Collaborating with Impact: Differentiating CMC Statistics in a Data-Driven Era

Enhancing business insights through analytics and collaboration

Stan Altan
October 20, 2025
IABS CMC Workshop

Johnson&Johnson

Abstract

CMC statisticians have often been viewed as service providers, engaged on an as-needed basis for specific tasks. Today, the rise of data-intensive roles and advanced technologies creates an opportunity to redefine this perception. By integrating statistical rigor with process understanding, regulatory insight, and lifecycle thinking, CMC statisticians can become strategic partners who drive drug development and enhance product quality.

This talk will explore how statisticians can differentiate themselves and deliver measurable business value through collaboration. Key contributions include:

- **Designed experimentation** to build robust, efficient methods and processes,
- **Bayesian, risk-based decision-making** aligned with regulatory expectations,
- **Multivariate approaches** that enable real-time release testing (RTRT) in continuous manufacturing and improved quality outcomes.

A real-world example involving strategies to enhance patient centricity will illustrate how these capabilities set CMC statisticians apart from other quantitative stakeholders. By leveraging these differentiators, statisticians can move beyond a support role to become equal partners in shaping the future of drug development.

Outline

- Introduction
- Some history - Data Science, Big Data impacting CMC Statistics
  - Roger Hoerl (2021) laying a path forward
- Key Differentiators
- Collaboration and Case Study
- Summary

Data Science, Big Data impacting CMC Statistics

- CMC statistics in the pharmaceutical industry
  - 1950s – Lilly, Merck, Smith-Kline (now GSK) hired the first nonclinical statisticians although they were not referred to as such until the early 1980s
- Data science traces its origins to 1957 Arthur Samuel coined the term "machine learning“, IBM develops Fortran
- 1997 IBM's supercomputer program, Deep Blue, beat the world chess champion, Gary Kasparov, in a six-game match
- 2010 – Statistics reaches its zenith (Andrew Gelman), Data Science takes off, Formation of Data Science academic departments
- 2013 – Marie Davidian - Aren’t we data science (“Statistics departments (and by extension, Statistics) could become obsolete if they fail to engage with data science.”) https://magazine.amstat.org/blog/2013/07/01/datascience/
- 2018 – David Donoho - 50 years of Data Science (“statisticians can lead by contributing rigorous thinking and foundational principles …. Failure to adapt… risks marginalizing statistics in the era of Big Data.”) https://www.tandfonline.com/doi/full/10.1080/10618600.2017.1384734#abstract
- 2021 NCB Conference - Roger Hoerl https://community.amstat.org/biop/events/ncb/past-presentations/ncb-2021-presentations

Data Science, Big Data impacting CMC Statistics

- 2025 NCB Conference
- 3/12 presentations = 25%
- 1/8 posters = 12.5%

- 2023 NCB Conference
- 1/14 = 7%

- 2021 NCB Conference
- 1/14 = 7%

- 2019 NCB Conference
- 1/11 = 0% (1 poster)

Presentations Title	Stat
Optimal Experimental Designs for Process Robustness	S
Bridging Pharma and Medical Device Technologies	S
To Dilute or Not to Dilute: Nominal Titer Dosing for Genetic Medicines	S
Shelf-Life Estimation and Internal Release Limits accounting for Excursions	S
Measurement System Analysis: Practical Considerations	S
Statistical Analytical comparability between pre- and post-change processes	S
Leveraging Experimental Data to Inform an OMARS designs	S
Bayesian Approach to address common data challenges in CMC	S
Differential Projection Pursuit on Flow Cytometry Data	S
SubVision: Deep Learning to Accelerate Formulation Development	ML
Illuminating the role of statistics in GenAI	ML
Application of Causal Machine Learning in Pharmaceutical Manufacturing	ML

Poster Title	Stat
Tree-based Machine Learning Techniques in cell-gene therapy experiments	ML
Is your Excipient Robustness Study Telling the Full Story?	S
Comparison of 4 methods in Method Agreement Analysis	S
Rethinking Scale Down Model Qualification: Some Statistical Evolutions	S
Automated Data Import Toolkits to Maintain Data Quality	S
A Comparison between Frequentist vs. Bayesian in Stability Data Analysis	S
Dealing with Non-Normality in Lot Uniformity Testing	S
Semi-parametric Regression Techniques Applied to Biologics Development	S

