White Paper

For reprint orders, please contact: reprints@future-science.com

Bioanalysis

2021 White Paper on Recent Issues in Bioanalysis: ISR for Biomarkers, Liquid Biopsies, Spectral Cytometry, Inhalation/Oral & Multispecific Biotherapeutics, Accuracy/LLOQ for Flow Cytometry (Part 2 – Recommendations on Biomarkers/CDx Assays Development & Validation, Cytometry Validation & Innovation, Biotherapeutics PK LBA Regulated Bioanalysis, Critical Reagents & Positive Controls Generation)

Sarah Hersey[†,1], Steve Keller[†,§,2], Joel Mathews[†,3], Lindsay King[†,13], Abbas Bandukwala[†,4], Flora Berisha[†,5], Mary Birchler[†,6], Joe Bower[†,7], Valerie Clausen[†,8], Jose Duarte[†,§,9], Fabio Garofolo*[,†,§,10], Shirley Hopper[†,11], Sumit Kar[†,§,12], Omar Mabrouk[†,14], Jean-Claude Marshall[†,15], Kristina McGuire[†,16], Michael Naughton[†,17], Yoshiro Saito[†,18], Imelda Schuhmann[†,19], Gizette Sperinde[†,20], Priscila Teixeira[†,§,9], Alessandra Vitaliti[†,§,19], Yow-Ming Wang[†,4], Richard Wnek[†,21], Yan Zhang[†,1], Sue Spitz[§,22], Vilma Decman[§,23], Steven Eck[§,24], Jose Estevam[§,25], Polina Goihberg[§,41], Enrique Gómez Alcaide[§,9], Christèle Gonneau[§,26], Michael Nathan Hedrick[§,1], Gregory Hopkins[§,27], Fabian Junker[§,9], Sandra Nuti[§,28], Ulrike Sommer[§,19], Nathan Standifer[§,29], Chad Stevens[§,13], Erin Stevens[§,30], Carrie Hendricks[#,31], Meenu Wadhwa[#,11], Albert Torri[#,16], Mark Ma[#,32], Shannon Harris[#,33], Seema Kumar[#,37], Michael A Partridge[#,16], Teresa Caiazzo[#,13], Shannon Chilewski[#,1], Isabelle Cludts[#,11], Kelly Coble[#,34], Boris Gorovits[#,35], Christine Grimaldi[#,34], Gregor Jordan[#,36], John Kamerud[#,13], Beth Leary[#,13], Meina Liang[#,29], Hanjo Lim[#,20], Andrew Mayer[#,6], Ellen O'Connor[#,24], Nisha Palackal[#,16], Johann Poetzl[#,38], Sandra Prior[#,11], Mohsen Rajabi Abhari[#,4], Natasha Savoie[#,12], Catherine Soo[#,39], Mark Ware[#,5], Bonnie Wu[#,5], Yang Xu[#,40], Tong-Yuan Yang[#,5] & Jad Zoghbi[#,31]

[1]Bristol-Myers Squibb, Lawrenceville, NJ, USA
[2]AbbVie, South San Francisco, CA, USA
[3]Ionis, Carlsbad, CA, USA
[4]US FDA, Silver Spring, MD, USA
[5]Janssen R&D, Spring House, PA, USA
[6]GlaxoSmithKline, Collegeville, PA, USA
[7]Precision for Medicine, Bethesda, MD, USA
[8]Alnylam, Cambridge, MA, USA
[9]F. Hoffmann-La Roche, Basel, Switzerland
[10]BRI – a Frontage Company, Vancouver, BC, Canada
[11]UK MHRA, London, UK
[12]WRIB, Montreal, QC, Canada
[13]Pfizer, Cambridge, MA, USA
[14]Biogen, Cambridge, MA, USA
[15]Moderna, Cambridge, MA, USA
[16]Regeneron, Tarrytown, NY, USA
[17]GlaxoSmithKline, Stevenage, UK
[18]Japan MHLW-NIHS, Tokyo, Japan
[19]Novartis, Basel, Switzerland
[20]Genentech, South San Francisco, CA, USA
[21]Merck, Kenilworth, NJ, USA
[22]Incyte Research Institute, Wilmington, DE, USA
[23]GlaxoSmithKline, Philadelphia, PA, USA
[24]AstraZeneca, Gaithersburg, MD, USA
[25]Takeda, Cambridge, MA, USA
[26]Labcorp, Meyrin, Switzerland
[27]Bluebird Bio, Cambridge, MA, USA
[28]GSK vaccines, Rockville, MD, USA
[29]AstraZeneca, South San Francisco, CA, USA
[30]Pfizer, Groton, CT, USA
[31]Sanofi, Framingham, MA, USA
[32]Alexion, New Haven, CT, USA
[33]HilleVax, Cambridge, MA, USA
[34]Boehringer Ingelheim, Ridgefield, CT, USA
[35]Sana Bio, Cambridge, MA, USA

[†]SECTION 1 – Biomarker & CDx Development and Validation (Authors in Section 1 are presented in alphabetical order of their last name, with the exception of the first 4 authors who were session chairs, working dinner facilitators and/or major contributors.)
[§]SECTION 2 – Cytometry Validation & Innovation (Authors in Section 2 are presented in alphabetical order of their last name, with the exception of the first 3 authors who were session chairs, working dinner facilitators and/or major contributors.)
[#]SECTION 3 – LBA Regulated Bioanalysis, Critical Reagents & Positive Controls (Authors in Section 3 are presented in alphabetical order of their last name, with the exception of the first 7 authors who were session chairs, working dinner facilitators and/or major contributors.)

Disclaimer: The views expressed in this article are those of the authors and do not reflect official policy of the US FDA, Europe EMA, UK MHRA, Austria AGES, Norway NoMA, Brazil ANVISA, Health Canada, Japan MHLW and WHO. No official endorsement by the FDA, EMA, MHRA, AGES, NoMA, ANVISA, Health Canada, MHLW and WHO is intended or should be inferred.

Note: The White Paper on Recent Issues in Bioanalysis [1–28] is issued yearly to ensure that the Global Bioanalytical Community is informed with the most updated Industry/Regulators' recommendations/consensus based on novel case studies and emerging approaches each year. Please always refer to the latest version of these recommendations/consensus since they may change and advance over the years due to continuous evolutions of Science, Regulations and Technologies.

*Author for correspondence: fabiogarofolo@hotmail.com

newlands press

[36]Roche Pharma Research & Early Development, Roche Innovation Center, Munich, Germany
[37]EMD Serono, Andover, MA, USA
[38]Sandoz, Munich, Germany
[39]Health Canada, Ottawa, ON, Canada
[40]Merck, West Point, PA, USA
[41]Pfizer, Andover, MA, USA

The 15th edition of the Workshop on Recent Issues in Bioanalysis (15th WRIB) was held on 27 September to 1 October 2021. Even with a last-minute move from in-person to virtual, an overwhelmingly high number of nearly 900 professionals representing pharma and biotech companies, contract research organizations (CROs), and multiple regulatory agencies still eagerly convened to actively discuss the most current topics of interest in bioanalysis. The 15th WRIB included three Main Workshops and seven Specialized Workshops that together spanned 1 week in order to allow exhaustive and thorough coverage of all major issues in bioanalysis, biomarkers, immunogenicity, gene therapy, cell therapy and vaccines. Moreover, in-depth workshops on biomarker assay development and validation (BAV) (focused on clarifying the confusion created by the increased use of the term "context of use" [COU]); mass spectrometry of proteins (therapeutic, biomarker and transgene); state-of-the-art cytometry innovation and validation; and critical reagent and positive control generation were the special features of the 15th edition. This 2021 White Paper encompasses recommendations emerging from the extensive discussions held during the workshop, and is aimed to provide the bioanalytical community with key information and practical solutions on topics and issues addressed, in an effort to enable advances in scientific excellence, improved quality and better regulatory compliance. Due to its length, the 2021 edition of this comprehensive White Paper has been divided into three parts for editorial reasons. This publication (Part 2) covers the recommendations on ISR for Biomarkers, Liquid Biopsies, Spectral Cytometry, Inhalation/Oral & Multispecific Biotherapeutics, Accuracy/LLOQ for Flow Cytometry. Part 1A (Endogenous Compounds, Small Molecules, Complex Methods, Regulated Mass Spec of Large Molecules, Small Molecule, PoC), Part 1B (Regulatory Agencies' Inputs on Bioanalysis, Biomarkers, Immunogenicity, Gene & Cell Therapy and Vaccine) and Part 3 (TAb/NAb, Viral Vector CDx, Shedding Assays; CRISPR/Cas9 & CAR-T Immunogenicity; PCR & Vaccine Assay Performance; ADA Assay Comparability & Cut Point Appropriateness) are published in volume 14 of Bioanalysis, issues 9 and 11 (2022), respectively.

First draft submitted: 12 March 2022; Accepted for publication: 28 April 2022; Published online: 17 May 2022

Keywords: bioanalysis • biomarkers • cell therapy • gene therapy • immunogenicity • vaccine • WRIB

Index

Abbreviations	
ABC:	Antibody binding capacity

ADA:	Anti-drug antibody
AI:	Artificial intelligence
ASO:	Antisense oligonucleotide
BAV:	Biomarker assay validation
BBB:	Blood–brain barrier
BLA:	Biologics license application
BLI:	Biolayer interferometry
BMV:	Bioanalytical method validation
CAP:	College of American Pathologists
CAR T:	Chimeric antigen receptor T-cell
CDR:	Complementarity-determining region
CDx:	Companion diagnostics
Companion diagnostic:	A companion diagnostic is a medical device, often an *in vitro* device, which provides information that is essential for the safe and effective use of a corresponding drug or biological product. The test helps a health care professional determine whether a particular therapeutic product's benefits to patients will outweigh any potential serious side effects or risks [126].
CLIA:	Clinical Laboratory Improvement Amendments
CLSI:	Clinical Laboratory Standards Institute
CMC:	Chemistry, manufacturing and controls
COU:	Context of use
CRISPR:	Clustered regularly interspaced short palindromic repeats
CRO:	Contract research organization
CSF:	Cerebrospinal fluid
CyTOF:	Time-of-flight mass cytometry
ddPCR:	Droplet digital polymerase chain reaction assays
DIG:	Digoxigenin
DoL:	Degree of labeling
DSP:	Digital special profiling
ECP:	Erythrocyte microparticle
ELISA:	Enzyme-linked immunosorbent assay
EMP:	Endothelial microparticle
eQC:	Endogenous quality control
ERF:	Equivalent number of reference fluorophores
EV:	Extracellular vesicle
FFP:	Fit-for-purpose
FFPE:	Formalin-fixed paraffin-embedded
FMO:	Fluorescence minus one
FMx:	Fluorescence minus one or three
FNA:	Fine needle aspiration
FTE:	Full-time equivalents
GalNAc:	N-acetylgalactosamine
GCP:	Good Clinical Practices
GCLP:	Good Clinical Laboratory Practices
GLP:	Good Laboratory Practices
GOI:	Gate of interest
GI:	Gastro-intestinal
HMW:	High molecular weight

HRMS:	High resolution mass spectrometry
IDE:	Investigational device exemption
IFU:	Indications for use
IHC:	Immunohistochemistry
IMC:	Imaging mass cytometry
IND:	Investigational new drug
IQ/OQ:	Installation qualification/operation qualification
IRB:	Institutional review board
ISR:	Incurred sample reproducibility
ISS:	incurred sample stability
IU:	Intended use
IVD:	*In vitro* diagnostic
KOL:	Key opinion leader
LBA:	Ligand binding assay
LCMS:	Liquid chromatography mass spectrometry
LDT:	Laboratory developed test
LLOQ:	Lower limit of quantitation
LMP:	Leukocyte microparticles
LOB:	Limit of blank
LOD:	Limit of detection
mAb:	Monoclonal antibody
MDB:	Multi-domain biotherapeutic
MdFI:	Median fluorescence intensity
MESF:	Molecules of equivalent soluble fluorochromes
mIF:	Multiplexed immunofluorescence
MMO:	Mass minus one
MOA:	Mechanism of action
MP:	Microparticle
MRD:	Minimum required dilution
MRE:	Minimal number of required events
MRM:	Multiple reaction monitoring
MRPS:	Microfluidic resistive pulse sensing
MS:	Mass spectrometry
MSD:	MesoScale Discovery
NAb:	Neutralizing antibody
NDA:	New drug application
NGS:	Next-generation sequencing
NTA:	Nitriloacetic acid
pAb:	Polyclonal antibody
PAD:	Peripheral artery disease
PBMC:	Peripheral blood mononuclear cells
PC:	Positive control, used in an immunogenicity assay
PCR:	Polymerase chain reaction assays
PD:	Pharmacodynamics
PET:	Positron emission tomography
PK:	Pharmacokinetics

PMP:	Platelet microparticles
POC:	Point of care
PTM:	Post-translational modifications
QC:	Quality control
qPCR:	Quantitative polymerase chain reaction assays
RMT:	Receptor-mediated transcytosis
RNA:	Ribonucleic acid
RO:	Receptor occupancy
RUO:	Research use only
SD:	Standard deviation
SEC-MALS:	Size exclusion chromatography-multi angle light scattering
SNR:	Signal to noise ratio
SOP:	Standard operating procedure
SPR:	Surface plasmon resonance
t-SNE:	t-distributed stochastic neighbor embedding
tAb:	Therapeutic antibody
TE:	Target engagement
Target engagement:	Interaction of ligands with their target biomolecules
Tfh:	T-follicular helper
TMDD:	Target-mediated drug disposition
UK NEQAS:	United Kingdom National External Quality Assessment Service
VALID:	Verifying accurate leading-edge IVCT development
WRIB:	Workshop on Recent Issues in Bioanalysis

Introduction

The 15th edition of the Workshop on Recent Issues in Bioanalysis (15th WRIB) was held on 27 September to 1 October 2021. Even with a last-minute move from in-person to virtual, an overwhelmingly high number of nearly 900 professionals representing pharma and biotech companies, contract research organizations (CROs), and multiple regulatory agencies still eagerly convened to actively discuss the most current topics of interest in bioanalysis. The 15th WRIB included three Main Workshops and seven Specialized Workshops that together spanned 1 week in order to allow exhaustive and thorough coverage of all major issues in bioanalysis, biomarkers, immunogenicity, gene therapy, cell therapy and vaccines.

Moreover, in-depth workshops on biomarker assay development and validation (BAV) (focused on clarifying the confusion created by the increased use of the term "context of use" [COU]); mass spectrometry of proteins (therapeutic, biomarker and transgene); state-of-the-art cytometry innovation and validation; and critical reagent and positive control generation were the special features of the 15th edition.

As in previous years, this year's WRIB continued to gather a wide diversity of international, industry opinion leaders and regulatory authority experts working on both small and large molecules to facilitate sharing and discussions focused on improving quality, increasing regulatory compliance, and achieving scientific excellence on bioanalytical issues.

The active contributing chairs included: Dr Eugene Ciccimaro (BMS), Dr Anna Edmison (Health Canada), Dr Fabio Garofolo (BRI), Dr Swati Gupta (AbbVie), Dr Shannon Harris (HilleVax), Dr Carrie Hendricks (Sanofi), Ms. Sarah Hersey (BMS), Dr Steve Keller (AbbVie), Dr Lina Loo (Vertex), Dr Mark Ma (Alexion), Dr Joel Mathews (Ionis), Dr Meena (Stoke), Dr Manoj Rajadhyaksha (Alexion), Dr Ragu Ramanathan (Vertex), Dr Susan Spitz (Incyte), Dr Dian Su (Mersana), Dr Matthew Szapacs (Abbvie), Dr Albert Torri (Regeneron), Dr Jian Wang (Crinetics), Dr Jan Welink (EU EMA) and Dr Yuling Wu (AstraZeneca).

The participation of major and influential regulatory agencies continued to grow at the 15th WRIB during its traditional Interactive Regulators' sessions including presentations and panel discussions on:

- **Regulated Bioanalysis and BMV Guidance/Guidelines:** Dr Seongeun Julia Cho (US FDA), Dr Arindam Dasgupta (US FDA), Dr Anna Edmison (Health Canada), Dr Elham Kossary (WHO), Mr. Gustavo Mendes Lima Santos (Brazil ANVISA), Dr Sam Haidar (US FDA), Dr Sandra Prior (UK MHRA), Dr Mohsen Rajabi Abhari (US FDA), Dr Diaa Shakleya (US FDA), Dr Catherine Soo (Health Canada), Dr Nilufer Tampal (US FDA), Mr. Stephen Vinter (UK MHRA), Dr Yow-Ming Wang (US FDA), Drs Jan Welink (EU EMA) and Dr Jinhui Zhang (US FDA)
- **Biotherapeutic Immunogenicity, Gene Therapy, Cell Therapy and Vaccines:** Dr Nirjal Bhattarai (US FDA), Dr Elana Cherry (Health Canada), Dr Isabelle Cludts (UK MHRA), Dr Heba Degheidy (US FDA), Dr Akiko Ishii-Watabe (Japan MHLW), Dr Susan Kirshner (US FDA), Dr Kimberly Maxfield (US FDA), Dr Joao Pedras-Vasconcelos (US FDA), Dr Mohsen Rajabi Abhari (US FDA), Dr Vijaya Simhadri (US FDA), Dr Dean Smith (Health Canada), Dr Therese Solstad (EU EMA/Norway NoMA), Dr Daniela Verthelyi (US FDA), Dr Meenu Wadhwa (UK MHRA), Ms. Leslie Wagner (US FDA), Dr Günter Waxenecker (Austria AGES), Dr Haoheng Yan (US FDA) and Dr Lucia Zhang (Health Canada)
- **Biomarkers/CDx and BAV Guidance/Guidelines:** Mr. Abbas Bandukwala (US FDA), Dr Soma Ghosh (US FDA), Dr Shirley Hopper (UK MHRA), Dr Kevin Maher (US FDA), Dr Yoshiro Saito (Japan MHLW) and Dr Yow-Ming Wang (US FDA)

All the traditional "working dinners" attended by both industry key opinion leaders (KOL) and regulatory representatives were held in a virtual format this year, and the extensive and fruitful discussions from these "working dinners" together with the lectures and open panel discussions amongst the presenters, regulators and attendees culminated in consensus and recommendations on items discussed in this White Paper.

A total of 66 recent issues ('hot' topics) were addressed and distilled into a series of relevant recommendations. Presented in the current White Paper is the background on each issue, exchanges, discussions, consensus and resulting recommendations.

Due to its length, the 2021 edition of this comprehensive White Paper has been divided into three parts for editorial reasons. This publication covers **Part 2** recommendations.

Part 1 – Volume 14 Issue 9 (May)

Regulated Mass Spectrometry of Large Molecules (six topics)

1. Regulated Bioanalysis/ICH M10 of Large Molecules by Mass Spectrometry
2. Internal Standard Selection in Hybrid Assays in Regulated Bioanalysis/ICH M10
3. Regulatory Feedback on Nanomedicines Bioanalysis by Mass Spectrometry
4. Bioanalytical Mass Spectrometry Strategies for CRISPR Quantification
5. Extracellular Vesicle Bioanalysis by Mass Spectrometry
6. Immunocapture Platform Considerations for Intact Mass LCMS

Endogenous Compounds & Complex Methods (five topics)

1. Chiral Methods for Method Development and BMV
2. ICH M10 Section 7.1 for Endogenous Compound Quantification
3. Tissue Analysis, Rare Matrices and Atypical Sample Collection in Regulated Bioanalysis
4. Quantitation of Intracellular Disposition of Oligonucleotides and Sensitivity/Specificity Challenges
5. Recent Developments of Urinary Endogenous Compounds and Fit-For-Purpose Validation

Regulated Bioanalysis for Small Molecule & Point of Care (six topics)

1. Dealing with GLP, GCP and GCLP Frameworks in Regulated Bioanalysis
2. Importance of Incurred Sample Stability in Regulated Bioanalysis
3. Challenges when Changing Platforms (LBA to LCMS) in Regulated Bioanalysis
4. Patient-Centric Approaches and Point of Care in Regulated Bioanalysis
5. Regulatory Standards to Perform Bioanalysis in China
6. Bioanalytical Challenges for Oncology Drug Development

Mass Spectrometry of Proteins (six topics)

1. Hybrid Assays to Quantify Therapeutics Proteins

2. PTM/Glycosylation Analysis for Biomarkers and Biotherapeutics
3. Hybrid Assays to Quantify Protein Biomarkers
4. Hybrid Assays for Target Engagement Assessment
5. Quantification of ADA by Hybrid Assays
6. Hybrid Assays to Quantify Transgene Proteins

Input from Regulatory Agencies on Regulated Bioanalysis & BMV

Input from Regulatory Agencies on Immunogenicity, Biomarkers, Gene Therapy, Cell Therapy & Vaccines

Part 2 – Volume 14 Issue 10 (May)

Biomarkers & CDx Development & Validation (nine topics)

1. Liquid Biopsy: Challenges and Opportunities with Extracellular Vesicles
2. ISR for Biomarker Assays, Parallelism & Biomarker Assay Validation Guidance
3. Clinical Biomarkers as Surrogate Endpoints or for Patient Segmentation
4. Quality Oversight of CLIA Laboratories for Companion Diagnostics
5. Emerging Trends and Impact on Diagnostic Development
6. Breath & Airway Biomarker Determination
7. High Sensitivity Platforms for Biomarkers and Companion Diagnostics
8. Exploratory and Target Engagement Biomarker Assays
9. PBMC Sample Collection for Pharmacodynamic Biomarkers

Cytometry Validation & Innovation (eight topics)

1. Recent Developments in Flow Cytometry Validation in a Bioanalytical Lab
2. Evaluation of Accuracy for Flow Cytometry in Regulated Laboratories
3. Sensitivity Determination in Flow Cytometry Validation
4. Clinical Biomarker Development, Validation and Interpretation by Cytometry
5. Spectral Cytometry Ultra-High Order/Dimensional Assays in Clinical Applications
6. Imaging Cytometry Quantitative Analysis of Target Engagement
7. Mass Cytometry in Clinical Biomarkers and "Clinical Trial Compatibility"
8. Multivariate Analytical Techniques and Multiparameter Flow Cytometry

LBA Regulated Bioanalysis, Critical Reagents & Positive Controls (nine topics)

1. Bioanalytical Challenges for Inhalation and Oral Delivery of Biologics
2. Free, Bound, Total, Active, Monoactive, Biactive, and Multiactive PK Assays
3. Implementing Free/Total PK Assays in Regulated Bioanalysis
4. Multi-Domain Biotherapeutic PK Assays in Regulated Bioanalysis
5. Bioanalytical Challenges to Study the Biodistribution of Biotherapeutics
6. Advanced Approaches in Critical Reagent Selection for PK Assays
7. Challenges with Positive Control Generation for ADA Assays
8. Critical Reagent Assay Comparability
9. Novel Critical Reagent Modalities: "Thinking out of the Box"

Part 3 – Volume 14 Issue 11 (June)

Gene Therapy, Cell Therapy & Vaccines (nine topics)

1. TAb/NAb & Anti-Viral Vector Antibody Companion Diagnostic Assays
2. Viral Vector Shedding Assays
3. Viral Vector Gene Therapy Immunogenicity & Pre-Existing Immunity
4. CRISPR/Cas9 Immunogenicity & Bioanalytical Challenges
5. CAR-T Immunogenicity & Cellular Kinetics
6. qPCR/ddPCR Assay Performance
7. Analyte Stability in Vaccine Serology Assays
8. Vaccine Critical Reagent Management and Bridging
9. Vaccine Bioanalytical Assays & Immune Monitoring

Immunogenicity of Biotherapeutics (eight topics)

1. NAb Assays – Drug and Target Interference
2. ADA Cut Points – Appropriateness and False Positive Rates
3. Circulating Immune Complexes – ADA/Drug Complexes
4. ADA Assay Comparability
5. Integrated Summary of Immunogenicity Harmonization
6. China NMPA Immunogenicity Guidance
7. MDB & Bispecific Immunogenicity
8. Biosimilar ADA Assay Validation & Harmonization

SECTION 1 – Biomarker & CDx Development & Validation

Sarah Hersey[1], Steve Keller[2], Joel Mathews[3], Lindsay King[13], Abbas Bandukwala[4], Flora Berisha[5], Mary Birchler[6], Joe Bower[7], Valerie Clausen[8], Jose Duarte[9], Fabio Garofolo[10], Shirley Hopper[11], Sumit Kar[12], Omar Mabrouk[14], Jean-Claude Marshall[15], Kristina McGuire[16], Michael Naughton[17], Yoshiro Saito[18], Imelda Schuhmann[19], Gizette Sperinde[20], Priscila Teixeira[9], Alessandra Vitaliti[19], Yow-Ming Wang[4], Richard Wnek[21] & Yan Zhang[1]

Authors are presented in alphabetical order of their last name, with the exception of the first 4 authors who were session chairs, working dinner facilitators and/or major contributors.

