in 1995[6] published an article on the potential use of TOC for cleaning agents, but focused on recovery and did not report any other method parameters.

In 1996, Jenkins, et. al.[7] published a comprehensive review of swab and rinse recoveries for a variety of swab and filter materials and briefly discussed the DL. In their article, DL was defined as "...the absolute value of the intercept plus three times the estimated standard deviation." The authors reported DLs that ranged from 1 ppm to 14 ppm.

Strege, et. al. in 1996[8] discussed the DL and stated, "A limit of detection and limit of quantitation were established at 9.2 µg/swab and 12.1 µg/swab." The authors did not provide details on how the DL was arrived at but wrote that "...a set of 10 swab blanks were prepared and analyzed." Data or calculations for the DL were not provided, but the authors included a glossary from the USP XXII that mentioned "...analyzing a number of blank samples and calculating the standard deviation of this response. The standard deviation multiplied by a factor, usually 3, provides an estimate of the limit of detection." Since the rest of this article will present DLs in ppb (parts per billion), these results need to be converted. Based on their description of the handling of other swab samples, it appears that the swab dilution volume was 5 mL, and this would translate to a DL of 1,840 ppb (9.2 µg/5 mL = 1.84 µg/mL = 1,840 ppb).

Holmes, Alison J. and A. J. Vanderwielen in 1997[9] reported using TOC for analysis of aspirin residues on several materials of construction and reported DLs of 3 to 15 ppm. These investigators included swabbing an unspiked coupon surface as part of the swab blank, which other investigators have not mentioned.

Guazzaroni, et. al. in 1998[10] discussed the use of TOC for a number of compounds (cleaning agents, endotoxin, biologic media, and PEG) and reported the DL as 50 ppb "...

as per the manufacturer's specification," but stated that the TOC background, including the swab and filter material, was about 2.5 ppm.

Kirsch in 1998[11] discussed the parameters important to the validation of methods used for cleaning and mentioned the applicability of TOC. Kirsch stated that the DL is "... most practically defined as approximately three times the standard deviation of the baseline noise level around the analyte peak." While this is applicable to HPLC and some TOC analyzers, it is not applicable to all TOC analyzers.

In 2000, Karen Clark[12] analyzed "swab blanks" as a means to calculate a DL for TOC.[5] In her study, a swab blank is defined as a vial containing low TOC water (<25 ppb) along with the head of one swab. Four replicate analyses were performed on each swab blank and the mean and standard deviation were calculated. Using a Student t-test analysis of 10 swab blanks, Clark found the DL for TOC to be 50 ppb.

In 2004, Wallace, et. al. of Teledyne Instruments[13] published a brief review of some factors to consider for implementing TOC, such as detergent selection and acceptance criteria, and discussed the choice of TOC technologies, but did not discuss analytical method parameters. They did provide a table comparing the two major technologies used to oxidize the carbon in the sample to CO_2: high temperature combustion (HTC) and UV/persulfate (UV/P). This table contained data on the reagent water used as a blank. From this data we can determine that the DL for HTC in this study is 51 ppb (12 ppb + 13 ppb x 3) and UV/P is 10 ppb (7 ppb + 1 ppb x 3). Both of these values are lower than they should be, as these were not "swab blanks" and did not contain any swab material. It is well known that swab material can contribute significant carbon background to the swab blank.

In 2006, Chris Glover[14] performed a study of TOC using albumin and included the DL. Glover used a different approach by evaluating the accuracy data. The DL was designated as the lowest albumin weight with acceptable accuracy results (no less than 50 percent recovery). Glover set the DL at 49 ppb, but the water blank was subtracted from this value. Glover provided a table containing the raw data for 15 swab blanks. From this data, a DL can be calculated as above to be 237 ppb (186 ppb + 17 ppb x 3).

Nieves and Strege[15] reported a study of the development of a test method for polysorbate 20, which was being used as a cleaning agent for vial closures. These authors used

Table 1:
Calculated TOC Averages from 10 Blank Vials[12]

Vial Number	Average TOC (ppb)
1	58
2	72
3	75
4	93
5	79
6	102
7	60
8	83
9	67
10	54
Average	74.3
Standar Deviation	15.5
MDL (Student t, n-10)	50 ppb
LOQ	151 ppb

Table 2:

DLs Reported or Calculated from the Literature

Authors	Year	DL Reported	DL Calculated
Baffi, et. al.	1991	100 ppb	-
Gavlik, et. al.	1995	NR*	-
Jenkins, et. al.	1996	1 - 14 ppm	-
Strege, et. al.	1996	NR	1,840 ppb
Holmes & Vanderwielen	1997	3 - 15 ppm	-
Kirsch	1998	NR	-
Guazzaroni, et. al.	1998	2.5 ppm	-
Clark	2000	50 ppb	-
Wallace, et. al.	2004	51 ppb / 10 ppb	-
Glover	2006	NR	237 ppb
Nieves & Strege	2007	660 ppb	-
Frey, et. al.	2007	NR	-
Bader, et. al.	2009	< 50 ppb (?)	-
Clifford & Tanaka	2012	NR	-
Xue Li, et. al.	2018	38 ppb	-

* NR = Not Reported | ? = Not Specifically Stated

the ICH Q2 3sigma/slope approach for calculating the DL. Their calculation yielded a DL value of 660 ppb for this study.

In their article on using TOC for cleaning validation of nutraceuticals, Frey, et. al.[16] mentioned the importance of method parameters including DLs, but did not provide any values obtained.

Bader, et. al.,[17] in their study of the use of online TOC, state that the instruments examined met the "instrumental limit of detection of 50 ppb TOC required by USP," but did not state how that was determined or what the actual results were.

In 2012 Clifford and Tanaka[18] published a study on six water soluble and water-insoluble compounds and compared the results for recovery by rinse sampling, swab sampling with a water extraction, and swab sampling using direct combustion. No analytical method parameters were discussed other than recovery.

Most recently, Xue Li, et. al.,[19] in a study on cleaning agents, reported a QL of 114 ppb based on the linearity data and defined the DL as QL/3, or 38 ppb.

Table 2 summarizes the DLs reported in the literature or calculated from the data provided in the articles.

While there were significant differences in the methods used to determine the DLs in these articles, it is more important to note the wide disparity in the DL values reported/calculated, which range over three orders of magnitude. As described in the visual inspection article,[1] if the DL of TOC is known, then it can be compared to the TOC limit for a compound to justify the use of TOC for that compound. It should be immediately obvious that the higher the DL of TOC, the harder it will be to justify its use for compounds with lower limits. Clearly, obtaining a low DL is a very important task for the analyst developing the method, and this is something that the analyst should be aware of and address during the TOC method development.

IMPACT OF 1/1,000TH DOSE AND 10 PPM LIMITS ON THE USE OF TOC

As described in the visual inspection article,[1] to demonstrate the undesirable impact that retaining the 1/1,000th or 10 ppm limits would have on the use of TOC, swab limits

Table 3:

Parameter Assumptions for TOC Limit Calculations

Parameter	Value
Batch Size	100 kg
Maximum Daily Dose	10 gm
Total Equipment Surface Area	25,000 cm^2
Swab Area	25 cm^2
Dilution Volume	50 mL
% Swab Recovery	100%
% Carbon	70%

Figure 2:

Comparison of HBEL, 1/1,000th and 10 ppm to TOC Detection Limits (100 ppb)

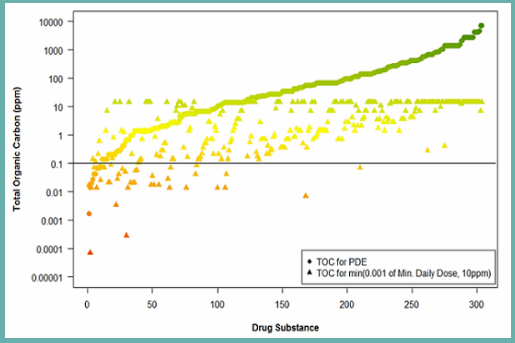
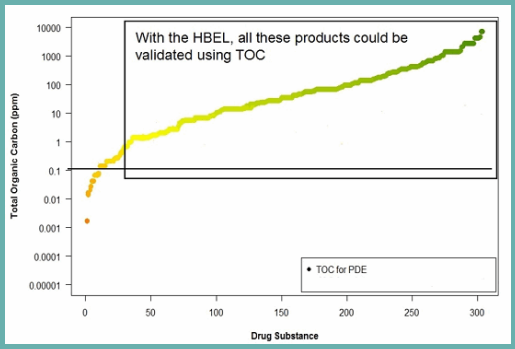


Figure 3:

Drugs where HBEL can meet TOC detection limits (100 ppb)



were calculated for 304 drug compounds based on their HBELs and their corresponding 1/1,000th or 10 ppm limits[21] using the assumptions in Table 3.

The data obtained was plotted using R statistical software on a log scale in order to visualize it all on one graph (Figure 2).

A line has been drawn in Figure 2 at a 100-ppb level (0.1 pbm) as the TOC DL for an example. Many of these compounds cannot be considered for evaluation by TOC as the TOC swab limits based on the 1/1,000th or 10 ppm are below this 100 ppb DL or too close to it to be justified as a safe method of analysis. If some of the DLs in Table 2 were shown, even fewer compounds could be justified.

However, if only the HBELs are used to calculate the TOC swab limits, many of these compounds could easily be considered for evaluation by TOC (Figure 3). It would seem that the compounds on the right side of the curve in the rectangle could easily be justified for evaluation by TOC. But as we move to the left and the TOC swab limits get closer and closer to the 100 ppb DL it would seem harder to justify using TOC. To appropriately evaluate the acceptable use of TOC, it would be helpful to have some way of judging how close the TOC swab limit is to the TOC detection limit.

USING THE DETECTION LIMIT OF TOC AS A MEASURE OF DETECTABILITY

In the same way as with visual inspection, we are trying to measure how close the TOC swab limit is to the detection limit of TOC so we can make a decision on whether we can use TOC. Again, a simple method to measure the relative distance is to look at the ratio of the two values. If we then take the log of this ratio we can obtain a logarithmic scale that equals zero when the values of the TOC swab limit and detection limit of TOC are equal and becomes negative when the detection limit of TOC is lower than the TOC swab limit and becomes positive when it is higher. This calculation can provide us with a carbon detection index that can be applied in all manufacturing cleaning situations (Equation 1).

$$CDI = \frac{log\ DL_{TOC}\ (in\ ppb)}{SL_{TOC}\ (in\ ppb)} \qquad \textit{(Equation 1)}$$

where:

CDI = Carbon Detection Index
DL_{TOC} = TOC Detection Limit
SL_{TOC} = TOC Swab Limit

TOC detection limits can also be converted into a scale by simply taking the logarithm of the ratio of TOC detection limit to the TOC swab limit as derived as described above in Equation 1 (Table 4).

In this example, any CDIs above zero are unacceptable and CDIs below -1.0 are acceptable. As can be seen comparing the three columns, as the DL increases, fewer and fewer swab limits can be met. Each company can select how close to a CDI of zero it is comfortable with. For example, one company may require its CDIs to be < -2.0, or at least 2 logarithms below the zero point.

DISCUSSION

As stated in the article on visual inspection,[3] the selection of methods for assessing cleaning should be science-based and risk-based. Key considerations for the risk assessment may include the hazard or risk of the process residue to be analyzed (toxicity score), level of detection required, applicability of existing methods, other quality and compliance risks, as well as risks to the business such as difficulty of implementation and the possible long-term maintenance of the method for ongoing monitoring programs. A reasoned and logical approach needs to be taken, as some methods may be unnecessarily expensive or difficult to implement for the process residues under consideration. Conversely, a simple inexpensive method may not be appropriate for all process residues. In general, the simplest techniques should be examined first and used if determined to be appropriate through an assessment based on science and risk. Ultimately, the goal should be to use the simplest technique that is appropriate and can be justified. TOC has proved to be one of the easiest analytical methods to implement, and is becoming a method of choice, for cleaning validation.

The scale reveals the two aspects of method development for cleaning that work in concert with each other. One is the HBEL, which drives the swab or rinse limits, which must be determined judiciously. Undue conservatism in calculating the HBEL

Table 4:
Detection Limit-based Scales for Detectability of Residues by TOC

SL_{TOC} (ppb)	CDI Log (DL_{TOC}/SL_{TOC})		
	DLTOC = 30 (ppb)	DLTOC = 100 (ppb)	DLTOC = 1000 (ppb)
0.1	2.5	3.0	4.0
1	1.5	2.0	3.0
3	1.0	1.5	2.5
30	0.0	0.5	1.5
100	-0.5	0.0	1.0
350	-1.1	-0.5	0.5
1000	-1.5	-1.0	0.0
10000	-2.5	-2.0	-1.0
100000	-3.5	-3.0	-2.0

through the excessive application of adjustment factors can easily result in swab or rinse limits that are so low as to be unachievable, which is in conflict with the long-standing guidance on cleaning from the FDA that limits should be "practical, achievable and verifiable."[22] This can lead to excessive cleaning efforts or unnecessary dedication, which conflicts with not only business goals but the intended purposes of Risk-MaPP.[23] The second aspect, the method DL, decides whether the method can be used based on the first aspect. Inattention to the DL when developing a TOC method can lead to high DLs (see Table 2) and result in TOC being restricted from use with many compounds. Conversely, lower DLs would allow the TOC method to be used more widely. If the TOC limits are set too low using arbitrary or non-health-based limits (as discussed above), then, again, TOC could not be justified. Regulators should ask to see scientifically justified swab limits (i.e., based on the HBEL), along with the corresponding DL when TOC or any other analytical methods are used for cleaning validation.

This article is intended to specifically address the use of TOC, but this scale is appropriate for any analytical method being developed for swab (or rinse) sample testing and can be applied to large biological molecules as well as small molecules. The principle simply informs the user whether a method for a given compound can be considered acceptable based on their swab (or rinse) limit. If the HBEL is very low ("Green Zone"), the corresponding swab limit will be very low, too, and will probably surpass the method's DL, and the scale will give a measure of how good that is. The user can simply clean and measure how well they have achieved that using the process capability scale. If they are in the "Red Zone" and the DL is equal to or above the swab limit, they cannot detect the compound at a level that assures meeting the HBEL-based limit. Such methods should not be considered completely useless; they can still be used to demonstrate that residues have been removed close to the limit. However, the user would need to pursue additional steps to provide assurance that the residues are at safe levels, such as demonstration of inactivation, degradation, or decontamination.[24]

It should be obvious that the DL for TOC is very important, and one of the main goals in swab method development should be to reduce the DL as much as possible. Previously, the limits on the applicability of TOC have been unclear, and this tool may be helpful in such assessments. Prior to this, careful consideration should be given to how DLs are

experimentally determined and a standard procedure for determining the DL for TOC should probably be established.

A subsequent article will discuss how these new detectability scales for TOC (or any other analytical method) and for visual inspection[3] can be used in conjunction with the HBEL-derived toxicity scale[1] and the Cpu-derived probability scale[2] as tools to evaluate the level of risk in cleaning.

PEER REVIEW

The authors wish to thank our peer reviewers, Bharat Agrawal; Sarra Boujelben; Gabriela Cruz, Ph.D.; Parth Desai; Ioanna-Maria Gerostathi; Jessica Graham, Ph.D., DABT; Miguel Romero Obon; Laurence O'Leary; John Leahy; and Ersa Yuliza, for reviewing this article and for their insightful comments and helpful suggestions.

REFERENCES

1. Walsh, Andrew, Thomas Altmann, Alfredo Canhoto, Ester Lovsin Barle, David G. Dolan, Mariann Neverovitch, Mohammad Ovais, Osamu Shirokizawa, and Kelly Waldron. "An MSSR-derived Scale for Assessing the Detectability of Compound-Carryover in Shared Facilities," Pharmaceuticalonline.com, December 2017
2. Walsh, Andrew, Ester Lovsin Barle, Michel Crevoisier, David G. Dolan, Andreas Flueckiger, Mohammad Ovais, Osamu Shirokizawa, and Kelly Waldron. "An ADE-Derived Scale For Assessing Product Cross-Contamination Risk In Shared Facilities," Pharmaceuticalonline.com, May 2017
3. Walsh, Andrew, Ester Lovsin Barle, David G. Dolan, Andreas Flueckiger, Igor Gorsky, Robert Kowal, Mohammad Ovais, Osamu Shirokizawa, and Kelly Waldron. "A Process Capability-Derived Scale For Assessing Product Cross-Contamination Risk In Shared Facilities," Pharmaceuticalonline.com, August 2017
4. ICH Q2 (R1): *Validation of Analytical Procedures: Text and Methodology* International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use [ICH], London, 2005.
5. Baffi, R. G. Dolch, R. Garnicka, Y. F. Huang, B. Mar, D. Matsuhiro, B. Niepelt, C. Parra, and M. Stephan "A Total Organic Carbon Analysis Method for Validating

ABOUT THE AUTHORS

THOMAS ALTMANN is RD&E global technical manager – life sciences at Ecolab. He has 18 years' experience in consulting in the pharmaceutical and biotechnology industry, as well as API production related to cleaning, validation, and disinfecting of product contact surfaces.

ALFREDO CANHOTO, PH.D., is a senior manager of quality assurance validation for Alnylam Pharmaceuticals and has expertise in cleaning development and validation.

DAVID DOLAN, PH.D., DABT is the toxicologist specializing in occupational, environmental, and product quality toxicology at Amgen Inc. over the past 12 years, after previously specializing in occupational toxicology and occupational epidemiology at Merck & Co., Inc.

ANDREAS FLUECKIGER, MD, was in charge of occupational health including occupational toxicology in the Roche Group for 32 years before his retirement in August 2018. His department also established Roche's health-based exposure limits values used in GMP cleaning.

IGOR GORSKY is a Pharmaceutical, Biopharmaceutical and Medical Device industry expert with over 35 years of experience in multiple disciplines including Quality Assurance. His started his career at Alpharma and continued it at Wyeth and Shire where he led global Pharmaceutical Technology Validation. For the last six years he has been consulting globally for Valsource, LLC.

ROBERT KOWAL is a retired cleaning validation subject matter expert (SME) for Johnson & Johnson.

ESTER LOVSIN BARLE, DVM, MSC, PHD, MScTox is the Head of Corporate Toxicology at Lonza and responsible for Health topics at Corporate Lonza EHS. Previously she has been the Head of Health Hazard Assessment in Novartis Global HSE & BCM.

MARIANN NEVEROVITCH MS Pharmacy Research Scientist at Bristol-Myers Squibb. Cleaning Validation Subject Matter Expert with 15+ years of experience in cleaning verification analytical method development and support. Since 2010 leading cleaning verification program in Support of Clinical Supply Operations.

Cleaning Between Products in Biopharmaceutieal Manufacturing," PDA J Pharm Sci and Tech 1991, 45 13-19

6. Gavlik, Walter K., Lane A. Ohlemeir and Herbert Kaiser Analytical Strategies for Cleaning Agent Residue Determination, Pharmaceutical Technology March 1995
7. Jenkins, K.M., A.j. Vanderweilen, J.A Armstronmg, L.M. Leonard, G.P. Murphy and N.A Piros "Application of Total Organic Carbon Analysis to Cleaning Validation," PDA Journal of Pharmaceutical Science and Technology Vol. 50, No. 1 January/February 1996
8. Strege, Mark A., Tery L Stinger, Brett T. Farrell and Avinash L. Lagu "Total Organic Carbon Analysis for Cleaning Validation of Bioprocess Fermentation Equipment," Biopharm International April 1996
9. Holmes, Alison J. and A. J. Vanderwielen "Total Organic Carbon Method for Aspirin Cleaning Validation," PDA J Pharm Sci and Tech 1997, 51 149-152
10. Guazzaroni, Maria, Boudin Yiin and Julia L. Yu "Application of total organic carbon analysis for cleaning validation in pharmaceutical manufacturing," American Biotechnology Laboratory September 1998
11. Kirsch, Robert Validation of Analytical Methods Used in Pharmaceutical Cleaning Assessment and Validation, Pharmaceutical Technology Analytical Validation 1998
12. Clark, Karen A., "Total Organic Carbon Analysis for Cleaning Validation in Pharmaceutical Manufacturing" October 2000
13. Wallace, Brian and Stephen Lawson "Detection Selection for TOC Analysis: Analytical Considerations for Cleaning Validation and Purified Water Testing," Teledyne Tekmar White Paper, www.teledynetekmar.com
14. Glover, Chris "Validation of the Total Organic Carbon (TOC) Swab Sampling and Test Method," PDA Journal of Pharmaceutical Science and Technology Vol. 60, No. 5, September-October 2006
15. Juarbe, Nieves and Mark Strege "Validation of a Method for the Determination of Polysorbate 20 Residues for the Support of the Cleaning of Pharmaceutical Vuial Closures," Journal of Validation Technology February 2007, Vol. 13, No. 2
16. Frey, Benjamin, David Koczan, and Kenneth Shrout "Implementing a Compliant Cleaning Program in the Dietary Supplement Industry," Contract Pharma/Nutraceuticals World Manufacturing Supplement, March 2007

ABOUT THE AUTHORS (CONT.)

MOHAMMAD OVAIS MPharm is a Cleaning Validation Consultant. He has more than thirteen years of hands-on experience in validating cleaning processes. He provides training and consulting services in process and cleaning validation.

OSAMU SHIROKIZAWA is President of Life Scientia Limited, a pharmaceutical engineering and consultancy firm.

KELLY WALDRON is a senior consultant with Valsource and a regulatory science researcher with the Pharmaceutical Regulatory Science Team at the Dublin Institute Technology.

ANDREW WALSH is president of the Center for Pharmaceutical Cleaning Innovation, a non-profit research and educational organization.

17. Bader, Keith, John Hyde, Peter Watler and Amber Lane "Online Total Organic Carbon (TOC) as a Process Analytical Technology for Cleaning Validation Risk Assessment" Pharmaceutical Engineering January/February 2009
18. Clifford, Robert and Minako Tanaka "Carbon Measurement Methods for Cleaning Validation: Comparing Direct Combustion with Rinse and Swab Sampling Methods," Pharmaceutical Technology Europe August 2012
19. Xue Li, Imad A. Haidar Ahmad, James Tam, Yan Wang, Gina Dao and Andrei Blasko " Cleaning verification: A five parameter study of a Total Organic Carbon method development and validation for the cleaning assessment of residual detergents in manufacturing equipment," Journal of Pharmaceutical and Biomedical Analysis 149 (2018) 33–39
20. Vander Heyden, Yvan and Ricard Boqué "The Limit of Detection," LCGC Europe 2009 Volume 22, Issue 2, pg 82–85
21. Walsh, Andrew, Michel Crevoisier, Ester Lovsin Barle, Andreas Flueckiger, David G. Dolan and Mohammad Ovais "Cleaning Limits—Why the 10-ppm and 0.001-Dose Criteria Should be Abandoned, Part II," Pharmaceutical Technology 40 (8), 45-55 (2016).
22. FDA Guide to Inspections: Validation of Cleaning Processes, July 1993, U.S. Food and Drug Administration (FDA), www.fda.gov.
23. ISPE, Baseline® Pharmaceutical Engineering Guide: Risk-Based Manufacture of Pharmaceutical Products: A Guide to Managing Risks Associated with Cross-Contamination. (ISPE, Tampa, FL, ed. First, 2010), vol. 7, pp. 186.
24. Walsh, Andrew "Cleaning Validation for the 21st Century: Can Alternative Approaches to the PDE-ADE be Justified?" Biopharm International October 2015

AN MSSR-DERIVED SCALE FOR ASSESSING DETECTABILITY OF VISUAL INSPECTION

ANDREW WALSH
THOMAS ALTMANN
ALFREDO CANHOTO, PH.D.
ESTER LOVSIN BARLE, PH.D.
DAVID G. DOLAN, PH.D.
ANDREAS FLUECKIGER, M.D.
IGOR GORSKY; ROBERT KOWAL
MARIANN NEVEROVITCH
MOHAMMAD OVAIS
OSAMU SHIROKIZAWA
KELLY WALDRON

Two previous articles discussed how the acceptable daily exposure (ADE) of a compound and the process capability (Cpu) of its cleaning process can be used to assess the level of risk associated with cross contamination in shared facilities.[1,2] This article will discuss how the maximum safe surface residue (MSSR) can be combined with the visual residue limit (VRL) to assess the acceptability of visual inspection for detecting the possibility of compound carryover in shared facilities. Combined with the ADE-derived toxicity scale[1] and Cpu-derived probability scale,[2] this new detectability scale can provide for a total measure of risk, and this new scale can also assist in determining whether visual inspection is acceptable for use in cleaning validation or verification.