J&J

Roger Hoerl (2021) Statistical Thinking and Engineering in a Big Data World

Surveying the impact of Big Data

- "…Big Data … has not produced the results that many of us were expecting"
- "…the fundamentals of sound statistical thinking have been overlooked, perhaps because they are no longer viewed as important in a Big Data world"
- "…the engineering aspect of statistics/data science, that is, how to link and integrate tools in a sequential manner to solve complex problems, is grossly under-developed"

A path forward for Statistics

- … the statistics/analytics profession has the potential to significantly enhance its impact on society…
- …statistical thinking is still relevant in a Big Data World (distinct CMC Statistics expertise)
- …need for an engineering mindset to balance a statistical/data science mindset
- Integrating the principles of statistical thinking, with statistical engineering, can produce results…"

Implications of Roger Hoerl's survey for CMC statisticians

- Recognize the state of affairs : Evolution vs Transformation,
- Cross-Functional Collaboration
- Maintaining Visibility and Influence

Key Differentiators

Key Differentiators of CMC Statisticians

Regulatory fluency
CMC statisticians navigate regulatory guidelines to ensure compliance in drug development and manufacturing.

DoE expertise
Use of designed experiments helps build robust and optimized processes in pharmaceutical production.

Process and Product Understanding
Deep knowledge of process-product interactions enables efficient modeling for knowledge building and improvement.

Risk Framing for Decisions
Statistical insights based on Bayesian modeling support actionable risk-based decision-making in drug development.

Lifecycle Stewardship
Continuous improvement is ensured from development through commercial manufacturing stages.

Case Study

Putting the key differentiators into practice

Setting Specifications

Company strategy

- Quality by Design
- Process capability leading to control limits
- Scientific/Clinical judgment

FDA strategy

- Assess empirical ranges of product used in phase 3 clinical trials
  - Example – Dissolution Guidance (1997) - … the specifications should be based on the dissolution characteristics of batches used in pivotal clinical trials …

Salient Points

- Clinical outcomes are related to batch parameters, not to individual dosage units
  - Conflation of product with analytical determinations (Parameter space vs Data Space) – Patients get dosage units, not analytical determinations
  - CMC quality focus on individual dosage units is contrary to clinical trials paradigm
- ***Specifications are intended to addressg customer requirements but no formal nonclinical (CMC) - clinical causal linkage is established***

Regulatory Context

- Sandra Suarez-Sharp. Establishing Clinically Relevant Drug Product Specifications: FDA Perspective. American Association of Pharmaceutical Scientists Annual Meeting and Exposition, October 16, 2012, Chicago IL (US).
- Cares Act 2017 mandating "Patient Focused" drug development
  - to accelerate the discovery, development, and delivery of 21st century cures…
  - Section 3002 Patient Focused Drug Development Guidance – Paragraph c

    (2) methodological approaches that may be used to develop and identify what is most important to patients with respect to burden of disease, burden of treatment, and the benefits and risks in the management of the patient's disease;