Author affiliations can be found at the beginning of the article.

HOT TOPICS & CONSOLIDATED QUESTIONS COLLECTED FROM THE GLOBAL BIOANALYTICAL COMMUNITY

The topics detailed below were considered as the most relevant "hot topics" based on feedback collected from the 14th WRIB attendees. They were reviewed and consolidated by globally recognized opinion leaders before being submitted for discussion during the 15th WRIB. The background on each issue, discussions, consensus and conclusions are in the next section and a summary of the key recommendations is provided in the final section of this manuscript.

Liquid Biopsy: Challenges & Opportunities with Extracellular Vesicles

What are the best ways to isolate extracellular vesicles (EVs)? What are the best methods of normalizing EV counts in samples? What are the potential approaches and the major limitations in the validation and clinical implementation of EVs? How can the gap in lack of controls and standard reagents be addressed to enable the use of a calibration curve and proper quantification?

ISR for Biomarker Assays, Parallelism & Biomarker Assay Validation Guidance

Do you interpret FDA bioanalytical method validation (BMV) [29] to suggest incurred sample reanalysis (ISR) should be conducted for biomarker assays? Is it within the scope of this guidance? Given the range of biomarkers and COU, ISR as currently defined for pharmacokinetic (PK) does not address biomarker assay needs. Is there a biomarker form of ISR that would be useful? If so what, when and why? Quality controls (QCs), incurred sample stability (ISS) and parallelism are more appropriate to address biomarker assay needs, therefore, are these being sufficiently leveraged? How should ISR be replaced? What acceptance criteria should be applied? What can be done if parallelism fails or cannot be assessed? How can confidence in the assay be assured without parallelism? Is it possible to justify a lack of parallelism for biomarker assays used for regulatory submission? If yes, what are the possible justifications? How will the minimum required dilution (MRD) be determined for a biomarker assay if lack of parallelism is expected? How is the sensitivity of an endogenous analyte measured in a biomarker assay when parallelism fails?

Clinical Biomarkers as Surrogate Endpoints or for Patient Selection/Stratification

To establish clinical utility of a biomarker as a surrogate endpoint, what are the requirements for the assay to be utilized? Can *in vitro* diagnostic (IVD)-cleared/approved assays or laboratory developed tests (LDT) be used to support surrogate endpoints? What are the best practices for the analytical and clinical validation of clinical biomarkers when used as surrogate endpoints?

Quality Oversight of CLIA Laboratories for Companion Diagnostics

Is the use of an LDT an acceptable substitute for an approved companion diagnostic (CDx)? Are current Clinical Laboratory Improvement Amendments (CLIA) requirements sufficient? If bridging is required to demonstrate test equivalency for premarket approval, should the same requirement be applied to an LDT or should the requirement for bridging CDx be lowered? What is the impact on laboratory cost/reimbursement when requiring the approved CDx? What is the likelihood of an inaccurate test using an LDT leading to inappropriate therapeutic assignment and/or patient safety issues? Should LDTs have adverse event review and reporting requirements? What laboratory auditing should be required for CLIA laboratories running CDx and/or LDTs? Is the infrastructure cost-prohibitive to auditing of laboratories performing CDx/LDT patient selection testing? How often and how thorough does the auditing need to be? If a clinical laboratory is running an approved CDx versus an LDT does this change the auditing requirements?

Emerging Trends & Impact on Diagnostic Development

What are the potential impacts of increasingly complex CDx testing? What creates the need for a CDx product? What is necessary to establish analytical and clinical validity? What quality systems should be used for biomarkers for each stage of development of a CDx? What are the most suitable next generation technologies for CDx assays and the strategies for their implementation?

Breath & Airway Biomarker Determination

Breath and airway biomarkers are an emerging but very promising field with broad applications outside of respiratory diseases (cardiovascular, metabolic, neurological). Many new sampling types have emerged from academic groups, small biotech companies, and pharma partnerships. What are the challenges and considerations to take into account with breath biomarkers in clinical trials? Where can bioanalytical scientists contribute to this early discovery field?

High Sensitivity Platforms for Biomarkers

Platforms such as Quanterix SIMOA, MesoScale Discovery (MSD) S-Plex, or MilliporeSigma SMC have provided improved assay sensitivity and generated data on important biomarkers with very low levels to enable study decisions. What are the challenges of these platforms? What strategies can be used to ensure quality reagents and the best assay sensitivity? When measuring endogenous analytes, how can it be ensured that the most appropriate calibrator is chosen? Does validation need to be unique from other immunoassays? Can acceptance criteria be relaxed or are different validation tests needed?

Exploratory & Target Engagement Biomarker Assays

Is broader adoption across the industry anticipated for tissue based pharmacodynamic (PD) assessments in early development clinical trials? Are on-treatment tissues for target engagement (TE)/PD being analyzed? What technologies are being used for sample analysis? How is the data being utilized by stakeholders? What are the most challenging aspects of performing tissue bioanalysis? What steps are critical in assay development, qualification and/or analytical validation: tissue heterogeneity, tissue stability, pre-analytics, other? During bioanalytical support for TE of a candidate molecule, is it common practice to offer free, total and complex assays? Who is currently making the decisions regarding what type of assays are needed to provide evidence of TE? How should matrix effects be managed while minimizing disruption to the drug-target complex in a free target assay? Parallelism will inform what dilution is required for the optimum analytical conditions, but dilution may result in the overestimation of free analyte in study samples. Is it best to prioritize maintaining the complex over the matrix effect? Which is worse? Are there any analytical solutions or platforms that could help?

PBMC Sample Collection for Pharmacodynamic Biomarkers

Peripheral blood mononuclear cells (PBMCs) are collected fairly frequently for biomarker analysis but there are significant logistical challenges that can severely impact the success of sample analysis. Is there enough attention paid to how PBMCs are being sampled during clinical trials? Is COU being considered when choosing the PBMC isolation strategy? How can on-site PBMC sample collection in multicenter clinical trials be further facilitated? PBMC isolation procedures and time of isolation are very dependent on the intended use (e.g., immunophenotyping, *in vitro* stimulation assays, etc.). What are the current recommendations?

DISCUSSIONS, CONSENSUS & CONCLUSIONS

Liquid Biopsy: Challenges & Opportunities with Extracellular Vesicles

Extracellular vesicles contain cell-specific proteins, lipids and nucleic acids that reflect physiological health and pathophysiological state. Previously overlooked as an analytical artifact or cell debris, it is increasingly clear that EVs play an important role in intercellular communication and growing evidence suggests that microparticles (MPs), one of the classes of EVs, are important mediators of physiological and pathological cellular processes [30]. EVs are found in all bodily fluids and they may provide important information about tissue-related changes in the context of drug response, disease activity and/or patient stratification while reducing the patient burden of invasive tissue sampling. For instance, EVs may be used in diagnostic assays for urologic malignancies and liver disease [31,32]. Despite their potential as biomarkers, variability in results may be caused by lack of standardized pre-analytics, including EV isolation and subsequent detection methods. Other challenges with developing and validating robust assays to confirm the potential of EVs as non-invasive biomarkers include EV heterogeneity (size and composition may depend on the cell of origin but also on stimulation), matrix complexity and heterogeneity, lack of reference material, EV size and method sensitivity/specificity.

There are several isolation and extraction methods in use for EVs [33–35]. One can use a simple polymer-based precipitation followed by a protein depletion column to obtain a highly enriched EV fraction. Other methods of purification include size-exclusion chromatography and ultracentrifugation. The isolation method chosen should be informed by the scientific question and downstream application.

EV isolation may not be comparable between independent preparations and each preparation may represent different and/or heterogeneous populations of EVs. One of the main problems with the standardization of isolation methods is the heterogeneity with different types of matrices. Several other biological particles can interfere in EV purity, such as lipoproteins, immuno-complexes, viruses, and cell debris. The level of the impurity is dependent on disease status, diet, and the time of day that the sample was collected, amongst other factors. In the case of human samples, recommendations like fasting prior to sample collection may be advisable.

These challenges are further confounded by studies often utilizing limited subjects; large clinical studies are needed due to the very heterogeneous population and lack of understanding of sample and biomarker stability. The EV field continues to evolve rapidly and new approaches for isolation and detection of EVs continue to advance.

For EVs to have the potential for use as biomarkers, independent confirmation of populations using identical or similar detection approaches may be needed for future studies as well as better reproducibility through more refined control of pre-analytics. Consideration should be given to the validation and implementation of EV methods.

Currently, there are gaps in the lack of controls and standard reagents for EV detection. Most of the materials used to develop EV assays are only for exploratory research and may not be suitable for use in clinical trials. Therefore, the first step in the development of EV assays is to determine whether a qualitative or quantitative method is needed. Qualitative methods can still answer important questions for clinical trials. For quantitative methods, some options for calibrators and reagents include EVs prepared from cell cultures, sample pools from those with high levels of the intended target EVs, or recombinant proteins. Negative controls should also be included. For diagnostic approaches, similar approaches to ExoDx can be followed [36].

Approaches that can be used to detect and quantify EVs include microfluidic resistive pulse sensing (MRPS), microscopy, scatter-based nanoparticle tracking analysis, Western blot or ELISA-based characterization, multifluorescence flow cytometry, global proteomic analysis using mass spectrometry (MS), fluorescence-based nanoparticle tracking analysis, and interferometric imaging coupled to fluorescence analysis.

More detailed definitions and descriptions of EV collection and isolation procedures are needed in published methodologies to allow the comparison of approaches and provide robust recommendations for specific EV populations. A standard for the minimal reporting of EV isolation and detection methods has been proposed by the International Society For Extracellular Vesicles and should be looked at as a source for aiding in the reporting of these methods in manuscripts to enable reproducibility of results.

Normalization may be another challenge in quantitative EV methods depending on the technology used. Quantification of biomarkers on organ-specific EVs should take into consideration two parameters in order to accurately evaluate differences in biomarker expression between samples: expression level of the biomarker on the EVs and the overall number of EVs. Because of these differences, normalization is needed to remove variations between isolations and to be able to access biology. Several means of normalization have been suggested, including nitriloacetic acid (NTA) and total protein. NTA analysis is not yet established in routine diagnostic settings

and may necessitate large sample volumes. The total protein content is not typically suitable for normalization and normalization via subpopulation ratios (biomarker expressing organ-specific EVs/overall organ-specific EVs) demonstrates potential but is currently not sufficiently robust. Thus, there is still work needed regarding data normalization approaches. Stability is also a complex parameter, as the stability of the exosome (or extracellular vesicles) is needed, as well as the stability of the marker of interest. The majority of EV assays are qualitative or semi-quantitative, although advances in the field may allow for more quantitative assays in the future.

The 2020 White Paper in Bioanalysis provided initial recommendations for fit-for-purpose (FFP) validations of EV methods, on the potential value of EV quantification and the methods that need to be developed [28]. It was agreed that more case studies were needed to demonstrate the clinical applications of EVs, especially clinical endpoints. Indeed, more case studies were discussed this year to better corroborate and confirm previous recommendations by focusing on the standardization of pre-analytical and analytical considerations for EVs.

A case study for potential roles of membrane MPs (a subclass of EVs) in the process of atherosclerosis and in particular in peripheral artery disease (PAD) was discussed. Membrane microparticles (or membrane microvesicles) are 0.1–1.0 μm fragments of membrane shed into the extracellular space from cells. MPs may be distinguished from other classes of EVs on the basis of size, content and mechanism of formation [37]. MPs are present in different bodily fluids (e.g., blood, urine, cerebrospinal fluid [CSF]) and can be measured in healthy individuals. Alterations in levels have been shown to be correlated with various diseases conditions, including diabetes, chronic kidney disease, and vascular injury. Circulating MPs can be identified as platelet microparticles (PMPs), endothelial microparticles (EMPs), leukocyte microparticles (LMPs), and erythrocyte microparticles (ECPs) [38–40].

MPs may have the potential to serve as disease, diagnostic, efficacy and safety biomarkers. However, due to the pre-analytical and analytical limitations already discussed, challenges remain in developing and validating robust assays to confirm the potential of MPs as non-invasive biomarkers of disease conditions.

Atherosclerosis occurs when arteries get narrow and stiffen due to a buildup of fatty deposits on the artery walls. In PAD, narrowed arteries reduce blood flow in the limbs. Endothelial dysfunction is a pivotal element in the development and progression of vascular diseases and it has been shown that EMP levels are increased in various vascular pathologies including PAD. A relationship between EMPs and endothelial dysfunction has been proposed [41]. The most widespread approach for MP enumeration and characterization is flow cytometry. The primary limitation in analyzing MPs is their submicron size which is close to the size limit of detection and resolution of the standard flow cytometry instrument. Using the next generation of highly sensitive instruments such as CytoFLEX can bring potentially unprecedented resolution and the needed sensitivity to expand to the nanoparticle range and thus allow for more precise measurement.

Three separate flow cytometry panels were designed for the measurement of PMPs and EMPs to address their potential role as mechanistic, disease and efficacy biomarker in PAD patients treated with an antioxidant agent as part of a clinical trail. Blood collection and sample processing of platelet-poor plasma were based on recommendations by Poncelet *et al.* [42] and adjusted for the clinical trial situation (e.g., centrifugation conditions, etc.). A bead-based approach was used for enumeration.

Validation using healthy volunteer plasma assessed intra-assay repeatability, inter-assay reproducibility, long-term stability, inter-operator reproducibility, and intra- and inter-donor reproducibility in 3 donors. The PMP assay was successfully validated with intra- and inter-assay variability (CV) <17%, fulfilling the acceptance criteria (≤25%), whereas because the EMP levels are very low, the EMP assay could not be fully developed and validated. Good quality data have nevertheless been generated for both PMP and EMP from samples collected across multiple clinical trial centers. Results indicate that PMP and EMP may serve as disease biomarkers; EMP may also be used as a biomarker to monitor the response to treatment evaluated for a new therapy targeting vascular oxidative stress.

Overall, other biological considerations need to be evaluated, such as correlation between the marker levels in the EVs and the tissue biopsy, factors that influence intra-donor variability, factors that influence the release of exosomes, expression of specific markers in specific tissues, correlation with outcomes and disease versus healthy conditions.

ISR for Biomarker Assays, Parallelism & BAV Guidance

ISR is a well-established approach and post-method validation parameter for PK assays [29,43]. It was originally proposed to ensure the precision of the measurement of the incurred samples was similar to that of the spiked samples that were used during validation [44]. It can also help to uncover unexpected metabolites that can be

back-converted to the parent species for small molecule drugs, unexpected protein binding, recovery issues, lack sample heterogeneity, as well as technical and human errors.

Biomarkers were included in the scope of BMV for the first time in 2018 and ISR is expected for biomarker studies in NDAs and BLAs [29] raising the question as to whether ISR should be considered for all biomarker assays as well. The 2018 BMV guidance states: "*When biomarker data will be used to support regulatory decision making, such as the pivotal determination of safety and/or effectiveness or to support dosing instructions in product labeling, the assay should be fully validated.*" However, the BMV 2018 guidance also states: "*The approach used for drug assays should be the starting point for validation of biomarker assays, although the FDA realizes that some characteristics may not apply or that different considerations may need to be addressed.*"

Thus, the applicability of ISR for biomarker assays, even in pivotal studies, has not been widely recognized across the broader biomarker community [15,19]. ISR for biomarkers was not discussed in the Critical Path White Paper [45] and discussion at industry biomarker meetings and in publications has been limited [27].

When evaluating the need for ISR for biomarkers, it is important to understand what questions ISR could answer in this context. The goal of ISR for PK assays is to characterize assay performance in-study with real samples, determine unexpected matrix effects, unexpected metabolites in the patient population, and unexpected analyte instability. Many biomarker assay validation parameters already address the questions that ISR is meant to answer for PK assays. Biomarker assay validation (BAV) requires the use of samples with measurable analyte levels from a similar subject population to that for which the assay is intended to be used in drug development and therefore generally does not rely only on spiked samples. Specifically, un-spiked individual lots of matrix samples are used for parallelism and stability assessments, which are important experiments for biomarker assays and may address some of the same questions as ISR.

Parallelism is a critical experiment for BAV and, aside from defining a MRD to remove matrix related non-specific binding, it is used to confirm the suitability of surrogate matrix used for calibrators and similarity between the calibrator and the analyte; similarity between the calibrator and the endogenous analyte is critical. It can also identify unexpected specific binding partners [46]. In addition, un-spiked matrix QCs are often used to monitor long-term assay performance independent of the kit lot, calibrators, and spiked QCs.

The parallelism assessment is not always straightforward. However, since the parallelism assessment for biomarker assays is a key consideration for biomarkers, previous recommendations on the parallelism evaluation for BAV were discussed in the context of when parallelism would be expected to fail [25,27,45]. This is different from situations where parallelism fails for unexplained reasons, undermining confidence in the assay, and prompting possible method redesign prior to study sample analysis.

In cases where non-parallelism may be expected, it is important to understand the scientific rationale for such cases. Specific recommendations were issued for the scenarios when a parallelism test is expected to fail. For example, non-parallelism may be expected when the analyte is a pool of heterogeneous proteins, exists in multiple forms, is in a complex with other endogenous proteins, or when diluting bound proteins in free target/ligand biomarker assays might shift equilibria. Lack of parallelism can also occur in antibody assays due to the diversity of the antibodies in endogenous matrix samples between subjects and the difference between endogenous antibodies and calibrators [47,48] as well as from the impact of protein aggregates on the analyte (e.g., neurofilament heavy measurements [49], where parallelism could be improved by breaking up neurofilament aggregates).

It is challenging to understand the impact of lack of parallelism and matrix interference on assays used to analyze antibodies and proteins. When non-parallelism occurs, it may not negate the use of the assay depending on the COU and the understanding of the underlying cause for failure; limitations can be placed on assay applications (e.g., only generating qualitative or semi-quantitative results, only quantifying high-affinity antibodies using high dilutions). Dilution strategies can also be implemented to mitigate matrix interference if appropriate for the biomarker. Disease-relevant samples should be used as QCs and compared to other validation tests to inform confidence in the assay.

It is recommended that performing ISR for biomarkers should not be automatic and should be evaluated on a case-by-case basis to determine if it would add value. The FDA concurred that the biomarker section of FDA BMV [29] and the Biomarker Qualification Program gives flexibility for biomarker assays.

The research group of the Japan Agency for Medical Research and Development recently recommended that ISR be performed for each matrix with samples from representative clinical studies, such as those using biomarkers for an important evaluation to characterize drugs (e.g., using biomarkers as surrogate endpoints in late-stage clinical

studies) [50]. It should be noted, though, that access to relevant matrix samples during method validation may be more difficult in Japan and some other countries, thereby requiring increased use of incurred samples.

Two case studies were discussed demonstrating successful ISR for small molecule biomarkers. The first was for cholesterol and its metabolite (4β-hydroxycholesterol), measured in plasma using a fully validated LCMS assay; ISR was successfully conducted prior to the release of the 2018 BMV guidance for both analytes based on an internal PK ISR standard operating procedure (SOP) [51]. The other case study was for a free target assay investigation where ISR was performed on a small subset of samples to assess robustness as part of an investigation into samples with unexpectedly higher levels of free analyte at a specific time when total analyte was very high and drug levels were rapidly declining.

If endogenous QCs (eQCs) representing the disease of interest are available, they are likely representative of study samples. As a result, other validation tests may be considered more appropriate. If eQCs are not available, there may be certain circumstances when ISR is appropriate, such as when an analyte is undetectable in validation but expected to increase or decrease with treatment. Even in these cases, other approaches such as ISS or in-study sample parallelism may provide more relevant information for the biomarker than just traditional ISR. These analyses have the potential to improve biomarker assay robustness by increasing the utilization of trend analysis for biomarkers.