SELECTION OF ANALYTICAL METHODS IN CLEANING

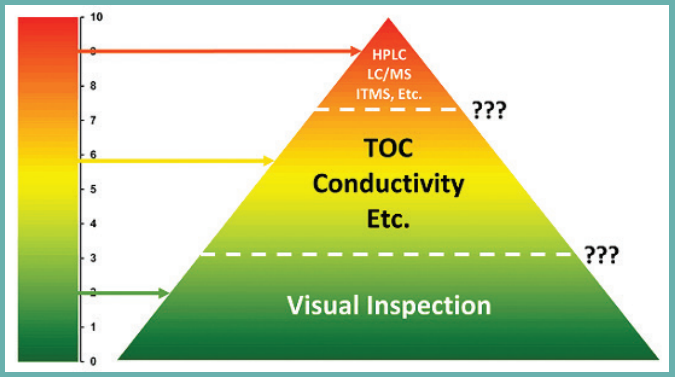
The analytical methods typically used in cleaning validation fall into two broad categories: specific methods and nonspecific methods. The decision to use a specific or nonspecific method should be science- and risk-based. For example, total organic carbon analysis (TOC) has been a method of choice for proteins due to its ease of use, the high carbon content of proteins, and the physiochemical difficulties in using specific methods such as enzyme-linked immunosorbent assay (ELISA).

In general, visual inspection would be considered a foundation method, and therefore always required. It would be supported next by TOC, conductivity, and other nonspe-

cific methods or by more specific methods as necessitated by the risk level. Figure 1 illustrates this risk-based hierarchy of analytical methods. Particularly for low-risk situations (e.g., low-hazard compounds and easily inspected equipment), visual inspection could be the sole method used for release of equipment after cleaning for return to manufacturing (which it typically is after successful cleaning validation studies). Visual inspection can also complement other methods such as TOC or conductivity to provide additional documentation for the release of equipment after cleaning. Finally, in high-risk situations, specific methods such as HPLC, LC/MS, etc., may need to be used along with, as always, a visual inspection.

Figure 1 presents a hierarchy for selecting analytical methods for cleaning based on the ADE-derived toxicity scale.[1] As the level of the hazard increases, the rigor required of the analytical method should increase. However, as indicated by the question marks, the transitions from using simple visual inspection to needing TOC, conductivity, etc., and from there to needing specific methods are not clear. The use of the scale based on visual detection limits discussed in this article may provide a tool to help resolve this question for visual inspection.

Figure 1:
Risk hierarchy of analytical methods
(Note: Toxicity Scale is based on –log(ADE) where ADE is the Acceptable Daily Exposure in grams)



DETECTION LIMITS FOR VISUAL INSPECTION

While the potential for using visual inspection in the validation of cleaning was proposed by Doug Mendenhall as long ago as 1989,[3] there have not been many actual studies performed or publications on using it. In 1993, Fourman and Mullen[4] published an article on cleaning validation acceptance limits where they stated:

> Spiking studies have determined that the active ingredients in most products are visible at approximately 100µg per 2 X 2 in. swab area.

This single sentence has somehow become a de facto industry standard in the minds of many industry workers, although no data or any information other than the sentence quoted above was provided by the authors. The "100µg per 2 X 2 in. swab area" translates into 4µg/cm^2, and it is the 4µg/cm^2 value that is most frequently quoted, although that's not what was actually written in the article. Shortly after, in 1994, Jenkins and Vanderwielen stated that the VRL (visual residue limit) could be reduced to 1µg/cm^2 if a light source is used during the inspection.[5] In 1998, two workers at Bristol-Myers

Squibb published an article referencing a study where they had used visual inspection as the sole method for solid dosage packaging equipment.[6] In 2004, Richard Forsyth and coworkers at Merck started publishing a series of articles in which they claimed that VRLs for APIs and common excipients could be seen down to 2µg/cm^2 and even as low as 1µg/cm^2. These studies used a "spotting" technique to apply the residues, which may have made the residues easier to see. However, based on the results of only four observers in their initial article, they concluded that visual inspection had many challenges making it difficult to justify.[7] More recently, Forsyth published an article on the logistical difficulties involved in qualifying a large group of personnel without knowing the VRLs of the products.[8] A statistical study by Mohammad Ovais in 2010 demonstrated that the VRL could be determined using logistic regression analysis of the inspection data.[9] Subsequently, a series of studies using large numbers of observers was performed at Stevens Institute of Technology from 2011 to 2013, employing an "even coating" technique that found VRLs for one particular product to range from 3 to7µg/cm^2, which were found to be dependent on training.[10]

While there is not a lot of supporting data at this point in time, it appears reasonable to say that the VRL for the majority of products, under the majority of viewing conditions, probably lies somewhere between 1 and 10µg/cm^2, depending on the product, type of surface (so far only stainless steel has been evaluated), the training of the inspectors, the preparation of the test coupons, and possibly the environmental conditions of the inspection.

If it is truly the case that the majority of drug product residues on equipment can be discerned by visual inspection in the range of 1 to 10µg/cm^2, then equipment could be released after cleaning, and cleaning procedures could even be validated, using only visual inspection as long as the MSSR levels were safely above this range. If the MSSRs are set too low using arbitrary limits, then visual inspection would be harder to justify, limiting its usefulness.

IMPACT OF 1/1,000TH DOSE AND 10PPM LIMITS ON THE USE OF VISUAL INSPECTION

A recent article examined 304 drugs from several companies and compared their ADEs to their corresponding 1/1,000th of the lowest therapeutic dose.[11] This article revealed

Figure 2:

Comparison of ADEs to the 1/1,000th dose

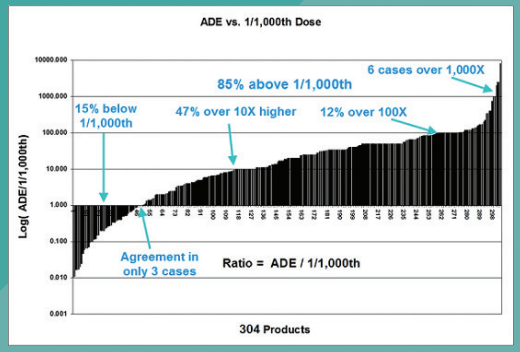


Figure 3:

Comparison of ADE, 1/1,000th and 10ppm to visual residue limits (10µg/cm²)

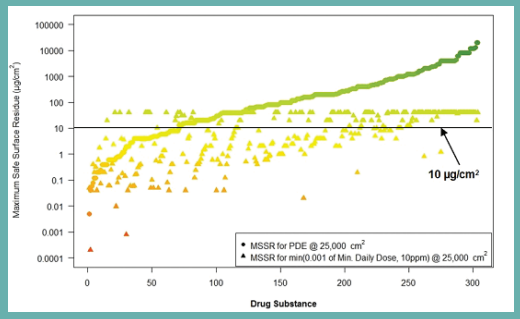
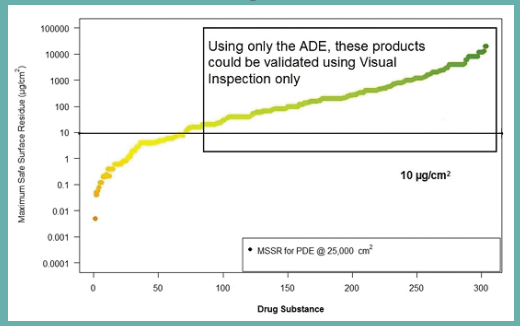


Figure 4:

Drug products where ADE can meet visual residue limits (10µg/cm²)



that the limits calculated by the 1/1,000th were not low enough in about 15 percent of the cases and were too low in 85 percent of the cases – by as much as 10x, 100x and, in a few cases, over 1,000x (Figure 2).

The cases where the ADEs were 85 percent higher are very significant to this discussion, as this is where continuing the use of the 1/1,000th or 10ppm will preclude the use of visual inspection in many of these cases, especially if both are used in combination.

To illustrate the impact that the 1/1,000th or 10ppm limits have on the use of visual inspection, the MSSRs were calculated for the 304 ADEs and their corresponding 1/1,000th limit using the following assumptions:

Table 1:

Parameter Values Used For Calculations

Parameter	Value for Example
Batch Size	100 kg
Maximum Daily Dose	10 grams
Total Surface Area	25,000 cm²

The MSSR (Maximum Safe Carryover/Total Equipment Surface Area) was calculated.

The data obtained was then plotted using R statistical software and, just as in the previous article,[11] the data had to be plotted on a log scale in order to visualize it all on one graph. The results can be seen in Figure 3.

A line has been drawn at the 10µg/cm² level for the VRL. Obviously, many of these compounds could not be considered for evaluation by visual inspection as the MSSRs based on the 1/1,000th dose or 10 ppm are below the VRL or too close to it to be justified as a safe method of analysis. However, if only the ADE is used to calculate the MSSRs, then many of these compounds could easily be considered for evaluation by visual inspection (Figure 4).

It would seem that the compounds on the right side of the curve in the rectangle could easily be justified for evaluation by visual inspection. But as we move to the left and

Figure 5:

Where is visual inspection appropriate to use?

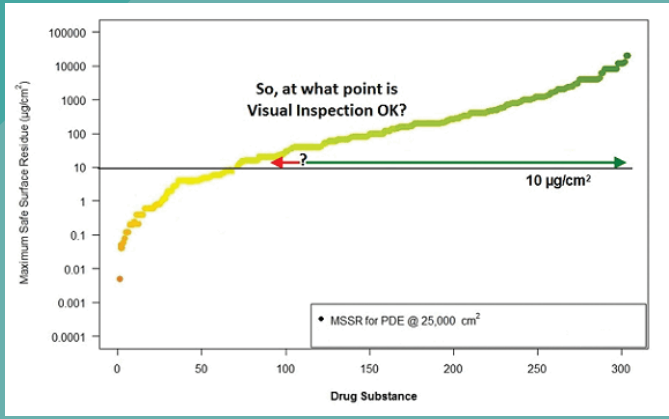


Table 2:

Example Of Visual Detection Index Scales For Different MSSRs And VRLs

MSSR (µg/cm²)	VDI Log (VRI/MSSR)	
	VRL = 5 (µg/cm²)	VRL = 10 (µg/cm²)
0.01	2.7	3.0
0.1	1.7	2.0
1	0.7	1.0
3.5	0.2	0.5
10	-0.3	0.0
35	-0.8	-0.5
100	-1.3	-1.0
1000	-2.3	-2.0
10000	-3.3	-3.0

the MSSRs get closer and closer to the 10µg/cm² VRL, it would seem harder to justify evaluation by visual inspection (Figure 5).

Would it not be helpful to have some way of judging how close the MSSR is to the VRL and whether it is acceptable to use?

USING MSSR AS A MEASURE OF DETECTABILITY

Since we are trying to measure how close the MSSR is to the VRL so we can make a decision on whether we can use visual inspection, a simple method to measure the relative distance regardless of what the MSSRs or the VRLs happen to be is to look at the ratio of the two values. If we then take the log of this ratio we can obtain a logarithmic scale that equals "0" when the values of the MSSR and VRL are equal and becomes negative when the VRL is lower than the MSSR and positive when it is higher. This calculation would provide us with a visual detection index that could be applied across all manufacturing situations (Equation 1):

$$VDI = \frac{log\ VRL}{MSSR} \qquad \textit{(Equation 1)}$$

where

VDI = Visual Detection Index
MSSR = Maximum Safe Surface Residue
VRL = Visual Residue Limit

Table 2 shows the range of VDIs for different MSSRs based on VRLs of 5ug/cm² and 10µg/cm².

In this example, any VDI above 0 is unacceptable, with VDIs below -1.0 being acceptable. A VDI of 0 can be considered the "vanishing point" where an MSSR equals its VRL and is about to pass below it. Consequently, each company could select how close to a VDI of 0 it believes is justifiable before allowing visual inspection to be used. For example, one company may require all its VDIs to be < -2.0, or at least 2 logs below this vanishing point.

DISCUSSION

The selection of methods for assessing cleaning should be science-based and risk-based.

Risks to consider in the risk assessment can include the hazard or risk of the process residue to be analyzed (toxicity score), level of detection required, applicability of existing methods, other quality and compliance risks, as well as risks to the business, such as difficulty of implementation and the possible long-term maintenance of the method for ongoing monitoring programs. A reasoned and logical approach needs to be taken, as some methods may be unnecessarily expensive or difficult to implement for the process residues under consideration. Conversely, a simple, inexpensive method may not be appropriate for all process residues. In general, the simplest techniques should be examined first and used if determined to be appropriate through an assessment based on science and risk. Ultimately, the goal should be to use the simplest technique that is appropriate and can be justified. Visual inspection could be the sole method if properly justified based on risk.

As noted in a previous article on visual inspection, U.S. regulation 21 CFR 211.67 (b)(6) has required the "inspection of manufacturing equipment immediately gbefore use" since 1979 and, in practice, pharmaceutical manufacturers have been releasing equipment based on a "visual" inspection for many years. Subsequently, the industry and even regulators have come to see this "inspection" as a "visual inspection" requirement. For example, the Pharmaceutical Inspection Co-operation Scheme (PIC/S) recommends "no quantity of residue should be visible on the equipment after cleaning procedures are performed." Similar statements can be found in many of the regulatory guidances on cleaning. PIC/S also mentions "spiking studies should determine the concentration at which most active ingredients are visible."

Recently, it appears that regulators are open to the possible use of visual inspection for cleaning validation. The draft of Annex 15 in Paragraph 9.2 originally stated:

> A visual check for cleanliness may form an important part of the acceptance criteria for cleaning validation however, ***it is not acceptable*** for this criterion alone to be used.

However, after comments from industry stakeholders, in the final version of Annex 15 Paragraph 10.2 this was changed to:

> A visual check for cleanliness is an important part of the acceptance criteria for cleaning validation. ***It is not generally acceptable*** for this criterion alone to be used.

ABOUT THE AUTHORS

THOMAS ALTMANN is RD&E global technical manager – life sciences at Ecolab. He has 18 years' experience in consulting in the pharmaceutical and biotechnology industry, as well as API production related to cleaning, validation, and disinfecting of product contact surfaces.

ALFREDO CANHOTO, PH.D., is a senior manager of quality assurance validation for Alnylam Pharmaceuticals and has expertise in cleaning development and validation.

DAVID DOLAN, PH.D., DABT is the toxicologist specializing in occupational, environmental, and product quality toxicology at Amgen Inc. over the past 12 years, after previously specializing in occupational toxicology and occupational epidemiology at Merck & Co., Inc.

ANDREAS FLUECKIGER, MD, was in charge of occupational health including occupational toxicology in the Roche Group for 32 years before his retirement in August 2018. His department also established Roche's health-based exposure limits values used in GMP cleaning.

IGOR GORSKY is a Pharmaceutical, Biopharmaceutical and Medical Device industry expert with over 35 years of experience in multiple disciplines including Quality Assurance. His started his career at Alpharma and continued it at Wyeth and Shire where he led global Pharmaceutical Technology Validation. For the last six years he has been consulting globally for Valsource, LLC.

ROBERT KOWAL is a retired cleaning validation subject matter expert (SME) for Johnson & Johnson.

ESTER LOVSIN BARLE, DVM, MSC, PHD, MScTox is the Head of Corporate Toxicology at Lonza and responsible for Health topics at Corporate Lonza EHS. Previously she has been the Head of Health Hazard Assessment in Novartis Global HSE & BCM.

MARIANN NEVEROVITCH MS Pharmacy Research Scientist at Bristol-Myers Squibb. Cleaning Validation Subject Matter Expert with 15+ years of experience in cleaning verification analytical method development and support. Since 2010 leading cleaning verification program in Support of Clinical Supply Operations.

It should be noted that Annex 15 is used by PIC/S, which has 49 health authorities as members, including the FDA, which joined PIC/S in January of 2011. So, it appears that most of the world's health authorities are ready to accept visual inspection for cleaning validation under the right circumstances and justification. The question, of course, is under what circumstances and justification.

This new detectability scale for visual inspection can be used in conjunction with the ADE-derived toxicity scale[1] and the Cpu-derived probability scale[2] as tools to evaluate the level of risk in cleaning validation. Going further, the toxicity scale could also help define the circumstances for visual inspection (a low hazard) and the detectability scale can provide the justification (easy to see at levels well below the safe limit for that hazard). This has the potential to help justify visual inspection in clinical manufacturing or R&D areas where there are limited amounts of API available for analytical method development and help speed and simplify the introduction of new compounds in a safe manner.

To move forward on implementing visual inspection as a method for cleaning validation, we believe it will be important to develop and maintain a formal visual inspection program, including justification through risk assessments, training and qualification of operators and inspectors, and with periodic assessments to ensure the integrity of the program. Future articles will examine the implementation of visual inspection in both clinical and commercial manufacturing environments.

A subsequent article will discuss how another detectability scale can be developed using total organic carbon detection limits and TOC swab limits, and a final article will discuss how all these new scales can be used together to create a cleaning risk dashboard and how they can also be used for scoring cleaning FMEAs (failure modes and effects analyses)/FMECAs (failure mode effects and criticality analyses) of cleaning processes.

PEER REVIEWERS:

Sarra Boujelben; Gabriela Cruz, Ph.D.; Mallory DeGennaro; Parth Desai; Ioanna-Maria Gerostathi; Miquel Romero Obon; Joel Young; and Ersa Yuliza

ABOUT THE AUTHORS (CONT.)

MOHAMMAD OVAIS MPharm is a Cleaning Validation Consultant. He has more than thirteen years of hands-on experience in validating cleaning processes. He provides training and consulting services in process and cleaning validation.

OSAMU SHIROKIZAWA is President of Life Scientia Limited, a pharmaceutical engineering and consultancy firm.

KELLY WALDRON is a senior consultant with Valsource and a regulatory science researcher with the Pharmaceutical Regulatory Science Team at the Dublin Institute Technology.

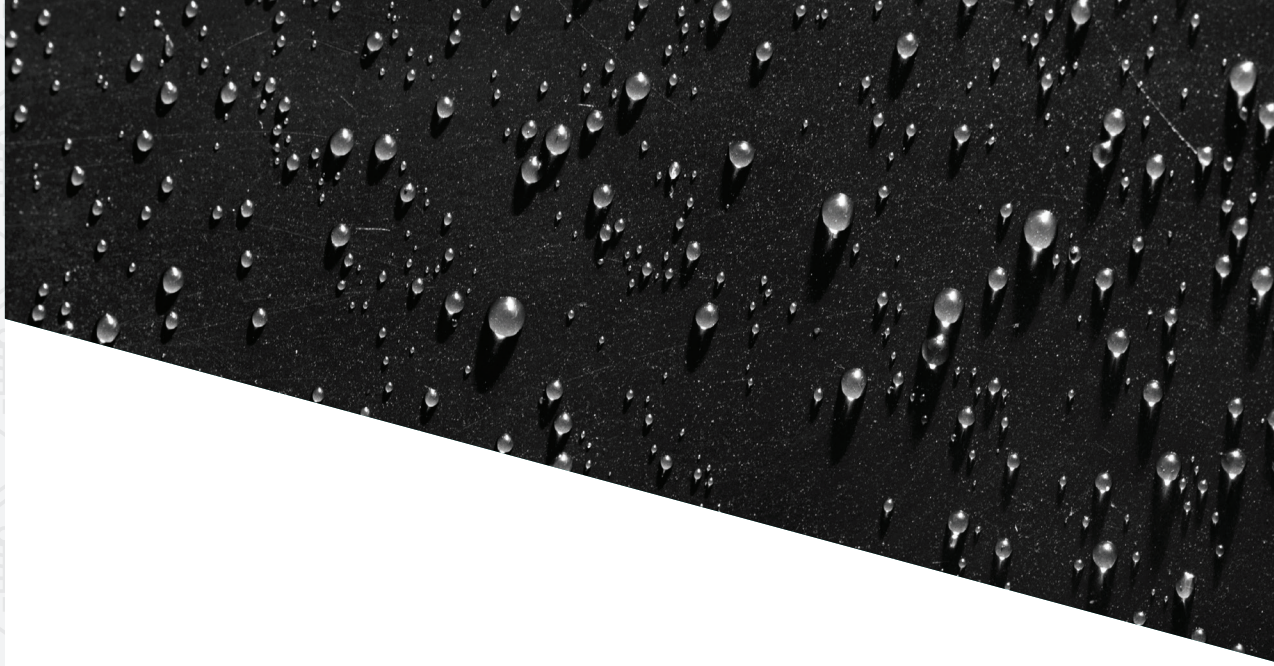
ANDREW WALSH is president of the Center for Pharmaceutical Cleaning Innovation, a non-profit research and educational organization.

REFERENCES:

1. Walsh, Andrew, Ester Lovsin Barle, Michel Crevoisier, David G. Dolan, Andreas Flueckiger, Mohammad Ovais, Osamu Shirokizawa, and Kelly Waldron, "An ADE-Derived Scale For Assessing Product Cross-Contamination Risk In Shared Facilities," Pharmaceutical Online, May 2017.
2. Walsh, Andrew, Ester Lovsin Barle, David G. Dolan, Andreas Flueckiger, Igor Gorsky, Robert Kowal, Mohammad Ovais, Osamu Shirokizawa, and Kelly Waldron, "A Process Capability-Derived Scale For Assessing Product Cross-Contamination Risk In Shared Facilities," Pharmaceutical Online, August 2017.
3. Mendenhall, D., "Cleaning Validation," *Drug Development and Industrial Pharmacy*, 15(13), pp. 2105-2114, 1989
4. Fourman, G.L., and Mullen, M.V., "Determining Cleaning Validation Acceptance Limits for Pharmaceutical Manufacturing Operations," *Pharmaceutical Technology*, 1993, Vol. 17, No. 4, pp 54-60.
5. Jenkins K.M. and A.J. Vanderwielen, "Cleaning Validation: An Overall Perspective", Pharm. Technol. 18 (4), 60–73 (1994).
6. Alvey, A. P., and Carrie, T.R. "Not Seeing is Believing — A Non-Traditional Approach for Cleaning Validation," *Journal of Validation Technology*, May 1998, Vol. 4, Issue 3.
7. Richard J. Forsyth, Vincent Van Nostrand, and Gregory P. Martin, "Visible-Residue Limit for Cleaning Validation and its Potential Application in Pharmaceutical Research Facility," *Pharmaceutical Technology*, Oct. 02, 2004.
8. R. Forsyth, "Qualifying Personnel to Visually Inspect Cleaned Equipment Part II: Small vs. Large Group Training," *Pharmaceutical Technology* 40 (6) 2016.
9. Ovais, Mohammad, "Statistically Justifiable Visible Residue Limits," *Pharmaceutical Technology*, 2010, Volume 34, Number 3.
10. Desai, Parth and Andrew Walsh, "Evaluation of Visual Inspection for use in Cleaning Validation - Part I," Pharmaceutical Online, September 2017.
11. Walsh, Andrew, Michel Crevoisier, Ester Lovsin Barle, Andreas Flueckiger, David G. Dolan, and Mohammad Ovais, "Cleaning Limits—Why the 10-ppm and 0.001-Dose Criteria Should be Abandoned, Part II," *Pharmaceutical Technology*, 40 (8), 45-55 (2016).

VALIDATION OF VISUAL INSPECTION AS AN ANALYTICAL METHOD FOR CLEANING VALIDATION

PARTH DESAI
ANDREW WALSH



U.S. Regulation 21 CFR 211.67 (b) (6) has required the "inspection of manufacturing equipment immediately before use" since 1979. In practice, pharmaceutical manufacturers have been releasing equipment based on a “visual" inspection for many years. Subsequently, the industry and even regulators have come to see this "inspection" as a "visual inspection" requirement. For example, the Pharmaceutical Inspection Co-operation Scheme (PIC/S) recommends “no quantity of residue should be visible on the equipment after cleaning procedures are performed.”[1] Similar statements can be found in many of the regulatory guidances on cleaning.[2-4] PIC/S also mentions "spiking studies should determine the concentration at which most active ingredients are visible.”