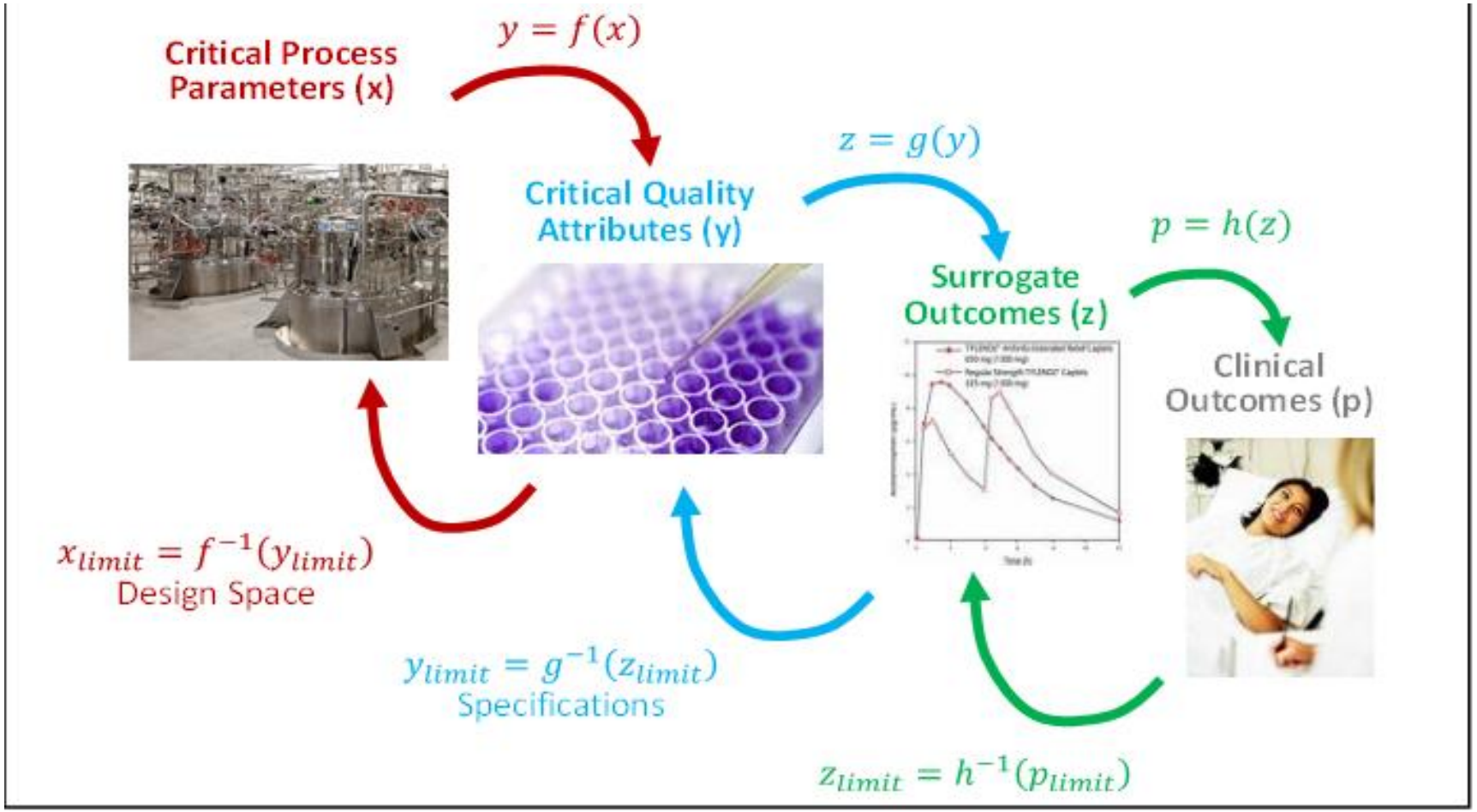
    - Patient-centric in the CMC space
- 2022 PRIME Toolbox Guidance
  - Section 4.4.3 : Setting of specifications
    - The justification of specification limits for CQAs should be linked to clinical performance rather than solely derived from statistical methods such as tolerance intervals.

A framework for a Paradigm Shift for defining Specifications

- Goal : Judicious causal linkages of CQA to clinical responses
  - "Clinically relevant" Critical Quality Attributes (CQAs) and/or Critical Material Attributes (CMAs) are linked to Critical Process Parameters (CPPs)
  - Critical Quality Attributes (CQAs) have impact on efficacy, safety, and/or product performance
- Acceptance Criteria (AC) are defined by multivariate relationships between CPPs and CQAs enabled by operational criteria to be contained within a Design Space
- Define patient response as a function of product quality attributes and functions derived from CPPs and CMAs.
- Assumes the existence of a physiologic variable as a surrogate for therapeutic activity eg. plasma concentration
- "Safety and Efficacy by Design" to achieves Patient driven specifications
- This will require a close collaborations between the stakeholders

Causal linkages from manufacture to clinical outcome

CMC CQAs intermediate linkages from manufacturing floor To Patient Outcome



From Tim Schofield

A QbD human pK Study

Nonclinical (CMC) – Clinical Linkage Design – QbD pK Study

Objective – Characterize formulation/process parameter effects on pK performance (AUC, Cmax, Tmax)

Example QbD study: Consider a CMC formulation/process study with following factors:

- ✓ Factor 1 – Disintegrant level (Low, High)
- ✓ Factor 2 – Particle Size level (Low, High)
- ✓ Tablet Hardness (Low, High) as a split plot factor
- ✓ 2^2 factorial + 1 split plot factor + Center Pts = design point 10

- Responses: Dissolution, Tablet properties are studied

QbD manufacturing design

Design Point	Factor 1	Factor 2	Factor 3	
			Low	High
1-2	Low	Low	A	B
3-4	Low	High	D	E
5-6	High	Low	F	G
7-8	High	High	H	I
9	Mid	Mid	C1	
10	Mid	Mid	C2	