It was noted that ISR can also address undocumented technical errors; however, ISR acceptance criteria is broad enough that these may remain unidentified. Hence, it is recommended to address unexpected results in biomarker assays through specific root cause investigations.

It is recommended to evaluate eQCs in normal, disease, and on-treatment ranges. It may be considered appropriate not to define acceptance criteria beforehand, especially if studies are performed after validation.

Clinical Biomarkers as Surrogate Endpoints or for Patient Selection/Stratification

A surrogate endpoint is often defined as a biomarker intended to substitute for a clinical endpoint and can be imaging-based, molecular assay readouts, or chemical measurements, for example. Use of surrogate endpoints can reduce the size, complexity or length of a clinical trial, which is particularly needed in rare diseases due to the limited number of patients. The FDA has stated its hope that increased use of biomarkers as surrogate endpoints will facilitate more efficient drug development while also highlighting their risks and limitations [52]. The 21st Century Cures Act mandated that the FDA publish a list of surrogate endpoints which may be used for approvals [53].

Potentially, surrogate endpoints have great utility in rare disease trials and/or when clinical outcomes may take extended time to study. These biomarkers may measure a change early on in the course of a disease or measure changes that occur as the disease progresses [54]. However, the potential surrogate biomarker may not be well characterized in a rare disease population; therefore, the analytical requirements of these biomarkers may be poorly understood. Such requirements could include dynamic range requirements, potential interference and/or sensitivity requirements of the assay. The 2020 White Paper [27] recommended that for biomarkers serving as primary endpoints or surrogate markers, extensive validation of all parameters is necessary to support the context of use.

Surrogate clinical endpoints may be measured with a variety of technologies including LBA, flow cytometry, and quantitative or droplet digital polymerase chain reaction assays (qPCR/ddPCR). Novel hybrid LCMS (nanoflow-LC and high-resolution MS (HRMS) was demonstrated to support patient treatment decisions [55]. Identification of the best biomarker, analytically and clinically, is a prerequisite for successfully proposing a biomarker as a surrogate clinical endpoint. This may require significant investments in understanding the natural history of a disease. Early conversations with regulatory agencies regarding the proposed use of a biomarker as a surrogate endpoint are essential in understanding potential concerns and issues that must be addressed. Analytical validation of the assay is essential but not in itself sufficient for the use of a clinical surrogate endpoint. Surrogate biomarkers in rare or ultra-rare disease populations may require additional natural history or non-interventional studies in these disease populations, especially when the biological role of the analyte in those disease populations is not well understood, or samples are difficult to acquire.

Two case studies were used to discuss the challenges of developing CLIA MS-based methods. The first was a case study for leucine rich repeat kinase (LRRK2), a promising disease modifying target for Parkinson's disease [55]. Use of novel hybrid LCMS assays to support patient treatment decisions in global clinical trials (i.e., inclusion/exclusion or treatment/dose changes) requires extensive validation. There was difficulty placing novel MS-based biomarker methods in a CLIA-certified lab due to lack of technology, capabilities, and relevant expertise. Two separate labs

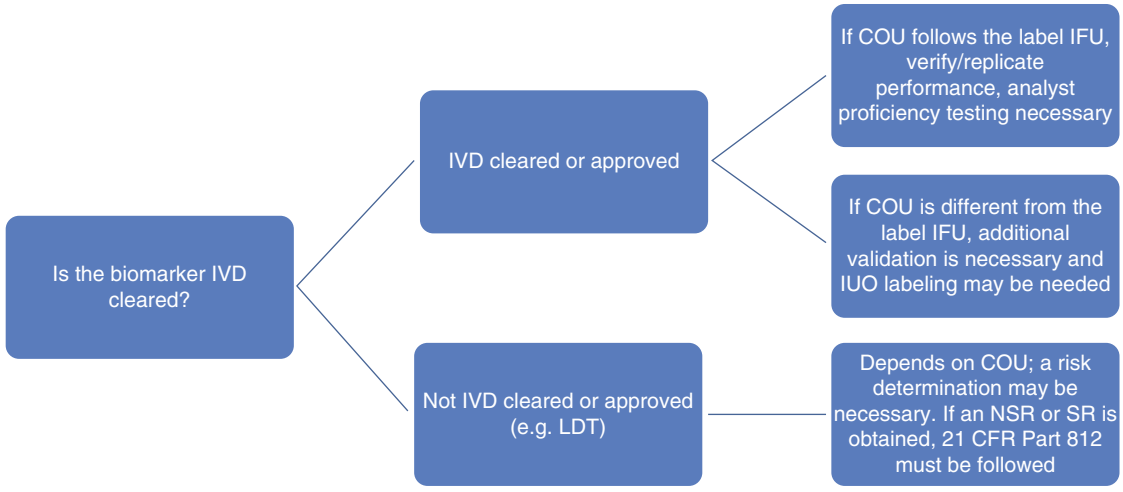

Figure 1. Recommendations for the use of IVD assays (cleared/approved or not cleared/approved).

with CLIA-accreditation could not achieve similar sensitivity. The solution was to establish partnerships with the CLIA-certified labs to grow their capabilities to support these technologies and associated novel methods.

Another case study was discussed regarding the bioanalytical strategy for a CLIA-accredited biomarker assay using MS. Pre-method validation considerations included instrument specificity, parallelism, signal to noise at the LOQ, system suitability, specificity, and the contamination analysis of multiple lots of sample storage tubes. Several lessons were learned from this case. Assay development time was significant, therefore analytical and financial resources should be budgeted accordingly. Large volumes of QCs should be prepared with aliquot volumes in tubes consistent with expected clinical volumes to ensure homogeneity and stability. Pre-validation mock experiments can be useful to detect challenges with the method. Judicious choice of multiple reaction monitoring (MRM) transition is necessary due to m/z implications on selectivity and specificity. Non-specific binding to tubes and plates should be evaluated. High purity reference and internal standards are essential, therefore it was recommended to secure chemistry, manufacturing and controls (CMC) resources for additional reference and internal standard synthesis and recertification.

The assay requirements and best practices for analytical validation for use of biomarkers as surrogate endpoints or for patient selection/stratification were discussed. It was recommended that appropriate bioanalytical validation must be in place before using the assay to clinically validate the biomarker as a surrogate endpoint or deploying the assay for selecting or stratifying patients for inclusion in a trial. If data is used to make a diagnosis, or prevent or treat an illness, in the USA, a CLIA environment must be used. However, because of variability in validation rigor amongst CLIA labs, validation packages may or may not be suitable for use to validate a biomarker as a surrogate endpoint. The assay sensitivity, specificity, precision, accuracy (if appropriate) should be evaluated at a minimum. In addition, the assay cut point should be established and locked down prior to use in the registrational trial. Further, validation should address changes in cut point, tissue type, or population (e.g., for an immunohistochemistry (IHC) assay, a validation in one tissue type or with a single cut point does not confer validation across cut points or disease indications). Regarding the use of IVD-cleared assays and LDTs for surrogate endpoints or patient selection/stratification, the validation package required depends on context of use; sometimes additional validation is required even for exploratory endpoints.

Figure 1 is an example of the USA recommended pathway. Each country may have specific regulatory requirements that must be met when the assay is intended to be used as a surrogate endpoint and/or for patient selection/stratification. In many cases, these requirements must be met prior to deployment onto the clinical trial. In the USA, it is recommended when an assay is to be deployed onto a clinical trial for patient selection/stratification, a risk determination be conducted to determine the requirements prior to initiating the trial. If the risk determination indicates a nonsignificant risk (NSR) or significant risk (SR), 21 CFR Part 812 must be followed. When running global trials, a thorough regulatory assessment of each country where the trial is being conducted is recommended.

Since analytical validation alone is not sufficient to demonstrate use of a biomarker as a surrogate endpoint, the best practices for clinical validation of biomarkers used as surrogate endpoints was discussed. Clinical validation

is a case-by-case decision based on what the endpoint is trying to replace (e.g., accelerated approval versus post-marketing versus a surrogate which can be used across trials, therapeutics as a surrogate endpoint, etc.). In other cases, randomized trials may also be necessary. The recommendation is to engage with regulatory agencies early to plan the appropriate validation for the specific context of use case and the subsequent validation package should be completed before use. A study risk determination may need to be submitted to US FDA CDRH or the institutional review board (IRB) to determine if an assay is required to meet 21 CFR 812 requirements prior to deployment onto the trial.

Quality Oversight of CLIA Laboratories for Companion Diagnostics

Companion diagnostics are IVDs that provide essential information for the safe and effective use of a corresponding therapeutic. In 2020, 39% of drug approvals were personalized medicines resulting in an increase of therapeutics including a companion diagnostic in their drug labeling [56]. The 2020 White Paper in Bioanalysis recommendation centered around biomarker assay requirements for patient selection/stratification for clinical trial purposes [27]. The next application of a CDx is in routine patient management. CDx test development, registration, and commercialization require significant investment and coordination for biotech and pharmaceutical sponsors as well as IVD test partners and is subject to extensive quality system regulation requirements. In the USA, drug labels often indicate "using an FDA approved test". An LDT does not meet the drug-labeling requirement for "an FDA-approved test" unless it has gone through a formal pre-market approval process. As such, usage of an LDT that does not have FDA pre-market approval to decide if a patient receives treatment would result in off-label usage of the drug. USA laboratories may use their own LDT version of an assay versus the FDA approved test to report results for biomarkers (e.g., EGFR and BRAF mutations as well as others, which may have an association with therapeutic usage). As a result, a potential discrepancy may be created in the health care setting in that a physician may use results from an LDT to prescribe a therapeutic. Key reasons for why an approved CDx may not always be used were the increased cost and reimbursement challenges along with the need for using a specific technology platform.

The differences between using an approved CDx versus an LDT for patient treatment decisions, both in the USA and globally, were discussed, as well as the impact on data quality. The goal was to create dialogue on testing barriers as well as future best practices to ensure patients are receiving the appropriate therapeutics for personalized medicine.

There may be inherent risks with not using the approved diagnostic for patient care. LDTs are not required to demonstrate concordance to the approved diagnostic; performance may be better, similar or worse for patient selection, resulting in the possibility that patients may receive the treatment they should not have or forego treatment they should have received. Clinical validity is not required for an LDT. Quality oversight of clinical laboratories, whether they are running an LDT or an approved CDx, may not be sufficient to ensure tests are being run per instructions or appropriate standards to ensure accurate results. Attention should be paid to the intended use/indications for use (IU/IFU) to ensure CDx is used as labeled. It is the responsibility of the IVD sponsor, the drug developer, the laboratory and the regulators to ensure accurate and reliable quality results for patient care.

IU/IFU should be carefully thought out by the IVD sponsor and reflect the intended patient population. Furthermore, validation should be conducted as appropriate to allow routine sample types representing the clinical trial population (e.g., formalin-fixed paraffin-embedded [FFPE] is a common sample type; however, fine needle aspiration [FNA] samples are often collected and submitted for testing). Pre-analytical variables such as slide cutting, temperature of slide baking, and shipping temperatures should also be carefully monitored to ensure an accurate result.

The merit of additional regulatory oversight of LDT/CLIA laboratories has been debated and in discussion for some time. Pending legislation may revise the framework for regulatory oversight of diagnostics in the US and many other countries. Regulators have expressed concerns about the increasing complexity of LDTs and the potential for misdiagnosis and inappropriate therapy use. LDTs may have less stringent regulatory requirements than IVDs, consequently, the LDT may have different performance characteristics. The Verifying Accurate Leading-edge IVCT Development (VALID) Act of 2021 is being evaluated to see if it is possible to address some of the discrepancies on testing oversight and usage [57]. It is currently structured to focus on the more complex, highest-risk diagnostics, where high-risk tests will still be subject to the same types of review requirements. There is a lengthy process for the VALID Act to become law, likely with revisions. Until then, adherence to existing regulations should be continued, in consultation with health authorities.

Of note the FDA Companion Diagnostic guidance document provides recommendations on study risk determination and framework for clinical trial assays [58]. In general, the performance characteristics of assays to be used as clinical trial assays should be adequately established. The assay cut point(s) should be established and locked down, and the assay should be appropriately validated prior to its use in the trial for patient enrollment. For example, for next-generation sequencing (NGS) assays to be used for patient selection for multiple biomarkers in a trial, the biomarker calling rules/variant classification rules should be pre-specified to distinguish between biomarker positivity versus negativity, as applicable; the assay cut point(s) and pre-analytical requirements (e.g., nucleic acid input, tumor purity, etc., for a tissue-based assay) should be established and locked, and the assay should be validated adequately for identification of the biomarker-defined patient population in the therapeutic clinical trial. The assay should not be altered after validation.

Emerging Trends & Impact on Diagnostic Development

A companion diagnostic assay is an *in vitro* diagnostic device that facilitates safe, effective use of a corresponding therapeutic product, and as such, requires a high level of analytical and clinical validation. Thus, it is critical to define the intended use of the biomarker and incorporate it into the trial design early on with a high-level validation which can easily transition to commercial assay (CDx). Alignment of CDx and clinical development followed by a global commercialization strategy is the key to success.

Pharmaceutical and diagnostic companies face significant challenges when commercializing their tests in a global environment often dominated by LDTs. Pharma needs to respond to these CDx commercialization challenges to minimize the failure of biomarker-eligible patients to receive appropriate targeted therapy. The 2020 White Paper in Bioanalysis provided a framework for biomarker analysis and CDx development for patient selection/stratification in registrational programs. FDA recommends use of Clinical Laboratory Standards Institute (CLSI) evaluation protocols to evaluate device performance. The clinical validity data requirement will depend on the IU/IFU of the assay [59]. Additionally, to address this challenge, pharmaceutical companies are improving and supporting the testing/CDx commercialization process through a range of approaches in order to ensure the right treatment.

The emerging trends in testing were discussed. Testing is becoming increasing complex with more deep, high-throughput sequencing for rare, emerging, and driver mutations, and to monitor for measurable/minimal residual disease. Non-invasive sampling types, such as liquid biopsies for non-invasive cancer screening and diagnostics, are developing rapidly. Faster and large-scale sequencing for infectious agent surveillance (emerging variants and epidemiology) is also an important use case that is increasing both in prevalence and importance. These trends have the potential to have one-stop screening, diagnosis and disease classifications versus traditional step-wise approaches. They also have the potential to enable effective non-invasive monitoring outcomes, potentially overcoming tumor heterogeneity issues.

To address these emerging trends, recommendations for diagnostic development were provided during the 15th WRIB. Lab directors are urged to review emerging changes to regulatory practice to assess compliance based on context of use and local regulations. Analytical expectations and processes for the addition of supplementary markers for laboratories and test developers should be streamlined. Several parameters should be assessed, such as reproducibility, limit of detection (LOD), linearity, cross-reactivity and matrix effects. Clinical performance should also be demonstrated, such that the CDx must have a clinical indication and must add value to clinical management. Clinical performance can be determined retrospectively only if the study supports the intended use of the test, samples were collected and stored appropriately in a manner that reflects current practice, and there is no sampling bias. Regulators support enabling new technologies and have provided clear expectations for assays with multiple analytical and clinical claims, therefore innovation can be facilitated. Entry costs are high and may become prohibitive therefore best practices should be carefully identified and analytical and manufacturing requirements reviewed with quality in mind. For example, all software and instrumentation used in CDx test systems must be fully validated as a total system, with demonstrable database integrity and cyber-security.

Every diagnostic development situation is unique and should be evaluated on a case-by-case basis. Regardless of the technology, it is important to clarify the intended use and select the right validation that best fits the clinical strategy. Early discussions with regulators are necessary. Centralized testing is preferred; however, if local testing is necessary, a risk assessment and mitigation plan prepared by a global diagnostic cross-functional team should be requested and the approach should be reviewed with the relevant health authorities. Finally, selecting and building a close relationship with the right partner that has the needed expertise to transition the assay into an investigator

use only kit and plan every stage of development including regulatory submission and marketing in alignment with drug development is the key to success.

Breath & Airway Biomarker Determination

Airway sampling can provide keen insights into biology. As this past year has demonstrated with SARS-Cov-2 infections (COVID-19), some biomarker levels can only be assessed through airway sampling. Many respiratory conditions cause an inflammatory response that is heterogeneous in nature that cannot be properly assessed by lung function assessments like spirometry alone. Surveying targeted tissue has a high relevance for understanding disease and drug mechanism. Airway sampling by methods such as induced sputum and bronchoalveolar lavage are not well tolerated, operationally demanding and difficult for bioanalysis and assay development due to high sample viscosity and the presence of proteases. Recently, novel airway sampling techniques have been introduced including breath, nasosorption and bronchosorption, which allow less invasive sampling of nasal lining fluid and bronchial lining fluid.

Breath bioanalysis has a promising future as a non-invasive matrix to analyze broad biomarker modalities including gas, volatile organic compounds, metabolites, proteins, electrolytes and microorganisms. Exhaled breath reflects lung pathophysiology and the metabolism of other organs. Measurement of biomarkers in exhaled breath has gained research interest in several diseases including lung cancer, and respiratory, liver and metabolic disorders. Breath research is yielding new insights into the development of standardized sampling approaches and detection technologies. However, only a few breath tests have been transitioned into clinical practice to date.

Even with easier sampling methods, there are many challenges in breath research; the breath community is exploring them from different angles as reflected by the increasing number of scientific publications [60–64]. Several factors should be considered in order to provide reproducible and reliable results. Robust study designs, validation across multicenter studies, in-depth knowledge of biomarker origin, and behavior and corresponding breath profile are all important factors. Standardized collection procedures depending on the type of analyte (protein, volatile organic compound, gas, etc.), require different sample collection and biomarker detection approaches. Standardized collection equipment and methodologies as well as control of potential environmental contamination are important factors for consistent sampling. Analytes present in low concentrations in the matrix very often requires highly sensitive analytical technologies. Selection of appropriate platforms (e.g., gas chromatography, LCMS, inductively coupled plasma-MS, high sensitivity ligand binding assays [LBA]) and sample processing is critical. Normalization is a key consideration, and each type of collection should be evaluated for if and how samples should be normalized including and not limited to normalizing by breath volume, breath particle amount and urea ratio. Analysis of metadata and their biological relevance is also important.

Designing validation to fit the detection method and technology is highly recommended for both targeted and untargeted biomarker analysis in breath and sorption studies. Method qualification/FFP validation considerations are similar to other biomarkers and include sensitivity, selectivity and specificity, intra- and inter-assay reproducibility, and carryover. As when working with other biological matrices, the study subject's physiological conditions, lifestyle, demography, and medication need to be evaluated and included in the data analysis to understand their influence on inter- and intra-subject variation. Thoughtful study design including discovery and confirmatory studies, with cross-site and study alignment, can enable proper comparison and compiling of results.

Consistent definitions and terminology across different collection types would be helpful for this developing field. Due to challenging collection protocols, small volumes, potential contamination, and often the need for high assay sensitivity, the accuracy of these methods is not well understood. However, useful qualitative questions could be answered with these methods.

High Sensitivity Platforms for Biomarkers

Immunoassays have come a long way since the earliest chromogenic ELISAs were developed decades ago. New technologies have greatly increased the sensitivity of immunoassays, often down to the fg/ml range. The small scale electrochemiluminescent platforms offered by MSD have been widely adopted across the industry. Newer platforms, including the Singulex Erenna, Quanterix Simoa, O-Link and ProteinSimple Ella, offer additional options, each with pros and cons to consider. For example, single molecule counting does offer better sensitivity but the improvement varies by analyte. Gains in sensitivity need to be balanced with increased time, higher cost, decreased throughput, and less precision. The "right" platform depends on the situation.

However, immunoassays are dependent not only on the platform, but the reagents used in each – in particular, the capture and detection antibodies, and the calibrator. These critical reagents often differ between commercial kit options, and these differences complicate the comparison of platforms. Different kits, intended to detect the same antigen, may differ in terms of interfering substances and ability to bind free or soluble receptor-bound forms. Platforms also differ in reproducibility, sensitivity, precision, and accuracy. Due to these differences, it is necessary for end-users to characterize reagents and/or kit performance for the context of use.

Since high sensitivity assays are targeted for low-level analytes, they are more susceptible to matrix interference and less tolerant of reagents with suboptimal quality. Significant optimization is needed to ensure assay performance and robustness, especially when utilizing in a multiplex format. So far, these newer multiplex platforms have mainly been used for exploratory biomarkers or target biomarkers for clinical support when sensitivity is the driver. Their use for CDx has not been documented.

Commercial kits sometimes introduce challenges with critical reagent lot changes. In the USA, kits labeled as IVD or reagents labeled as analyte specific reagents are subject to quality system regulations, and therefore are required to be validated to a higher standard than those labeled research use only (RUO), which are not bound by this quality standard. As a result, lot-to-lot differences may be more pronounced among RUO reagents and kits. Often, calibrators may be in a different matrix than sample buffers, artificially lowering the lower level of quantification (LLOQ). Challenges are further amplified with multiplexing where a single analyte can be problematic out of several in the kit.

Strategies to ensure quality reagents and the most appropriate calibrator for the best assay sensitivity were discussed. Reagent quality and choice of calibrators should be a part of the development strategy for high-sensitivity immunoassays. Consideration should be given to vendor qualification or direct involvement with vendor release, since kit lot batches can change without notice and different lot numbers for reagents may be derived from the same production batch. Users can attempt to contractually require early notification of lot changes. It is recommended to use calibrators and controls from different manufacturers to better identify assay variability due to lot changes and to procure large batches, if possible. In addition, the use of eQCs and trending analysis is necessary to monitor kit lot performance, matrix interference, and stability concerns.

For high-sensitivity immunoassays, validation may need to be fit for purpose, unique from other immunoassays. In addition, plate position bias is a common issue with these platforms. It was recommended to work with statisticians to apply randomization to samples and experiments.