There have been only a few studies on visual inspection (VI). In 1993, Gary Fourman and Dr. Mike Mullen of Eli Lilly published an article in which they mentioned that spiking studies indicated many compounds were visible at approximately 100 µgrams/4 inch^2 (or 4 µgrams/cm^2).[5] No details were provided on how the studies were performed or on what compounds. A year later, an article by Jenkins and Vanderwielen claimed to be able to see residues down to 1 µgrams/cm^2 by using an additional light source. Six years later, in 2000, Frank Buscalferri et al, published an article where they claimed to see residues of several compounds down to approximately 0.4 µgrams/cm^2.[6] Then, from 2004 to 2014, Richard J. Forsyth, along with several others, published a series of articles that found a range of 0.4 to >10 µgrams/cm^2 for several different compounds.[7-11]

The basic approach for conducting these VI studies has been “spotting” a known amount of residue onto a "coupon" or surrogate material surface. Analysts then observed the spiked surface under various specified viewing conditions, and the lowest amount of residue that all analysts could see was considered the visual detection limit (VDL) for that compound/product.

However, it has never been clear whether VI is reliable since it has never actually been validated. If VI can be validated, or qualified, then the use of this technique for release of equipment would be justified and might even be used in place of traditional swab methods, which take substantial amounts of time and resources to develop, validate, and perform.

As mentioned, the above approach used a “spotting” technique that may result in an easily visible deposit. There has been a question of whether actual residues typically deposit all in one spot. Another possible approach is to apply an even coating of residue on the material surface and see if analysts can see these residues at the same levels. If analysts could, that could lead to the validation of VI as an analytical method.

Three separate studies on VI were conducted at the Stevens Pharmaceutical Research Center at Stevens Institute of Technology from 2011-2013 to explore the accuracy, precision, linearity, and detection limits of VI and the possibility that VI could be validated as an analytical method for cleaning validation. The studies used techniques to achieve an even coating of a product residue.

MATERIALS AND METHODS

- 6 cm by 6 cm "coupons" of 316L stainless steel with a #8 surface finish were prepared and supplied by Environmental Resource Associates, Inc. (Golden, CO).
- The product selected and spiked onto the coupons was an OTC sunscreen, as low-risk products would be most amenable for use with VI.
- Two techniques were used to evenly coat the coupon surface with a known amount of sunscreen: a technique using a syringe to manually apply an even coating to a 5 cm by 5 cm area and another technique using a spray-coating device. The manual technique was used in the first study, and the spray coater was used in the second and third studies.

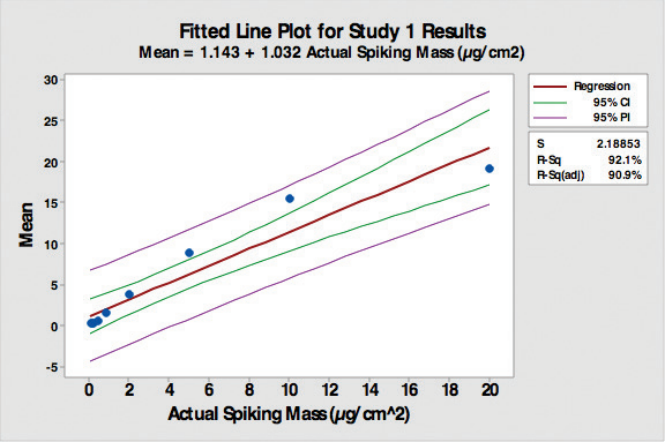
Table 1:

Standard Coupon Residue Levels for Study 1

Standard Coupon Number	Surface Residue Level (µgram/25 cm²)	Surface Residue Level (µgram/cm²)
1	10	0.4
2	25	1
3	50	2
4	86	3.4
5	131	5.2
6	186	7.4
7	251	10
8	325	13
9	410	16.4

Figure 1:

Fitted line plot for Study 1



- Two different sets of coupons were prepared, one being a standard set and the other an unknown set. Analysts had to compare the sets under defined viewing conditions and find the closest match. The potential benefit of using a visual standard is it provides a basis for visual measurement, since the analyst may not know what they are specifically looking for, nor what level they are seeing. This would be similar to comparing samples analyzed on an analytical instrument with an external standard or an acquired linearity curve. All the following studies used this type of approach.
- An Omega HHLM1337 handheld light meter was used to measure light levels.

STUDY 1

Sixty-eight graduate students in the Stevens Pharmaceutical Manufacturing and Engineering Program volunteered to participate in the first study. A set of nine coupons was prepared with known amounts of residue evenly deposited (manually) at increasing concentrations to be used as "standards." An additional six sets of coupons were prepared at each of these same concentrations to be used as "unknowns," for a total of 54 "unknown" coupons. The values of the residue levels are shown in Table 1.

The participants were provided minimal training before beginning the exercise to see how well they would do without specific instructions. The fitted line plot for the results can be seen in Figure 1.

The R-Sq (adj) was 90.9 percent despite the non-linearity of the first four coupons, which had significantly less residue spiked on them and were harder to see. Many students had difficulty distinguishing among these first four coupons. The accuracy of the VI increased as the concentration of residue increased.

There was fairly wide variation in the reported results. We made the following observations from our experience with this study:

- Since residues were deposited manually, there was some non-uniformity in the deposited residue, and this created some problems for analysts at the higher concentrations.

- Insufficient training of students was an issue. Students were not aware what the residue was supposed to look like (e.g., some students confused dust particles or scratches on coupons with residues). (Note: It is important to distinguish between student and industry-worker populations who have different levels of understanding; an industry worker should already know what visually clean equipment should look like.)
- A number of students reported questionable results and were requested to repeat the study. These students returned and repeated the study with different results, which were consistent with the expected values. These students offered plausible explanations for how they might have made errors in their results (e.g., a sloppily written "3" that looked like a "5").
- A small number of students reported impossible results: highest scores for lowest standards and lowest scores for highest standards. These students were also requested to repeat the study and would not, so this data was discarded. (Note: This is an important observation as the results of a visual inspection in the industry are a matter of trust.)
- Like any set of standards, these coupons require a certain degree of care to prevent the residues on them from being damaged. These coupons require proper storage and handling to prevent dust, fingerprints, or other contamination from accumulating on them and altering their values and the analysts' responses.

Due to these issues, and the variation in results, it was decided that another study would be performed with the following modifications:

1. A similar set of coupons would be manufactured using a spray-coating device in place of the manual method to deposit the residues. This was intended to standardize appearance, eliminate inter-coupon variability, and provide a more consistent presentation to the analysts.
2. Provide documented training to the analysts on how to perform the visual inspection (e.g., coupon handling, inspection angle/distance, difference between residue and dust/scratches).
3. Reduce the number of analysts to approximately 30.

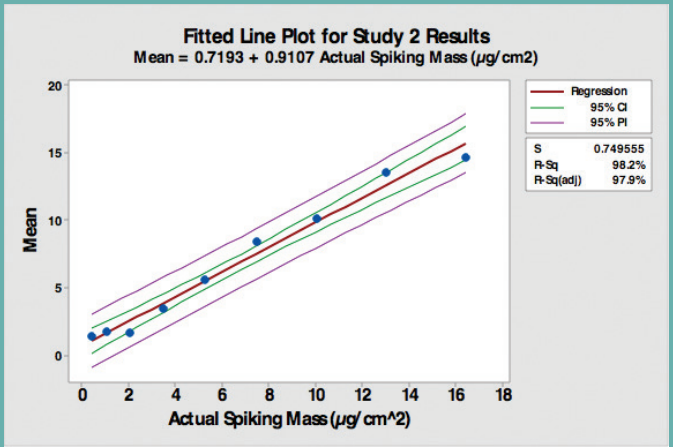
Table 2:

Standard Coupon Residue Levels for Study 2

Standard Coupon Number	Surface Residue Level (µgram/25 cm²)	Surface Residue Level (µgram/cm²)
1	10	0.4
2	25	1
3	50	2
4	86	3.4
5	131	5.2
6	186	7.4
7	251	10
8	325	13
9	410	16.4

Figure 2:

Fitted line plot for Study 2



STUDY 2

A second study was conducted with a population of 32 graduate students using the same number of "standard" and "unknown" coupons. However, this time, all coupons were prepared using a spray-coating device that applied a very uniform level of residue on the coupon surfaces. The values of the residue levels are shown in Table 2.

An SOP on how to handle the coupons and perform the visual inspection was created, and all participants were given documented training. All coupons were stored in a box between studies to minimize contamination (e.g., dust). The fitted line plot for the results can be seen in Figure 2.

The following observations were made from this study:

- The R-Sq (adj) increased to 97.9 percent despite the continued non-linearity of coupons 1 to 3.
- The precision and accuracy of the VI increased significantly compared to the first study including coupon number 4.
- However, it appeared analysts' responses accumulated at the ends of the standard range. It was realized that the study design had restricted the students to reporting only 1-9 even if they thought a coupon was cleaner than 1 or dirtier than 9. It also forced them to select one of the standard values (no extrapolation), and this created a bias in the reporting of data. It should be noted that the bias on the low end results in "cleaner" coupons being reported as "dirtier."

STUDY 3

A third study was conducted allowing the participants to report higher or lower than the standard coupons based on what they observed, and a 10th coupon was added to the standards as a "blank" (i.e., no product was applied). Therefore, in this study a total of 60 coupons were used as "unknowns" for analysis. All coupons were again prepared using a spray device that applied residue uniformly on the coupon surfaces. The values of the residue levels are shown in Table 3.

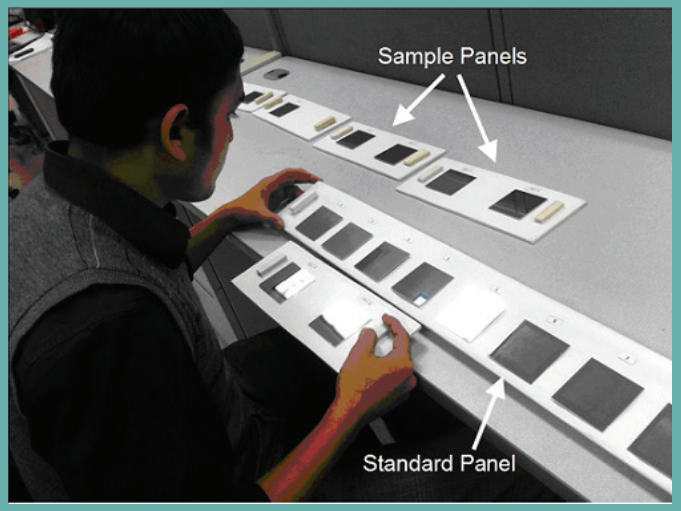
Table 3:

Standard Coupon Residue Levels for Study 3

Standard Coupon Number	Surface Residue Level (µgram/25 cm²)	Surface Residue Level (µgram/cm²)
0	0	0
1	10	0.4
2	25	1
3	50	2
4	86	3.4
5	131	5.2
6	186	7.4
7	251	10
8	325	13
9	410	16.4

Figure 3:

Student evaluating a pair of "unknowns" against the standard panel



For this study, the 10 "visual standard" coupons were affixed to a white panel and labeled 0-9. The 60 "unknown" coupons were affixed to 30 separate panels, each containing two "unknown" coupons. The "sample panels" containing the "unknown" coupon residue were compared to a "standard panel" containing the "standard" coupon residue for students to use to identify the values of the unknown coupon residues (Figure 3).

INSPECTION PROCEDURE

An SOP was written providing training on residue appearance, how to perform the visual inspection, and how to handle both the standard and unknown coupon panels. Thirty students were then trained on this SOP. The panels were laid out on a flat surface under normal indoor lighting to simulate typical manufacturing inspection conditions that might occur (for example, when inspecting cleaned packaging equipment on a wash cart).

The level of illumination during these experiments was determined and found to range from 63 to 93 foot-candles (680 to 1,000 lux). (Note: It was observed that light intensity varies considerably, even from one coupon position to another, so it should be understood that it is not possible to specify that a particular area, including a manufacturing area, is all at one lux level). During the evaluation of coupons, students were directed to sit in an upright position to make the visual angle between 30 to 90 degrees, similar to previously published conditions.[6-10]

During analysis, the students observed only the "unknown" coupon. If no residue was identified on the "unknown" coupon, it was reported as "0" meaning "clean". If any residue was seen, the analyst then compared the amount of residue present on the "unknown" coupon with the amount present on the "visual standard" coupons. A value of 1-9 was then selected for the unknown coupon by choosing the closest match with the visual standard coupons; it was then recorded in the spreadsheet. If any residue was seen that did not match a standard, meaning a concentration greater than #9 (410 microgram/25 cm²), the analyst reported it as a 10 in the spreadsheet. All the data was collected in an Excel spreadsheet and then analyzed using Minitab 16 for accuracy, precision, linearity, and detection limit. The fitted line plot for the results can be seen in Figure 4.

The average responses of each "unknown" closely matched the standard values with an R-Sq (adj) value of 93.1 percent, showing a reduction from Study 2 but still indicating

Figure 4:

Fitted line plot for Study 3

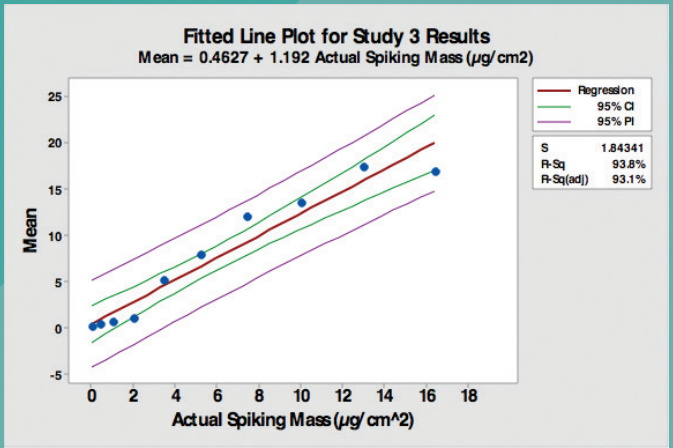
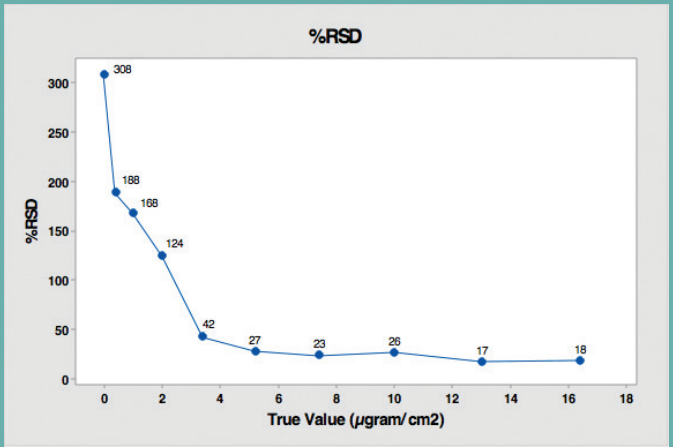


Figure 4:

% RSD for Study 3



a very linear response. The first four responses (0.4 to 3.4 µgram/cm^2) were essentially flat, indicating analysts in general did not "see" any residue on the first four coupons so they reported them as clean (0). Therefore, the amount spiked onto coupon number 4 (3.4 µgram/cm^2) could be considered as a possible detection limit (DL) in this visual inspection study.

Accuracy and precision were evaluated for all "unknown" coupon levels by determining the mean and standard deviation and calculating the bias and the percent relative standard deviation (RSD).

Table 4:

Bias and % RSD for Study 3

Coupon #	0	1	2	3	4	5	6	7	8	9
True Value (µgram/cm^2)	0	0.4	1	2	3.4	5.2	7.4	10	13	16.4
Mean	0.1	0.4	0.6	1.0	5.1	7.9	12.0	13.5	17.4	16.8
STD	0.4	0.8	0.9	1.3	2.1	2.1	2.8	3.5	3.0	3.0
Bias	+0.1	0.0	-0.4	-1.0	+1.7	+2.7	+4.6	+3.4	+4.4	+0.4
%RSD	308%	188%	168%	124%	42%	27%	23%	26%	17%	18%

The data was also evaluated for percent RSD. The standard deviation is very high for the 4 lowest standards (see Figure 5) but then decreased substantially as concentration increased. The lowest standards were the most difficult to see and were under-reported. There was lower, but still significant, variation on the higher standards. These coupons had more drug residue present, and analysts could see them more easily but not distinguish between them very well.

All of the higher standards (coupons 5 to 9) were reported higher then their actual values. For the purposes of cleaning verification, this would be a preferred bias.

This graph also reveals that from coupon number 5, the standard deviation markedly decreased. This indicates analysts were starting to more precisely see the residues on these coupons. So these results indicate that about 85 µgram/25 cm^2, or 3.4 µgram/cm^2 could be considered the DL for visual inspection in this study.

Table 5:

Calculated Detection Limits based on ICH Q2 for All Studies

Study	Slope	Residual Standard Deviation	Detection Limit (µgram/cm^2)
1	1.032	2.19	7
2	0.9107	0.75	3
3	1.192	1.84	5

Calculation of Detection Limit

The DL was calculated based on ICH Q2 using the standard deviation of the response and the slope. The detection limit may be expressed as:

$$\mathrm{DL} = 3.3\ \sigma/\mathrm{S}$$

where:

σ = the standard deviation of the residuals and
S = the slope of the calibration curve

The estimate of slope (S) and the standard deviation (σ) were carried out from the regression equation and residual standard deviation obtained by plotting fitted line plot of the average response versus actual spiking mass (µgram), respectively. The slope, residual standard deviation, and calculated detection limits for each study are shown in Table 3.

In the first study, students were given great latitude in how they performed the inspection, and the DL was the highest. With the second study, the students were provided much better instructions for their inspection, and the DL was more than cut in half. However, they were very restricted in how they reported their results, and when given more freedom the DL almost doubled. But despite what were considered significant differences in the execution of these studies, these calculated DLs are actually fairly close to each other and could be considered equivalent. So the DL for these studies could be the average, or about 5 µgrams/cm^2.

DISCUSSION

From the above analyses, comments from the analysts, and observation of the studies, we found there were four main sources of variation in visual inspection: the preparation of the coupons, the analysts, the inspection method, and the type of residue.

The Coupons

In our first study, coupons were spiked manually. Examination of all of the coupons and comments from the analysts indicate there was some obvious variation in the preparation of the coupons. Residues were noticed to be deposited differently, with some having

clear thick edges and some exceeding the boundaries of the inscribed areas. Scratches on the coupon surfaces, specks of dust dried onto the coupon, and unevenness of residue deposits also caused questions with some analysts. In the second and third studies, the coupons were spiked using an automated spraying device resulting in a uniform coating of residue. The coupons were also stored in a box to avoid deposition of dust. Preparation, storage, and handling of coupons are important parameters in visual inspection.

The Analysts

The total number of analysts in the first study was 72. Examination of data showed a few results that were not believable. There were a small number of cases where coupons with a reference value of "1" were designated as "9," and vice versa. The differences between coupon sets 1 and 9 were too dramatic to accept these results as correct. These students would not confirm their results, so this data was discarded. Some analysts had data entry errors (e.g., illegible, gave two answers) and could not be contacted to rectify the entries, so their results were also removed from the study, reducing the analyst population to 68. The analysts were made up of students in the Pharmaceutical Manufacturing Program attending class, and some may not have clearly understood the purpose of the study or taken it as seriously as they should have. Consequently, training is a very important aspect of visual inspection. We found that without training some analysts did not know what to look for or even what a residue looked like which implies that newly hired operators in industry should be trained so as to know what a residue looks like.

The Inspection Method

In the first and second studies, the analysts were instructed to report the best possible match with the standards. Because of this restriction, there was positive bias on the low end and negative bias on the high end. So in the third study that restriction was removed and the protocol was set up in a way that the analyst could report the residues as they saw them.

The Residue

The visualization of residues can be very dependent on the nature and property of residue. For example, some residues may reflect colors, some are crystal in nature and are easier to see, and some may have a lighter or a darker color than the coupon surface. These properties can make visualizing residues of one compound on a surface different from another compound.

SUMMARY

This study demonstrated the VI of stainless steel surfaces for residues to be linear with R^2 values greater than 90 percent, fairly accurate but with a positive bias, and not very precise with percent RSDs of around 20 percent at levels greater than the detection limit. The detection limits were found to vary from about 3 µgram/cm^2 to about 7 µgram/cm^2 based on how the residue was deposited on the surface and how the inspection was performed. So, detection limits found in literature may not be completely depended on. However, from these studies it does appear there may be a fairly narrow range where residues begin to be readily visible which may lie somewhere between 1 and 10 µgram/cm^2 depending on the type of residue, the training of the inspectors, and how the inspection is performed.

Based on the results of these studies, we highly recommend a training program for VI in order to increase its reliability. Such training may qualify pharmaceutical operators and quality inspectors to see drug residues down to demonstrated levels for confident release of equipment to production after cleaning. The use by the industry of a "standard panel," as employed for these studies, does not seem feasible as it is cumbersome for normal inspection use and has potential issues with ongoing storage and maintenance. A subsequent article will discuss how such qualification of VI for operators and inspectors may be more simply accomplished.

With the advent of the ADE/PDE (acceptable daily exposure/permitted daily exposure), the safe levels of residues on equipment surfaces can now be scientifically determined, and many low-hazard products whose MSSRs (maximum safe surface residues) are well above their visual detection limits may become candidates for VI as a sole criterion for validation.

ACKNOWLEDGEMENT

The authors wish to thank Paul Battle of Environmental Resource Associates for technical assistance and preparation of coupons for these studies.

PEER REVIEWERS

Thomas Altmann, Mallory DeGennaro, Ioanna-Maria Gerostathi, Mohammad Ovais, Osamu Shirokizawa, and Sarra Boujelben

ABOUT THE AUTHORS

PARTH DESAI, MS, is Validation Supervisor at Nostrum Laboratories, Inc. and worked as a research assistant on the investigation of visual inspection, development of total organic carbon and conductivity analysis, and the application of design of experiments in cleaning process development in the Stevens Pharmaceutical Research Center (SPRC) at Stevens Institute of Technology.

ANDREW WALSH, MS, CLSSBB, is president of the Center for Pharmaceutical Cleaning Innovation, a non-profit research and educational organization that supports pharmaceutical, biologics, nutraceutical, cosmetic, and medical device industries through its research into developing new technologies and training on the implementation of science, risk, and statistics-based approaches to cleaning process development and validation. Andrew was also an industry professor in the Pharmaceutical Manufacturing and Engineering program at Stevens Institute of Technology and founded and directed the Stevens Pharmaceutical Research Center.