BIBD design for 10 formulations in 15 subjects (4 periods)

BIB Design for 12 Formulations 15 Subjects in 4 Periods				
Subject	**Period 1**	**Period 2**	**Period 3**	**Period 4**
1	E	I	A	G
2	C1	E	A	D
3	I	G	D	B
4	D	C2	F	G
5	F	H	I	C1
6	C2	C1	D	I
7	G	B	C1	H
8	E	G	C1	F
9	A	H	G	C2
10	F	I	B	A
11	B	F	E	C2
12	D	E	B	H
13	C1	A	C2	B
14	H	D	F	A
15	I	C2	H	E

This is the final BIBD design of the 10 factor combinations in as few human subjects as possible.

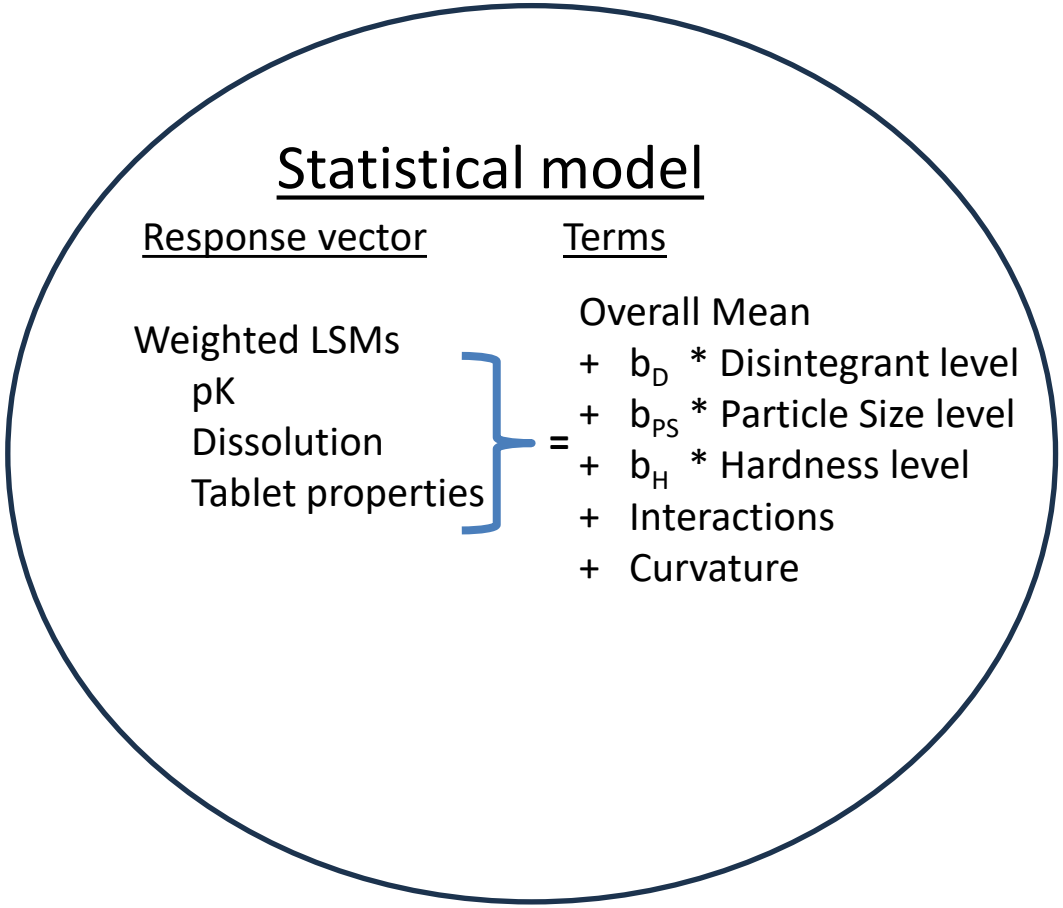
Parameters of the BIBD

v = 10 treatments (formulations)
r = 6 replications
b = 15 blocks (subjects)
k = 4 treatments/block
λ = 2 number of times each pair occurs together

pK responses are measured for each subject across 4 periods

NB more than 4 periods is not practical Wash out at least 5x$t_{1/2}$

QbD pK Design



- Therapeutic effect and Product Quality attributes (Dissolution, Tablet properties) as a multivariate response are linked to the CMC formulation /process parameters