Exploratory & Target Engagement Biomarker Assays

The discussions and recommendations on exploratory biomarker development and validation were inspired by a recent blog entitled "Call me a heretic" [65] regarding the increased use of the term "context of use" in biomarker method development and the confusion and challenges it has created. This blog has been considered by many experts in the field as an important reminder for the Global Bioanalytical Community of the basic method development strategies for any bioanalytical assay independent of whether it is a biomarker assay or not:

> "*. . . Whether you call it 'Context of Use' or its predecessor 'Fit for Purpose', a bioanalytical assay must be designed to answer specific questions within a biological system based on the intended use of the data; be it a biomarker, a replacement drug for an endogenous molecule or a new drug entity. These are not new considerations. We've had to consider these issues for a long time. . . The challenges today are similar to those of the past in developing a biomarker assay. The tools we use may have changed, but we are still focused on measuring an endogenous molecule to understand a disease and, in some cases, measure a change in response to a drug therapy. The quality of the assay, with well-characterized limitations, informs those who use the data of the reliability of the decisions that can be made. This hasn't changed, we're just calling it something different.*" [65]

The utility and importance of biomarkers and biomarker data to help accelerate drug development and inform on clinical decisions has created excitement throughout the industry. But despite all the excitement, there is still a very long process for identifying and validating biomarkers. The recent discussions and confusion around the term COU to develop and validate biomarkers versus just using the traditional back-to-basics, science-based method development strategies seems to have complicated this process. This is particularly problematic given the high number of exploratory biomarkers developed in the industry, as opposed to the far fewer biomarkers used for regulatory submissions. Therefore, it is increasingly important to have a unified approach for implementing a successful exploratory biomarker strategy during drug development. A focus back to bioanalytical method

development strategies is necessary along with guidance on how to properly define and develop biomarkers for internal decision making based on the "*what, why, where, who*, and *how*" method development basic questions that any good scientists should always ask before starting development and validation of any bioanalytical assay.

The 2018 White Paper in Bioanalysis recommended distinguishing between the predictive biomarkers that are eventually targeted for decision making at the patient level and the biomarkers used for internal decision making by drug developers [21]. However, it is crucial that the biomarker community move away from applying the BMV guidance documents [29,43,66] to BAV in a prescriptive or overly rigid fashion. Biomarkers of interest that are not intended for regulatory decision-making should use FFP BAV [67,68]. There is a current lack of industry focus on bioanalytical method development strategies for biomarker assays and a lack of guidance on how to approach biomarker assays for internal decision-making [69,70]. There is also a lack of comfort to decide what constitutes FFP. The 2021 discussions focused on using a back-to-basics approach for method development strategies with regards to exploratory biomarkers beyond the concept of COU.

It is crucial to understand the biology of the biomarker, as well as what limitations have been identified from the methods and platforms being used for measurement to help assess if the assay is able to provide meaningful data based on the expected change in responses. In order for any biomarker data to be meaningful, it is important to understand the expected changes (increase, decrease) and magnitude of change (twofold, fourfold, etc.). The limitations of the methodology being used as well as the technology platform are both critical to ensuring confidence in internal decision-making when assessing exploratory biomarkers. It is also crucial to anticipate pre-analytical concerns for sample collection and sample handling (stability, matrix type). Collection of samples for biomarker analysis is not often the main priority in clinical trial study planning; therefore, samples for biomarkers may not be drawn at optimal time points. Too often, there are requests to measure multiple different biomarkers with only a single available sample aliquot and likely stability is unknown.

To establish a successful exploratory biomarker strategy, it is recommended to have a standard approach and common vocabulary within the company. A clear biomarker strategy should be established for each study outlining which biomarkers are of interest, their priority and the defined expectations. The use of the data generated should be defined in the context of the clinical trial and inform the appropriate assay platform, kit type, assay panel and level of validation required; any limitations of the assay should be known. Finally, life cycle maintenance of all assays is critical throughout the drug development process; the methods should be refined and improved when possible.

Sequential tissue biopsy collections are increasingly integrated into early phase clinical trial design to more wholly understand the PD effect of novel therapies at the site of pathology and to offset costly attrition in late-stage development. However, their impact to enable decisions on dose selection and/or to yield veritable proof of mechanism remains largely unarticulated. The 2014–2017 White Papers in Bioanalysis indicated quantitative tissue bioanalysis can provide critical information on target validation, target coverage for dose projections, occupancy, and precision medicine for patient stratification [26]. Yet, tissue procurement and handling, analyte recovery, and normalization were highlighted as scientific challenges compared to serum or plasma.

The science has evolved since 2017, with current and emerging technologies necessary to accommodate tissue sampling for the assessment of target engagement. The performance of immunoassays, LCMS, flow cytometry, single cell RNAseq, IHC, and genomics for tissue analysis have significantly improved. These advancements can enable first in human dose level predictions directly in tissue without relying on blood target expression. Site of action tissue-based PD assessments are proposed to enhance or increase confidence in program decision making using a multi-phase workflow from sample acquisition to tissue dissociation and analysis with optimization and quality control recommendations at each stage. For example, flow cytometry assays for tissue target engagement can involve developing a cytometry panel in fresh tissue followed by single cell cytometry protocols.

Discussions highlighted challenging aspects of performing tissue bioanalysis and the critical steps for method development. The challenges depend on tissue type (e.g., non-oncology versus tumor tissue). Biological variability associated with tissue heterogeneity is possibly the most challenging issue to be addressed for tissue bioanalysis. IHC approaches are often the first choices for most tissue protein biomarker analyses to obtain both signal intensity and spatial expression information. However, IHC results from non-FDA-approved methods are semi-quantitative and sometimes the reagents may not be available. LBA and/or LCMS approaches have been used to analyze tissue homogenates in the cases when intended biological or PD responses were higher than biological variability caused by tissue heterogeneity. Normalization of measured values with the selected marker(s) can mitigate issues caused by heterogeneity to some extent. The data have been used as supportive evidence of PD activities, relative target levels, and drug distribution at the intended tissues. Tissue analysis data has been used to support the feasibility of

surrogate markers in body fluids for the intended purposes. There may be a growing need for gene therapy-based therapeutics too. Although LCMS- or LBA-based tissue bioanalysis are tools to address specific questions, they have not been broadly adopted due to challenging method development, data variability, cost and availability of tissue biopsies outside of oncology.

Pre-analytical factors, tissue stability, and sensitivity are important drivers of development and validation. Sampling type (core, excision, aspirate) should inform assay considerations and the effect of fixation (fresh, frozen, FFPE) on analyte and stability should be characterized. Tissue preservation, enzymatic digestion, and post-processing are also key considerations [71]. Tissue heterogeneity needs to be considered when assessing the analytical performance of the assay. Homogenization of tissue and contamination from blood may affect analyte response and result in incomparability to the original tissue. Autopsy samples are commonly used for validation but analyte stability in autopsy tissue compared to patient samples should be understood.

Target engagement data is used for decision-making, PK/PD modeling for efficacy and dose prediction, target validation and mechanism of action. The three types of TE assays for body fluid samples are free (unbound active form of target), total (free and bound fraction of target), and complex assays. Dynamic equilibrium is a common challenge in TE assays (particularly free assays) as the assay process may interrupt complex formation which is dependent on free drug/target concentrations [16,26,27].

Numerous choices for platform selection, advances in dynamic range, sensitivity, automation and commercial immunoassay kits make it difficult to know where to begin. The rise of high sensitivity immunoassays has enabled more free TE assays by keeping dilution to a minimum. Add to this the choice of reagent, buffer and assay conditions, and development can be a daunting task. It is therefore essential to ensure the target levels, biology and use of data drive the analytical decisions.

Target engagement of a candidate drug may require independently developed assays to generate data for the free, total and complex fractions of the target. Each of these assays presents their own set of challenges (e.g., sensitivity for free target assays and reagent availability for total assays) [72]. Consideration for each assay regarding the anticipated analytical range, knowledge of target (presence of binding partners, different forms/multimerization), the biology of the disease, expected levels of target in pre-dose and dosed individuals, matrix type, sample volume and sourcing of reagents will have an impact on the decision-making process for platform selection, what parameters need to be assessed and the final assay conditions. Agreement was reached that it is not common practice to offer three types of assays (free, total, and complexes) for a TE program. Typically, only 1–2 forms are measured depending on many factors including the target biology, target-drug affinity, whether the soluble form is the intended target or just a surrogate target, and assay feasibility. A decision is usually made from discussions between representatives from various functional teams (biomarkers/biology, modeler, and bioanalysis).

Measurement of the free target requires a full understanding of binding kinetics to determine whether this is at all feasible. Modeling and simulation can be used in combination with experimental data to do this and assess sample dilution, the impact of time and reagent binding, etc. Rapid separation of a bound and free target is preferred when there are concerns about complex dissociation. As overestimation of free target is inevitable, it is important to define how much overestimation and analytical variability can be tolerated without impacting the study decisions. If levels of free target overestimation are similar between preclinical and clinical or between disease populations, some levels of overestimation may not affect dose projection that much. It is important to meet the precision criteria at the MRD especially for the samples with biomarker levels close to LOQ. The optimization of buffer, capture reagents, and other conditions may help.

Parallelism must be done without the drug present to evaluate the use of surrogate matrix for calibrators, confirm the MRD and the assay buffer used for dilution to mitigate matrix effects and nonspecific binding and demonstrate the similarity between the calibrator and the analyte.

PBMC Sample Collection for Pharmacodynamic Biomarkers

PBMCs comprise the cell population of interest for the majority of immune studies and certain other biomarker assays. After separation by density gradient centrifugation, the PBMC fraction should contain only the lymphocytes (T, B and NK cells) and monocytes; granulocytes and non-nucleated cells are excluded. The use of PBMCs is usually driven largely by logistics and sample stability rather than the superior quality of PBMCs relative to native specimens due to limitations caused by cell loss, antigen loss, etc. Alternatives to PBMC assays include whole blood methods such as flow cytometric methods or induced cytokine studies with cryopreserved culture supernatants. Significant

effort must be made to demonstrate that any biomarkers identified in PBMCs retain their predictive/prognostic characteristics when used on fresh samples, if that is the final goal.

Early phase clinical trials often have an exhaustive biomarker strategy in place, and the isolation of PBMCs is becoming ever more common in Phase I and II trials. PBMC samples can be used for a multitude of biomarker assays, from genetics and genomics to cellular-based assays, being therefore an excellent matrix to support such complex biomarker strategies. The advantage of PBMC versus fresh blood for flow cytometry-based assays is that they can be stored for several years, allowing prospective and retrospective analysis based on the clinical trial outcomes and new questions that may arise.

Nonetheless, the collection of PBMC samples in the context of clinical trials poses substantial operational and logistical challenges, as they require expert handling skills for isolation, specialized laboratory equipment and fast turnaround times for cryopreservation; these are often not compatible with clinical site set-up. There are many preanalytical considerations for PBMC isolation that must be characterized and prioritized during method development. The need to control PBMC isolation variables should be informed by the COU. More specifically, the analytes and assay type will determine what controls need to be placed on PBMC isolation, shipment, and storage; the solutions should be determined by the sponsor. Considerations to take into account include isolation method, shipping conditions, stabilizers, buffers, resuspension volume, fresh versus frozen versus cryopreservation, time from collection to processing (sample aging), and time and conditions of long-term storage. Sample disease status can also affect PBMC isolation. Cancer patients showed increased contamination of granulocytes in PBMC preparations, a decrease in CD3ζ (TCRζ) expression and decreased cytokine production (IFN-γ, IL-4, IL-2, TNF-α) [73].

PBMC isolation procedures in clinical trials tend to be set up in central laboratories of specialized CROs. This option guarantees adequate operator skills and laboratory equipment but often fails to deliver the necessary turnaround times for optimal sample conditions, especially for multicenter clinical trials. Central labs can be uncooperative in terms of processing time and introduce complex courier and shipping challenges. Ambient shipping sample temperature is heavily influenced by seasonal temperature changes; exposure to extreme high or low temperatures negatively affects PBMC viability, yield and immune function [74].

Sample aging during shipment of blood to a central lab for PBMC isolation should also be considered since this leads to granulocyte activation and density changes that result in contamination of PBMC preparations. Granulocyte contamination leads to an overestimation of PBMC seeding densities that can affect biomarker assay readouts, interactions with detection reagents that can lead to the decrease of signal intensity, and downregulation of CD3ζ chain that impairs T cell proliferation capacity [75]. Some downstream assays for PBMCs are particularly sensitive to sample aging (e.g., analyte stability in fresh matrix is low) and thus the impact of time should be tested in a dependent manner.

Another option is to set up PBMC isolation directly at the clinical sites. Samples isolated on-site can be readily cryopreserved, with the potential to maximize sample viability, yield and functionality. Logistically this is usually a much more complex operation that will demand increased resources on the sponsor, central lab and investigator and can introduce higher variability. Proficiency testing, training protocols with videos, and harmonization of reagents have proven useful. Site feasibility can be performed with the following criteria evaluated for site selection: existing laboratory equipment, staff proficiency in PBMC isolation, aseptic work in BSL2 laminar flow hoods, number of available operators for training, sample storage capabilities (liquid N_2, -80°C freezers), willingness to participate in site operator qualification processes, and local restrictions on reagent imports. Other factors to consider are that some countries require specific protocols and approvals for sample collection for testing and training purposes.

Other considerations for PBMC isolation include the appropriate positive/quality controls for a PBMC-based PD biomarker. Such controls could serve multiple purposes, including ensuring acceptable recovery following freeze-thaw where viability could be an issue. Standards for successful PBMC isolation (e.g., minimum standards for viability) may be useful as well. Finally, more matrix stabilization solutions compatible with downstream functional assays are needed in order to facilitate PBMC sample collection. These questions require further experience and future discussion.

RECOMMENDATIONS

Below is a summary of the recommendations made during the 15th WRIB:

Liquid Biopsy: Challenges & Opportunities with Extracellular Vesicles

- More case studies are needed to demonstrate the clinical applications of EVs, especially as clinical endpoints.

- Isolation, extraction and pre-analytical methods for EVs need to be tightly controlled to minimize heterogeneity and obtain different EV populations across samples:
  - Simple polymer-based precipitation followed by a protein depletion column
  - Size-exclusion chromatography
  - Ultracentrifugation
- Key technologies used to detect and quantify EVs:
  - MRPS
  - Microscopy
  - Scatter-based nanoparticle tracking analysis
  - Western blot or ELISA-based characterization
  - Multifluorescence flow cytometry
  - Global proteomic analysis using mass spectrometry
  - Fluorescence-based nanoparticle tracking analysis
- More detailed definitions and descriptions of isolation procedures in published methodologies are needed to compare approaches and provide robust recommendations for specific EV populations.
- Pre-analytical variables such as sample volume, buffers used, freeze-thaw stability, time of collection, and serum versus plasma are factors that could affect the total and tissue-specific particles isolated and should be well controlled and defined.
- The isolation method chosen should be informed by the scientific question and downstream application.
- For quantitative methods, some options for calibrators and reagents include EVs prepared from cell cultures, sample pools from those with high levels of the intended EVs, and recombinant proteins. Negative controls should also be included.
- Quantification of biomarkers on organ-specific EVs has to consider two parameters: biomarker expression on EVs and the overall number of EVs in order to accurately evaluate differences in biomarker expression between samples. Normalization is also needed because of variations between isolations.
- Assay parallelism and spike recovery should be demonstrated with immunoassay approaches.
- MPs belong to the family of EVs, their size range from 100 to 1000 nm. MPs are important mediators of physiological and pathological cellular processes and can be detected in different bodily fluids (e.g., blood, urine, CSF) in patients and in healthy individuals.
- The most widespread approach for MP enumeration and characterization is flow cytometry. The primary limitation in analyzing MPs is their submicron size. Using the next generation of highly sensitive instruments can bring potentially unprecedented resolution and the needed sensitivity to expand to the nanoparticle range and thus allow for more precise measurements.
- Other biological considerations need to be evaluated such as: correlation between the marker levels in the EVs and the tissue biopsy, factors that influence intra-donor variability, factors that influences the release of exosome, expression of specific markers in specific tissues, correlation with outcomes and disease versus healthy conditions.

ISR for Biomarker Assays, Parallelism & Biomarker Assay Validation Guidance

- eQCs, ISS, parallelism, and incurred sample parallelism may preclude the need for ISR of biomarkers as in many cases, they may provide more relevant information for biomarker assays than ISR alone.
- Performing ISR for biomarkers should not be automatic; and should be evaluated on a case-by-case basis with a biomarker specific approach and rationale.
- eQCs may be obtained from normal or disease samples. Value assignment and acceptance criteria require a sufficiently sized data set, especially for new eQCs added after validation.
- It is important to understand the scientific rationale for cases where non-parallelism may be expected.
- Non-parallelism may be expected
  - When the calibrator is not recognized the same way as the endogenous analyte
    - In antibody assays due to the diversity of the antibodies in endogenous matrix samples between subjects and the differences between endogenous antibodies and the calibrator
    - From the impact of aggregates on the measurement
  - When there are endogenous specific binding partners
    - When diluting bound proteins in free target/ligand biomarker assays due to shifting equilibria.

  - When matrix effects have not been adequately managed or when an imperfect surrogate matrix is used
- When non-parallelism occurs, provided the source and rationale for the non-parallelism is understood, it may not negate use of the assay depending on the COU.

Clinical Biomarkers as Surrogate Endpoints or for Patient Selection/Stratification

- It is critical to understand the behavior of a surrogate biomarker in the actual population prior to using it in a pivotal clinical trial.
- Identification of the best biomarker, analytically and clinically, is a prerequisite for successfully proposing a biomarker as a surrogate clinical endpoint.
- Early conversations with regulatory agencies around the proposed use of a biomarker as a surrogate endpoint are essential for understanding potential concerns and issues that must be addressed.
- For CLIA MS-based biomarker methods:
  - Large volumes of QCs are needed with aliquot volumes in tubes consistent with expected clinical volumes;
  - High-purity reference and internal standards are essential; secure CMC resources for additional reference and internal standard synthesis and recertification;
  - Pre-validation mock experiments can be helpful;
  - Judicious choice of MRM transition is necessary due to m/z implications on selectivity and specificity;
  - Nonspecific binding to tubes and plates should be evaluated.
- Appropriate bioanalytical validation must be in place before using the assay to clinically validate the biomarker as a surrogate endpoint. If data is used to make a diagnosis, treat or prevent illness, in the USA, a CLIA environment must be used.
- Suggested requirements for analytical validation for use of biomarkers as surrogate endpoints:
  - Sensitivity, specificity, precision, stability and accuracy (if appropriate) should be evaluated;
  - For IHC, validation will need to be repeated if there are changes in cut point, tissue type, or population. This is applicable to other technologies as well. The cutoff for a specific biomarker/analyte may not be the same across different tissue types.
- Clinical validation:
  - Requirements should be assessed by COU and case-by-case;
  - Engage with regulatory agencies early to plan for the specific context of use case and required validation package;
  - The validation package should be in place before use;
  - A study risk determination may need to be submitted to US FDA CDRH or the IRB to determine if an assay is required to meet 21 CFR 812 requirements prior to deployment onto the trial if assay is being used for prospective patient selection/stratification.

Quality Oversight of CLIA Laboratory for Companion Diagnostics

- Indications for use of the CDx should be carefully thought out by the IVD sponsor and reflect the intended patient population.
- Analytical validation should be conducted as appropriate using intended use clinical specimens as far as possible, i.e., specimens that represent the clinical trial population or the CDx intended use population.
- Pre-analytical variables such as specimen collection, specimen processing and shipping temperatures should be carefully monitored to ensure a quality result.
- Quality review of CLIA labs and the extensive LDT testing menu will take resources, time, and implementation processes. Pending legislation may revise the framework for regulatory oversight of diagnostics in the US and many other countries. Until then, adherence to existing regulations should be continued, in consultation with health authorities.

Emerging Trends & Impact on Diagnostic Development

- When transforming a clinical assay into a CDx assay, define the intended use of the assay and define the target biomarker (including the biomarker definition, and biomarker classification and interpretation rules, as needed) and have a development plan in place in alignment with clinical development and global commercialization prior to starting a registrational drug trial and prior to finalizing the diagnostic sponsor

and the assay to be developed as the final market-ready assay for identifying patients who may benefit from an investigational drug.

- A companion diagnostic is a medical device, often an *in vitro* device, which provides information that is essential for the safe and effective use of a corresponding drug or biological product.
- The CDx must show a drug is safe and effective in certain patients and be analytically and clinically validated.
- CDx development:
  - Lab directors are urged to review emerging regulatory changes compared against the current practice to assess compliance;
  - Analytical and clinical validation to support a CDx assay should be robust. Streamlined approaches may be developed for adding supplementary markers to the device IU/IFU, and be discussed with the regulatory agency;
  - FFP manufacturing requirements, IVDR requirements and international harmonization should be considered;
  - Clarify the intended use and select the right validation strategy that best fits the clinical claims;
  - Early discussions with regulators are necessary;
  - If local testing is used, request a risk assessment and mitigation plan prepared by a global diagnostic cross-functional team;
  - Evaluate and select the right partner capable to develop, file and commercialize a successful IVD product.

Breath & Airway Biomarker Determination

- Novel airway sampling techniques have recently been introduced, such as breath, nasosorption and bronchosorption, which allow non-invasive sampling of nasal lining fluid and bronchial lining fluid.
- Breath bioanalysis has a promising future as a non-invasive matrix to analyze broad biomarker modalities including gas, volatile organic compounds, metabolites, proteins, electrolytes and microorganisms.
- Breath bioanalysis requires robust study design and validation across multicenter studies, in-depth knowledge of biomarker origin, behavior and corresponding breath profile, standardized collection procedures, optimal sample processing and analysis, highly sensitive and analyte relevant technology, and analysis of metadata and their biological relevance.
- Breath sampling considerations:
  - different sample collection and biomarker detection approaches are required depending on the type of analyte;
  - standardize collection equipment and methodologies;
  - control potential environmental contamination (clean materials and working environments, pure chemicals and reagents);
  - selection of appropriate platforms and sample processing is critical due to low amounts of analytes in the matrix;
  - each type of collection should be evaluated to determine whether samples should be normalized.
- Appropriate validation levels for the detection method and technology performance are highly recommended for both targeted and untargeted biomarker analysis.
- Method qualification/FFP validation considerations are similar to other biomarkers:
  - Sensitivity;
  - Selectivity and specificity;
  - Intra- and inter-assay reproducibility;
  - Quality control/calibration;
  - Carryover.
- The subject's physiological conditions, lifestyle, demography and medication need to be evaluated and included in the data analysis to understand their influence on inter- and intra-subject variation.
- Consistent definitions/terminology across different collection types would be helpful for this developing field.

High Sensitivity Platforms for Biomarkers

- Characterize kit performance and applicability for the context of use.