REFERENCES:

1. PIC/S "Recommendations on Validation Master Plan, Installation And Operational Qualification, Non-Sterile Process Validation, Cleaning Validation," PI 006-3 25 September 2007
2. FDA Guide to Inspections Validation of Cleaning Processes, July 1993, U.S. Food and Drug Administration (FDA)
3. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH Harmonised Tripartite Guideline, Pharmaceutical Development – Q7, November 2000
4. World Health Organization Supplementary Guidelines on Good Manufacturing Practices (GMP): Validation
5. Jenkins, K.M. and Vanderwielen, A.J. "Cleaning Validation: An Overall Perspective." Pharmaceutical Technology Vol 18, No. 4, 1194 pp. 60-73
6. Frank Buscalferri, Dr. Sabine Lorenzen, Dr. Michael Schmidt, Dr. Hans-Martin Schwarm, Dr. Ehrhard Anhalt, Dr. Reinhard Herzog, and Prof. Ronald Ziegler, "Bestimmung der Sichtbarkeitsgrenzen von pharmazeutischen Feststoffen auf Edelstahloberflächen," Pharm. Ind. 62, No 6, 2000
7. Richard J. Forsyth, Vincent Van Nostrand, and Gregory P. Martin, "Visible-Residue Limit for Cleaning Validation and its Potential Application in Pharmaceutical Research Facility," *Pharmaceutical Technology* - Oct 02, 2004
8. Richard J. Forsyth and Vincent Van Nostrand, Using Visible Residue Limits for Introducing New Compounds into a Pharmaceutical Research Facility *Pharmaceutical Technology* - Apr 02, 2005
9. Richard J. Forsyth and Vincent Van Nostrand, "Application of Visible-Residue Limit for Cleaning Validation," *Pharmaceutical Technology* - Oct 02, 2005
10. Vincent Van Nostrand, Richard J. Forsyth, "Application of Visible-Residue Limit for Cleaning Validation," *Pharmaceutical Technology* - Oct 02, 2005
11. Jeffrey Hartman, Vincent Van Nostrand, Richard J. Forsyth, "Risk-Management Assessment of Visible-Residue Limits in Cleaning Validation" *Pharmaceutical Technology* - Sep 02, 2006

JUSTIFICATION & QUALIFICATION OF VISUAL INSPECTION FOR CLEANING VALIDATION IN A LOW-RISK, MULTIPRODUCT FACILITY

ANDREW WALSH
DONGNI (NINA) LIU
MOHAMMAD OVAIS

Proposals for the use of visual inspection (VI) as an analytical method for cleaning validation have been rising for several years now.[1] This article discusses regulatory views on the use of VI as a sole criterion in cleaning validation, presents a case study on how inspectors can be qualified for VI, recommends the use of statistical techniques, and suggests how VI could be implemented as part of a control strategy in a cleaning validation program based on the level of risk.

CURRENT REGULATORY VIEWS

In its 2015 update to Annex 15, the European Medicines Agency (EMA) indicated the possibility of using visual inspection alone in cleaning validation where it stated:

> "*A visual check for cleanliness is an important part of the acceptance criteria for cleaning validation. It is not* ***generally*** *acceptable for this criterion alone to be used.*"[2]

Implied in the wording of this guidance is that VI may be acceptable under certain conditions, as the original wording in the 2014 draft stated, "It is ***not*** *acceptable for this criterion alone to be used.*"

Based on industry interest, and considering this wording change, the possibility of using VI as a sole acceptance criterion was included in the newly released American Society

for Testing and Materials (ASTM) E3106 "Standard Guide for Cleaning Process Development and Validation" under the following conditions:

> "*Using visual inspection alone for validation may be acceptable only when the risk is low and 100 percent of the equipment surface can be inspected under appropriate viewing conditions.*"[3]

This wording was selected to acknowledge that VI may be acceptable within these two criteria. However, E3106 does not provide guidance for determining when a risk is low enough or whether VI might still be used when somewhat less than 100 percent of equipment surfaces can be inspected by VI (e.g., 95 percent). This standard was meant to provide the first defined criteria that could be acceptable to regulators. (Note: Two of this article's authors as well as several peer reviewers were co-authors of the E3106 standard.)

On April 16, 2018, the EMA posted an update to its Q&A on the Guideline for Setting Health Based Exposure Limits (HBELs).[4] In this Q&A, there are two new Q&As (#7 and #8) that are directly applicable to the use of VI. These Q&As state:

> *Q7. Is analytical testing required at product changeover, on equipment in shared facilities, following completion of cleaning validation?*
>
> *A: Analytical testing is expected at each changeover unless justified otherwise via a robust, documented quality risk management (QRM) process. The QRM process should consider, at a minimum, each of the following:*
>
> - *the repeatability of the cleaning process (manual cleaning is generally less repeatable than automated cleaning);*
> - *the hazard posed by the product;*
> - *whether visual inspection can be relied upon to determine the cleanliness of the equipment at the residue limit justified by the HBEL.*
>
> *Q8. What are the requirements for conducting visual inspection as per Q&A 7?*
>
> *A. When applying visual inspection to determine cleanliness of equipment, manufacturers should establish the threshold at which the product is readily visible as a residue. This should also take into account the ability to visually inspect the equipment, for example, under the lighting conditions and distances observed in the field.*

Visual inspection should include all product contact surfaces where contamination may be held, including those that require dismantling of equipment to gain access for inspection and/or by use of tools (for example mirror, light source, boroscope) to access areas not otherwise visible. Non-product contact surfaces that may retain product that could be dislodged or transferred into future batches should be included in the visual inspection.

Written instructions specifying all areas requiring visual inspection should be in place and records should clearly confirm that all inspections are completed.

Operators performing visual inspection require specific training in the process including periodic eye sight testing. Their competency should be proven through a practical assessment.

From these Q&As, there is now sufficient regulatory guidance for industry to begin determining acceptable approaches to implementing VI as a control strategy in cleaning validation programs. This article proposes a systematic and comprehensive approach to address these points that are:

1. Science-based
2. Risk-based and
3. Statistics-based

We believe these are minimum requirements for successfully implementing such programs. The following case study will illustrate how these three aspects were combined using the concepts in ASTM E3106 to implement successful usage of VI at the facility as a sole acceptance criterion for cleaning validation.

CASE STUDY

A pharmaceutical facility was instituting a new cleaning validation program in response to a 483 observation. A risk assessment and cleaning validation studies were performed following the concepts in ASTM E3106.[3] In accordance with ASTM E3106 and ICH Q9, this approach included the following four quality risk management (QRM) steps:

Table 1:

List of APIs and their ADEs

Compound	ADE
API-1	80 µg/day
API-2	2,800 µg/day
API-3	8,300 µg/day
API-4	2,100 µg/day

Table 2:

MSSRs and TOC Swab Limits for All Tanks for All ADEs

Tanks & Kettles	API	Max. Safe Surf. Residue for API (µg/cm²)	TOC Swab Limits for API (ppm)
Tank-1	API-2	991	531
Tank-1	***API-1***	***28***	***17***
Tank-1	API-3	2939	1202
Tank-1	API-4	744	176
Tank-2	API-2	1983	1062
Tank-2	API-1	57	34
Tank-2	API-4	1487	351
Tank-3	API-3	7442	3044
Tank-3	API-4	1883	445
Tank-4	API-2	1475	790
Tank-4	API-1	42	25
Tank-4	API-4	1106	261
Tank-5	API-3	7442	3044
Tank-5	API-4	1883	445
Tank-6	API-3	7442	3044
Tank-6	API-4	1883	445
Tank-7	API-3	5871	2401
Tank-7	API-4	1485	351
Tank-8	API-3	5871	2401
Tank-8	API-4	1485	351
Tank-9	API-3	5871	2401
Tank-9	API-4	1485	351
Tank-10	API-1	54	32

1. Hazard (risk) identification
2. Risk analysis
3. Risk evaluation
4. Risk control

HAZARD (RISK) IDENTIFICATION

All raw materials used for manufacturing the 107 products at the facility were reviewed to determine the level of risk posed by any of these substances to patient safety. The cleaning agent components were also reviewed, as residues of the cleaning agent may have safety implications for patients as well. Over 350 excipients, 10 active pharmaceutical ingredients, and three cleaning agent components were included in this review. Only four of the active pharmaceutical ingredients were identified as posing a potential hazard to patient safety, so acceptable daily exposure (ADE) values were determined for them by a qualified toxicologist/pharmacologist (Table 1).

The other six APIs already had existing safety assessments documenting satisfactory safety profiles at product use levels. All of the cleaning agent components were on the FDA's Select Committee on GRAS [generally recognized as safe] Substances (SCOGS) Database.[5] Therefore, these compounds were considered safe and cleaning validation studies for them were deemed unnecessary based on the level of risk.

RISK ANALYSIS

The risk analysis included selection of analytical methods, calculation of maximum safe carryovers (MSRs), calculation of maximum safe surface residues (MSSRs), and calculation of analytical swab limits from the MSSRs. As these APIs contained organic carbon, total organic carbon (TOC) was chosen as the analytical method. Although TOC is becoming the method of choice for cleaning validation, it is not applicable to APIs with no organic carbon content.

Table 2 shows the lowest possible TOC swab limits calculated for all the tanks in the manufacturing area. The manufacturing tanks have the largest surface areas in the facility and, since limits are inversely proportional to the equipment surface area, only the limits in the manufacturing area were calculated. All TOC swab limits for the other areas (e.g., packaging) would be substantially higher.

Figure 1:

Control chart for all TOC results

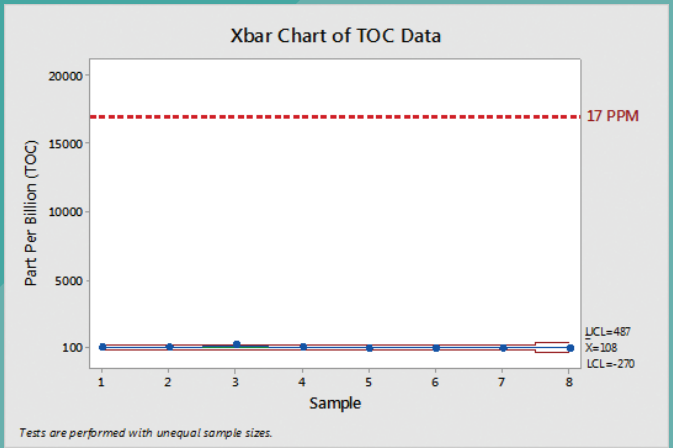
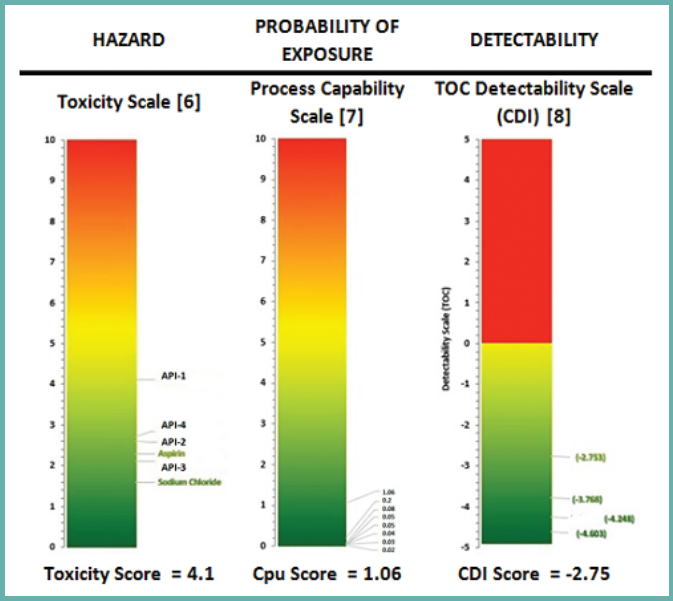


Figure 2:

Cleaning risk dashboard for facility



The lowest possible TOC swab limit was for Tank-1 (largest surface area) for ***API-1*** (lowest ADE). So the TOC swab limit for ***API-1 in Tank-1*** was chosen for setting the acceptance limits for this study.

TOC data was then collected for the cleaning procedures for five kettles, two packaging lines, and the raw material preparation area. The statistical analysis of these data showed these cleaning procedures are ***controlled well below*** the TOC swab limit for ***API-1 in Tank-1*** of 17,000 ppb (17 ppm). The control chart in Figure 1 shows the data for these cleanings compared to the ***API-1 in Tank-1*** (17 ppm). The upper control limit (UCL) for all of the TOC swab data was only 487 ppb, meaning that 99.87 percent of all the TOC swab data fell below this value.

Note: The data for runs #6-8 (packaging lines and raw material preparation area samples) use the Tank-1 limits in this chart, but it should be realized that their actual TOC swab limits would have been much higher than 17 ppm. These data should clearly satisfy the first criterion listed by EMA in its Q&A #7.[4]

CLEANING RISK DASHBOARD

Figure 2 presents a "cleaning risk dashboard" showing the level of relative risk for the cleaning processes at this facility based on three main risk factors associated with cleaning. These risk factors are:

1. The Toxicity Score of **API 1**,
2. The Process Capability Score for **API 1**,
3. The Detectability Score of **API 1** for VI.

It should be noted that all of these scores were derived directly from actual objective data.

The results of the risk assessment and the cleaning validation studies demonstrated that the products and cleaning procedures presented a ***very low risk to patient safety*** from cross contamination due to product carryover after cleaning. This analysis should clearly satisfy the second criterion listed by EMA in Q&A #7. These points above, combined with the ICH Q9[9] principle that the amount of validation effort, the formality, and the level of documentation should be commensurate with the level of risk, made this facility a strong candidate for converting to VI only.

Table 3:

Coupon Data

Compound	ADE
Blank	0
Low Residue Level	0.02 mcg/cm^2
High Residue Level	0.2 mcg/cm^2

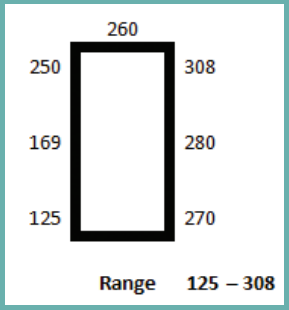
Table 4:

Survey of Light Levels in Facility (lux)

Area	Site	Lux Level
Manufacturing	Tank 1	220
Manufacturing	Tank 2	130
Manufacturing	Tank 3	230
Manufacturing	Tank 4	220
Packaging	Filler	234
Packaging	Filler	280
Packaging	Filler	280
Packaging	Filler	290
Packaging	Filler	236
Raw Material Preperation	Wash Area	130
	Range	***130-290***

Figure 3:

Light levels at coupon evaluation area (lux)



VISUAL INSPECTION QUALIFICATION CASE STUDY

Based on this risk assessment and supported by cleaning validation studies, the manufacturing site decided to institute a VI program. A qualification study that included 33 personnel from four different departments was developed and performed.

MATERIALS AND METHODS

Materials

A total of 30 316L/#4 finish stainless-steel coupons were prepared for the qualification study by "spiking" three different levels of ***API 1*** on to them as shown in Table 3. These coupons were individually numbered from 53 to 82 and randomly assigned to one of the three groups to help prevent inspectors from remembering coupons.

Methods

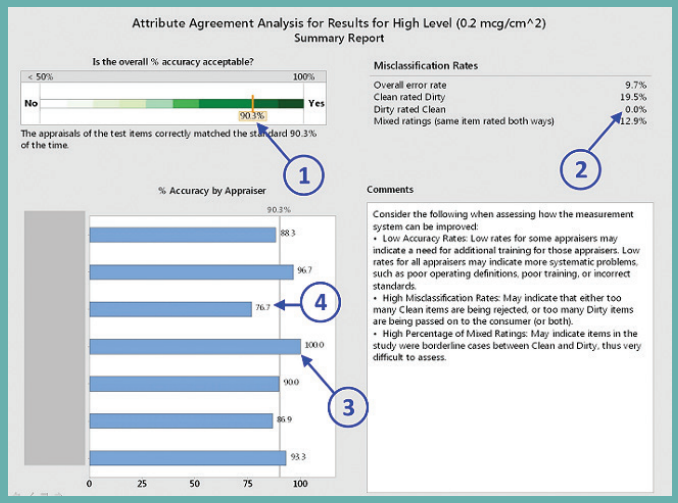
A list of all manufacturing personnel who were to participate in the qualification study was provided to the Center for Pharmaceutical Cleaning Innovation (CPCI). From the list provided, CPCI created a Measurement Systems Analysis for Attribute Data using Minitab 18 and generated data sheets for each inspector. This attribute analysis study was set up for three inspections for each of the inspectors.

Prior to the qualification study, a survey of light levels (in lux) in the manufacturing areas was taken using an Omega HHLM1337 Digital Illuminance Meter, and these results are shown in Table 4.

The coupons were randomly arranged along the edge of a large table in a manufacturing area and the light levels around the edge of the table were recorded (Figure 3) and were equivalent to those found in the other manufacturing areas (Table 4).

All personnel were provided no additional instructions other than to inspect the coupons for product residue as they normally inspect equipment after cleaning and to designate the coupons as either "Dirty" or "Clean" on data sheets provided. Theoretically, the inspectors should identify the blank coupons as clean and the low residue level and high residue level coupons as dirty. All personnel performed these inspections three

Figure 4:
Attribute agreement summary report



times over the course of three days, with the exception of two individuals who had been absent and performed them over two days, with two of the inspections on one day (one in the morning and one in the afternoon). Before each inspection, the coupons were rearranged to prevent the inspectors from remembering the coupons.

In addition to the inspection data, metadata was also collected about the inspectors to examine whether these factors played any role in the results of the study. The metadata collected included the department of the inspector, inspector age, inspector years of service, inspector gender, and whether the inspector wore glasses or not.

RISK EVALUATION (EVALUATION OF RESULTS AGAINST ACCEPTANCE CRITERIA)

The completed data sheets, including the metadata, were sent to the CPCI laboratory in Hillsborough, NJ for analysis. The acceptance criterion for the VRL was the level at which all inspectors could identify the dirty coupons correctly 100 percent of the time. All inspectors were able to correctly identify all of the dirty coupons at the 0.2 mcg/cm^2 level 100 percent of the time but could not do so at the 0.02 mcg/cm^2 level (approximately 90 percent did so).

A data analyst at CPCI analyzed and graphed the data using Minitab 18, R and SAS statistical software. The collected attribute data (Dirty/Clean) was analyzed using Minitab™ 18. Minitab 18 can analyze up to 10 inspectors for attribute analysis at a time. However, each department contained 10 personnel or fewer, so the inspection data was evaluated by department. Figures 4 through 6 show three graphs generated by Minitab that give extensive information on the analyzed data.

1. **Summary Report** - This graph is an overall summary of the results including the overall misclassification rates and the overall percentage accuracy of each inspector. The first graph in the report shows the overall percentage accuracy on a scale that rates the acceptability of the results ①. The misclassification rates show the rates that the clean coupons were rated dirty and the dirty coupons were rated clean. In this case, none of the dirty coupons were rated clean, ② while 19.5 percent of the clean coupons were rated dirty. This can be considered acceptable, since clean equipment that is suspected of being dirty will sim-

Figure 5:

Attribute agreement accuracy report

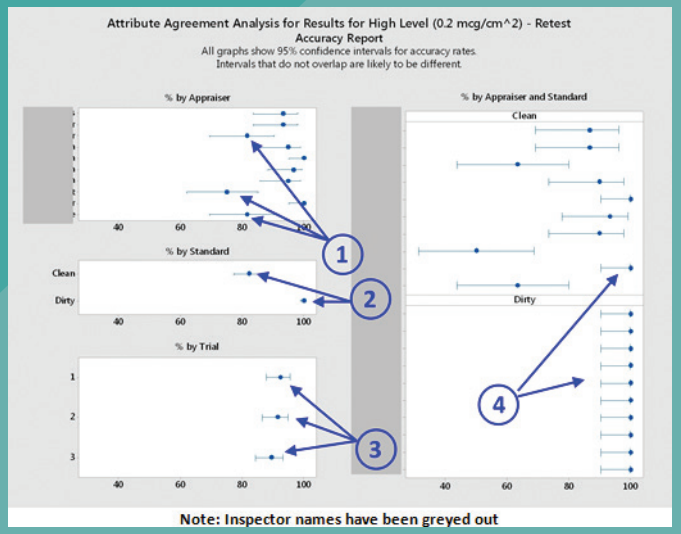
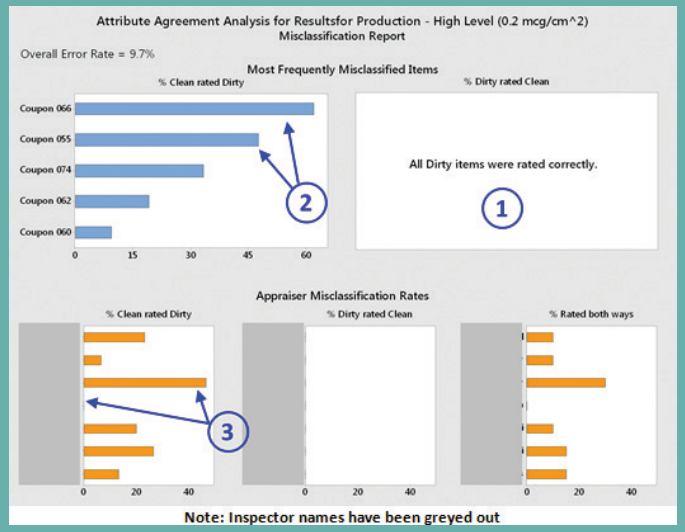


Figure 6:

Attribute agreement misclassification report



ply be cleaned again. The percentage accuracy by appraiser is also shown, and the differences among the appraisers can be seen in this graph. One appraiser was accurate 100 percent of the time, ③ while another appraiser was accurate only 76.7 percent of the time, indicating a possible need for training ④.

2. **Accuracy Report** - This graph shows the percentage accuracy for each inspector by appraiser, by standard, by trial, and by appraiser and standard collectively. Percentage accuracy by appraiser shows that three appraisers are significantly lower in accuracy than the others ①. The percentage accuracy by standard type (dirty or clean) indicates how well the two types of standards were identified. Here we see that the dirty coupons were all identified correctly, while the clean coupons were correctly identified >80 percent of the time ②. The percentage accuracy by trial shows whether there was any decrease over time. It can be seen that the accuracy seemed to decrease slightly as the qualification process went on ③. However, the results for all three trials are within the 95 percent confidence intervals, so the trials can be seen as equivalent at this confidence level. The final graph on the right shows the percentage accuracy of each inspector for both standard types. This is important for this study as it indicates whether all inspectors can correctly identify a dirty coupon at that level. For this group, all appraisers correctly identified the dirty coupons 100 percent of the time and two appraisers correctly identified the clean coupons 100 percent of the time ④. One appraiser incorrectly identified the clean coupons as dirty about half of the time, indicating that this person may require additionawl training.

3. **Misclassification Report** - This graph provides details on the misclassification rates for both the coupons and the inspectors. For the coupons, the percentage of dirty rated clean was 0 ①. For the percentage clean rated dirty, the graph reveals that two of the coupons (#066 and #055) were misclassified at a very high rate ②. These coupons were examined to determine why they had such high misclassification rates. Coupon #066 was found to have a slight discoloration on its surface that was not noticed while preparing the coupons, which many inspectors mistakenly identified as residue. Coupon #055 was found to have a spot on its surface that was not present at the time of preparation and must have occurred during the qualification. This report not only helps explain

the rate of misclassifying the clean coupons as dirty but also points to the need for increased scrutiny of the clean coupons prior to the qualification process and careful monitoring of the coupons during the qualification to prevent errors from occurring. The graphs in the appraiser misclassification rates provide insight into the inspectors. For example, for this group we see that one inspector correctly identified all clean coupons, while another inspector misclassified the clean coupons as dirty over 40 percent of the time ③.

RISK EVALUATION (DETECTABILITY OF CLEANING PROCESS FAILURE)

Table 1 showed the MSSRs calculated for all the tanks in the manufacturing area. The tanks have the largest surface areas in the facility and, since limits are inversely proportional to the equipment surface area, only the limits in the manufacturing area were calculated. As with the TOC swab limits, all the MSSRs for the other areas would be much higher. The lowest possible MSSR was for Tank-1 (largest surface area) for ***API-1*** (lowest ADE). Therefore, ***API-1*** was chosen for this evaluation.

At the 0.2 µg/cm^2 level for ***API-1***, all inspectors correctly identified all of the dirty coupons, while at the 0.02 µg/cm^2 level for ***API-1***, inspectors correctly identified all of the dirty coupons only 90 percent of the time. Therefore, the 0.2 µg/cm^2 level for ***API-1*** level was selected to be the VRL for ***API-1***. A scale for evaluating detectability based on the VRL has been described in a previous article[10] and was used to calculate a detectability score, or visual detection index (VDI), based on a VRL of 0.2 µg/cm^2 for ***API-1*** that was determined in this study.

$$\mathrm{VDI} = \log_{10}\left(\frac{\mathrm{VRL}}{\mathrm{MSSR}}\right) \qquad \textit{(Equation 1)}$$

where

VDI = Visual Detection Index
MSSR = Maximum Safe Surface Residue
VRL = Visual Residue Limit

Calculations for API-1 where:

MSSR = 28 µg/cm²

VRL = 0.2 µg/cm² (based on the results of this study)

$$\text{VDI} = \log_{10} \frac{0.2\ \mu g/cm^2}{28\ \mu g/cm^2}$$

$$\text{VDI} = \log_{10} 0.007\ \mu g/cm^2$$

$$\text{VDI} = -2.15$$

The VDI can be used in conjunction with the ADE-derived toxicity scale[6] and the Cpu-derived probability scale[7] as tools to evaluate the level of risk in cleaning validation. Going further, the toxicity scale could also help define the circumstances for VI (low hazard) and the detectability scale can provide the justification (easy to see at levels well below the MSSR for that hazard).