- Final deliverable is a statistical relationship describing effective Patient Exposure
  - Risk control strategies and common Bayesian model to describe risk

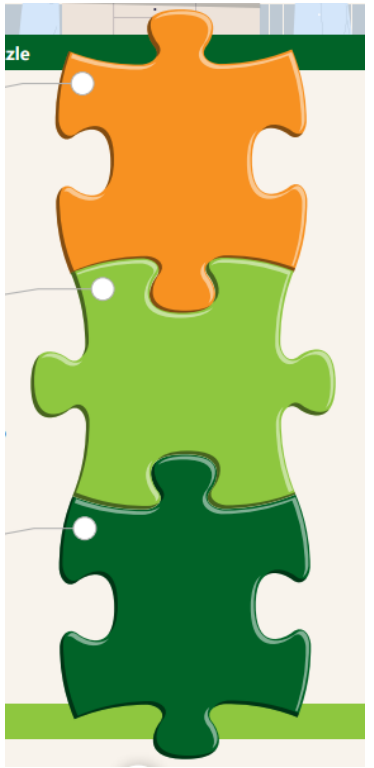
Case Study features

- Key differentiators highlighted
  - Regulatory fluency – awareness of a changing regulatory landscape imply a new paradigm for the development of “patient centric” specifications, integrating with the vision of QbD
  - Close collaboration with key stakeholders
  - Application of statistical designs to achieve efficiency, maintaining statistical rigor
  - Risk estimation through a Bayesian model leading to a design space linked to physiologic response
  - Improved process and product understanding, leading to lifecycle management , direct business impact
- Statistical engineering concepts to establish a framework for pursuing patient centric specifications that identify causal linkages
- Cannot overemphasize the importance of close collaboration between the stakeholders.

Summary

- The “Big Tent” include Big Data, Data Science, and CMC Statistics, how we synergize is the challenge
  - Evolution means learning about Data Science and using it when appropriate for company benefit
  - Maintain a holistic “engineering” perspective
- The value proposition of CMC statistics is enhanced through the key differentiators
  - Collaboration
  - Knowledge building
  - Exploiting core strengths
- The future is bright for CMC statisticians

Enhancing the value proposition of CMC Statistics to help connect the pieces of the puzzle



https://www.americanpharmaceuticalreview.com/Industry-Expert-Hub/619193-Building-the-Drug-Development-Puzzle-Key-Analytical-Tests-and-Practices-for-Success/

Thank you for your attention

Johnson&Johnson

Back up slides

Marie Davidian (2013) Key Points

- **Disconnect Between Statistics and Data Science**

- **Data Science Landscape**
  - Businesses and universities are creating data science centers and programs without significant input from statistics departments.

- **Opportunities for Statisticians**
  - Big Data needs statistical thinking: sampling, design, causal inference must scale up.
  - Statisticians can lead in developing "experimental design for the 21st century."
  - Many current data science practitioners lack deep understanding of these principles.

- **Skills for Modern Statisticians**
  - Strong foundation in statistical theory and methods plus:
    - Programming (R, Python, scripting)
    - Parallel computing and handling large, messy data
    - Data visualization (e.g., D3.js)
    - Machine learning techniques
    - Communication and real-world problem-solving

- **Call to Adapt**
  - Data science is here to stay; statisticians must evolve to remain relevant and address foundational issues.
  - Statistics departments (and by extension, Statistics) could become obsolete if they fail to engage with data science.

David Donoho (2018) Key Points

- **Historical Context**
  - John Tukey's 1962 vision of "The Future of Data Analysis"
  - Data science is not new, an evolution of ideas Tukey championed

- **Gap Between Statistics and Data Science**
  - Traditional statistics focuses heavily on theory and inference
  - Modern data science demands skills in programming, data cleaning, visualization,

- **Core Components of Data Science**
  - Data Gathering, Preparation, and Cleaning
  - Data Exploration and Visualization
  - Data Representation and Transformation
  - Computing with Data
  - Data Modeling
  - Science about Data Science (meta-analysis, reproducibility)

- **Educational Reform**
  - Calls for curricula that teach how to handle messy, large-scale data

- **Philosophical Point**
  - Data science is not replacing statistics—it's a broader umbrella that includes statistical thinking.
  - Statisticians risk irrelevance if they fail to adapt and claim their role in this expanded ecosystem.