- Since high sensitivity assays are targeted for low level analytes, they are more susceptible to matrix interference and less tolerant of reagents with suboptimal quality. Significant optimization is needed to ensure assay performance and robustness.
- Reagent quality and choice of calibrators should be a part of the development strategy for high-sensitivity immunoassays.
- Consider vendor qualification or be involved in vendor's lot release process to better monitor the impact of kit and reagent lot changes.
- Use of external calibrators and controls from a different manufacturer and procure large batches is recommended, where possible to control for kit lot variability.
- The use of eQCs and trending analysis is necessary to monitor kit lot performance, matrix interference, and stability concerns.
- In some instances, validation may need to be unique from other immunoassays with relaxed acceptance criteria.
- Work with statisticians to apply randomization to samples and experiments due to plate position bias.

Exploratory & Target Engagement Biomarker Assays

- A focus back to bioanalytical method development strategies is necessary to avoid confusion created by COU.
  - Scientists should always ask themselves the "*what, why, where, who* and *how*" method development basic questions before starting the development and validation of any bioanalytical assay beyond the concept of COU.
  - The limitations of the methodology being used as well as the technology platform are both critical to ensuring confidence in internal decision-making when assessing exploratory biomarkers.
  - It is also crucial to anticipate concerns for sample collection and sample handling (stability, matrix type).
  - It is recommended to have a standard approach and common vocabulary within the company.
- A clear biomarker strategy should be established for each study outlining which biomarkers are of interest, their priority and the defined expectations.
- The intended use of the data generated should be defined in the context of the clinical trial and inform the appropriate assay platform, kit type, assay panel and level of validation required.
- Life cycle maintenance of all assays is critical throughout the drug development process; the methods should be refined and improved when possible.
- Biological variability associated with tissue heterogeneity is possibly the most challenging issue to be addressed for tissue bioanalysis.
- Tissue stability, pre-analytics, and sensitivity are important drivers of development and validation.
- Sampling type (core, excision, aspirate) should inform assay considerations and the effect of fixation (fresh, frozen, FFPE) on analyte and stability should be characterized. Tissue preservation, enzymatic digestion, and post-processing are also key considerations.
- Autopsy samples are commonly used for validation but analyte stability in autopsy tissue compared to patient samples should be understood.
- Target engagement of a candidate drug may require independently developed assays to generate data for the free, total and complex fractions of the target.
- It is not common practice to offer three types of assays (free, total, and complexes) for a TE program. For TE support, it is recommended that priority is given to development of total assays, then complex and free. Total assays are less susceptible to overestimation of target when complex dissociation occurs.
- Measurement of the free target requires a full understanding of binding kinetics to determine feasibility using modeling and simulation in combination with experimental data.
- Rapid separation of the bound and free target is preferred when there are concerns about complex dissociation.
- It is important to define how much overestimation of free target and analytical variability can be tolerated without impacting the study decisions.
- Meeting the precision criteria at the MRD is important especially for the samples with biomarker levels close to LLOQ.
- Parallelism must be done without the drug present to evaluate the use of surrogate matrix for calibrators, confirm the MRD and the assay buffer used for dilution to mitigate matrix effects and nonspecific binding and demonstrate the similarity between the calibrator and the analyte.

PBMC Sample Collection for Pharmacodynamic Biomarkers

- The most critical consideration is whether the PBMCs reflect the *in vivo* state
- The analyte and assay type will determine what controls need to be placed on PBMC isolation, shipment, and storage and the solutions should be determined by the sponsor.
- PBMC sampling considerations to address:
  - isolation method.
  - shipping conditions.
  - stabilizers and buffers.
  - resuspension volume.
  - fresh versus frozen versus cryopreservation.
  - time from collection to processing sampling.
  - sample disease status.
- Sample aging during shipment of blood to a central lab for PBMC isolation should be considered since this leads to granulocyte activation and density changes that result in contamination of PBMC preparations.
- PBMC isolation can be performed using a central lab or directly at the clinical sites. Both options have their pros and cons.
- If performing isolation at the clinical site:
  - demands increased resources on the sponsor, central lab and investigators.
  - can introduce higher variability.
  - proficiency testing and training protocols with videos have proven useful.
  - regulatory aspects and local restrictions need to be evaluated and can cause significant delays in the study.
- Site feasibility can be performed with the following criteria evaluated for site selection: existing laboratory equipment, staff proficiency in PBMC isolation, aseptic work in BSL2 laminar flow hoods, number of available operators for training, sample storage capabilities (liquid N_2, -80°C freezers), willingness to participate in site operator qualification processes, and local restrictions on reagent imports.

SECTION 2 – Cytometry Validation & Innovation

Sue Spitz[22], Steve Keller[2], Vilma Decman[23], Jose Duarte[9], Steven Eck[24], Jose Estevam[25], Fabio Garofolo[10], Polina Goihberg[41], Enrique Gómez Alcaide[9], Christèle Gonneau[26], Michael Nathan Hedrick[1], Gregory Hopkins[27], Fabian Junker[9], Sumit Kar[12], Sandra Nuti[28], Ulrike Sommer[19], Nathan Standifer[29], Chad Stevens[13], Erin Stevens[30], Priscila Teixeira[9] and Alessandra Vitaliti[19]

Authors are presented in alphabetical order of their last name, with the exception of the first 3 authors who were session chairs, working dinner facilitators and/or major contributors.

Author affiliations can be found at the beginning of the article.

HOT TOPICS & CONSOLIDATED QUESTIONS COLLECTED FROM THE GLOBAL BIOANALYTICAL COMMUNITY

The topics detailed below were considered as the most relevant "hot topics" based on feedback collected from the 14th WRIB attendees. They were reviewed and consolidated by globally recognized opinion leaders before being submitted for discussion during the 15th WRIB. The background on each issue, discussions, consensus and conclusions are in the next section and a summary of the key recommendations is provided in the final section of this manuscript.

Recent Developments in Flow Cytometry Validation in a Bioanalytical Lab

How is the validation of unplanned analytes handled in flow cytometry? How is stability in patient samples and tumor samples assessed? How is assay validation handled over the duration of a trial? How is a PD endpoint approached when it is not clearly intended as a biomarker/CDx but when study entrance or release criteria may be desirable? Must such an assay be CLIA or equivalently compliant?

Evaluation of Accuracy for Flow Cytometry in Regulated Laboratories

What progress has been made in developing standards for flow cytometry? When there is no relevant reference material, what are the alternative approaches to assess accuracy? What are the challenges associated with these alternative approaches?

Sensitivity Determination in Flow Cytometry Validation

Can isotype controls be used to create LLOQ validation samples? When the FMO (fluorescence minus one) may not be the right choice for some assays, how is the validation of rare events/assay sensitivity dealt with in high dimensional flow cytometry? How are the data reported when they are below the sensitivity acceptance criterion? Is this based on COU or is a strict rule needed? For example, if results are above LLOQ but the minimal number of required events (MRE) is not reached, should the results be flagged or not reported? Conversely, if MRE is reached but %positive cells are below the LLOQ, should the results be flagged or not reported? Is there consensus for experimentally assessing the MRE in the gate of interest (GOI) as one integral part of defining the assay sensitivity? When is the LLOQ assessment needed for validation? For example, is the LLOQ assessment necessary when 'overexpression' of an analyte is expected in specific patient populations or as a result of the drug MOA?

Clinical Biomarker Development, Validation & Interpretation by Cytometry

The 2020 White Paper in Bioanalysis recommended the need for algorithm-based analysis platforms to be evaluated for the likelihood of gating an incorrect population. What has been the experience with clinical data analysis requiring cloud-based data storage? What criteria should be met so that dimensionality reduction and algorithm-based unsupervised or supervised gating can be incorporated into the clinical data analysis work stream? What is required to ensure data security to enable use of these emerging platforms? Supervised algorithm-aided analysis for clinical samples presents both operational advantages and reduced subjectivity associated with manual gating. How can algorithms be 'educated' to analyze clinical samples that have atypical profiles? How does the algorithm manage cases that are unusual? How does one select in-study validation samples that are reflective of the patient population for the trial, for example, when early clinical trial patient samples may not have the 'expected' profile of a sample that has received multiple doses of the drug? Is there a need for both manual and unsupervised data analysis? Would novel populations discovered in unsupervised analyses be used on an exploratory basis only?

Spectral Cytometry Ultra-High Order/Dimensional Assays in Clinical Application

How is spectral cytometry used for drug development? At which level of the clinical trial objectives (secondary or exploratory) or which phase of clinical trials (dose escalation vs expansion phase) does it add additional value? Will ultra-high parameter flow cytometry assays ever be fit for more than exploratory endpoints? What are the technical considerations for clinical trials? How do we choose the right technology to address assay development challenges (e.g., fixatives, fluorochromes, etc.)? What are the strategies for the validation of spectral cytometry assays? What are the QCs and in-study quality considerations for use of spectral cytometry? What controls should be used? Are criteria being set for the number of events? What sample handling considerations are needed for spectral cytometry? How is data analysis handled for spectral cytometry?

Imaging Cytometry Quantitative Analysis of Target Engagement

How is imaging cytometry being used for drug development? What are the technological and sample handling considerations? What are the strategies for validation and method transfer of imaging cytometry assays?

Mass Cytometry in Clinical Biomarkers & "Clinical Trial Compatible"

How is mass cytometry being used for drug development and at which stages? What are the technological and sample handling considerations? What are the strategies for validation and method transfer of mass cytometry assays? What are the necessary controls for longitudinal studies where batch effects can be introduced? Is mass cytometry (being used for PD biomarker purposes in the clinic and for what objectives (exploratory?) How does one stabilize sample as well as panels for large studies in order to decrease batch effects and allow for method transfer and run-to-run consistency?

Multivariate Analytical Techniques & Multiparameter Flow Cytometry

The advent of technologies enabling monitoring of greater than 25 parameters/cell has resulted in the ability to monitor hundreds of cell populations in a single clinical sample. What is the path forward for the validation of such high-dimensional flow-based assays? How is data analysis handled for multiparameter flow cytometry? What considerations need to be taken for the security of patient information and metadata? How can consistency and run-to-run variability be controlled for studies?

DISCUSSIONS, CONSENSUS & CONCLUSIONS

Recent Developments in Flow Cytometry Validation in a Bioanalytical Lab

Previous White Paper recommendations [21,25,27] were specifically designed as a practical guide for bioanalytical laboratories to complement and expand more general flow cytometry guidelines such as the CLSI H62 [76] and the British Pharmacopoeia [77]. The latter recommendations use a tiered approach to provide minimal standard examples applicable to common assay development situations in different regulatory contexts. However, while this is possibly the most specific flow cytometry guideline, it is no substitute for a library of examples that reduce their recommendations to practice, especially considering it readily admits that the guidance provided is subject to FFP considerations.

The iterative FFP approach largely uses increasing numbers of samples and experiments to ensure that the validation study adequately characterizes the performance of the assay. While there is ample discussion around the general concept of appropriate FFP approaches, they are, by their nature, flexible to specific case needs. The 2021 discussions and recommendations were focused on providing such examples in the form of practical case studies to the bioanalytical cytometry community.

The first case study discussed was the validation for prospective hypothesis testing and the considerations needed for validating receptor occupancy and related assays. Special considerations are needed in monitoring depletion and repletion of a rare population and these assays can be used as part of enrollment criteria to ensure that patients are evaluable for changes if depletion occurs. They can be exploratory or secondary endpoints. Required precision and sensitivity can be evaluated by a healthy donor partial reference range study.

Another case addressed the validation of unplanned analytes in flow cytometry. Strategies were provided to evaluate poorly-characterized analytes present within well-characterized assays in a FFP manner. It can be valuable to anticipate future analyses, particularly where there is a high cost for repeating studies on precious samples. The best practice for handling unplanned analytes, those revealed by association to other parameters (e.g., clinical outcome), is to develop a validation plan capturing all potentially relevant data. A data analysis plan is needed that clearly describes the calculation of all variables to establish the markers of interest in the research/development phase (e.g., establishes cutoff in training set) followed by validation/confirmation in a second, independent, intended-use population. Statistical justification for the methods applied should be pre-established. If up-front analysis is not possible, capturing the data in list-mode may be prudent, so that critical analyses can be performed at a future date if needed. If results will be used to direct patient care, then validation should be sufficient for that purpose. If results are from a test that is not validated and intended as exploratory only, they should be reported separately and clearly marked as exploratory in order to ensure that the results are not misconstrued and used in patient care.

Stability in patient samples and tumor samples was another validation concern for flow cytometry assays. Tumor samples (and other tissue samples) typically cannot be evaluated until 24 hours post collection. Since tumor biomarker expression can be affected by time, all tumor samples should be assessed within the validated stability time-frame, which should typically be at least 24 hours to accommodate for sample collection durations. Samples tested beyond stability should be clearly discernable as such and used only in exploratory analyses (usually run only in supplement to analyses that exclude such samples). It is important to understand the biology of the biomarker and the effect of tissue processing (e.g., enzymes, mechanical disruption, washing buffer, etc.). Due to marker instability, it may be necessary to consider on-site processing/testing, if feasible, based on use of the result.

The collection of high dimensional datasets has led to using study outcomes to identify additional analytes of interest and, retrospectively, assessing their analytical performance. Therefore, assay validation over the duration of a trial was discussed. Regulators recommended well-characterized cells in order to detect markers which may allow for correlation to efficacy. Validation should be refined with early clinical samples while taking into consideration the disease state when identifying cell populations. Since the CLSI H62 guidance [76] is not recognized at present by the FDA and other Health Authorities, it was recommended to speak to the regulators early. The regulators appreciate that methods have limitations, hence there is flexibility with the approach and specimen types used.

A final concern was the necessity of meeting CLIA requirements to support a PD endpoint that is not clearly intended as a biomarker. Such an analyte may still be highly desirable, especially as an enrollment or release criterion. For these situations, if the assay is being used for patient selection/stratification, it was recommended to engage with the FDA for a pre-IND study risk determination to establish if the intended use of the assay confers significant risk or non-significant risk and requires adherence to 21 CFR Part 812 for an investigational use only device. If the data are used to make a diagnosis, prevent illness, or confer risk to a patient in the US, the testing should

be performed in a CLIA-certified laboratory. However, an assay controlling study enrollment and release that has no known health associations and was not used to direct treatment decisions nor to identify or enrich for clinical responses does not need to conform to CLIA requirements.

Evaluation of Accuracy for Flow Cytometry in Regulated Laboratories

Over the past two decades, flow cytometry has become a critical technology supporting all stages of drug development in various areas such as oncology, vaccines, and cell and gene therapies. The use of flow cytometry methods in various regulated environments requires robust method validation with defined validation parameters and performance criteria, all of which should be tailored to the intended use of the data. Though validation criteria such as precision, robustness, or sample stability can be relatively straightforward to evaluate, other parameters such as accuracy/trueness or linearity are difficult to address in the traditional sense, due to the general lack of certified reference material for specific cell populations and biomarkers. Accuracy is defined as the closeness of a result to the true value. For flow cytometry, this requires the knowledge of known quantities of the analyte that is being measured with reference standards. The challenge in flow cytometry is that the availability of reference standards is very limited: cell populations of interest are very diverse and complex and can sometimes be unique to specific disease indications. Thus, nearly all flow cytometry methods fall into the quasi-quantitative category rather than being fully quantitative.

The 2019 and 2020 White Paper in Bioanalysis demonstrated the challenges in accuracy assessment for flow cytometry and recommended that alternative approaches can be used with justification to mitigate risks in an assay [25,27]. The 2021 discussion focused on providing specific recommendations for these alternative strategies in accuracy assessment.

Depending on the intended use of the data and its associated regulatory requirements, some level of accuracy assessment may be required. The preferred method for accuracy assessment is to use reference standards. In cases where such reference standards are not available, alternative approaches may be taken. Additional accepted approaches may include the use of proficiency testing survey materials, the use of biological specimens from patients with a confirmed diagnosis, inter-laboratory comparisons, and comparisons with current reference methodologies.

Flow cytometry generally lacks reference standards because it is traditionally a ground up/user-driven technology. However, there has been some progress in developing standards for flow cytometry. In some cases, biological standard QC material can be used. For some common analytes like $CD3^+$ lymphocytes, $CD3^+$ lymphocyte target counts and acceptable ranges in QC material can be provided by manufacturers. However, the set of evaluable antigens is small and acceptable ranges can vary widely in these samples to accommodate multiple platforms and methods. Once adopted, users should develop their assay specific acceptance criteria for those QC materials. Such QC materials include stabilized blood, lyophilized PBMCs, or custom cell lines.

There are, however, several alternative approaches that can be used when reference standards are not available. These include use of disease patient samples with a confirmed diagnosis, proficiency testing samples from proficiency testing programs, inter-laboratory comparisons, and comparison to gold standard approaches. Multiple organizations such as the United Kingdom National External Quality Assessment Service (UK NEQAS) for Leucocyte Immunophenotyping or the College of American Pathologists (CAP) provide proficiency testing programs for enrollment [78]. These are available for multiple analytes and disease indications using patient samples. Another approach is to exchange samples during validation between a reference laboratory and the validating test laboratory to confirm the comparability of the data. One may also consider using orthogonal methods (e.g., NGS) to determine if both methods trend the same. A recent publication by Medina *et al.* compared NGS to flow cytometry for assessment of minimal residual disease in multiple myeloma [79]; however, there is little guidance on how to approach correlations and which correlations are meaningful. Regardless of the method, accuracy may be evaluated during validation and monitored during study conduct. For critical assays, early discussions with regulators on plans to fulfil accuracy requirements are recommended.

While some advances are being made, the widespread use of different specimen types, analytes, antibodies, fluorochromes, processing methods, instruments, analyses and reporting methods has made standardizing reference materials difficult, and it is a certainty that flow cytometry will always need to deal with reference material gaps. In the absence of standardized reference materials, one alternative means to standardize the field is to utilize standard nomenclature and methods, e.g., CLSI guidance [76]. In addition, the NIST Flow Cytometry Standard Consortium is a recently-launched consortium focused on developing reference standards including reference

materials, reference data, reference methods, and measurement service for assigning the Equivalent Number of Reference Fluorophores (ERF) to calibration microspheres and assessing the associated uncertainties.

Sensitivity Determination in Flow Cytometry Validation

Highly sensitive analysis of rare cell populations is increasingly demanded in drug development. Flow cytometry is an ideal platform for the analysis of rare events which can include low-abundance cell populations such as Tfh (T-Follicular Helper) cells or cellular subsets (e.g., memory/naive B or T cells), analysis of cell depletion triggered by a drug and analysis of the downregulation of a proximal PD marker (e.g., reduced phosphorylated proteins triggered by a drug). Advances in instrumentation, fluorophores, antibody affinity and specificity provide the technical prerequisites for this high-sensitivity analysis. In addition to the technical abilities to generate data from rare events, assay sensitivity becomes a critical parameter that must be carefully considered.

Previous recommendations provided guidance for high-sensitive flow cytometry assay design and development for rare event analysis [21,25,27].

Developing high-sensitivity flow cytometry assays requires applying the best practice principles of panel design as well as paying attention to other factors such as the frequency of the population of interest and background signal. The establishment of assay sensitivity during assay development and validation is multifaceted. A staggered approach is highly recommended including assessment of limit of blank (LOB), LOD and LLOQ, as well as the impact of a minimal number of events in the parent gate and in the GOI [80].

Initial method evaluations should determine the biology of the population and marker of interest (phenotype, activation state, prevalence in healthy vs disease specimens), drug mechanism of action (will a population or marker be merely reduced or fully depleted upon treatment?), and application of the assay. A critical aspect of high-sensitivity flow cytometric assays is acquiring a sufficient number of events in the parent gate as well as in the GOI. The stage of development defines the requirement for establishing analytical sensitivity (LOB and LOD) and assesses the MRE in the GOI [76,127]. A minimum of ten data points is typically used to calculate the LOB (signal in the absence of the detection of the analyte) and LOD (specific signal above LOB which could not be reliably detected). A minimum of three donors and at least two analytical runs are recommended [76,127]. The LOB is typically the mean value in the GOI (number of events and percent positive events) obtained from all donors and all runs plus 1.645 standard deviations (SD). By definition, the LOD must be equal or greater than the LOB. The simplest and most common way to establish the LOD is by using the same data from the LOB experiment and adding 3SD (mean + 3SD) [76,127]. The MRE is an empirical/experimental determination of the relationship between event count and assay precision. This is done by acquiring a large amount of cells (reference sample) and then acquiring stepwise fewer cells (stopping gate), whether that is on the instrument or by data analysis [76,127].

Discussions focused on FFP approaches for validation of rare events and assay sensitivity in high dimensional flow cytometry. If median fluorescence intensity (MdFI), antibody binding capacity (ABC) or molecules of equivalent soluble fluorochrome (MESF) values will be used instead of %-positive, additional measures should be evaluated in order to ensure that comparable MdFI could be generated between experimental days, runs, etc., especially if samples will be tested at multiple time points. When markers show a continuity of expression and are not easily discerned from negative, pre-blocking of the target with excess unconjugated antibody (so-called isoclonal controls), followed by incubation with the conjugated antibody (at final concentration) may be useful for distinguishing specific from non-specific binding. Additionally, in cases of multi-center studies, MdFI comparability across instruments must be confirmed.

Validation should include assessing the LLOQ (functional sensitivity) on rare event populations. When establishing the LLOQ, it is recommended to create a minimum of five levels from each of three donors. There are several ways to create samples at different levels and the approach used should be considered as to whether it is fit for purpose. One common approach is diluting partially stained samples (e.g., using FMx or unlabeled antibody blocking for the marker of interest) with fully stained samples at various ratios. Alternatively, one can sometimes dilute stained disease state sample into stained healthy ones to obtain titration effect of the disease population of interest. Titration effect should be verified to determine if it is consistent with the dilution scheme, and the characteristic population clustering should remain intact. The precision (%CV) for the reportable results should be evaluated. The LLOQ will be the lowest level at which the precision and bias are acceptable for all donors evaluated based on the predefined acceptable precision criteria. For analysis of expression levels such as MdFI or MESF, data should not be reported, except perhaps for exploratory research use, when the cellular population in the GOI is below the established MRE and LLOQ.

Isotype controls cannot be used for LLOQ [81]. Instead, the precision and bias of the measured analyte should be assessed (see CLSI EP17A-2 [81]). However, in some cases, properly titrated isotype controls (e.g., with fluorescence intensity matched to that of the specific analyte-specific antibody that are used at the proper dilution) may be preferable to FMOs for the assessment of LOB, because FMOs do not include non-specific binding of the critical reagent, and this can be an important component of background noise that should be captured in the LOB assessment.