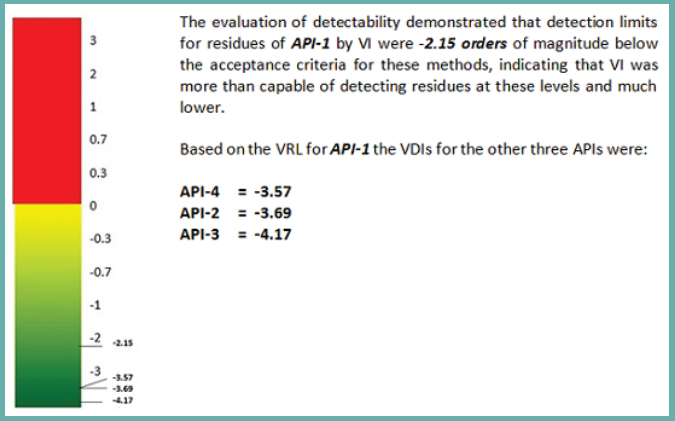
For analysis of the metadata collected during the study (department, age, years of service, gender, and use of glasses), click here.

DISCUSSION OF RESULTS

The qualification study for performing VI for the manufacturing equipment demonstrated that all personnel can identify the presence of ***API 1*** residues on manufacturing equipment surfaces at a level of **0.2 µg/cm²** correctly 100 percent of the time.

- This **0.2 µg/cm²** level is more than 100 times lower than the lowest ADE-based residue limit of ***28 µg/cm²*** for ***API 1*** in ***Tank-1***. Considering that all other API limits are more than 10 to more than 1,000 higher than ***API 1***, these compounds would also be clearly visible at these levels. Since all the visible residue limits are much higher for all other equipment, this qualification is considered to apply to these products as well.
- The evaluation of detectability also demonstrated that detection limits for residues by TOC and VI were 2.75 and 2.15 orders of magnitude, respectively, below the acceptance criteria for these methods, indicating both of these methods were more than capable of detecting residues at these levels.

Figure 7:
Detectability score for API-1



- No noticeable differences were found with VI regarding the department, age, years of service, or the gender of the personnel or whether they wore glasses or not and these are not factors that affect VI.

The results of this qualification study, including the analysis of the metadata, should satisfy the third criterion listed by EMA in its Q&A #7.

Therefore, it was recommended that this facility could move from TOC swab testing to VI for future studies and to qualify all operators and inspectors on VI. Unless a new product is introduced with a lower ADE than ***API-1***, VI will be ***considered acceptable as the sole criterion for the cleaning validation acceptance limit for all products***, current and future, manufactured at this facility.

Implementing Visual Inspection As Part Of A Cleaning Control Strategy

Based on the data collected, their analysis, and the experiences in this case study, the following observations and recommendations on implementing VI programs can be made.

Coupons

During this study, a number of observations were made about the coupons used. As also noted in the previous article,[1] coupons can be easily damaged or contaminated and this could affect the results of the study, so storage, handling, and maintenance of coupons are important. Coupons for this study were kept in a storage box and each coupon had "feet" added to each corner of the underside of the coupon to facilitate handling. All the coupons were labeled as to the material of construction (316L SS/#4 Finish), with the date (month/year) of manufacture, and individually numbered. (Figure 8).

Coupon Preparation

For VI qualification studies to be valid, the coupons must be prepared in a manner that leaves a residue on the coupon that is the same in appearance as will be encountered in the manufacturing area. Evaporative drying has been studied for many solvents, including water, and there are significant differences in the deposition patterns of residues depending on the solvent.[11] Consequently, the improper preparation of coupons may lead to erroneous conclusions. Some workers have been using solvents (e.g., methanol) to deposit the compounds and drying them under conditions not encountered in operations

Figure 8:
Coupons and storage boxes

(e.g., under a nitrogen stream). Such techniques are not recommended. The coupons for this study were spiked and dried in a manner that simulated the actual conditions in the facility's manufacturing area. ***API-1*** was dissolved in purified water, spiked onto the coupons and then dried in an oven at 90°C. This procedure simulated the actual conditions in operations, that is, a hot purified water rinse with hot equipment surfaces for the ***API-1*** residue to dry quickly on. After preparation, all coupons should be examined to ensure they have been prepared correctly, including verifying that the blank coupons do not have stains, scratches, or fingerprints that may mislead the inspectors and confound the qualification study as has been pointed out above. Also, for one product to represent other products in a VI study, the residues of the other products must be similar in appearance (e.g., a white residue may not be similar to a blue residue).

Inspector Training

Most inspectors throughout the industry have been inspecting and releasing manufacturing equipment for many years and most know through experience what product residues look like and can accurately identify them. However, some may have not seen, or have not been formally shown, product residues so they may not truly be sure what product residue looks like. This study revealed that a few inspectors can miss product residues that should be seen, and some misidentified a discoloration on a clean coupon as product residue. In this study, no training was provided deliberately to see how well the inspectors would do without training. While the majority did not require specific training, it was clear that training of inspectors prior to the study on what to look for is both beneficial and necessary. However, this training should be provided as a means to identify product residues on manufacturing surfaces and not as a means to identify product residues on coupons. Therefore, inspectors are best trained using residues on actual manufacturing equipment and not with coupons. The same training for identifying product residues on ***manufacturing surfaces*** should be sufficient for inspectors to accurately identify product residues on ***coupons***. This would further help legitimize the results of a VI qualification study. While SOP training is necessary and required, it must go beyond training the inspectors to look at equipment and fill out inspection forms.

Viewing (Lighting) Conditions

Light levels are typically suspected by most industry workers to be critical parameters, but experiments and experience has not held this to be true. Some studies have been

performed showing no differences in inspection when light levels are between 200 and 1,400 lux.[12] This should not be considered unusual. The human eye is capable of rapid adaptation to changing light levels over a very wide range of intensities and the eye adapts to minor differences in light levels almost instantaneously and unnoticeably.[13] Therefore, minor changes in light levels, or minor changes in distance or the angle of viewing during inspection, may have little impact on the ability to inspect successfully. As mentioned, the human eye is very compensating; regardless, workers performing VIs should still be trained to correctly identify product residues and ensure that an appropriate inspection is performed.[14]

Documentation Review

A means of documenting VI on a continuous basis for routine monitoring should also be implemented. The documentation level for VI should be commensurate with the level of risk and the complexity of the equipment being inspected. In addition, documentation should be reviewed for trends and anomalies as a part of knowledge management program.

Attribute Analysis

The use of attribute analysis for statistically analyzing the data collected in this study greatly increased the amount of information about the inspection, while being relatively easy and simple to perform. Since there were only two levels used in the study (0.2 and 0.02 µg/cm^2), with the 0.2 level achieving 100 percent accuracy and the 0.02 only achieving 90 percent accuracy, the designation of the VRL was set to 0.2 µg/cm^2. However, it may well be the case that the accuracy could be 100 percent at the 0.15 µg/cm^2 or even at the 0.1 µg/cm^2 level. Logistic regression is a powerful statistical analysis that could be used to determine the VRL exactly by using several levels.[15] A subsequent study is planned to explore the use of logistic regression while still being designed to be simple and easy to implement.

VI Qualification Programs

Most cleaning validation workers are well aware of the time and resources involved in developing and validating swab methods, the sampling of equipment, analyzing the samples, and releasing the equipment for use. While the idea of using VI only may seem simple and very attractive, the process of qualifying a large group of inspectors for VI should not

consume an equal or greater amount of time and resources. For VI to become desirable, valuable, and accepted, the qualification of VI must also be an easy program to implement, document, and maintain (requalification), accepted by regulators, and ultimately valid.

The approach described in this article was relatively simple to set up, execute, and analyze. The coupons were prepared in less than one day. The set up and collection of data took only three days and only a total of about 15 minutes of each inspector's time (three inspections x five minutes each). The statistical analysis of the data took approximately three days by one analyst. This is a reasonable amount of effort and time that yielded a great deal of process knowledge and understanding.

We believe that, going forward, the products selected for VI must be low hazard products based on their HBELs (science-based), should have demonstrated reliable cleaning processes that do not present any significant concerns for patient safety (risk-based), the VI data collected must be analyzed appropriately to demonstrate that the VI is valid (statistics-based), and the VRL should be shown to be well enough below the MSSR to be legitimate to use. As shown above, statistical analysis of VI data is very revealing and can identify issues with inspectors and problems with coupons and provides significant insight into the inspection process, so it should not be considered optional.

SUMMARY

As stated in the introduction, we believe that this study met all the criteria provided by the EMA's Annex 15 Guideline and the new Q&As #7 and #8. The authors hope that the study described in this article will satisfy regulatory concerns, increase the science, risk analysis, and use of statistics behind qualifying VI, and help other companies to implement VI on science-based and risk-based foundations.

PEER REVIEW

The authors wish to thank our peer reviewers Bharat Agrawal; Thomas Altmann; James Bergum, Ph.D.; Alfredo Canhoto, Ph.D.; Gabriela Cruz, Ph.D.; Mallory DeGennaro; Parth Desai; Kenneth Farrugia; Ioanna-Maria Gerostathi; Igor Gorsky; Miquel Romero Obon; Laurence O'Leary; and Osamu Shirokizawa for reviewing this article and for their insightful comments and helpful suggestions.

REFERENCES

1. Desai, Parth and Andrew Walsh, "*Validation Of Visual Inspection as an Analytical Method For Cleaning Validation*", Pharmaceutical Online, September 2017
2. EudraLex, Volume 4 – Guidelines for Good Manufacturing Practices for Medicinal Products for Human and Veterinary Use, Annex 15: Qualification and Validation, available from: https://ec.europa.eu/health/documents/eudralex/vol-4_en.
3. American Society for Testing and Materials (ASTM) E3106-17 "*Standard Guide for Science and Risk-Based Cleaning Process Development and Validation*"
4. Questions and answers on implementation of risk-based prevention of cross-contamination in production and 'Guideline on setting health-based exposure limits for use in risk identification in the manufacture of different medicinal products in shared facilities' (EMA/CHMP/CVMP/SWP/169430/2012)
5. Food and Drug Administration Generally Recognized As Safe compounds data base https://www.fda.gov/Food/IngredientsPackagingLabeling/GRAS/SCOGS/default.htm
6. Walsh, Andrew, Ester Lovsin Barle, Michel Crevoisier, David G. Dolan, Andreas Flueckiger, Mohammad Ovais, Osamu Shirokizawa, and Kelly Waldron, "*An ADE-Derived Scale For Assessing Product Cross-Contamination Risk In Shared Facilities*", Pharmaceutical Online, May 2017
7. Walsh, Andrew, Ester Lovsin Barle, David G. Dolan, Andreas Flueckiger, Mohammad Ovais, Osamu Shirokizawa, and Kelly Waldron "*A Process Capability-Derived Scale For Assessing Product Cross-Contamination Risk In Shared Facilities*", Pharmaceutical Online, August 2017
8. Walsh, Andrew, Thomas Altmann, Alfredo Canhoto, Ester Lovsin Barle, David G. Dolan, Andreas Flueckiger, M.D., Igor Gorsky, Robert Kowal, Mariann Neverovitch, Mohammad Ovais, Osamu Shirokizawa and Kelly Waldron. "*A Swab Limit-Derived Scale For Assessing The Detectability Of Total Organic Carbon Analysis*", Pharmaceutical Online, January 2018
9. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH Harmonised Tripartite Guideline, Quality Risk Management – Q9, Step 4, November 2005, www.ich.org.

10. Walsh, Andrew, Ester Lovsin Barle, David G. Dolan, Andreas Flueckiger, Mohammad Ovais, Osamu Shirokizawa, and Kelly Waldron "*An MSSR-Derived Scale For Assessing the Detectability of Product Cross-Contamination Risk In Shared Facilities*", Pharmaceutical Online, December 2017
11. Xuefeng Xu and Jianbin Luo "*Marangoni flow in an evaporating water droplet*" American Institute of Physics, Applied Physics Letters 91, 124102, 2007
12. Forsyth, R.J., and Hartman, J., "*A Risk-based Approach to Cleaning Validation using Visible Residue Limits*," Pharmaceutical Engineering, May/June 2008, Vol. 28 No. 3, pp. 8-22, www.pharmaceuticalengineering.org.
13. Levin, L.A., et al., *Adlers' Physiology of the Eye: Expert Consult*, Elsevier Health Sciences, 11th Edition, 2011, www.us.elsevierhealth.com.
14. Melchore, J.A., "*Sound Practices for Consistent Human Visual Inspection*," AAPS PharmSciTech, March 2011, Vol. 12, No. 1, pp. 215-221, http://link.springer.com/journal/12249.
15. Ovais, M., "*Statistically Justifiable Visible Residue Limits*," Pharmaceutical Technology, March 2010, Vol. 34, Issue 3, pp. 58-71, www.pharmtech.com.

ABOUT THE AUTHORS

DONGNI (NINA) LIU is a Cleaning Validation Specialist/Data Analyst at the Center for Pharmaceutical Cleaning Innovation

MOHAMMAD OVAIS MPharm is a Cleaning Validation Consultant. He has more than thirteen years of hands-on experience in validating cleaning processes. He provides training and consulting services in process and cleaning validation.

ANDREW WALSH is president of the Center for Pharmaceutical Cleaning Innovation, a non-profit research and educational organization.

MEASURING RISK IN CLEANING: CLEANING FMEAS AND THE CLEANING RISK DASHBOARD

ANDREW WALSH
THOMAS ALTMANN
ALFREDO CANHOTO, PH.D.
ESTER LOVSIN BARLE, PH.D.
DAVID G. DOLAN, PH.D.
ANDREAS FLUECKIGER, M.D.
IGOR GORSKY
JESSICA GRAHAM, PH.D.
ROBERT KOWAL
MARIANN NEVEROVITCH
MOHAMMAD OVAIS
OSAMU SHIROKIZAWA
KELLY WALDRON, PH.D.

This article discusses the concept and measurement of risk as it applies to the cleaning of pharmaceutical products. Four previous articles discussed how science-based data-derived scales could be created using compound HBELs (health-based exposure limits), from the process capability (Cpu) of the products' cleaning processes and from the detection limits for visual inspection or for total organic carbon (TOC) analyses of these compounds.[1-4] This article continues the discussion about the potential use and application of these new scales in cleaning failure modes and affects analysis (cleaning FMEA) to assist in measuring the risk of cleaning process failures as well as how these scales can be applied to develop a cleaning risk dashboard. The article will also discuss how these new scales can be utilized to accelerate new product introductions.

Note: This article uses the term health-based exposure limit (HBEL), which is synonymous with the terms acceptable daily exposure (ADE) and permitted daily exposure (PDE).

WHAT IS RISK AND WHY IS MEASURING IT IMPORTANT IN CLEANING?

Most people will tell you they know what risk is, and they can give clear examples of risks in their lives. But if asked, they will not know, or will have difficulty identifying, what the underlying components of risk are. This is probably because most people have come to understand risk through personal experience and not through any formal study

of risk or its measure. Historically, risk has not been very well understood or evaluated properly.[5] For example, many people consider all snakes to be dangerous and a risk although only some snakes are actually poisonous and many are harmless and even beneficial. Similarly, while some drugs may be hazardous, that does not mean all of them should be considered a high risk. While risk management has been in use in various industries for many years, it has been seriously misconstrued.[6, 7] These problems also apply to the consistency of hazard classification and risk assessment of chemicals.[8]

In 2005, risk was defined for the pharmaceutical industry in the International Council on Harmonization Quality Risk Management Guideline (ICH Q9), which was formally adopted by the FDA in 2006.[9] As stated in ICH Q9:

> "It is commonly understood that risk is defined as the combination of the ***probability of occurrence*** of harm and the ***severity*** of that harm."

and further on:

> "The ability to detect the harm (***detectability***) also factors in the estimation of risk."

In ICH Q9 we see risk deconstructed into two subparts: severity and probability, and a third element of possible prevention, detectability. If we could measure these two (or three) subparts as they apply to the cleaning of healthcare products, we could then determine what the level of risk is for cleaning validation and ultimately for a cleaning process. Why would measuring risk be important for cleaning validation? Most importantly because of a regulatory concern of ICH Q9 asserting that the two primary principles of quality risk management are:

- "The evaluation of the risk to quality should be based on scientific knowledge and ultimately link to the protection of the patient; and
- The level of effort, formality, and documentation of the quality risk management process should be commensurate with the level of risk."

From these two primary principles it can be understood that if we can determine the level of risk to a patient from cleaning, then the level of cleaning validation effort, its formality, and its documentation can be adjusted based on that risk. More simply, cleaning validation for low-risk situations should not require the same level of effort as for high-

risk situations. This is quite logical. The level of effort, formality, and documentation of cleaning validation should be scaled to the level of risk, as well as the available knowledge of a cleaning process. ICH Q9 clearly states that these principles are applicable to validation (in Annex II.6). Moreover, they apply to cleaning, including setting acceptance limits for cleaning processes (in Annex II.4). So, cleaning validation efforts, formality, and documentation should be adjusted based on the level of risk(s) identified in a risk assessment (RA) and managed through a quality risk management system.

While that may be good news, an article in 2015 by Kevin O'Donnell of the Health Products Regulatory Authority asserted that the implementation of quality risk management in the pharmaceutical industry may have been riddled with misunderstandings.[10] One of the issues with risk management he identified was a lack of sound scientific principles being used in that the "probability of occurrence estimates are not based on any kind of historical data, preventative controls, or on modeling data," and that there have been "assumptions regarding risk severity and detection that are totally unsound." Another issue was making "important decisions based on Risk Priority Number (RPN) values which fail to recognize that those values are derived only from ordinal scale numbers" and "are not mathematically meaningful" and that these RPNs are often "associated with high levels of subjectivity, uncertainty and guesswork."[10] Other recent articles have explored the weaknesses of the use of risk matrices to derive RPNs.[11-18]

Clearly, it would be very helpful if the pharmaceutical industry had the means to measure these elements of risk based on sound scientific principles. The scales presented in the first four articles[1-4] offer science-based answers to these issues – specifically with regard to cleaning – that can be readily utilized in meaningful, measurable, and practical risk-based approaches.

Going back to ICH Q9, we see risk can be formally expressed as:

$$Risk = f\,(Severity\ of\ Hazard,\ Level\ of\ Exposure\ to\ Hazard,\ Detectability\ of\ Hazard)$$

Now, if the hazard is ***intrinsic to an active pharmaceutical ingredient (API) and the risk being considered is harm to a patient from exposure to residues of that API after cleaning***, then this equation can be further refined to:

$$Cleaning\ Risk = f\,(Toxicity_{API\ residue},\ Level\ of\ Exposure_{API\ residue},\ Detectability_{API\ residue})$$

Since the scales presented in the previous four articles are all based on good science and derived from actual data, they would consequently make good choices to use for evaluating the risk in cleaning.

MEASURING CLEANING RISK: CLEANING FMEAS

One of the most commonly used tools for risk assessment, widely used in the pharmaceutical industry, is the FMEA. The FMEA is considered a systematic, comprehensive, and powerful tool for performing risk management and has also been adapted for the evaluation of processes, so it fits well into the assessment of cleaning processes. The FMEA was developed by the U.S. military shortly after World War II and published as MIL-P-1629.[19] It was adopted for use by NASA and the aviation industry in the early 1960s, then in the 1970s by the automotive industry. It was adopted later by many other industries, eventually making its way into international standards such as ASTM and ISO, but only in recent years has it been implemented in the pharmaceutical industry.

FMEAs typically use three criteria in their evaluation of failure modes or hazards that fit well in the ICH Q9 definition of risk:

1. Severity (of the hazard)
2. Occurrence (probability of the hazard)
3. Detectability (of the hazard)

Once a failure mode is identified, the severity of the effect of the failure, the likelihood of its occurrence, and the ability to detect this failure are then determined. In the FMEA, these three criteria are normally evaluated using ordinal scales that can range from 1-10, 1-5, 1-3 (Low/Medium/High), or other combinations, with 1 being the lowest score and 3, 5, or 10 being the highest. Table 1 shows some general rating scores used in FMEAs.[20]

After the values are selected from the three categories, they are subsequently multiplied to arrive at an RPN, which is typically used to rank failures and prioritize them for any needed actions (e.g., when the identified number is above a specified RPN, remedial actions must be taken, and when the number is below a specified RPN, no remedial actions are required). For scales that use 1-10 scoring, the possible range of RPNs is

Table 2:

General Rating Scales for FMEA[20]

Rating	Degree of Severity	Likelihood of Occurrence	Ability to Detect
1	Customer will not notice the adverse effect or it is insignificant	Likelihood of occurrence is remote	Sure that the potential failure will be found or prevented before reaching the next customer
2	Customer will probably experience slight annoyance	Low failure rate with supporting documentation	Almost certain that the potential failure will be found or prevented before reaching the next customer
3	Customer will experience annoyance due to the slight degradation of performance	Low failure rate without supporting documentation	Low likelihood that the potential failure will reach the next customer undetected
4	Customer is made uncomfortable or their productivity is reduced by the continued degradation of the effect	Occasional failures	Controls may detect or prevent the potential failure from reaching the next customer
5	Warranty repair or significant manufacturing or assembly complaint	Relatively moderate failure rate with supporting documentation	Moderate likelihood that the potential failure will reach the next customer
6	Warranty repair or significant manufacturing or assembly complaint	Moderate failure rate without supporting documentation	Controls are unlikely to detect or prevent the potential failure from reaching the next customer
7	High degree of customer dissatisfaction due to component failure without complete loss of function. Productivity impacted by high scrap or rework levels.	Relatively high failure rate with supporting documentation	Poor likelihood that the potential failure will be detected or prevented before reaching the next customer
8	Very high degree of dissatisfaction due to the loss of function without a negative impact on safety or governmental regulations	High failure rate without supporting documentation	Very poor likelihood that the potential failure will be detected or prevented before reaching the next customer
9	Customer endangered due to the adverse effect on safe system performance with warning before failure or violation of governmental regulations.	Failure is almost certain based on warranty data or significant DV testing	Current controls probably will not even detect the potential failure
10	Customer endangered due to the adverse effect on safe system performance without warning before failure or violation of governmental regulations.	Assured of failure based on warranty data or significant DV testing	Absolute certainty that the current controls will not detect the potential failure.

therefore from 1 to 1,000 (S*O*D). So, for example, if the Severity Score = 5, the Probability Score = 9, and the Detectability Score = 8, the resulting RPN would be 360.

A review of the descriptions and definitions in Table 1 will quickly reveal that these factors do not directly translate to many pharmaceutical operations. The consequences of manufacturing failures affecting pharmaceutical products, such as a cleaning failure, are substan-

Figure 1:
Distribution of RPN results (used with permission of the author)

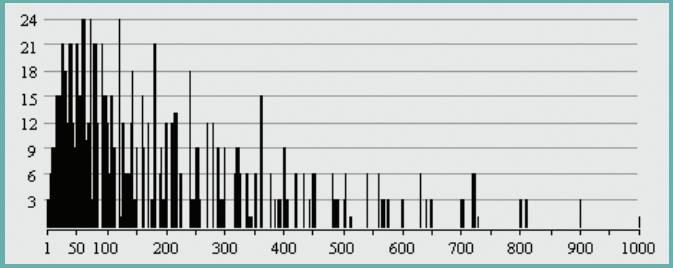


Figure 2:
Fifteen "equivalent" problems having an RPN = 360 (used with permission of the author)

	Severity of Problem		Likelihood of Occurrence		Likelihood of Detection	
1	Hazardous	10	Very High	9	Mod. High	4
2	Hazardous	10	Moderate	6	Low	6
3	Hazardous	10	Moderate	4	Very Remote	9
4	Hazardous	9	Very High	10	Mod. High	4
5	Hazardous	9	High	8	Moderate	5
6	Hazardous	9	Moderate	5	Remote	8
7	Hazardous	9	Moderate	4	Impossible	10
8	Hazardous	8	Very High	9	Moderate	5
9	Hazardous	8	Moderate	5	Very Remote	9
10	Moderate	6	Very High	10	Low	6
11	Moderate	6	Moderate	6	Impossible	10
12	Low	5	Very High	9	Remote	8
13	Low	5	High	8	Very Remote	9
14	Very Low	4	Very High	10	Very Remote	9
15	Very Low	4	Very High	9	Impossible	10

tially different from the failures that might affect other unrelated industries. There is therefore a need for pharmaceutical companies to establish more appropriate definitions and descriptions for each of these values within their organizations that are truly reflective of the realities of their operations. Compounding this challenge are issues with different stakeholders, such as QA, technical services, and operations, having widely different opinions on what is a correct score, since most definitions are general, subjective, and debatable.

Beyond these difficulties and the issues mentionwed above,[11-18] there are other issues with the traditional FMEA approach that have been identified and described by Donald J. Wheeler.21 In his article, Wheeler points out that while the possible RPNs range from 1 to 1,000, an actual calculation of these RPNs results in a very skewed distribution of only 120 possible actual results (Figure 1).

Wheeler goes on to show that there are no fewer than 15 combinations that could result in an RPN of 360, some of which could be considered critical and others, perhaps, not so much. So the RPN numbers derived using these subjective scales have the potential to be very misleading (Figure 2).

Wheeler further explains that the ordinal scales typically used in FMEAs cannot be multiplied legitimately. Looking at the definitions of the scores in Table 1 and the example results in Figure 2, it quickly becomes obvious that the RPN values from their multiplication have no particular or practical meaning.

Wheeler goes on to suggest that instead of multiplying them, these scores should remain as they are and the severity (S), occurrence (O), and detectability (D) scores could simply be expressed as a numerical string -- SOD. For example, SOD = 937, or SOD = 396. This approach would maintain the integrity of the original scores, which could allow for more appropriate ordering. This also enables a reviewer to see where quantitative improvements were made after any recommended actions were taken. For example, if a failure mode had an SOD of 978, and the new score was 965, it would be clear that a small decrease was made in the occurrence and a greater improvement made in the detectability. However, when the scores are converted to RPN values, they would be 504 and 270, which would seem to be a significant overall improvement, while in reality there only was a small improvement. Therefore, the magnitude of calculated numbers is very misleading and the actual "how it happened" is unclear.

SCIENCE- AND RISK-BASED SCALES FOR SEVERITY, OCCURRENCE, AND DETECTABILITY

The subjectivity of the FMEA scales typically used, and the lack of a scientific/statistical basis for their RPN numbers, make both these scales and their RPNs unacceptable for use in the pharmaceutical industry. If pharmaceutical manufacturing is to advance to a science- and risk-based approach, the scales for severity, occurrence, and detectability used in FMEAs must be scientifically justified using scientific principles, process knowledge, and statistics. These scales should be derived from, and based on, empirical data. Such data exists for cleaning and is readily available in pharmaceutical manufacturing production. As stated in the introduction, scales already exist that can be used for the following criteria:

1. HBEL-derived Toxicity Scale for Severity of Process Residues[1]
2. Cpu-derived Scale for Occurrence of Exposure to Process Residues[2]
3. Visual Detectability Index for Detectability of Process Residues[3]
4. TOC Detectability Index for Detectability of Process Residues[4]

For example in a cleaning process, if a failure mode could result in residues of an API remaining on equipment, then the HBEL-derived toxicity score of that API would replace the severity score. Furthermore, if the process capability of the cleaning process is known, then its Cpu-derived score could replace the occurrence score (as the cleaning process effectiveness and the probability of residues are known). Finally, if either the visual detectability index[3] or the TOC detectability index[4] is known, one or both of these could replace the detectability score. Since these scores are derived directly from empirical data, their values are specific, objective, and nondebatable.

For a refresher on these scales, please see the following articles:

- An ADE-Derived Scale For Assessing Product Cross-Contamination Risk In Shared Facilities
- A Process Capability-Derived Scale For Assessing The Risk Of Compound Carryover In Shared Facilities
- An MSSR-Derived Scale For Assessing Detectability Of Visual Inspection
- A Swab Limit-Derived Scale For Assessing The Detectability Of Total Organic Carbon Analysis

Figure 3:
Risk hierarchy of analytical methods [Note: Toxicity scale is based on −log(HBEL) where HBEL is the health-based exposure limit in grams]

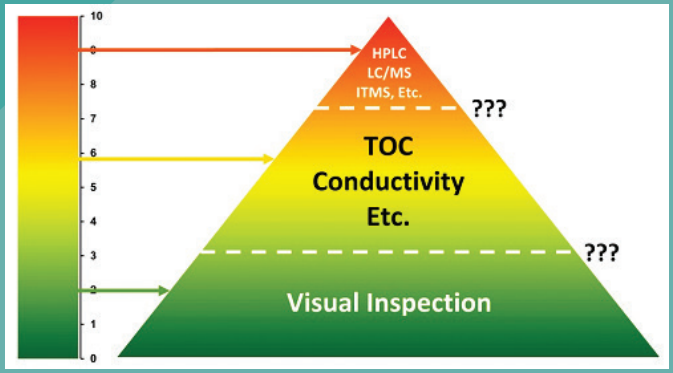


Table 2:
Method Selection Based on Detectability Scores

Compound	CDI	VDI	Method Selection
Drug 3	3.2	1.3	HPLC/other selective method
Drug 2	1.2	1.9	HPLC/other selective method
Drug 8	0	3.2	HPLC/other selective method
Drug 9	-1.4	-1.6	TOC
Drug 5	-1.6	-0.7	TOC
Drug 4	-1.8	-2.2	TOC/Visual
Drug 7	-2.1	-2.4	TOC/Visual
Drug 1	-3.2	-4.4	Visual
Drug 10	-3.5	-3.7	Visual
Drug 6	-4.1	-5.2	Visual

Note: In all cases a visual inspection will still be done.

USING THE DETECTABILITY SCALES FOR METHOD SELECTION

In the previous articles on detectability scales,[3, 4] it was suggested that the selection of the analytical methods used in cleaning validation studies should be based on the level of risk. These articles showed a diagram (Figure 3) that linked the selection of analytical methods to the toxicity scores of compounds. Compounds of low toxicity (lower risk) might only use visual inspection, while compounds of high toxicity (higher risk) might require advanced selective methods. However, when to transition from one group of methods to another is unclear from this figure, and these articles presented detectability scales for visual inspection and TOC that could guide the selection process based on actual data.

Table 2 shows how detectability scores derived using the calculations from the detectability articles[3, 4] could be used to determine the most advisable risk-based approach for 10 drugs.

For the 10 drugs in Table 2, the hypothetical criteria used for selecting a TOC method was at least 1 log below zero and for using visual inspection was at least 2 logs below zero. (Note: Companies will need to select their own criteria based on their level of risk acceptance.) So for Drugs 2, 3, and 8, selective methods are necessary as they are well above zero. For Drugs 5 and 9, TOC is acceptable, but visual inspection is not, and for Drugs 4 and 7, both TOC and visual inspection are acceptable. Visual inspection alone would be acceptable for Drugs 1, 6, and 10, as they are well below -2 logs.

USING THE TOXICITY AND CPU SCALES TO MEASURE CLEANING RISK

As both the HBEL-based toxicity scale for severity of hazard and the Cpu-based process capability scale for probability of exposure (occurrence) are not arbitrary values, they consequently have real significance. The toxicity and probability of exposure may be evaluated first, and then detectability can be considered for prioritization when the toxicity and probability of exposure of two hazards are equal. Table 3 shows the toxicity and process capability scales side by side from the highest to the lowest possible values.

In the article on Cpu-based process capability scale,[2] a table was shown (Table 4) asking the reader to select the risk ranking for 10 hypothetical drugs based on these SO scores.

Table 3:

Calculating Cleaning Risk Using the Toxicity and Cpu ScalesScores

Toxicity Score	Process Capability Score	SO (Risk) Scores	
10	10.0	10	10
9	9.0	9	9
8	8.0	8	8
7	7.0	7	7
6	6.0	6	6
5	5.0	5	5
4	4.0	4	4
3	3.0	3	3
2	2.0	2	2
1	1.0	1	1

Table 4:

Example Drug Scores and RPNs in Shared Equipment Facilities

Compound	SO Score		Traditional RPN Score	Risk Rank? (Rank 1-10)
Drug 1	2	9	18	?
Drug 2	9	2	18	?
Drug 3	7	3	21	?
Drug 4	5	5	25	?
Drug 5	10	4	40	?
Drug 6	1	10	10	?
Drug 7	4	6	24	?
Drug 8	8	1	8	?
Drug 9	6	8	48	?
Drug 10	3	7	21	?

The following considerations are proposed to answer the question in that article:

- Drug 1 and Drug 2 have the same RPN scores, but the cleaning procedure for Drug 1 needs considerable improvement to assure that any residues after cleaning are at safe levels, while Drug 2 does not. However, the traditional RPNs assign them an equal level of risk.
- The traditional RPN method puts Drug 9 as the highest risk (RPN = 48), but it is not highly toxic, although its cleaning process is not very effective. Based on its high RPN, it is followed by Drug 5, which is highly toxic, although its cleaning process is very effective.
- Conversely, Drug 6, with a low toxicity, has a very poor cleaning process that is assured to leave residues, but it has the second lowest RPN score.

It should be evident that multiplying these scores obscures the important information found in the individual scores. More importantly, it can lead to poor risk analysis and decisions. So, keeping the raw scores is appropriate. The remaining question is how the risk is objectively analyzed. One possible way is to give priority to the toxicity scores. Table 5 shows the same data as Table 4 sorted from the highest toxicity score to the lowest.

- Now we see that Drug 5 is ranked as the highest risk, as it has the highest toxicity score, but its cleaning procedure is very effective and the risk of patient exposure to residues is very low.
- Drug 2 has the next highest toxicity score, but its cleaning procedure is more effective than Drug 5 (refer to Table 4) and the risk of patient exposure to residues is even lower.
- Drug 9 has a moderate toxicity score, but the cleaning procedure is much worse than both Drugs 5 and 2 and has a high probability of leaving residues leading to cross contamination and patient exposure.

Table 5:

Ranking Level of Risk by Toxicity Score

Compound	Tox Score	CPU Score	Risk Rank?
Drug 5	10	4	10
Drug 2	9	2	9
Drug 8	8	1	8
Drug 3	7	3	7
Drug 9	6	8	6
Drug 4	5	5	5
Drug 7	4	6	4
Drug 10	3	7	3
Drug 1	2	9	2
Drug 6	1	10	1

Table 6:

Ranking Level of Risk by Cpu Score

Compound	Tox Score	CPU Score	Risk Rank?
Drug 6	1	10	10
Drug 1	2	9	9
Drug 9	6	8	8
Drug 10	3	7	7
Drug 7	4	6	6
Drug 4	5	5	5
Drug 5	10	4	4
Drug 3	7	3	3
Drug 2	9	2	2
Drug 8	8	1	1

- Drugs 1 and 6 present low hazards, but their cleaning procedures will definitely leave residues leading to cross contamination and therefore have high risks for patient exposure. It becomes apparent that simply ranking compounds by their toxicity scores is not a suitable way to measure cleaning risk.

Table 6 shows the same data as Table 5 but sorted from the highest process capability (Cpu) score to the lowest.

- Now we see that Drug 6 is the highest risk since it has the worst probability score due to poor cleaning process capability and will leave residues. Although Drug 6 is not very hazardous, it clearly poses the highest risk for cross contamination.
- Drug 1 has the next highest cleaning process capability score. Although Drug 1 is slightly more hazardous than Drug 6, its cleaning procedure is more capable of reducing residues than Drug 6. This example shows that while Drug 1 is not very hazardous, it poses a high risk for cross contamination due to poor process cleaning.
- Drug 9 is next as its cleaning procedure is not very good and, although Drug 9 has a moderate toxicity score and is likely to leave residues and pose a high risk for cross contamination, the probability of residues is lower than for Drugs 6 or 1.

Table 7:

Example Risk Evaluation Based on Cpu Score and Toxicity Score

Compound	Tox Score	CPU Score	Risk Analysis		Risk Evaluation
Drug 6	1	10	Very Low Hazard	Very Poor Cleaning	Very High Risk
Drug 1	2	9	Very Low Hazard	Very Poor Cleaning	Very High Risk
Drug 9	6	8	Moderate Hazard	Poor Cleaning	High Risk
Drug 10	3	7	Low Hazard	Fair Cleaning	Moderate Risk
Drug 7	4	6	Low Hazard	Good Cleaning	Low Risk
Drug 4	5	5	Moderate Hazard	Good Cleaning	Low Risk
Drug 5	10	4	Very High Hazard	Excellent Cleaning	Low Risk
Drug 3	7	3	High Hazard	Excellent Cleaning	Low Risk
Drug 2	9	2	Very High Hazard	Excellent Cleaning	Low Risk
Drug 8	8	1	High Hazard	Exceptional Cleaning	Very Low Risk

These drugs are now ordered from 10 to 1 based on their risk of cross contamination. It appears that ranking by cleaning process capability followed by toxicity is a promising approach to risk management in cleaning. Detectability scores for visual inspection and TOC can be added into the analysis for more refinement of the level of risk.

But what of the ICH Q9 promise of the quality risk management process being commensurate with the level of risk? Can these Cpu and toxicity scores be used for managing cleaning programs and developing a control strategy based on the risk? Table 7 shows a proposed high-level evaluation of the 10 drugs in the above example that may be classified into different risk levels based on these scores. (Note: The reader should understand that the toxicity and Cpu scales are continu-

Table 8:

Possible Action Plans Based on the Level of Risk

Compound	Tox Score	CPU Score	Risk Evaluation	Possible Actions
Drug 6	1	10	Very High Risk	Cleaning Process Improvements Continued Monitoring Release after Sampling
Drug 1	2	9	Very High Risk	Cleaning Process Improvements Continued Monitoring Release after Sampling
Drug 9	6	8	High Risk	Cleaning Process Improvements Continued Monitoring Release after Sampling
Drug 10	3	7	Moderate Risk	Cleaning Process Improvements Periodic Monitoring (TOC?)
Drug 7	4	6	Low Risk	Visual Inspection Only
Drug 4	5	5	Low Risk	Visual Inspection Only
Drug 5	10	4	Low Risk	Cleaning FMEA required to ensure cleaning performance
Drug 3	7	3	Low Risk	Cleaning FMEA required to ensure cleaning performance
Drug 2	9	2	Low Risk	Cleaning FMEA required to ensure cleaning performance
Drug 8	8	1	Very Low Risk	Cleaning FMEA required to ensure cleaning performance

ous scales and can have intermediate values [e.g., 6.3, 4.7, etc.], so these classifications are for example only and should not be considered definitive in any way.)

Based on the example evaluations shown in Table 7, an action plan for each drug could be put in place to reduce risk or to mitigate the unacceptable risks or, if the risk is determined to be acceptable, to develop a control plan to maintain that acceptable level of risk.

Table 8 indicates that the cleaning procedures (SOPs) for Drugs 5, 3, 2, and 8 require a formal cleaning FMEA to ensure the continued cleaning performance for these drugs. While these drugs have excellent and highly effective cleaning procedures, a failure in one of the steps in these cleaning procedures could have catastrophic consequences since their hazard levels are so high. So, for these drugs, performing formal cleaning FMEAs as part of a continued quality risk management program and identifying possible failure modes and proactively implementing corrective actions, such as error-proofing (e.g., poka-yoke), improving cleaning procedures and methods, etc., are the most appropriate actions before any possible failure has a chance to take place. For example, the Viracept situation may not have happened if a formal cleaning FMEA had been performed before that incident occurred.22 However, the cleaning procedures for the other six drugs should also have formal cleaning FMEAs, but not until after any recommended cleaning process improvement activities are completed. Since many drugs share a common cleaning procedure, their formal cleaning FMEAs could be combined into one exercise.

Table 9 shows an example formal cleaning FMEA using the scales in this article. In this hypothetical, a number of basic possible cleaning failures are listed, such as “cleaning solution concentration too low.” While the listed product has a toxicity score of 7.7, the cleaning process is very effective and residues can be easily detected visually and by TOC. This detectability should be included in the risk analysis of these failures and then become part of the control strategy. However, the cleaning process capability shown may not always be the same if the cleaning agent solution is not made correctly. Similar concerns can arise about the cleaning agent contact time not being long enough or the temperature being too low. How should this be addressed? Such questions can be answered using data from design of experiments combined with Monte Carlo analysis and will be discussed in the next article.

Table 9:

Hypothetical Example of a Formal Cleaning FMEA

CLEANING PROCESS FMEA

Date:	1/1/18			
Cleaning SOP:	SOP-02101	Equipment:	Tank 1	
Product:	Drug 1	Product ADE	0.000000018	grams
Product Cpu:	12	Product MSSR	11	mcg/cm^2
		Product VRL	1	mcg/cm^2
		TOC Swab Limit	150	ppb
		TOC DL	30	ppb
Cleaning Agent:	Cleaner 2	Cleaning Agent ADE	0.12	grams
Cleaning Agent Cpu:	25.3	Cleaning Agent MSSR	40	mcg/cm^2
		Cleaning Agent CRL	2	mcg/cm^2
		TOC Swab Limit	500	ppb
		TOC DL	30	ppb

Cleaning Process Step	Step Description	Potential Failure Mode	Potential Effect(s) of Failure	Toxicity Score	CPU Score	VDI	CDI	Criticality
1	Pre-rinse Tank with potable water	Pre-rinse Time too short	Equipment cannot be cleaned completely	7.7	0.83	-1.0	-0.7	1
2	Prepare 2% Cleaning Solution	Cleaning Time too short	Product Residue Remaining	7.7	0.83	-1.0	-0.7	1
2	Prepare 2% Cleaning Solution	Cleaning Temperature too low	Product Residue Remaining	7.7	0.83	-1.0	-0.7	1
3	Clean Tank with 2% cleaning Solution for 1 hour at 80°C	Cleaning Time too short	Product Residue Remaining	7.7	0.83	-1.0	-0.7	1
3	Clean Tank with 2% cleaning Solution for 1 hour at 80°C	Cleaning Temperature too low	Product Residue Remaining	7.7	0.83	-1.0	-0.7	1
4	Rinse Tank with 90°C USP Purified Water for 10 minutes	Rinse Temperature too low	Cleaning Agent Residue Remaining	0.9	0.40	-1.3	-1.2	3
4	Rinse Tank with 90°C USP Purified Water for 10 minutes	Rinse Time too short	Cleaning Agent Residue Remaining	0.9	0.40	-1.3	-1.2	3
5	Inspect Tank for cleanliness	Visual Inspection not performed	Residues on equipment not observed	7.7	0.83	-1.0	-0.7	2

THE CLEANING RISK DASHBOARD

Dashboards are widely used in business to provide simple "at-a-glance" tools that can quickly show visual representations of complex relationships among many business metrics, key performance indicators (KPIs), or any other data important to making decisions about a business process. Dashboards communicate knowledge efficiently and simplify the decision-making process in business and other endeavors by making multiple sources of data and their relationships easy to visualize. Ultimately, a critically important process such as QRM would benefit from a dashboard that could easily present the multiple sources of data so decisions concerning risk can be made efficiently and with confidence.

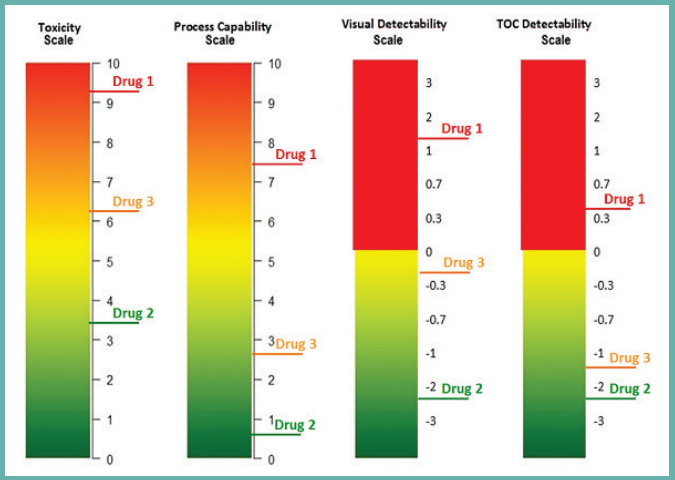
The scales discussed in this article and in the previous four articles can be used to develop such a dashboard. Figure 4 shows an example of how new compounds can be quickly and easily evaluated to determine whether the current cleaning process and analytical methods allow these compounds to be manufactured in a shared equipment facility. Their HBELs are determined and evaluated against the facility's existing cleaning data that compares its cleaning process capability against the known detection limits to determine if the existing methods are capable of detecting these new compounds.

Note: Excel spreadsheets for creating these scales can be downloaded for free:

- *Spreadsheet to Create a Toxicity Scale from HBELs*
- *Spreadsheet to Create a Process Capability Scale from Cpu Data*
- *Spreadsheet to Create Detectability Scales from TOC and Visual Inspection Detection Limits*

Immediately, it can be seen that Drug 1 is a very toxic compound and that the current cleaning process cannot adequately clean it to prevent cross contamination issues. (Note: Process capability can be evaluated based on existing cleaning data compared to the limits required by the new compound). In addition, residues cannot be detected at a safe level, visually or even by TOC. Introducing this drug would require substantial improvements in both the cleaning process and analytical methodologies. Most likely, a manufacturer would need to dedicate equipment or an entire facility to the manufacture of this drug.

Figure 4:
Examples of using the cleaning risk dashboard



Drug 2, on the other hand, is not highly toxic, and the current cleaning process can easily clean it to prevent cross contamination issues and any residues can be easily detected visually or by TOC. Introducing this drug would not require any improvements and would potentially require evaluation of initial manufacturing by visual inspection only.

Drug 3 is somewhat toxic, but the current cleaning process could adequately clean it to prevent cross contamination issues, and while residues cannot be detected visually, the TOC method is acceptable for detection. Introducing this drug would also not require any improvements.

There are other issues to consider in introducing a new product; however, this dashboard provides an effective screening tool for making decisions on whether cleaning process development is needed, what analytical methods can be used, and if analytical method development is needed to justify the introduction of new products. Such a dashboard also provides an easy, high-level view of manufacturing operations for rapid measurement of risk in a facility, department, or manufacturing line.

CONCLUSION

One of the stated goals of the ASTM E3106-17 Standard Guide for Science and Risk Based Cleaning Process Development and Validation was to provide a framework for a scientific risk- and statistics-based approach to cleaning processes and validation based on ICH Q9 and the FDA's 2011 Process Validation Guidance. Again, the benefit of such an approach would be the ability to scale the level of effort, formality, and documentation of the cleaning validation process commensurate with the level of risk, while providing a visual tool for communicating these risks. Objective tools to measure risk in cleaning can focus cleaning validation efforts where the risks are the greatest based on: the science behind the HBEL score, which informs us which products are the most hazardous; the Cpu score of the cleaning process, informing us what the probability of residues are; and, as we saw in Table 2, the detectability scores, which can determine the appropriateness of analytical methods and guide their selection.

Table 10 offers an example of how the toxicity score and the Cpu score could be used to make decisions on whether additional cleaning process development is necessary, whether continued or periodic monitoring or simple visual inspection may be appropriate, and even when product dedication may be necessary.