The reporting of data was also discussed, especially when results are below the sensitivity criteria. If reporting is based on COU, it was discussed if a rule should be used for flagging results if values exceed the LLOQ but the MRE is not reached or is reached but %-positive cells are below the LLOQ. It was concluded that when using the MRE approach and a sample is below LLOQ, but above LOD, it should be reported as such and another sample draw may be warranted. Otherwise, the result may be reported as "analyte present but below the limit of quantification". A numerical value may be reported for research use or exploratory endpoint (e.g., data trending or assay performance evaluation). Samples below LOD are unreportable. In a clinical setting, where results are critical to the patient and the samples are not easily replaceable (e.g., bone marrow), results are often reported with limitations. For positive samples that are below LLOQ, it was suggested that setting it as 0.5xLLOQ may be an option, but this can result in over-interpretation by statisticians. Instead, the sample may be flagged with a value <X, but this approach may also make data analysis challenging.

While percent of nucleated cells is often used to describe sensitivity for MRD, the use of percent values to define LLOQ may be confusing. The percent value is not an absolute measure of quantity but is measured relative to the abundance of other events in the sample. Therefore, the use of percent values to define the LLOQ should always be subject to other criteria such as adequacy of event count. It may be possible to establish percentage values that, for a fixed method, will flag specimens not likely to attain the MRE. Increasing the number of events collected by pooling sample preps of the same patient can permit meeting the MRE with no change to the percent.

Recommendations were given for the feasibility and need for validation of rare events/assay sensitivity in high-dimensional flow cytometry. Validation requirements should be FFP, not designed around challenges to validation. If adequate validation is not possible due to logistical, financial or specimen limitations, then the test is not yet ready for clinical application. If the analyte is only present at very low levels, and demonstrates the promise of utility at that level, attempts should be made to develop ways to overcome limitations (e.g., enrichment) or justify why a missing or less robust assessment of performance is acceptable. For example, if the LLOQ is very low but difficult to establish, and the clinical decision point is much higher, it may be acceptable to set the LLOQ at a level that is below the clinical cutoff, but analytically attainable. The MRE collected for the final reportable results should be validated so that the user of that information can appropriately interpret analytical results from the high-sensitivity assay. It is essential to determine MRE under conditions of acceptable accuracy and precision if used in a clinical study. For research purposes, while it may be permissible to consider out-of-specification results in some analyses, focusing on valid results can help prevent significant waste of time and resources as well reduce the possibility of false conclusions and unintended consequences.

Experimental approaches for determining the MRE in the GOI as one integral part of defining the assay sensitivity were further discussed. The minimum number of events collected is often referenced against the Poisson statistic, which indicates that the imprecision of a sample based on 25 events will have a %CV of 20%. The %CV increases as the number of events collected decreases. However, this calculation does not incorporate the analytical imprecision, including human and environmental factors, and therefore indicates that MRE should be established empirically in a manner like LLOQ. The imprecision of multiple runs of negative specimens spiked with dilutions of the population of interest may permit the identification of the minimum number of rare events that meets the acceptance criteria for imprecision. This number can be defined by dilutional recovery to calculate a measure of bias (deviation from expected) to define, along with the imprecision, LLOQ, which defines the lowest concentration where imprecision and bias are acceptable.

Finally, it is important to perform an LLOQ assessment even for an analyte that is expected to be downregulated in specific patient populations or as a result of the drug treatment. LLOQ is needed when reporting a continuous measure to establish the lower limit of the observed range. However, when a test has a numerical output, and the endpoint is categorical i.e. whether the value is above or below the cutoff, then the precision and bias/accuracy around that cutoff are important. In such cases, the LLOQ is not required. From a pragmatic point of view, LLOQ can be assessed when specimens for other validation studies are available to provide a lower barrier to completion, saving bioanalysts from the need for duplicate efforts. If the assay then needs to be applied to a completely different

population where the concentrations are expected to be closer to the LLOQ, the LLOQ will be better assessed on that population, and it may be preferable to wait to establish the LLOQ and expand the measuring range when we expand the COU to include that indication.

Clinical Biomarker Development, Validation & Interpretation by Cytometry

Flow cytometry clinical biomarker assays are often customized to address drug binding, mechanism of action (MOA) and changes in disease biomarkers. During the assay development phase, it is critical to confirm the specificity of reagents on the intended sample type and to ensure that matrix or drug interference will not impact assay performance. Appropriate sample collection and storage protocols should be established during development, as sample stability is critical to determining assay feasibility. Flow cytometry panels should be designed, and antibody clones should be assessed, to ensure the greatest dynamic range of the assay. Validation parameters to be assessed are assay-dependent, but should at a minimum include intra-assay precision, inter-assay precision, and sample stability. Additional parameters, including biological variability, LOD, LOB and LLOQ may be determined, but the level of validation should be dependent on the purpose of the biomarker [82,83].

When executing clinical sample analysis, the cytometers need to be under change control and an installation qualification/operation qualification (IQ/OQ) must be performed to ensure the instrument and associated software are performing correctly [84]. Standard QC protocols to ensure instrument performance must be performed daily and QC samples may be run to confirm assay performance. Cytometer settings need to be optimized or confirmed for each assay and application settings established to ensure comparable data are generated over time. When studies are run globally and samples are processed at different sites, instruments should be harmonized across sites and backup instruments should be available. Throughout the course of the study, instrument QC associated data should be immediately analyzed, and the resulting data trended for outliers to identify errors and enable immediate corrective action.

Data analysis represents an area of opportunity, with auto-gating programs minimizing human bias, increasing data reproducibility, and reducing analyst burden. A limitation to high-dimensional analysis using non-biased gating strategies that has not been addressed in prior white papers [25,27] is the requirement to move clinical data onto the cloud, which has raised concerns over data security and accessibility. With the emergence of auto-gating programs that are FDA CFR 21 Part 11 and ISO compliant, higher dimensional assays to maximize data generation from clinical trials are possible, which will increase our understanding of disease biology and the impact of our therapeutics.

2020 White Paper in Bioanalysis recommendations described the practical validation framework for clinical biomarkers for biological variability validation, algorithm-based analytics and gating, and receptor occupancy assays [27]. 2021 discussions built on these recommendations with additional experience and details.

For example, the 2020 White Paper stated algorithm-based analysis platforms need to be evaluated for the likelihood of gating an incorrect population [27]. Discussions included experience with clinical data analysis requiring data storage on the cloud, and criteria to be met so that unsupervised dimension reducing platforms and AI based supervised gating can be incorporated into the clinical data analysis work stream. There is a need to ensure data security to enable use of these emerging platforms. A secure platform should be used for transferring and storing data. This can be a long undertaking for companies; however, some vendors also provide their platforms on company servers, reducing the need to use a cloud infrastructure. The main advantage of this setup is that the third party software (code) is on the company server with firewalls and security that is company proprietary. As a result of this, companies are becoming ISO and FDA CFR 21 Part 11 compliant which includes establishing an audit trail.

The need for data infrastructure also drives the need for standardizing assays and approaches. Supervised analysis for clinical samples presents both operational advantages and reduces subjectivity associated with manual gating. Strategies were provided for 'educating' algorithms to analyze clinical samples that have atypical profiles and manage cases that are unusual. AI for automated gating can be useful but it takes time before users become comfortable with the output. It is also likely that some patient populations may not be representative; typical representative disease samples are needed. Currently, it is used in discovery to answer basic scientific questions. If automated gating is to work in lieu of manual gating, then the process should have similar checks and balances. For "discovery" mode, this should be used only for profiling approaches and controls may still be needed. To see advantages in cost for automated gating, sample numbers need to be greater. There are certain performance advantages, such as "limited" bias and reduced analysis time since gating analysis is done in real time, typically 12 weeks or more. For populations that are outside normal gating, these unusual samples are flagged; this may necessitate manual

gating or other investigations such as QC of the data to help with the analysis of sample data. There is currently insufficient experience to provide recommendations on how these approaches could be validated and should be a topic for future discussions.

A final topic of discussion was in-study validation to obtain validation samples that are reflective of the patient population for the trial. It was noted that early clinical trial patient samples may not have the expected profile of a sample that has received multiple doses of the drug. It was recommended that the scope for validation sample selection and data interpretation must be carefully defined during validation plan design. Establishing the range in cell populations and longitudinal variability may be done in patient baseline samples and through longitudinal measures in placebo-treated patients.

Spectral Cytometry Ultra High Order/Dimensional Assays in Clinical Application

Part of the power of flow cytometry comes in the ability to measure multiple markers to define specific cell types or make multiple measures on a particular cell type at the same time in the same sample. This is the driver to make ever more powerful instruments capable of making ever-increasing number of measurements from a single sample. Mass cytometry was a game-changer in this respect with upwards of 40 markers in a single assay. However, mass cytometry has limitations that do not make it amenable to most of the assays supporting clinical trials. The majority of the assays used for making such assessments must be performed in near real-time with samples to accurately measure changes in enrolled patients, which is one of the major hurdles for mass cytometry due to its low throughput. A solution to access these high-parameter flow cytometry assays in the clinic exists today in the form of spectral cytometry.

Spectral cytometry differs from conventional flow cytometry in the optical configuration and data deconvolution/unmixing. This method captures a fluorophore's spectral profile throughout the visible and near-IR spectrum as opposed to the use of optical bandpass filters. Consequently, fluorophores with very similar emission spectra can be distinguished in some cases. By analyzing the full spectrum of light emitted by excited fluorophores, instead of just a narrow bandwidth, flow cytometry can now achieve a similar number of markers to mass cytometry while using commercially available antibodies labeled with common fluorophores.

While nascent, spectral cytometry has been successfully implemented in clinical assays, and has replaced PBMC based mass cytometry assays with potential for use in the recent portfolio increase in immunomodulatory drugs [85,86]. The benefits of high/ultra-high parameter flow cytometry in the clinic include reduced patient material demand, reduction in the number of assays and costs, enabling reverse translational research, and robust results.

Clinical trial objectives (secondary or exploratory) and phases of clinical trials (dose escalation versus expansion phase) using high parameter flow cytometry were discussed. They are used primarily in research/translational studies looking for a select set of markers. Its use in clinical trials has included exploratory endpoint Phase I trials, deep immunophenotyping (30+ colors), receptor occupancy, and it has potential for rare tissues (looking at many lineages).

High/Ultra-high parameter spectral cytometry brings new challenges not only with regards to panel design, use of proper unmixing controls, definition of proper channel co-factors, but also new approaches to data analysis. Within the scope of clinical trials, some of these challenges become even more relevant due to the nature of clinical trials where samples are collected at different time points and run under quasi-identical settings of the instruments. The tremendous amount of data generated in clinical trials makes the current approaches for data transfer and data review inefficient. Consequently, optimized and agile workflows need to be developed to benefit high/ultra-high parameter assays. Due to the complexity and the singular conditions of performing such assays in clinical trials, it is important to keep in mind the question to be answered, and evaluate which assay allows an answer with high robustness and stability.

Spectral cytometry does have its own unique challenges in assay implementation. It remains challenging to implement in clinical trials as the instruments are not 510(k)-cleared or have a CLIA level of validation and systems may require in-house full computer system validation. In addition, method development may sometimes be longer than traditional flow cytometry due to more complex panel design which in turn may increase the risk for unwanted interferences between staining antibodies. Staining antibodies need to be titrated for use in spectral cytometry specifically. It is important to evaluate principal markers with standard antibodies; for less important markers, antibodies paired to less common fluorochromes can be used. Care with staining conditions should be taken due to possible unwanted interactions between fluorochromes, fixatives and functional monocyte Fc-

gamma receptor blockers. Data unmixing can present some challenges that should also be taken into consideration by avoiding specific combinations of fluorochromes which may be problematic (e.g., BUV805 vs BV785). A spreading error that can result in artificial populations should be prevented by using principles of good panel design to match the most appropriate fluorophore with each marker of interest and appropriate instrument setup. Despite these challenges, data and instrument performance is generally reliable with little variability over time.

The number of fluorophores and large amount of data generated can be a bottleneck for validation, and several strategies were provided. When the number of parameters is low, traditional validation can be used. If the results of the assay are to be used in decision-making or confer risk to the patient (defined in Study Risk Determination to be Significant Risk), then an investigational device exemption (IDE) would be required and analytical validation would be reviewed for safety. The regulators should be engaged as early as possible. Generally, validation would similarly include precision (intra-assay, inter-assay, and inter-operator/instrument reproducibility, etc.), sensitivity (LLOQ for quantitative assay, LOD, LOB), analytical specificity, specimen stability (processed/unprocessed), and possibly other validation studies as needed. If the assay is used only as an exploratory endpoint, then implementation could occur anytime during drug development. If the assay will be used in regulatory decision-making, used in clinical decision-making or patient selection, then the early introduction of the validation into the process is best; discuss with regulators and consider a pre-submission program. Gating strategies and algorithms should be defined as part of the study protocol and finalized prior to the start of the trial. Quality definitions (e.g., what defines the population, how the gate is set) and intervention points (i.e., at what level of deviation from the expected manual override allowed) should be clearly defined. Data collection, acceptance and rejection, should be objective and pre-defined before trial commencement. Modification to the assay, such as by adding markers, requires a new validation.

It is important to define the proper quality assurance strategy, where both the assay and instruments are quality controlled over time. QC materials that match the profile of the clinical samples and reference controls that are as bright as or brighter than a fully stained sample should be used. The instrument may be monitored using daily QC beads, however, similarly to traditional flow cytometry, frozen PBMCs or lyophilised samples can be used.

There was no standard recommended approach for setting criteria for the number of required events as this is dependent on the staining resolution. Determining the number of cells may be a better approach. For validation of low event numbers, an approach like LLOQ for flow cytometry can be used. This is recommended for the most important reported parameters in the assay, but not possible to do for all parameters. Understanding the purpose of the assay is required in order to tailor a fit for purpose approach. A customised approach may start with 50 events, although to validate a population, at least 100 events in a gate may be needed.

A major source of variation in flow cytometry originates from data analysis. Maecker *et al.* demonstrated that "instrument setup and statistical counting errors accounted for only a very minor proportion of the variability; individualized gating methods accounted for the vast majority of the inter-laboratory variation [87]."

Manual gating is usually performed by sequential gating in two-dimensional plots by looking at different combinations of parameters on the x and y axes, and as such, is affected by analyst specific decisions on gate placement. These user specific decisions on gate placement are the primary cause of subjectivity and variability in manual gating; they are heavily influenced by previous training, quality of staining and number of events analyzed. Despite observed subjectivity and, for high-parameter panels (20+ colors), significant time and labor investments, the majority of data analysis is still performed manually. Consequently, reproducibility and quality of data analysis have been identified as a major driving force behind data analysis automation.

Automated gating approaches, as an alternative to manual gating, have been explored in the past decade but are yet to be applied widely, especially in clinical data analysis. Application of automated data analysis should improve operational effectiveness and data consistency. Use of high-parameter flow cytometry panels results in the generation of complex multidimensional data sets that require substantial time investment in data analysis and may suffer from reproducibility issues when gating is done by different analysts. Different machine learning and clustering algorithms have been developed to aid this process. Based on their use, these approaches could be divided into tools developed for the detection of unknown cell populations (unsupervised learning) versus those developed for targeted analysis (supervised learning) [88]. Unsupervised learning methods include data clustering and dimensionality reduction. They do not require training data; the goal is to identify and quantify cell populations without *a priori* knowledge. For example, t-distributed stochastic neighbor embedding (t-SNE) and self-organized maps (e.g., FlowSOM) have all been utilized successfully for discovery purposes and/or as an additional aid in data visualization. Supervised learning methods require manually labeled training data sets to build a model, perform

predictions and include classification and regression such as support vector machines, artificial neural networks, random forests, etc. In addition, automated preprocessing tools that remove fluorescence anomalies (e.g., flowClean, flowAI), and perform transformation (e.g., flowCore) and normalization (e.g., flowStats) can be used to clean raw data, although the use of unvalidated software can pose a risk of calculation errors.

Development of automated data analysis approaches and their rigorous testing and validation will aid in their standardization and reproducibility and ultimately result in reduced bias, better efficiency and cost savings during clinical data analysis.

Issues with the implementation of automated data analysis approaches are preventing its widespread use. These include lack of trust in automated approaches, lack of understanding how to perform automated approaches, lack of computer programming knowledge, lack of understanding in how to best validate the software, lack of resources to do automated data analysis, and the increased regulatory requirements for clinical flow cytometry data analysis.

Limitations, recommendations and best industry practices for establishment and validation of supervised automated data analysis were discussed. These include but are not limited to direct comparison of automatically to manually gated data sets, the use of visual outputs as both quality control of automated data analysis and aid in understanding and interpretation of the data.

Based on the opportunities and challenges, recommendations were developed for data analysis for spectral cytometry which may also generally apply to mass cytometry and multiparameter flow cytometry. First, it is important to consider the intended use of the data. Frequently, the collection of information to evaluate responders and non-responders is required and the setup of a cut-off is also desired. Retrospective analysis simplifies logistics when assessing data from a clinical trial. A large number of analytes can be used for discovery purposes, and then the method can be adjusted to include fewer analytes to improve sensitivity for key parameters. This technique can be utilized for finding unique novel populations where the biology is not well understood. For automated data analysis, it is important to understand the feasibility in clinical trial settings for spectral cytometry. Supervised data can be useful to see PD, drug effects and predictive effects but development of a supervised data analysis approach requires representative manually gated FCS files that mimic clinical samples. Data quality is key, and additional mechanisms are needed to ensure the data are real readouts not artifacts. For unsupervised data analytics, evaluate different companies offering this service. Evaluate available unsupervised and supervised approaches on representative data sets, highlighting the advantages, disadvantages and limitations of each approach. Finally, validate supervised approaches using large clinical data sets.

Future areas of development for high/ultra-high parameter flow cytometry include developing workflows for implementation of semi-automated supervised gating strategies, for data management, standardization of assays across studies for more robust results, reverse translational research, and application of the quality of results in specific samples by removing allele frequencies (AF) (tumor biopsy material).

Imaging Cytometry Quantitative Analysis of Target Engagement

Imaging cytometry combines the throughput of flow cytometry with the capabilities of fluorescence light microcopy to visualize samples, interrogating numerous characteristics of cell morphology and fluorescence labeling, which generates information on cellular phenotype, signaling and function. Therefore, quantitative image analysis performed with imaging cytometry instruments allows collection of unique measurements informing on the molecule internalization from the cellular environment, co-localization with various cellular compartments, nuclear translocation, cellular synapse formation, and many others, serving as a valuable tool in drug development. Imaging cytometry can support the preclinical development of biotherapeutics, with a focus on quantitative biomeasures and biomarkers in support of translational modeling for assessing target feasibility and human dose prediction.

The 2020 White Paper in Bioanalysis started the discussion on uses and validation strategies for image cytometry-based assays [27]. The benefit of these technologies for blood and rare matrices is to provide semi-quantitative results on top of morphological and spatial characterization of analytes. The status of imaging cytometry use in industry was discussed together with current gaps in technology and barriers for broader implementation. The vision for addressing unmet needs in bioanalysis by expanding high-dimensional imaging cytometry capabilities was also discussed. A primary focus was to establish standardized practices and strategies for assay validation.

Although there are cases where imaging cytometry is being used for drug development, there should be a reason for its use over traditional flow cytometry. Examples for use include intracellular localization, receptor internalization, translocation, and cell shape (e.g., RBC abnormalities). It is currently being employed for receptor occupancy

(RO) (binding, internalization) and TE (signaling) outside of clinical trials. Other pre-clinical uses include PK/PD modeling and dose simulations.

Measurements of antibody-induced target internalization using imaging cytometry are conducted *in vitro* with a variety of modalities and cellular systems depending on the study goals. Therefore, the method must be established for each assay type for optimal performance. Assay reagents including target specificity of antibodies, compatible fluorophores, and matrix (whole blood vs isolated cells) must be optimized. Experimental design (cell activation, adherent vs suspended, time courses, controls), sample acquisition, and data analysis should also be evaluated. Imaging cytometry has the potential to be used for exosomes with options from Luminex with sensitivity designed for small particles.

Despite the benefits, there are technological challenges to consider. Chip cytometry requires PBMC isolation which can affect cell shape and population representation. Throughput is also lower. There are references for its use in clinical trials supporting traditional flow cytometry, however, experience is limited and typically exploratory fit-for-purpose. Validation approaches are similar to conventional flow cytometry, such as instrument QC, sample storage stability, critical reagent control. Aspects unique to imaging cytometry include validation of image segmentation and quantitative analysis as well as the use of biological controls. An important aspect is to ensure comparability and compatibility of TE and RO parameters such as sensitivity and reproducibility across different assays. Instrument harmonization is also a key consideration although experience is limited; potential strategies are to use instrument bead QCs, spike in controls or lyophilized controls. The shelf life of buffers should also be validated.

Finally, sample handling considerations were discussed. Sample preparation for imaging cytometry is often complex. For rare matrices, samples can be applied to slides with cytospin. As mentioned previously, PBMC isolation must be evaluated.

Mass Cytometry in Clinical Biomarkers & "Clinical Trial Compatibility"

The advent of mass cytometry has enabled high dimensional and unbiased examination of the immune system to simultaneously interrogate a large number of parameters and gain a better understanding of immunologic data from clinical trial samples. Mass cytometry is a leading platform for high dimensional analysis of cell phenotype and functional responses, facilitating characterization of complex biological systems and uncovering processes that take place in the course of diseases and with treatments, thus significantly advancing early-stage discovery. It has applications in intracellular cytokine detection, RO, cell viability discrimination, cell cycle identification, immune profiling and many others [89]. However, while its use in clinical research has increased, the capabilities of mass cytometry for biomarker analysis in pre-clinical and clinical settings are not fully utilized.

Mass cytometry technology, while being very flexible in its panel size accommodation, is relatively new to use in clinical research efforts and less extensively tailored towards regulatory compliance. Other cytometry technologies, like traditional flow cytometry, are commonly used to characterize samples from clinical trials but are limited in their panel size [90]. For mass cytometry, to maximize use as a patient characterization tool in clinical trials, its challenges and possible solutions should be assessed.