Table 10:

Example of Possible Actions Based on Toxicity and Cpu Scores

Cpu Score	Risk Evaluation	Toxicity Score 10	9	8	7	6	5	4	3	2	1
10	Very High	✓ Not Acceptable Consider Dedicated Facility or Single-Use Equipment				✓ Cleaning Process Improvement ✓ Continued Monitoring ✓ Release After Sampling ✓ Formal Cleaning FMEA					
9	Very High										
8	High	✓ Cleaning Process Improvement ✓ Formal Cleaning FMEA ✓ Continued Monitoring ✓ Release After Sampling				✓ Cleaning Process Improvement ✓ Continued Monitoring					
7	Moderate										
6	Low	✓ Formal Cleaning FMEA ✓ Continued Monitoring				✓ Periodic Monitoring		✓ Visual Inspection Only for Day-to-day Control (Note)			
5	Low										
4	Low	✓ Formal Cleaning FMEA ✓ Periodic Monitoring									
3	Low										
2	Very Low										
1	Very Low										

Note: For all cases, a visual inspection must still be done.

Table 10 offers a road map for a decision-making process for selecting cleaning validation activities and developing an ongoing control strategy based on data. However, this is just an example of how choices could be decided, and each company would need to decide how to implement this. In his book "Against the Gods: The Remarkable Story of Risk," Peter Bernstein[5] notes that:

> "The essence of risk management lies in maximizing the areas where we have some control over the outcome while minimizing the areas where we have no control over the outcome and the linkage between effect and cause is hidden from us."

We can maximize the cleaning process capability to reduce residues to the lowest practical levels while focusing on those parameters that lower our detection limits. Since the toxicity of APIs is intrinsic and cannot be influenced, we can minimize the likelihood for toxic compounds to cross contaminate other products. But this is only possible if we truly understand where the risks are. The recent requirement for all companies to determine HBELs for their compounds[23] provided a data-based measure of a compound's toxicity for determining cleaning limits and set the stage for the measurement of risk in cleaning based on scientific principles. In this article we have presented science- and data-based visual tools to advance the scientific rigor in the cleaning of healthcare products, such as pharmaceuticals, biopharmaceuticals, cosmetics, and medical devices.

PEER REVIEW:

The authors wish to thank our peer reviewers Bharat Agrawal; James Bergum, Ph.D.; Sarra Boujelben; Gabriela Cruz, Ph.D.; Mallory DeGennaro; Kenneth Farrugia; Ioanna-Maria Gerostathi; Miquel Romero Obon; Laurence O'Leary; Joel Young; Ersa Yuliza; and Mark Zammit for reviewing this article and for their many insightful comments and helpful suggestions.

REFERENCES:

5. Walsh, Andrew, Ester Lovsin Barle, Michel Crevoisier, David G. Dolan, Andreas Flueckiger, Mohammad Ovais, Osamu Shirokizawa, and Kelly Waldron. "An ADE-Derived Scale For Assessing Product Cross-Contamination Risk In Shared Facilities" Pharmaceutical Online, May 2017
6. Walsh, Andrew, Ester Lovsin Barle, David G. Dolan, Andreas Flueckiger, Igor Gorsky, Robert Kowal, Mohammad Ovais, Osamu Shirokizawa, and Kelly Waldron. "A Process Capability-Derived Scale For Assessing Product Cross-Contamination Risk In Shared Facilities" Pharmaceutical Online August 2017
7. Walsh, Andrew, Thomas Altmann, Alfredo Canhoto, Ester Lovsin Barle, David G. Dolan, Mariann Neverovitch, Mohammad Ovais, Osamu Shirokizawa and Kelly Waldron. "An MSSR-derived Scale for Assessing the Detectability of Compound-Carryover in Shared Facilities" Pharmaceutical Online December 2017.
8. Walsh, Andrew, Thomas Altmann, Alfredo Canhoto, Ester Lovsin Barle, David G. Dolan, Andreas Flueckiger, M.D., Igor Gorsky, Robert Kowal, Mariann Nev-

ABOUT THE AUTHORS

THOMAS ALTMANN is RD&E global technical manager – life sciences at Ecolab. He has 18 years' experience in consulting in the pharmaceutical and biotechnology industry, as well as API production related to cleaning, validation, and disinfecting of product contact surfaces.

ALFREDO CANHOTO, PH.D., is a senior manager of quality assurance validation for Alnylam Pharmaceuticals and has expertise in cleaning development and validation.

DAVID DOLAN, PH.D., DABT is the toxicologist specializing in occupational, environmental, and product quality toxicology at Amgen Inc. over the past 12 years, after previously specializing in occupational toxicology and occupational epidemiology at Merck & Co., Inc.

ANDREAS FLUECKIGER, MD, was in charge of occupational health including occupational toxicology in the Roche Group for 32 years before his retirement in August 2018. His department also established Roche's health-based exposure limits values used in GMP cleaning.

IGOR GORSKY is a Pharmaceutical, Biopharmaceutical and Medical Device industry expert with over 35 years of experience in multiple disciplines including Quality Assurance. His started his career at Alpharma and continued it at Wyeth and Shire where he led global Pharmaceutical Technology Validation. For the last six years he has been consulting globally for Valsource, LLC.

JESSICA GRAHAM, PH.D., DABT is a toxicologist specializing in occupational and product quality toxicology at Bristol-Myers Squibb.

ROBERT KOWAL is a retired cleaning validation subject matter expert (SME) for Johnson & Johnson.

ESTER LOVSIN BARLE, DVM, MSC, PHD, MScTox is the Head of Corporate Toxicology at Lonza and responsible for Health topics at Corporate Lonza EHS. Previously she has been the Head of Health Hazard Assessment in Novartis Global HSE & BCM.

MARIANN NEVEROVITCH MS Pharmacy Research Scientist at Bristol-Myers Squibb. Cleaning Validation Subject Matter Expert with 15+ years of experience in cleaning verification analytical method development and support. Since 2010 leading cleaning verification program in Support of Clinical Supply Operations.

erovitch, Mohammad Ovais, Osamu Shirokizawa and Kelly Waldron. "A Swab Limit-Derived Scale For Assessing The Detectability Of Total Organic Carbon Analysis" Pharmaceutical Online January 2018

9. Bernstein, Peter L., "Against the Gods: The Remarkable Story of Risk" John Wiley & Sons, Inc. 1998.
10. Hubbard, Douglas W. "The Failure of Risk Management: Why It's Broken and How to Fix It" Wiley; 2009, 1 edition
11. Taleb, Nassim Nicholas "Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets" Random House 2008
12. Hansson Sven Ove and Christina Rud´en "Evaluating the risk decision process" Toxicology 218 (2006) 100–111
13. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH Harmonized Tripartite Guideline, Quality Risk Management – Q9, Step 4, 9 November 2005, http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q9/Step4/Q9_Guideline.pdf.
14. O'Donnell, Kevin "QRM in the GMP Environment: Ten Years On — Are Medicines Any Safer Now A Regulators Perspective" Journal of Validation Technology December 21, 2015.
15. Ball, D. J., & Watt, J. (2013). Further thoughts on the utility of risk matrices. Risk Anal, 33(11), 2068-2078.
16. Baybutt, P. (2014). Allocation of risk tolerance criteria. Process Safety Progress, 33(3), 227-230.
17. Baybutt, P. (2016). Designing risk matrices to avoid risk ranking reversal errors. Process Safety Progress, 35(1), 41-46.
18. Cox, L. A., Jr. (2008). What's wrong with risk matrices? Risk Anal, 28(2), 497-512.
19. Cox, L. A., Jr., Babayev, D., & Huber, W. (2005). Some limitations of qualitative risk rating systems. Risk Anal, 25(3), 651-662.
20. Levine, E. S. (2012). Improving risk matrices: the advantages of logarithmically scaled axes. Journal of Risk Research, 15(2), 209-222.
21. Li, J., Bao, C., & Wu, D. (2018). How to Design Rating Schemes of Risk Matrices: A Sequential Updating Approach. Risk Anal, 38(1), 99-117.
22. Waldron, Kelly ,"Risk Analysis and Ordinal Risk Rating Scales—A Closer Look" Journal of Validation Technology December 21, 2015.

23. MIL-STD-1629A - Procedures for Performing a Failure Mode, Effects and Criticality Analysis. U.S. Department of Defense. 1949. (MIL–P–1629).
24. Textron, Inc. (https://www.slideshare.net/handbook/qm085design-andd-process-fmea)
25. Wheeler, D., "Problems with Risk Priority Numbers – Avoiding More Numerical Jabberwocky," Quality Digest, June 2011, www.qualitydigest.com/inside/quality-insider-article/problems-risk-priority-numbers.html
26. Killilea, M.C., "Cleaning Validation: Viracept, 2007," Journal of Validation Technology, November 2012, Vol. 18, Issue 4, www.ivtnetwork.com/jvt-journal.
27. "Guideline on setting health based exposure limits for use in risk identification in the manufacture of different medicinal products in shared facilities" European Medicines Agency 20 November 2014

ABOUT THE AUTHORS (CONT.)

MOHAMMAD OVAIS MPharm is a Cleaning Validation Consultant. He has more than thirteen years of hands-on experience in validating cleaning processes. He provides training and consulting services in process and cleaning validation.

OSAMU SHIROKIZAWA is President of Life Scientia Limited, a pharmaceutical engineering and consultancy firm.

KELLY WALDRON is a senior consultant with Valsource and a regulatory science researcher with the Pharmaceutical Regulatory Science Team at the Dublin Institute Technology.

ANDREW WALSH is president of the Center for Pharmaceutical Cleaning Innovation, a non-profit research and educational organization.

THE SHIROKIZAWA MATRIX: DETERMINING THE LEVEL OF EFFORT, FORMALITY, & DOCUMENTATION IN CLEANING VALIDATION

ANDREW WALSH
THOMAS ALTMANN
RALPH BASILE
JOEL BERCU, PH.D.
ALFREDO CANHOTO, PH.D.
DAVID G. DOLAN PH.D.
PERNILLE DAMKJAER
ANDREAS FLUECKIGER, M.D.
IGOR GORSKY
JESSICA GRAHAM, PH.D.
ESTER LOVSIN BARLE, PH.D.
OVAIS MOHAMMAD
MARIANN NEVEROVITCH
SIEGFRIED SCHMITT, PH.D.
OSAMU SHIROKIZAWA

The International Congress on Harmonization Quality Risk Management Guidance (ICH Q9) lists both cleaning (in Annex II.4) and validation (in Annex II.6) as potential areas for the application of quality risk management.[1] This clearly implies that the ICH Q9 principle for adjusting the level of "effort, formality, and documentation" based on the level of risk could be applied to cleaning and its validation. Previous articles discussed how science-based and data-derived scales could be created from HBELs (health-based exposure limits), from the process capability (Cpu) of cleaning processes, from the detection limits for total organic carbon (TOC) analyses of these compounds, or from visual inspection.[2–5] Another article discussed how these scales could be used to measure the level of risk in cleaning validation.[6] This article builds on these discussions and shows how these HBEL-based and process capability-based scales can be combined into a matrix that provides a clear visual guide for adjusting the level of effort, formality, and documentation for cleaning validation based on the level of risk.

QUALITY RISK MANAGEMENT UNDER ICH Q9

As previously discussed,[6] ICH Q9 describes two primary principles of quality risk management (QRM):

- "The evaluation of the risk to quality should be based on scientific knowledge and ultimately link to the protection of the patient; and
- The ***level of effort, formality and documentation*** of the quality risk management process should be commensurate with the level of risk".

From these two primary principles, it can be understood that if we can determine the level of risk to a patient from cleaning, then the level of cleaning validation effort, its formality, and its documentation could be adjusted accordingly. Stated simply, cleaning validation efforts for low-risk products should not require the same level of effort as for high-risk products. This is perfectly logical. The level of effort, formality, and documentation for cleaning validation should correspond to the level of risk, which includes the available knowledge of a cleaning process and the nature of the product.

ICH Q9 defines risk (in general) as:

> "It is commonly understood that risk is defined as the combination of the ***probability of occurrence*** of harm and the ***severity*** of that harm".

and further on:

> "...the ability to detect the harm (***detectability***) also factors in the estimation of risk".

We also see risk can be more formally expressed as:

> Risk = ***f*** (Severity of Hazard, Level of Exposure to Hazard, Detectability of Hazard)

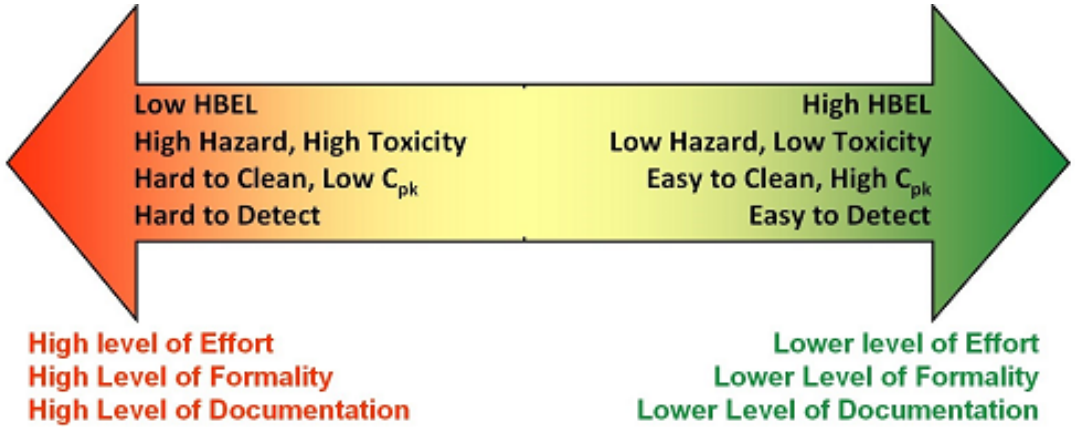


Figure 1: Continuum of risk in cleaning based on hazard, exposure, and detectability

Now, if the hazard is ***intrinsic to an active pharmaceutical ingredient (API) and the risk being considered is harm to a patient from exposure to residues of that API after cleaning***, then this equation can be further refined to:

$$\text{Cleaning Risk} = \boldsymbol{f}\,(\text{Toxicity}_{\text{API residue}}, \text{Level of Exposure}_{\text{API residue}}, \text{Detectability}_{\text{API residue}})$$

So, if we can measure these parameters of toxicity, exposure, and detectability we should be able to locate our position on the continuum shown in Figure 1.

SCIENCE- AND RISK-BASED SCALES FOR TOXICITY, EXPOSURE, AND DETECTABILITY

The scales presented in the preceding four articles[2-5] offer scientifically based methods to measure these risk parameters using actual data. Since the scales presented in the four articles are all based on science and derived from actual data, they would consequently make good choices for evaluating the risk in cleaning, including patient safety.

These scales were originally developed as replacements for the typical scales used in failure modes and effects analysis (FMEA). The subjectivity of the scales typically used

in FMEAs, and the lack of a scientific/statistical basis for their risk priority numbers (RPNs), often makes both the scales and their RPNs inappropriate for use in the pharmaceutical industry, as discussed in an earlier article.[6] If pharmaceutical manufacturing is to advance to a science- and risk-based approach, the scales for severity, occurrence, and detectability used in FMEAs must be scientifically justified using science, process knowledge, and statistics. Such scales should be derived from, and based on, empirical data. Data such as this exists for cleaning and is readily available in pharmaceutical manufacturing production. As stated in the introduction, scales already exist that could be used for the following criteria:

1. HBEL-derived Toxicity Scale → Severity of Process Residues[2]
2. Process Capability-derived Scale → Occurrence of Exposure to Process Residues[3]
3. TOC Detectability Index → Detectability of Process Residues [4]
4. Visual Detectability Index → Detectability of Process Residues [5]

For example, in a cleaning process, if a failure mode could result in residues of an API remaining on equipment, then the HBEL-derived toxicity score of that API could be used as a severity score. Furthermore, if the process capability of the cleaning process is known, then its process capability-derived score could be used as an occurrence score (as the effectiveness of the cleaning process and the probability of residues remaining are known). Finally, if either the visual detectability index4 or the TOC detectability index5 is known, then one or both of these could be used as a detectability score. Since these scores are derived directly from empirical data, their values are specific and objective and should not be subject to debate, as happens frequently with traditional FMEA scales. It should be noted that detectability indexes may also be compiled from other available analytical test methods used, such as UV, HPLC, etc.

As a refresher on the previous articles,[2-5] the HBEL-derived Toxicity Scale[2] is based on converting HBEL (ADE/PDE) values to a scale in a manner similar to that used to create the pH (log-based) scale. By converting the HBEL value into grams and taking its negative logarithm, a continuous scale from 0 to 10 (as is typical of the severity scales used in FMEAs) can be generated for HBELs, ranging from a high value of 1 gram/day (low hazard) to a low value of equal to or less than 1 nanogram/day (high hazard).

The Process Capability-derived Scale[3] depicts the process capability (Cpu) by converting actual cleaning data into a scale from 1 to 10 by taking the reciprocal of the Cpu and multiplying by 10. This results in a scale that has a high value (i.e., 10) associated with a high probability of failure and a low value (i.e., 1) associated with a low probability of failure, as is typical of an occurrence scale used in FMEAs. Note: The process capability for a Six Sigma cleaning process (Cpu = 2) was chosen as the midpoint of this scale.

THE TOXICITY/CLEANING CAPABILITY MATRIX

Figure 2 shows an example of how the Toxicity Scale and Process Capability Scale can be combined into a matrix and groupings created based on the risk. This is called the Shirokizawa Matrix, and it can be used as a guide to determine the level of effort, formality, and documentation necessary for cleaning validation activities.

The matrix in Figure 2 is separated into eight groups based on the coordinate positions in the matrix from the toxicity and process capability scores. The upper left quadrant contains the highest risk situations: high hazard compounds with poor cleaning processes. Compounds whose scores fall into this quadrant would require the highest levels of effort, formality, and documentation in their cleaning validation programs. The lower right quadrant contains the lowest risk situations: low hazard compounds with excellent cleaning processes. Compounds whose scores fall into this quadrant would require lower levels of effort, formality, and documentation in their cleaning validation programs.

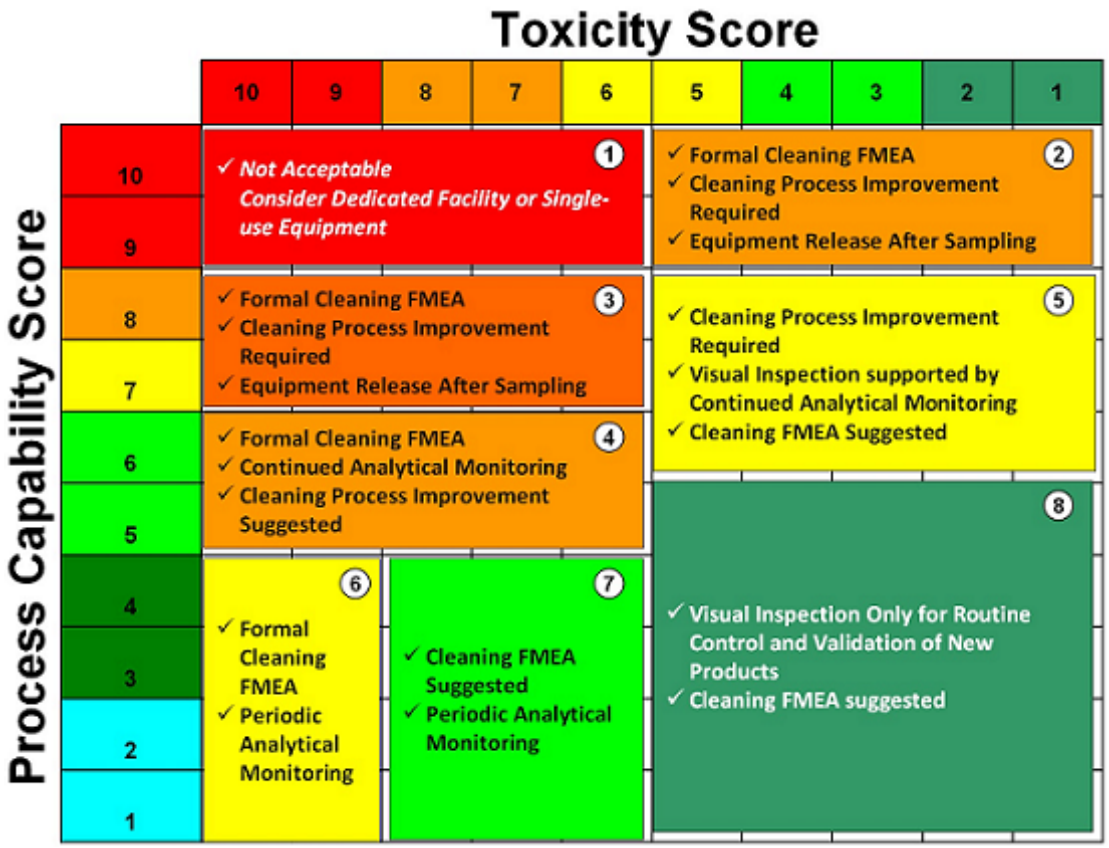


Figure 2: Shirokizawa Matrix of Cleaning Validation Effort – *This matrix is an example and a potential model of how toxicity and process capability scores can be used together to determine an acceptable level of effort, formality, and documentation in cleaning validation programs.*

Table 1 provides more detailed descriptions of the levels of effort, formality, and documentation that might be required for those compounds whose toxicity and their cleaning process capability positioned them in a particular group.

Table 1: Suggested Levels of Effort, Formality, and Documentation Required for Products Whose Toxicity and Process Capability Scores Place Them in One of the Eight Groups

Group	Level of Effort	Formality	Documentation
1	**Summary:** • Hazard level is significant (Tox score = 6 - 10) • The cleaning procedure is inadequate to insure removal of residues to safe levels (Cpu Score = 9 - 10) • Residue Detection Methods may not be adequate or acceptable Since the risk level is the highest, this group may require dedication of a facility and/or single-use equipment. Deactivation/denaturation steps should be considered. Therefore, significant validation efforts may not be necessary; however, batch to batch or campaign to campaign carryover and removal of impurities, degradants, and microbial contaminants, including endotoxin, should be considered.	**Formal RM**	• HBEL Monograph • Cleaning RA[a] • Cleaning Process Development Report • Evaluation of Historical Data (if exists) • Detailed Cleaning SOP and Record • Validation Protocols/Reports developed from QRM Report • Method Validation Report • Visual Inspection Documentation
2	**Summary:** • Hazard level is low (Tox score = 1 - 5) • The cleaning procedure is seriously inadequate to ensure removal of residues to safe levels (Cpu Score = 9 - 10) • Residue Detection Method must be capable of reaching low levels of detection. Specific methods may be necessary The Hazard (Toxicity) level is medium to low, but the risk of residues remaining is high due to significantly low repeatability and control of the cleaning process, so this group will require Cleaning Process improvement or dedication. A statistically significant number of Cleaning Performance Qualification runs will be necessary as significant data is required to assess whether the process is repeatable and in control. A continued monitoring plan must be established. Must establish sampling plan after each manufacturing lot or campaign. Dedication of equipment or parts of equipment may be considered. Detailed documented justification of the Cleaning/Cleaning Validation program outlining control for all three stages[b] of the Cleaning Validation life cycle may be required. Modification of equipment design may be necessary; highly detailed cleaning SOPs and Cleaning Process Records and documented operator training programs are necessary.	**Formal RM**	• HBEL Monograph • Cleaning RA • Evaluation of Historical Data (if exists) • Validation Master Plan • Cleaning Process Development Report • Method Development and Validation Reports • Validation Protocols/Reports developed from QRM • Cleaning Process Record • Operator Training Program • Visual Inspection Documentation
3	**Summary:** • Hazard level is significant (Tox score = 6 - 10) • The cleaning procedure is inadequate to ensure removal of residues to safe levels (Cpu Score = 7 - 8) • Residue Detection Method must be capable of reaching low levels of detection. Specific methods may be necessary The Hazard (Toxicity) level is high and the risk level is high, so this group will require Cleaning Process improvement. Since cleaning process failures pose serious risk to the patients, a formal cleaning RA (such as Cleaning FMEAs) on the cleaning procedure is required. A statistically significant number of Cleaning Performance Qualification runs will be necessary as significant data is required to assess whether the process is repeatable and in control. A continued monitoring plan must be established. Establishing a sampling plan after each manufacturing lot or campaign is recommended. Dedication of equipment or parts of equipment may be considered. Cleaning procedure and cleaning record that would allow monitoring of the process success and detecting issues should be established. Detailed documented justification of the Cleaning/Cleaning Validation program outlining control for all three stages of the Cleaning Validation. Highly detailed cleaning SOPs and Cleaning Process Records, as well as documented operator training programs, are necessary.	**Formal RM**	• HBEL Monograph • Cleaning RA • Evaluation of Historical Data (if exists) • Validation Master Plan • Cleaning Process Development Report • Method Development and Validation Reports • Validation Protocols/Reports developed from QRM • Cleaning Process Record • Operator Training Program • Visual Inspection Documentation
4	**Summary:** • Hazard level is significant (Tox score = 6 - 10) • The cleaning procedure is borderline adequate to insure removal of residues to safe levels (Cpu Score = 5 - 6) • Residue Detection Method should be capable of reaching low levels of detection. Non-specific methods may be acceptable The Hazard (Toxicity) level is high, but the risk is more acceptable due to the lower occurrence level. Cleaning Process Improvement plan should be considered. Since cleaning process failures pose serious risk to the patients, a formal cleaning RA (such as FMEAs) on the cleaning procedure is required.	**Formal RM**	• HBEL Monograph • Cleaning RA • Evaluation of Historical Data (if exists) • Validation Master Plan • Cleaning Process Development Report • Method Development and Validation Reports • Validation Protocols/Reports developed from QRM

Group	Level of Effort	Formality	Documentation
	Statistically significant data is required to assess variability. A continued monitoring plan must be established. Detailed cleaning procedures that would allow monitoring of the process success and detecting issues are required. This group may require a statistical process control program. A detailed documented justification of the Cleaning/Cleaning Validation program outlining control for all three stages of the Cleaning Validation is required. Highly detailed cleaning SOPs and/or Cleaning Process Records are necessary. Documented operator training programs are necessary.		Cleaning Process Record Operator Training Program Visual Inspection Documentation
5	**Summary:** • Hazard level is low (Tox score = 1 - 5) • The cleaning procedure is inadequate to insure removal of residues to safe levels (Cpu score = 6 - 8) • Residue Detection Method should be capable to reach the level of detection. Non-specific methods are acceptable The Hazard (Toxicity) level is low but the risk of residues is high due to a relatively poor reliability of the cleaning process, so this group will require Cleaning Process improvement. A statistically significant number of Cleaning Performance Qualification runs will be necessary as significant data is required to assess variability and the ability of the process to consistently remove residuals to acceptable levels. Statistically significant data is required to assess variability and ability of the process to consistently remove residuals to acceptable by the protocol levels. A cleaning process improvement plan should be considered. Detailed cleaning procedure that would allow monitoring of the process success and detecting issues. A detailed documented justification of the Cleaning/Cleaning Validation program outlining control for all three stages of the Cleaning Validation is recommended. Establishing a Continued Verification plan is required. Highly detailed cleaning SOPs and Cleaning Process Records and documented operator training programs are necessary.	**Informal RM**	HBEL Monograph Cleaning RA Evaluation of Historical Data (if exists) Validation Master Plan Cleaning Process Development Report Method Development and Validation Reports Validation Protocols/Reports developed from QRM Cleaning Process Record Operator Training Program Visual Inspection Documentation
6	**Summary:** • Hazard level is high (Tox score = 9 - 10) • The cleaning procedure is very capable of ensuring removal of residues to safe levels (Cpu Score = 1 - 4) • Residue Detection Methods may not be adequate. Specific methods may be necessary The Hazard (Toxicity) level is high but the risk level is very low, so this group does not require a Cleaning Process improvement plan. Since cleaning process failures could pose serious risk to the patients, a formal cleaning RA (such as FMEAs) on the cleaning procedure is required. As significant data is not required to assess variability and ability of the process to consistently remove residuals to acceptable levels this group may only require one (1) Cleaning Performance Qualification run. A periodic monitoring plan should be established. Cleaning procedures that would allow monitoring of the process success and detecting issues. A detailed documented justification of the Cleaning/Cleaning Validation program outlining control for all three stages of the Cleaning Validation is required. Establishing a periodic monitoring plan is required. Detailed cleaning SOPs and Cleaning Process Records and operator training programs are necessary. A visual inspection monitoring program may be possible, but a periodic monitoring plan would still be required.	**Informal RM**	HBEL Monograph Cleaning RA Evaluation of Historical Data (if exists) Validation Master Plan Cleaning Process Development Report Method Development and Validation Reports Validation Protocols/Reports developed from QRM Cleaning Process Record Operator Training Program Visual Inspection Documentation
7	**Summary:** • Hazard level is moderate (Tox score = 6 - 8) • The cleaning procedure is very capable of ensuring removal of residues to safe levels (Cpu score = 1 - 4) • Residue Detection Methods should be adequate. Non-specific methods may be acceptable, including visible inspection The Hazard (Toxicity) level is moderate but the risk level is very low, so this group does not require a Cleaning Process improvement plan. A significant number of Cleaning Performance Qualification runs may not be necessary. As significant data is not required to assess variability and ability of the process to consistently remove residuals to acceptable levels, this group may require only a Cleaning Verification using documented and monitored using analytical verification or by visual inspection (if qualified). Data is required to assess the	**Minimal RM (Use SOPs)**	HBEL Monograph Cleaning RA Evaluation of Historical Data (if exists) Validation Master Plan Cleaning Process Development Report Method Development and Validation Reports Validation Protocols/Reports developed from QRM Cleaning Process Record Operator Training Program Visual Inspection Documentation

Group	Level of Effort	Formality	Documentation
	variability and the ability of the process to consistently remove residuals to acceptable by the protocol levels. A periodic monitoring plan must be established. If used, a visual inspection monitoring program must be established.		
8	**Summary:** • Hazard level is low (Tox score = 1 - 5) • The cleaning procedure is very capable of insuring removal of residues to very safe levels (Cpu Score = 1 - 5) • Residue detection methods should be adequate. Non-specific methods are acceptable including visible inspection The Hazard (Toxicity) level is very low and the risk of residues remaining is very low, so this group does not require Cleaning Process improvement. A significant number of Cleaning Performance Qualification runs may not be necessary. As significant data is not required to assess variability and ability of the process to consistently remove residuals to acceptable levels, this group may require only a Cleaning Verification using documented and monitored visual inspection (if qualified) at a minimum. Data (visual inspections) is required to assess variability and ability of the process to consistently remove residuals to acceptable by the protocol levels. A periodic monitoring plan must be established. A visual inspection monitoring program must be established.	**Minimal RM (Use SOPs)**	HBEL Monograph Cleaning RA Evaluation of Historical Data (if exists) Validation Master Plan Cleaning Process Development Report Method Development and Validation Reports Validation Protocols/Reports developed from QRM Cleaning Process Record Operator Training Program Visual Inspection Documentation

Table 1 - Suggested Levels of Effort, Formality, and Documentation required for products whose Toxicity and Process Capability Scores place them in one of the eight groups.

a - Cleaning RA (Risk Assessment), for example, FMEA at the Formal Level

b - Stage 1 – Cleaning Process Design, Stage 2 – Cleaning Process Qualification and Stage 3 – Continued Cleaning Process Verification

DISCUSSION

When the Risk-MaPP Guide[7] first introduced the acceptable daily exposure (ADE) concept in 2010, one of the first reactions was that all the cleaning limits would become so low that companies would not be able to pass any cleaning validations. This concern was not based on any actual knowledge and was addressed by an article published in 2016 that evaluated 304 pharmaceutical products and compared their ADEs to 1/1,000th of their therapeutic dose limits.[8] This article showed that in 85 percent of these cases, the limits were higher, even substantially higher, not lower, so this concern was unfounded. The next concern that was raised was that these higher HBELs would allow companies to "relax their cleaning efforts." This is another mis-guided assumption.

An original goal of Risk-MaPP, and subsequently of the ASTM E3106 Standard Guide,[9] has been to implement the ICH Q9 principle that the "level of effort, formality, and documentation" of the cleaning validation process should be based on science and risk. A quick examination of the matrix in Figure 2 and the details in Table 1 will reveal that any "relaxation" of cleaning effort will lead to poor cleaning process capability and result in an ***increase*** of validation efforts. Only very good cleaning efforts can

allow a company to reduce the levels of validation effort — even to the possibility of only a visual inspection program. The EMA indicated in Annex 15 that visual inspection could be used alone in certain cases[10] and recently provided guidance on how a visual inspection program can replace analytical testing for release of equipment after cleaning.[11] Such programs could provide significant operational benefits to companies that can successfully implement them. But the benefits for companies moving to the HBELs can only be realized if their cleaning processes are shown to be ***effective, repeatable, and safe***.

Readers should understand that the groupings shown here are our initial recommendations and should undergo updating as experience is obtained using them. Further, their boundaries should not be considered rigid or fixed. For example, scores that put the risk at the intersection of 5 and 6 on the toxicity scale and 8 and 9 on the process capability scale could equally fall into groups 1, 2, 3, or 5. These groupings and their levels of effort should not be applied blindly, and serious consideration should be put into what efforts are truly necessary for each particular situation. These groupings are only meant to help guide the decision process. Therefore, practitioners are urged to exercise significant QRM efforts in which all elements of the firm's cleaning and cleaning validation programs are carefully identified, analyzed, and evaluated.

The previous articles cited[2–6] show how the application of scientific principles and statistical tools can be used to measure the level of risk in cleaning. Moving into the future, there should be a shift in the focus of cleaning validation programs to the derivation of HBELs and the development of cleaning risk assessments. The HBELs and the cleaning risk assessments will be used to inform the master plans and protocols for the level of cleaning validation that is necessary (Figure 3). The practice of simply using master plans and protocols from previous validations as templates should come to an end. The contents of master plans and protocols should be determined based on science and risk analysis and not on what was done the "last time."

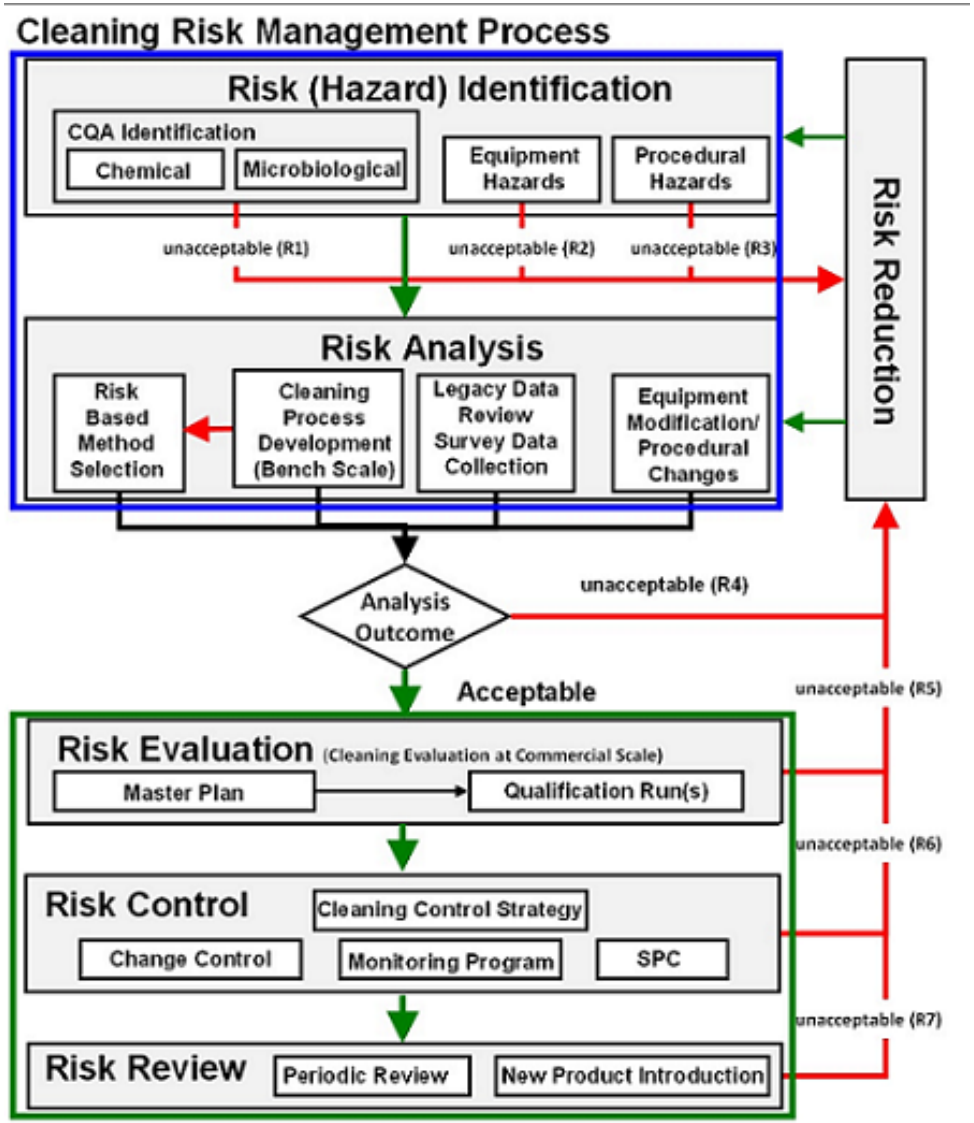


Figure 3: The Cleaning Quality Risk Management Process – *The activities in the* ***BLUE*** *zone are where full cleaning QRM efforts are applied. When these cleaning QRM efforts are completed, and the risk is now acceptable, the process can move into the* ***GREEN*** *zone. The* ***GREEN*** *zone is where the cleaning QRM level of effort, formality, and documentation can be reduced based on the results of the cleaning QRM efforts in the* ***BLUE*** *zone. This is where the grouping criteria from the Shirokizawa Matrix are applied.*

LEVEL OF EFFORT, FORMALITY, AND DOCUMENTATION

The three concepts of "effort, formality, and documentation" may appear to be separate ideas, but in practice they are closely connected. For example, it is hard to imagine how a process that requires very little effort with very little documentation could have a very high level of formality or how a process that requires a very high level of effort with a very high level of formality can have very little documentation. Clearly, these aspects scale up or down with each other. Therefore, discussing them separately is not really possible or even useful. It is more informative to look at them together at each step of the risk management process (Figure 3).

The most important documentation in the cleaning validation process is found at the risk (hazard) identification stage, and these are the HBEL monographs for chemical hazards. At this stage, the levels of effort, formality, and documentation must remain at their highest, since all subsequent activities, decisions, and calculations flow from the information found in the HBEL monographs. There can be no "shortcuts" in the derivation of HBELs. An ASTM Standard Guide is currently being finalized on the derivation of HBELs[12] that will be of immense help for companies at this stage.

The next most important documentation is the cleaning risk assessment. With the HBEL monograph in hand, the risk assessment process can begin. While the ASTM E3106 standard considers the HBEL at the risk (hazard) identification stage, microbiological hazards, equipment design hazards, and procedural hazards must also be considered. For a new product and a new facility, these other hazards may not be well understood, so a full risk (hazard) identification may be necessary. This may trigger risk reduction activities (Figure 3). On the other hand, for an established product in an established facility, many of these other hazards may have already been identified, considered, and mitigated. This prior knowledge should be leveraged and the level of effort during the risk (hazard) identification may be reduced. The level of effort, formality, and documentation at this stage will be high, but it can clearly be adjusted.

At the risk analysis stage, there may be no historical data to consider and analyze for a new HBEL in a new facility. This may require collecting substantial cleaning data to satisfy concerns about the level of risk. Cleaning process development studies are required and may be extensive. The level of effort, formality, and documentation may

remain high. The lack of historical knowledge may likely trigger significant risk reduction activities. Alternatively, there may be substantial historical data on the cleaning process and an analysis may show that there is very little risk of cross contamination from this new product and, depending on the analyzed risk, only one verification run or even a simple visual inspection may be all that is necessary. A single cleanability test may be all that is needed to confirm that the cleaning process can be considered adequate and reliable.[13] Consequently, the subsequent levels of effort, formality, and documentation may be reduced.

At the risk evaluation stage, where there is no historical data to consider and analyze, the master plan and protocols may require the collection of substantial data and performance of several cleaning performance qualification runs. Alternatively, if there is substantial historical data and the risk analysis showed that there was little risk of cross contamination from this new product, the protocol (or SOP) may require only one verification run or even just a visual inspection. Such an approach is very appropriate for the cosmetics and personal care industries, where there are many low-hazard compounds that could also be demonstrated to be low risk.

At the risk control stage, based on the preceding risk analysis and risk evaluations, continued monitoring using swab sampling after every changeover may be required, and if the remaining risk is considerably high, then equipment may be released only after acceptable test results. Or, the preceding risk analysis and risk evaluations may indicate that just visual inspection is adequate.

Finally, in the risk review stage, when a new product is being introduced, the level of effort, formality, and documentation should follow from the knowledge and understanding from the preceding risk analysis, risk evaluation, and risk control activities. This is an important subject, and a more detailed article on the introduction of new products is currently in development that will provide a flow chart for how a new product should be introduced within a cleaning QRM program.

One final observation: When it comes to “formality,” ICH Q9 does not go into any details about how this factor is affected by a risk assessment; in fact, the word “formality” is only used twice in ICH Q9 and both are simply mentioned. Therefore, ICH Q9 provides no real guidance on this aspect. The term “formality” has many definitions

and uses, but for the pharmaceutical and related industries, it can be defined as “an established procedure or set of specific activities which need to be followed.” Many companies already have detailed procedures for their cleaning validation programs, such as formal protocol templates, and they may wish to maintain these as they are. However, companies with lower-risk products and operations may want to simplify their cleaning validation process and move to SOPs and checklists or cleaning records. While there is no risk to patient safety from a company maintaining a high level of formality, there is the operational risk of performing excessive and unnecessary work and being slow to introduce new products. The higher the risk, the higher the level of formality, and the lower the risk, the lower the level of formality. Each company will need to decide for itself what level of formality to apply.

CONCLUSION

One of the goals of the ASTM E3106-18 Standard Guide was to provide a framework for a scientific risk- and statistics-based approach to cleaning processes and validation within an ICH Q9 framework and based on the FDA’s 2011 Process Validation Guidance. The benefit of such an approach would be the ability to scale the level of “effort, formality, and documentation” of the cleaning validation process commensurate with the level of risk. The current ability to measure risk in cleaning provides an objective tool to focus cleaning validation efforts on the risks that are the most significant, based on the science behind the HBEL.

The authors know that most industry workers would agree that cleaning validation efforts for low-risk products (e.g., low toxicity, demonstrably easy to clean) should not require the same level of effort as for high-risk products (e.g., high toxicity, hard to clean). At the same time, we recognize that most industry workers prefer specific guidance on what they need to do, and simply stating “based on the level of risk” is not helpful or even useable. We have shown previously how the level of risk can be measured,6 but the question of what should be done at a particular level of risk still needed to be answered. The Shirokizawa Matrix described in this article is our first attempt at providing specific guidance on what levels of “effort, formality, and documentation” could be used. In this article we provide a science-based and data-driven approach to guide the level of effort, formality, and documentation for the cleaning of

many healthcare products, including pharmaceuticals, biopharmaceuticals, nutraceuticals, cosmetics, and medical devices.

PEER REVIEW

The authors wish to thank our peer reviewers: Bharat Agrawal, James Bergum, Ph.D., Sarra Boujelben, Gabriela Cruz, Ph.D., Mallory DeGennaro, Parth Desai, Kenneth Farrugia, Angela Garey, Laurence O'Leary, Tri Chanh Nguyen, Miquel Romero Obon, Prakash Patel, Stephen Spiegelberg, Ph.D., Basundhara Sthapit Ph.D., and Joel Young for reviewing this article and for providing many insightful comments and helpful suggestions.

REFERENCES

1. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH Harmonized Tripartite Guideline, Quality Risk Management – Q9, Step 5, 9 September 2015, http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q9/Step5/Q9_Guideline.pdf.

2. Walsh, Andrew, Ester Lovsin Barle, Michel Crevoisier, David G. Dolan, Andreas Flueckiger, Mohammad Ovais, Osamu Shirokizawa, and Kelly Waldron. "An ADE-Derived Scale For Assessing Product Cross-Contamination Risk In Shared Facilities" Pharmaceutical Online May 2017

3. Walsh, Andrew, Ester Lovsin Barle, David G. Dolan, Andreas Flueckiger, Igor Gorsky, Robert Kowal, Mohammad Ovais, Osamu Shirokizawa, and Kelly Waldron. "A Process Capability-Derived Scale For Assessing Product Cross-Contamination Risk In Shared Facilities" Pharmaceutical Online August 2017

4. Walsh, Andrew, Thomas Altmann, Alfredo Canhoto, Ester Lovsin Barle, David G. Dolan, Andreas Flueckiger, M.D., Igor Gorsky, Robert Kowal, Mariann Neverovitch, Mohammad Ovais, Osamu Shirokizawa and Kelly Waldron. "A Swab Limit-Derived Scale For Assessing The Detectability Of Total Organic Carbon Analysis" Pharmaceutical Online January 2018

5. Walsh, Andrew, Thomas Altmann, Alfredo Canhoto, Ester Lovsin Barle, David G. Dolan, Mariann Neverovitch, Mohammad Ovais, Osamu Shirokizawa and Kelly Waldron. "An MSSR-derived Scale for Assessing the Detectability of Compound-Carryover in Shared Facilities" Pharmaceutical Online December 2017

6. Walsh, Andrew, Thomas Altmann, Alfredo Canhoto, Ester Lovsin Barle, David G. Dolan, Andreas Flueckiger, Igor Gorsky, Jessica Graham, Ph.D., Robert Kowal, Mariann Neverovitch, Mohammad Ovais, Osamu Shirokizawa and Kelly Waldron "Measuring Risk in Cleaning: Cleaning FMEA and the Cleaning Risk Dashboard" Pharmaceutical Online April 2018

7. *ISPE Baseline® Guide: Risk-Based Manufacture of Pharmaceutical Products: A Guide to Managing Risks Associated with Cross-Contamination.* (ISPE, Tampa, FL, First Edition, 2010), Vol. 7.

8. Walsh, Andrew, Michel Crevoisier, Ester Lovsin Barle, Andreas Flueckiger, David G. Dolan, Mohammad Ovais (2016) *"Cleaning Limits—Why the 10-ppm and 0.001-Dose Criteria Should be Abandoned, Part II,"* Pharmaceutical Technology 40 (8)

9. American Society for Testing and Materials E3106-18 *Standard Guide for Science-Based and Risk-Based Cleaning Process Development and Validation*, www.astm.org.

10. EudraLex, "Volume 4 – Guidelines for Good Manufacturing Practices for Medicinal Products for Human and Veterinary Use, Annex 15: Qualification and Validation," https://ec.europa.eu/health/documents/eudralex/vol-4_en.

11. European Medicines Agency: Questions and answers on implementation of risk-based prevention of cross-contamination in production and "Guideline on setting health-based exposure limits for use in risk identification in the manufacture of different medicinal products in shared facilities," 19 April 2018 EMA/CHMP/CVMP/SWP/246844/2018

12. American Society for Testing and Materials Work Item WK59975 - "Standard Guide for the Derivation of Health Based Exposure Limits (HBELs)," www.astm.org.

13. Song, Ruijin, Alfredo Canhoto, Ph.D., and Andrew Walsh, "Cleaning Process Development: Cleanability Testing and "Hardest-To-Clean" Pharmaceutical Products" Pharmaceutical Online January 2019

ABOUT US

Pharmaceutical Online's mission is to facilitate connections and foster collaborations in the small molecule drug development and manufacturing space. We deliver exclusive, actionable information to help industry professionals tackle the challenges they face in bringing high-quality therapies to market quickly and efficiently. These insights, analysis, and best practices come from interviews with — and contributed articles from — recognized experts in the field. Topics covered include: Pharma 4.0 (AI, Big Data, automation, continuous manufacturing, etc.); regulations/cGMPs; facility design; process development; technology/equipment (filling, material handling, isolation, inspection, etc.); qualification, validation, and verification; quality management; supply chain; packaging/serialization; and logistics.

PHARMACEUTICAL ONLINE

PharmaceuticalOnline.com
info@PharmaceuticalOnline.com
814.897.9000
5340 Fryling Rd., Ste 300
Erie, PA 16510