The 2020 White Paper in Bioanalysis recommendations [28] advocated for the potential for mass cytometry and panel validation parameters and acceptance criteria were discussed. In 2021, additional case studies were discussed to describe strategies for its use in clinical trials. For example, mass cytometry identifies phenotypes and markers associated with peak CAR-T expansion, and may indicate association with response [91,92].

Mass cytometry is most typically used in a preclinical exploratory or discovery manner to identify phenotypic effects from treatments that can be funneled down to panels for other higher throughput flow cytometry methods. Despite its benefits (multiplexing, ease of use of CD45 barcoding, minimal signal spillover, internal references to monitor staining), there still exists some opinions that the technology is not ready for clinical use. There are several challenges with mass cytometry being fully clinical trial compatible, mainly due to sources of assay variability over time and low throughput. Clinical trials often take place over several months to years, which can lead to many separate experimental batches of samples being stained at different time points. Sources of batch-to-batch variability can come from differences in operator/analyst, reagents, and the mass cytometry instrument, among other factors. For these reasons, batch effects in data have been observed and can confound cross-batch analysis, especially in the context of multivariate clustering algorithms. There is no consensus for batch effect correction yet (e.g., picking baseline time point). Therefore, it is necessary to confirm comparisons are appropriate. Data generation takes time and therefore it is generally not amendable for end of cohort decisions when rapid turnaround of results is needed.

Custom and commercial solutions are both possible for reducing variability. Over the course of a clinical trial, these factors should be assessed and controlled wherever possible. Custom panels benefit from using internal references to score data quality. Commercial panels and analysis solutions have these quality measures partially built in. These can be applied to both automated and more manual workflows when automation is not yet an option [93]. Variability introduced into mass cytometry assays also has several possible solutions that have been practiced in the field. Antibody variabilities can be controlled by having antibody cocktails that are cryopreserved and used over the length of large portions if not the entirety of a clinical trial. These antibody cocktails can be preserved for long periods of time and shipped to other laboratories for sample staining as needed. Staining variability may also be reduced by automating many or all of the parts of a staining protocol.

Reliable mass cytometry data can be generated if the entire mass cytometry workflow is optimized and a FFP validation is performed. Process controls are critical to generating reliable data but can be challenging to manufacture and maintain lot-to-lot consistency. Positive and negative controls to monitor isotope impurities and '+1' spillovers should be used while trying to minimize lot changes. Mass minus one (MMO) controls help assess isotope impurity versus non-specific binding. Controls should be included in more runs; particularly for machine learning exercises. Initially only critical assay reportable should be reported to minimize the analysis complexity and generate data in a timely manner that can be used for clinical decision-making. Additional (biased or unbiased) analysis can be done later in the study as part of an exploratory analysis. Stained and unstained sample stability is important and should be studied before commencing validation. One successful strategy to minimize batch effects is to manufacture a mixture of cell types or cell lines including stimulated and unstimulated cells in a large quantity as controls. Lyophilized cell lines are also an option for use as controls. Identify a list of cell types that are necessary for decision-making early on and carefully monitor the quality of the data throughout the duration of the study.

In regards to sample handling considerations for mass cytometry, if remote clinical sites will be used for the study, specimen stability and transport studies should be performed before analytical validation and certainly prior to trial commencement. Ensuring the clones in your panel are compatible with the sample matrix, sample stabilizer, as well as buffers used for processing of samples. The use of barcoding and innovations on the ability to expand barcoding past 20 samples also allows for decreased variability and minimizes batch effects from staining as well as instrument runs.

Multivariate Analytical Techniques & Multiparameter Flow Cytometry

New technologies (e.g., mass cytometry or spectral cytometry) can enable the analysis of 40+ parameters per cell resulting in greater than 100 populations per analysis. The magnitude of such immunophenotyping datasets coupled with intrinsic response heterogeneity and correlated lymphocyte biology make the use of traditional flow cytometry analysis methods inefficient and impractical. Therefore, more appropriate analytical techniques are needed to assess large, heterogeneous immunophenotyping datasets to identify associations with dose, efficacy and/or safety read-outs. Multivariate analysis techniques offer a potential solution to these problems by enabling for dimensionality reduction and leveraging covariance among populations to identify underlying associations with therapeutic dose and clinical response.

Multivariate analysis has been used to identify therapy- and dose-dependent PD changes associated with best objective response from a Phase II study of two immuno-oncology therapeutic molecules [94]. Another use case is the evaluation of immune responses from clinical trial samples for critical information regarding the immunogenicity of vaccine candidates. Antibody and T-cell responses are commonly assessed as measures of vaccine candidates' immunogenicity.

One challenge of multiparameter flow cytometry is antibody steric hindrance which occurs when two large molecules compete for similar or overlapping epitopes of a defined antigen at the cell surface. This could result in false-negative reporting. Examples include a therapeutic antibody binding to a target which subsequently needs to be detected using an immunophenotyping antibody. Steric hindrance can be observed after *in vivo* drug dosing, for instance in PBMC preparations from cancer immunotherapy patients and hence affect immunomonitoring. As an example, the flow cytometric detection of B cells via anti-CD20 may be affected in patients that were treated with rituximab or similar CD20-targeting antibodies. It may also occur *in vitro* in functional assay design where it could affect pharmacology assay readouts if several antibodies compete for overlapping cell surface receptors [95,96]. There are routine workflows that can be used to systematically assess steric hindrance in a bioanalytical PD biomarker setting for assay development and subsequent data analysis in checkpoint inhibitor-experienced and -naïve patients. These involve monitoring of residual target engagement in previously treated patients, the introduction of new and

more specific flow cytometry detection reagents and the implementation of quasi-quantitative target assessment methods. A screening workflow should assess detection monoclonal antibody (mAb) resolution by a comparison with well-described mAbs, if available, detection mAb specificity by co-labeling and soluble target titration experiments, and co-labeling of cells with the detection mAb in the presence of the therapeutic drug mAb. Requirements for robust flow cytometry detection antibodies should also be followed to be highly target-specific, bind with good affinity, allow for high-resolution target detection and possess the ability to bind to target even in the presence of large molecule drugs.

Regarding sample stability of whole blood preparations, a case study was discussed characterizing the activation status of innate immune cell populations in stabilized human whole blood by high parametric flow cytometry. Smart tubes which preserve all leukocyte populations without cell purification were used. Assay specificity was confirmed from the FMO setup. Stability was determined for the phenotypic markers to be over 6 months. The mean repeatability CV and mean intermediate precision CV were within 40%. The functionality validation was shown by upregulation of activation markers in response to TLR agonist stimulation detected in monocytes, mDCs, pDCs and B cells.

The advent of technologies enabling for monitoring 25+ parameters/cell has resulted in the ability to monitor hundreds of cell populations in a single clinical sample. Fit-for-purpose validation strategies for high-dimensional flow-based assays were discussed and it was recommended to validate only parent populations, if it is known that the parent populations will be the relevant populations monitored during the trial. Any populations that might be used in a manner that confers significant risk to the patient should be validated prior to the commencement of the trial. If they are not validated and data is lost because it was not collected properly or specimen stability is limited, the success of the trial could be jeopardized.

Data analysis for multiparameter flow cytometry was a key topic. Previous discussions of large data sets for other flow cytometry technologies apply. Data analysis depends on the quality of the data generated using harmonized procedures, quality reagents, appropriately maintained instruments, etc. Algorithms will find correlations, therefore, the key is to take an active role to verify using secondary data sets and statistical testing in relation to the biology. Translational scientists should engage with bioinformatics scientists to assess trends in data. Sharing information and the scientific meaning of trends to determine clinical utility is important. Another topic to consider in the use of multiparameter flow cytometry is that the complexity of data analysis can bring significant delays with data delivery, so the decision to use this approach also needs to consider turnaround time requirements.

RECOMMENDATIONS

Below is a summary of the recommendations made during the 15th WRIB:

Recent Developments in Flow Cytometry Validation in a Bioanalytical Lab

- The best practice for handling unplanned analytes, those revealed by association to other parameters (e.g., clinical outcome), is to develop a validation plan capturing all potentially relevant data.
- A data analysis plan is needed that automates the calculation of all variables to establish the markers of interest in the research/development phase (e.g., establishes cutoff in training set) followed by validation/confirmation in a second, independent, intended-use population. Statistical justification for the methods applied should be pre-established.
- If up-front analysis is not possible, capturing the validation data in list-mode may be prudent, so that critical analyses can be performed at a future date if needed.
- If results are from a test that is not validated and intended as exploratory only, care should be taken so that the suitability of the results cannot be misconstrued and used in patient care or presented as analytically validated.
- Since tumor expression can be affected by time and stability, all tumor samples must be run within the validated stability, which should typically be at least 24 h to accommodate for sample collection durations. It may be necessary to consider on-site processing/testing, depending on the use of the assay.
- Regulators recommended well-characterized cells in order to detect markers that may allow for correlation to efficacy.
- Validation should be refined with early clinical samples while taking into consideration the disease state when identifying cell populations. It was recommended to speak to the regulators early. The regulators appreciate that methods have limitations, hence there is flexibility with the approach and specimen types used.

- Regarding PD endpoints used as enrollment or release criteria but that are not intended for consideration in clinical decisions or as a biomarker, it was recommended to engage with the FDA for a study risk determination to establish if the intended use of the assay confers significant risk or non-significant risk when using for patient selection/stratification.
- An assay does not need to conform to CLIA requirements if the assay controlling study enrollment and release has no known health associations and the assay was not used to direct treatment decisions nor to identify or enrich for clinical responses.

Evaluation of Accuracy for Flow Cytometry in Regulated Laboratories

- Depending on the intended use of the data and its associated regulatory requirements, some level of accuracy assessment may be required.
- The preferred method for accuracy assessment is to use reference standards, although some additional accepted approaches are:
  - Use of proficiency testing survey materials;
  - Use of biological specimens from patients with a confirmed diagnosis;
  - Inter-laboratory comparisons;
  - Comparisons with current reference methodologies.
- There has been some progress in developing standards for flow cytometry. In some cases, biological standard QC material can be used. For some standard analytes like $CD3^+$ lymphocyte counts, expected values range in QC material can be provided by manufacturers.

Sensitivity Determination in Flow Cytometry Validation

- Flow cytometry is an ideal platform for the analysis of rare events.
- Assay sensitivity is a critical parameter for flow cytometry methods for the analysis of rare events.
- A staggered approach to the establishment of assay sensitivity is highly recommended including assessment of LOB, LOD and LLOQ, as well as the impact of a minimal number of events in the parent gate and in the GOI.
- Acquiring enough cells in the parent gate and in the gate of interest is crucial for high-sensitive flow cytometric assays.
- For sensitivity determination is recommended to use at least ten data points to calculate the LOB and LOD, three donors and two analytical runs:
  - LOB is defined as the mean value in the gate of interest (i.e., number of events and percent positive events) attained from all donors and all runs plus 1.645 SD and LOD should be greater than the LOB [76,127].
- Establish the LOD using the same data from the LOB experiment and add 3SD (mean + 3SD).
- The MRE determination is determined by acquiring a large amount of cells (reference sample) and then acquiring stepwise fewer cells (stopping gate).
- If MdFI or ABC values will be used instead of % positive, additional measures should be evaluated in order to ensure that comparable MdFI could be generated between experimental days/runs, etc., especially if samples will be tested at multiple time points.
- For LLOQ determination recommended to generate a minimum of five levels from each of three donors by diluting partially stained samples,for instance, using FMx for the marker of interest [76,127].
- The titration effect should be verified to determine if it is consistent with the dilution scheme, and the characteristic population clustering should remain intact.
- LLOQ is defined as the lowest level at which the precision is acceptable for all donors evaluated based on the predefined criteria [76,127].
- For analysis of expression levels such as MdFI or MESF, data should not be reported when the cellular population in the gate of interest is below the established MRE and LLOQ.
- Properly titrated isotype controls (e.g., with fluorescence intensity matched to that of the specific analyte specific antibody that is used at the proper dilution) are preferable to FMOs for the assessment of LOB.
- When using the MRE approach and a sample is below LLOQ but above LOD, it should be reported as such and another sample draw may be warranted. Otherwise, the result may be reported as analyte present but below the limit of quantification. Samples below LOD should be reported as undetectable. Actual values

in these cases may be evaluated for exploratory research purposes but should not be considered as validated results.

- Validation requirements for rare events/assay sensitivity in high-dimensional flow cytometry.
  - Should be FFP, not designed around challenges to validation.
  - If adequate validation is not possible, then the test is not yet ready for clinical application beyond exploratory use.
  - If the analyte is never present at more than very low levels, but demonstrates the promise of utility at that level, attempts should be made to develop ways to overcome the limitations (e.g., enrichment) or justify why a missing or less robust assessment of performance is acceptable.
  - MRE collected for the final reportable results should be validated under conditions of acceptable accuracy and precision if used in a clinical study.
- Experimental approaches for determining the MRE in the gate of interest:
  - imprecision of multiple runs of negative specimens spiked with dilutions of the population of interest may permit the identification of the MRE that meets the acceptance criteria for imprecision.
  - this number can be defined by dilutional recovery.
- It is important to perform an LLOQ assessment for an analyte that is expected to be overexpressed in specific patient populations or as a result of the drug.

Clinical Biomarker Development, Validation & Interpretation by Cytometry

- During the assay development phase of flow cytometry clinical biomarker assays, it is critical to confirm the specificity of reagents on the intended sample type and to ensure that matrix or drug interference will not impact assay performance.
- Appropriate sample collection and storage protocols should be established during development, as sample stability is critical to determining assay feasibility.
- Validation parameters to be assessed should at a minimum include intra-assay precision, inter-assay precision, and sample stability.
- Additional parameters, including biological variability, LOD, LOB and LLOQ may be determined, but the level of validation should be dependent on the purpose of the biomarker.
- For clinical sample analysis, cytometers need to be under change control and an IQ/OQ must be performed.
- When studies are run globally and samples are processed at different sites, instruments should be harmonized across sites and back-up instruments should be available.
- Standard QC protocols to ensure instrument performance must be performed daily and QC samples may be run to confirm assay performance.
- A secure algorithm-based analysis platform may be used for transferring and storing data. One suggestion is to host third-party software on the company server, with firewalls and security that is company proprietary.
- Supervised analysis for clinical samples presents both operational advantages and reduces subjectivity associated with manual gating.
- Validation samples that are reflective of the patient population for the trial during in-study validation:
  - Early clinical trial patient samples may not have the expected immunophenotyping profile due to disease and prior drug treatment.
  - Scope for validation sample selection and data interpretation must be carefully defined during validation plan design.
  - Establishing the range of cell populations and longitudinal variability may be done in patient baseline samples and through longitudinal measures in placebo treated patients.

Spectral Cytometry Ultra-High Order/Dimensional Assays in Clinical Application

- Ultra-high parameter flow cytometry is used primarily in research/translational studies looking for a select set of markers. Its use in clinical trials has included exploratory endpoint Phase I trials, deep immunophenotyping (30+ colors), receptor occupancy, and it has potential for rare tissues.
- The benefits of high parameter flow cytometry in the clinic include reduced patient material demand, reduction in the number of assays and costs, enabling reverse translational research, and increased robustness of results.

- Optimized and agile workflows need to be developed for programs to benefit from multiparametric flow assays.
- Spectral cytometry implementation in clinical trials can be challenging as instruments are not currently 510(k)-cleared and validation to CLIA and similar standards may require in-house full computer system validation.
- It is important to evaluate important markers with standard antibodies; for less important markers, antibodies paired with less common fluorochromes can be used.
- Care with staining and buffers should be taken due to interactions between fluorochromes, fixatives and monocyte blockers. Specific combinations of fluorochromes should be avoided (e.g., BUV805 vs BV785).
- Spreading error that can result in artificial populations should be prevented by using good panel design to match the most appropriate fluorophore with each marker of interest and appropriate instrument setup.
- Method validation strategies:
  - When the number of parameters is low, traditional validation can be used.
  - If the results of the device are to be used in decision making or confer risk to the patient, then IDE would be required and analytical validation would be reviewed for safety.
  - Based on the purpose, validations may include:
    - Accuracy and precision: intra-assay, inter-assay, and inter-operator/instrument reproducibility, etc.;
    - Sensitivity: LLOQ for quantitative assay, LOD, LOB;
    - Analytical specificity;
    - Specimen stability: processed, unprocessed.
  - If the assay is used only as an exploratory endpoint, then implementation could occur anytime during drug development.
  - If the assay will be used in regulatory decision making, used in clinical decision making or patient selection, then early introduction of the validation into the process is best; discuss with regulators and consider a pre-submission program.
  - Gating strategies/algorithms should be defined as part of the study protocol and finalized prior to the start of the trial.
  - Quality definitions and intervention points should be clearly defined.
  - Modification to the assay (e.g., adding markers) requires a new validation.
  - to validate a population, a suitable minimal number of events in the gate (e.g., 100 events) are desirable.
- QC material and reference controls:
  - It is important to define the proper QC material and reference controls.
  - QC material should match the profile of the clinical samples.
  - Reference controls should be at least as bright as the full stained sample.
  - Use longitudinal controls.
  - Use beads for instrument daily QC and in sample acquisition, however, frozen PBMCs or lyophilized samples (similar to classic flow cytometry) can be used.
- There is no standard recommended approach for setting criteria for the number of events.
- Development of a supervised data analysis approach:
  - Requires representative manually gated FCS files that mimic clinical samples;
  - Direct comparison of auto to manually gated data sets;
  - Use of visual outputs as both quality control of automated data analysis and aid in understanding and interpretation of the data.
- Recommendations were developed for data analysis for spectral cytometry which may also apply to mass cytometry and multiparameter flow cytometry:
  - Consider the use of the data;
  - Frequent collection of information to evaluate responders and non-responders to determine a cut off;
  - For automated data analysis, it is important to understand the feasibility of automated data analyses in clinical trial settings;
  - Utilize unsupervised data analytics for finding unique novel populations where biology is not well understood;
  - Supervised data can be useful to see PD, drug effects and predictive effects;

- Evaluate available unsupervised and supervised approaches on representative data sets, highlighting the advantages, disadvantages and limitations of each approach;
- Validate supervised and automated data analysis approaches using large clinical data sets.

Imaging Cytometry Quantitative Analysis of Target Engagement

- Imaging cytometry is currently being used for RO and TE outside of clinical trials.
- The imaging cytometry method must be established for each assay type for optimal performance. Optimization of the following is important:
  - Assay reagents including target specificity of antibodies, compatible fluorophores, and matrix (whole blood versus isolated cells);
  - Experimental design (cell activation, adherent versus suspended, time courses, controls);
  - Sample acquisition;
  - Data analysis.
- Several validation approaches are similar to conventional flow cytometry, such as instrument QC, sample storage, critical reagent control. Aspects unique to imaging cytometry include validation of image segmentation and quantitative analysis as well as the use of biological controls.
- Ensure comparability and compatibility of TE and RO information across different assays (including sensitivity and reproducibility).
- Instrument harmonization is a key consideration; potential strategies are to use instrument bead QCs, spike in controls or lyophilized controls. The shelf life of buffers should also be validated.
- For rare matrices, samples can be applied to slides by a cytospin method.
- PBMC isolation and other sample manipulation *ex vivo* can affect cell populations and cell shape, and the impact must be evaluated.

Mass Cytometry in Clinical Biomarkers & "Clinical Trial Compatibility"

- Mass cytometry is most typically used in a preclinical exploratory or discovery manner to identify phenotypic effects from treatments that can be funneled down to panels for other higher throughput flow cytometry methods.
- Batch effects in data have been observed and can confound cross-batch analysis, especially in the context of multivariate clustering algorithms. There is currently no consensus for batch effect correction (e.g., picking baseline time point) therefore it is necessary to confirm comparisons are appropriate.
- Data generation takes time and therefore this technology is not appropriate for end of cohort decisions.
- Commercial panels and analysis solutions can reduce variability.
- Antibody variabilities can be controlled by having antibody cocktails that are cryopreserved and used over the length of large portions, if not the entirety of a clinical trial.
- Staining variability may also be reduced by automating many or all of the parts of a staining protocol.
- Positive and negative controls to monitor isotope impurities and '+1' spillover should be used while trying to minimize lot changes.
- Only critical assay parameters should be reported to minimize the analysis complexity and generate data that can be used for clinical decision-making.
- Sample stability is important (stained vs unstained) and should be studied before commencing validation.
- To minimize batch effects, prepare a mixture of cell types or lines (stimulated, unstimulated) in a large quantity as controls.
- Identify a list of cell types that are necessary for decision-making early on and carefully monitor throughout the study.
- Stability and transport studies should be performed before analytical validation and certainly prior to trial commencement.
- Ensure the panel is compatible with sample type, sample stabilizers, and buffers used for processing samples.

Multivariate Analytical Techniques & Multiparameter Flow Cytometry

- Newer analytical techniques are needed to assess large, heterogeneous immunophenotyping datasets to identify associations with dose, efficacy and/or safety read-outs.

- Multivariate analysis techniques enable dimensionality reduction and leverage covariance among populations to identify underlying associations with therapeutic dose and clinical response.
- Assess steric hindrance in a bioanalytical PD biomarker setting by monitoring residual target engagement in previously treated patients, introducing new and more specific flow cytometry detection reagents and implementing quasi-quantitative target assessment methods.
- Requirements for robust flow cytometry detection antibodies:
  - Highly target specific;
  - Binding with good affinity;
  - Allowing for high-resolution target detection;
  - Ability to bind to target even in the presence of large molecules.
- For high-dimensional flow-based assays, it was recommended to validate only parent populations, only if it is known that the parent populations will be the only relevant populations that will be used in the trial.
- Data analysis depends on the quality of the data generated using harmonized procedures, quality reagents, appropriately maintained instruments, etc.
- Algorithms will find correlations, therefore, the key is to take an active role to verify using secondary data sets and statistical testing in relation to biology.
- Translational scientists should engage with bioinformatics scientists to assess trends in data. Sharing information and scientific meaning of trends to determine clinical utility is important.

SECTION 3 – LBA Regulated Bioanalysis, Critical Reagents & Positive Controls

Carrie Hendricks[31], Meenu Wadhwa[11], Albert Torri[16], Mark Ma[32], Shannon Harris[33], Seema Kumar[37], Michael A Partridge[16], Teresa Caiazzo[13], Shannon Chilewski[1], Isabelle Cludts[11], Kelly Coble[34], Boris Gorovits[35], Christine Grimaldi[34], Gregor Jordan[36], John Kamerud[13], Beth Leary[13], Meina Liang[29], Hanjo Lim[20], Andrew Mayer[6], Ellen O'Connor[24], Nisha Palackal[16], Johann Poetzl[38], Sandra Prior[11], Mohsen Rajabi Abhari[4], Natasha Savoie[12], Catherine Soo[39], Mark Ware[5], Bonnie Wu[5], Yang Xu[40], Tong-Yuan Yang[5] & Jad Zoghbi[31]

Authors are presented in alphabetical order of their last name, with the exception of the first seven authors who were session chairs, working dinner facilitators and/or major contributors

Author affiliations can be found at the beginning of the article.

HOT TOPICS & CONSOLIDATED QUESTIONS COLLECTED FROM THE GLOBAL BIOANALYTICAL COMMUNITY

The topics detailed below were considered as the most relevant "hot topics" based on feedback collected from the 14th WRIB attendees. They were reviewed and consolidated by globally recognized opinion leaders before being submitted for discussion during the 15th WRIB. The background on each issue, discussions, consensus, and conclusions are in the next section and a summary of the key recommendations is provided in the final section of this manuscript.

Bioanalytical Challenges for Inhalation & Oral Delivery of Biologics

Should anti-drug antibodies (ADA) be determined in the lung as well as in serum? What is unique when evaluating immunogenicity (i.e., ADA) of inhaled and orally delivered biologics? What is the assessment strategy for ADA of inhaled biologics? Is ADA isotyping required, e.g., IgA? Is tissue PK assessment required? If so, how many tissues are considered sufficient? What are the technologies used for ultra-sensitive PK methods and have those technologies been used successfully for regulatory submissions? Is metabolite quantification required for drug products and excipients?

Free, Bound, Total, Active, Monoactive, Biactive, & Multiactive PK Assays

Which form of drug needs to be measured to support multispecific products: biactive, multiactive? Are there any best practices to overcome bioanalytical challenges? Are there potential interferences impacting free drug and total biactive PK assays? Are there any multiplex technologies to facilitate PK characterization (e.g., free/bound measurement) for multispecific biotherapeutics?

Implementing Free/Total PK Assays in Regulated Bioanalysis

Is free drug determination essential for PK/PD analysis? If modeling predicts equivalence between expected free and total drug concentrations, is free assay the default? Is the reverse argument true as well (i.e., can free be modeled

from total drug)? What are the recommendations regarding when to use a free assay vs total drug measurement? Could LCMS be a viable alternative for the total drug assay?

Multi-Domain Biotherapeutics PK Assays in Regulated Bioanalysis

Is PK assessment done differently for multi-domain biotherapeutics (MDB) than for single target biologics? How are target interferences in PK assays inferred for difficult to express recombinant target proteins? Should soluble targets without fusion to Fc or other frameworks be used for target interference evaluation? What are the pros and cons of alternate animal and non-animal-derived reagent generation strategies? What is the required extent of critical reagent characterization – fit for purpose or extensive (e.g., neutralizing capability of anti-idiotypes (anti-ids), stability of labeled reagents in assay buffer or in the biological matrix)? How are the neutralizing capability of anti-ids assessed when such target proteins are not available? What are the challenges with assay validation on analytical platforms with minimal vendor technical support?

Bioanalytical Challenges to Study the Biodistribution of Biotherapeutics

What is the validation strategy for clinical PK assays using non-liquid tissue matrices? What is the justification for using surrogate animal tissues? What is the minimal sampling volume for non-liquid tissue matrices? What is the industry experience using MS imaging and mass cytometry on studying biodistribution in animal studies? Are there any novel imaging technologies for investigating biodistribution in nonclinical and even clinical studies? How can PK data from non-liquid tissue and biofluid be more comparable? Are there any novel technologies for tissue sampling and drug extraction from tissue?

Advanced Approaches in Critical Reagents Selection for PK Assays

What most closely predicts antibody pair selection when developing a sensitive method (k_{on} and/or k_{off} rates, signal to noise ratios (SNR) in LBA platforms, other)? Are there new ideas on high throughput reagent pair testing (automation, increased full-time equivalents [FTEs], biolayer interferometry [BLI], other) and predictability to LBA method development? When and how many resources should be invested in early clinical development? Is there experience using LCMS for characterization of reagent Degree of Labeling (DoL) and translating changes in DoL to assay performance (especially impacts on PK assay sensitivity)? What changes matter (distribution of label incorporation and % unlabeled)? How is DoL used as part of bridging of new lots of labeled reagents in validated methods?

Challenges with Positive Control Generation for ADA Assays

What scientific criteria should be applied to select an appropriate ADA positive control (PC)? Are there best practices for selection and application of the ADA PC (e.g., select a PC with appropriate binding affinity to achieve recommended drug tolerance and sensitivity; use a polyclonal PC as "representative" ADA and use mAb PC as plate control)? For multi-domain molecules, when using a PC, is it necessary to characterize binding to all domains? As an industry, should there be more effort with characterizing clinical ADA responses in order to understand the relevance between the assays and clinical outcome? If so, what are the recommendations? What is the best practice for polyclonal PC purification to obtain the antibodies that are specific to the antibody variable regions? What characterization assessments should be done for PCs? Is there any correlation or trend between affinity and ADA assay sensitivity and drug tolerance? Is there experience using one characterized PC as "a single platform positive control" for several molecules, e.g., addressing one mutation in the Fc part of the molecules? When is a difference with the use of different purified positive controls expected? What should be the focus of immunogenicity assay development (drug tolerance or sensitivity)? What are the selection criteria for ADA positive controls (binding properties or performance in the assay)?

Critical Reagent Assay Comparability

What is the industry experience with challenges of assay variability despite having process controls in place? What are the strategies for successive bridging of new lots (e.g., new lot vs current lot or new lot vs original reference lot)?

Novel Critical Reagent Modalities: "Thinking out of the Box"

Are there situations where animal-derived reagents are thought to be required (i.e., non-animal-derived antibodies would not be fit for purpose)? What are some best practices when setting up *in vitro* antibody generation campaigns?

Can the scientific community agree with supporting the limited use of animals for antibody generation (e.g., use of generic PC recognizing a specific epitope of platform-drug)? Does the scientific community believe in the flexibility to utilize animals for antibody generation on a case-by-case basis? What repercussions are there to the use of alternative immunoreagents for novel vaccine targets? What would be the appropriate control for an immunoassay designed to measure a desired polyclonal immune response?

DISCUSSIONS, CONSENSUS & CONCLUSIONS

Bioanalytical Challenges for Inhalation & Oral Delivery of Biologics

The common route of drug administration for biotherapeutics is subcutaneous or intravenous, largely due to poor oral bioavailability. Recent advancements in technologies enable novel delivery modes for biotherapeutics, for example, nasal, inhalation, oral, transdermal. Non-invasive delivery improves patient experience and could target tissue delivery for better efficacy and safety profiles. While these novel delivery approaches are desirable, there are unique bioanalytical challenges that require innovative approaches.

Oral delivery is well-accepted and convenient to the patient, which significantly improves patient experience in chronic diseases. However, since the biotherapeutic administered orally must pass through the gastro-intestinal (GI) tract, the encountered environment poses many delivery challenges (e.g., stomach acidity, bacterium, enzymes, and tight epithelium junction). Therefore, excipients are required to enhance oral bioavailability, which could complicate drug quantification. Excipient PK profiles may need to be characterized to select an appropriate dose, increasing demands for bioanalytical methods. A case study was discussed [97] where the potential for oral delivery of an oligonucleotide was evaluated due to its enhanced potency and live bioavailability even though the oligonucleotide was typically administered subcutaneously or intravenously. Bioanalytical methods for PK in plasma and tissue samples, as well as PD and ADA analysis, required development in up to six species, two+ tissues, three antisense oligonucleotides (ASOs), two PD analytes and one excipient. Low concentrations in plasma PK samples were overcome using dual probe hybridizations; however, the methods could not differentiate N-acetylgalactosamine (GalNAc) conjugation status. Although the LCMS method could differentiate GalNAc conjugation status, its sensitivity may not be adequate for ASO quantification in plasma. As such, an LCMS method had primarily been used for quantification of ASO in tissues where ASO concentration is expected to be higher. Tissue PK sample processing required extra caution to avoid blood contamination. The challenge for clinical bioanalysis is to relate plasma exposure to tissue exposure.

Inhaled delivery is an attractive non-invasive route of administration for biotherapeutics for respiratory diseases. It offers direct delivery to the target tissue, avoids the harsh conditions of GI tract and first-pass metabolism, and thus enhancing drug concentration in disease tissue while maintaining low systemic exposure and potentially leading to enhanced efficacy and reduced safety risks [128,129]. As low circulating drug level is anticipated, quantification of drug will require ultrasensitive detection. Although drug level in tissue provides a direct measure of effective exposure, accurate quantification of drug in tissue is challenging due to sample variability and blood contamination. Approaches for data normalization are necessary to ensure data quality.

Bioanalytical methods used to support studies using non-invasive delivery systems can be complex due to the challenges discussed above. Poor bioavailability and variable exposure can significantly increase the requirement for quantification sensitivity, resulting in the demand for new detection technologies. Samples, whether blood-based or tissues, must be managed to account for excipients, biotransformations and blood contamination in the tissue samples. Furthermore, depending on the animal model and matrix, interference can be a challenge, as well as the availability of the appropriate matrix for method validation. Small sample volumes from certain matrices can limit the ability to assay samples more than once.

Immunogenicity potential could be affected by the route of administration and its assessment strategy should be carefully evaluated when supporting novel delivery of drug administration [98]. Discussions during the 2021 WRIB indicated that inhaled delivery is perceived to result in higher immunogenicity in humans. While IgA is considered a potential isotype with inhaled biotherapeutics, it was agreed that there is no additional value in characterizing this isotype in the absence of a specific clinical concern. For vaccine and cell therapies, bioanalysts may want to consider the cytotoxic T-cell response.

Discussions focused largely on the bioanalysis of lung tissue. Consensus was reached that ADA determination in serum is sufficient for inhaled biotherapeutics and that evaluation of ADA in lung tissue is not required as the presence of total ADA in serum would be indicative of ADA in the lung. However, it was acknowledged that the onset of ADA in the lung may be different than the onset in circulation. PK evaluation in lung tissue presents

similar challenges; tissue collection can be difficult during a clinical study due to the invasiveness of sampling, and PK profiles in tissue cannot be assessed because repeat sampling is not possible. As such, PK profiles in serum remain the primary source of data, which may require technologies for high sensitivity quantification.

Regarding metabolites, it was recommended to use the MIST guidance [99] to determine if quantification of metabolites is needed. Consideration should be given to metabolite identification, in particular if the metabolites are active and with high prevalence, providing value in following the metabolite.

Many future opportunities were identified regarding the discussion of this subject at future WRIB meetings, including the discussion of other routes of administration (e.g., ocular, nasal), and sharing experiences using the newer ultra-sensitive platforms – utility, limitations (e.g., impact of ADA on assay readout), challenges, ways to ensure consistency in reagents, etc. The focus for future discussions should also include if IHC is more informative than traditional bioanalytical platforms for tissue homogenate and what the requirements for IHC should be.

Free, Bound, Total, Active, Monoactive, Biactive, & Multiactive PK Assays

Multi-domain and multi-specific refer to same class of biotherapeutics and they present added benefits over traditional therapeutics by binding multiple targets, thus offering better tissue specificity, increased efficacy, and potentially lower toxicities. Additionally, low dosages may be sufficient to achieve their biological effects. These advantages can translate into the further development of novel therapies for diseases that are currently untreatable [100,101].

Despite the added benefits that multi-specific therapeutics offer, bioanalytical support for these biotherapeutics has become much more complicated than more traditional protein biotherapeutics. Drug concentration measurements may be approached through the detection of different forms of the biotherapeutic molecule such as free, bound or total, in addition to the need to distinguish between the active form towards one or all of the therapeutic targets (i.e., monoactive, biactive). The decision about which form to measure should be driven by the appropriate stakeholders (pharmacokineticist, modelers, bioanalytical scientists) and should consider other available assays or data (e.g., PD, target measurement, target engagement assay, ADA, etc.). Certain assay designs present potential for interference from target or neutralizing antibodies, therefore it is important to carefully consider the choices of capture and detection reagents.

As per the newly issued guidance [102], effort should be put into the design of bioanalytical assays to ensure their implementation is suitable. Initially different formats may be evaluated, with designs changing over time depending on the availability of reagents or the stage of development. For example, domain-specific PK assays can be developed initially as investigational tools but only one PK assay would be validated as the primary method for regulatory study support. The discussions on assay format and design are often ongoing and can change over time based on reagent availability, design limitations, timelines and resources. Incorporating multiplexing where technically feasible may increase throughput, but it also adds bioanalytical complexity and, for some biotherapeutics, there may be challenges with sensitivity.

In addition to the challenges of implementing ultrasensitive PK assays to support multi-specifics with very low starting doses where relevant, soluble target interferences should also be assessed to evaluate the potential impact from one or all targets in a formed complex. In addition, higher baseline levels of circulating targets can be expected in certain disease indications. Detection of the actual target in any form (free or bound) may be complicated by the presence of the drug and/or the other target. Thus, it is necessary to determine if the target method is sufficiently sensitive, specific and drug tolerant. In case of target engagement or free drug assays, sample manipulation (e.g., sample dilution) should not negatively impact the assay readout [103]. The method must also allow detection in the absence and presence of drug; in some cases, different assays may be needed to measure the target in baseline samples versus post-dose samples, to complement a single PK method for study support.

Implementing Free/Total PK Assays in Regulated Bioanalysis

Many biotherapeutics are designed to bind and neutralize soluble targets. Even in cases where the target is bound to a cell membrane, there often exists soluble or shed forms. Thus, any sample taken post-dose is likely to contain a mixture of free drug plus drug bound to its target. Various assay formats can be designed to measure only free drug, only bound drug or both (total drug) [72]. At the WRIB, recommendations for the most relevant PK assay were first discussed in 2014 [9] in the context of PEGylated proteins based on Lee *et al.* paper [72], continued in 2016 and 2017 [15,18] and were last updated in 2020 [27]. Consensus remains that the PK assay should be selected depending on the MOA, availability of critical reagents and the intended use of the data.

The scientific understanding behind both free and total PK assays has evolved. Both assay types have their own limitations that need to be considered. For example, the equilibrium that exists in the original sample will eventually shift if the sample is heavily diluted. The use of assay reagents with high affinity or avidity may also impact the drug-target binding equilibrium. Additionally, assays using anti-human IgG detection can suffer from high background issues. These challenges could result in the inaccuracy of free drug determination in the PK assay. Conversely, total PK assays can be equally difficult. For example, total assays that use generic anti-human IgG reagents generally cannot be used for clinical samples. In addition, non-inhibitory anti-ids (i.e., anti-ids that don't compete with target for drug binding) could be difficult to generate and target-drug complex formats are prone to artifacts. LCMS assays, utilizing a generic capture (such as Protein A), have proven reliable in measuring total drug levels, but ultra-sensitivity is not always achievable with this platform. Hybrid LCMS assays may be used to improve sensitivity and/or analyze free drug concentrations [104].

Modeling and simulations may help predict the concentrations of free and total drug as it relates to dose, drug affinity and the target turnover rate. For membrane bound targets, circulating drug is mostly, if not exclusively, free and thus the choice of assay format is not critical. If the target is in circulation, the difference between total and free drug becomes smaller as the dose, half-life, Kd or target-mediated drug disposition (TMDD) increases, or the target concentration decreases. Modeling can also be used to predict the proportions of free and bound drug after dilution. Models have indicated that at lower doses, typical in first in human studies, there may be significant differences between total and free forms. At doses typical of exploratory toxicology-specific studies, total and free drugs are generally within 20% for most targets. Targets with a slow half-life have less difference in total/active compared to targets with fast turnover. Finally, targets with TMDD show virtually no difference between total/active, assuming there is no target accumulation. Therefore, these models can inform our recommendations about which assay format to use.

Assay conditions can also impact free versus total drug quantitation. The Law of Mass Action predicts that when a sample is diluted, equilibrium will favor a higher concentration of the free drug. High affinity assay capture reagents (e.g., anti-id and high density of the coating reagent) may produce an avidity effect favoring target–drug complex dissociation. Finally, the duration and temperature of the incubation steps (e.g., sample incubation on the plate) may also impact the stability of the target–drug complex.

It was concluded that the measurement of total or free drug is a discussion with the PK/modeling team, who typically want to understand the MOA. The free drug exposure is important for PK/PD analyses, and total drug exposure is important for evaluating safety. If the antigen/target level is much lower than the drug level, then free and total are essentially the same with the exception of the terminal phase when drug concentration decreases. If it is possible to confirm whether the assay measures free or total drug, the PK scientist can make PK/PD predictions based on that information. Consensus was reached that the 2020 recommendations for free/total PK assays are still valid: free assays are usually preferred if the target is soluble and especially if present in elevated concentrations (e.g., diseased individuals); however, total assays are usually preferred in early discovery [27]. A final consideration for assay type is that target or anti-id capture assays can be impacted by neutralizing ADA. In the presence of ADA, the assay may therefore only detect "active" drug.

When evaluating the appropriateness of the PK assay format i.e., free vs total, regulators confirmed that they look at the totality of evidence that is presented along with the justification for what is being measured and the rationale of the assay used.

Multi-Domain Biotherapeutic PK Assays in Regulated Bioanalysis

MDBs contain complex structures, multiple functional domains, and often multi-step pharmacological MOA. The multi-functional domain structure and dynamic nature of these complex biological entities require additional scientific considerations for selecting an appropriate bioanalytical strategy. Additionally, MDBs sometimes present unique situations during assay validation and sample analysis that may require additional scientific discussions.

Recommendations on MDB PK assays were initially issued in the 2016 White Paper in Bioanalysis [15] when these modalities were considered new generation biologics. It was concluded at that time that multiple assays could be used to analyze MDBs based on the availability of the appropriate critical reagents. A complexity-based approach to method development using knowledge of the MOA was recommended. Target biology, the needs of the pharmacokineticists and the affinity of each arm of the molecule should also be considered to inform method development. Finally, it was recommended that bioanalytical assays demonstrate intact target binding for the individual domains.

The 2016 recommendations [15] were clarified to allow more than one assay early in development to determine which assay is preferred as drug development progresses. However, consensus was reached that one or more PK assays can be used early in the development, but whenever possible, a single PK assay measuring intact molecule should be used as the primary method for regulated study support. Furthermore, measurement of the most relevant target binding domain shall be the main driver for bioanalytical strategies supporting MDB; the team should decide which target binding domain is most meaningful. Understanding target biology and solubility (at different levels systemically) is necessary for effective method development.

With the evolution of science and regulations, the recommendations were updated in the 2020 White Paper [27], in the context of bispecific monoclonal antibodies and bispecific T-cell redirectors. Consensus was reached that the confirmation of molecular integrity and assay specificity for all functional domains was needed.

There is no single bioanalytical strategy that works for all MDBs; rather, the strategy should be established on a case-by-case basis depending on the requirements of assay sensitivity and specificity, molecular structure, endogenous target biology, the PK or PK/PD-related questions that need to be addressed, availability of high-quality critical reagents, integrated data analysis and availability of the technical platforms. The typical approaches for PK assays are currently ligand binding assays.

In order to achieve the high sensitivity requirements of some MDB bioanalytical assays, bioanalysts are turning towards ultrasensitive technologies. These platforms can be used to increase sensitivity and improve drug tolerance (i.e., for target measurement assays), as well as theoretically reduce variability due to automation. However, there has been minimal experience with these methodologies for regulatory submissions, in part because of challenges encountered with long-term critical reagent support. For example, large scale labeling may be needed to avoid potential batch-to-batch variability of the protein-conjugated beads used in the SIMOA platform; however, stability indicators for these beads have yet to be established. Similarly, a single large order of custom cartridges used in the Ella platform may be necessary to avoid potential lot-to-lot variability during clinical development. It is anticipated that with more experience, improved vendor support, and better characterization of reagents, these platforms may be used successfully to support long-term regulated studies.

When determining a critical reagent strategy, many options present themselves, depending on the complexity of the MDB. Conventional reagents include recombinant targets, engineered Fc-specific antibodies, and anti-ids using hybridoma technology. In the early stage of drug development, commercial recombinant targets can be used, although they may not always be a dose representative of the physiological protein. Some alternative approaches to reagent generation include phage display or yeast display for anti-id generation (although there can be challenges with specificity and lower affinities), synthetic peptide immunization versus full-length immunization, and immunization using full MDB versus single-domain biotherapeutic (SDB), or immunization using MDB/SDB with murine Fc and screening using MDB/SDB with human Fc. For cell surface receptors or difficult to express targets, screen them in the assay format that would be used; it is preferred to keep targets in the "native" linkage. The draft ICH M10 guideline [66] outlines proposed requirements for critical reagent characterization, stability and documentation, which includes identity, purity, and concentration amongst other details. Typically, the characterization using surface plasmon resonance (SPR) methods, such as Biacore™, is adequate. While polyclonal antibody (pAb) reagents are not usually recommended for regulated PK assays, if that is what is available, they can be used as long as they are characterized and a long-term supply is available.

Bioanalytical Challenges to Study the Biodistribution of Biotherapeutics

The 2021 discussions on biodistribution of biotherapeutics focused on the several scenarios under which tissue biodistribution of protein therapeutics will be needed. Traditionally the *in vivo* PK/PD relationship for protein therapeutics in clinical studies is established using systemic drug exposure/concentration (PK) with one or multiple downstream PD markers or surrogate biomarkers. In the pre-clinical stage, the biodistribution of the lead molecules can be used to assist in candidate selection. For bispecific and multi-specific protein therapeutics in particular, establishment of the PK/PD relationship at the site of action becomes a needed measurement for accurately projecting an efficacious dose. Biodistribution can detect potential toxicological issues in the target tissues versus effects in the off-target tissues. In the clinical stage, drug accumulation in target tissues is of interest (e.g., tears, aqueous humour, skin, bone marrow, GI tissues), if sampling is feasible. It was agreed that all these requirements may ask for drug concentration measurements from the target tissues which can be both an operational and bioanalytical challenge.

Several factors may affect biodistribution of therapeutic antibodies. The first is the site of administration. Subcutaneous injections result in a more complex absorption of the molecule than a direct intravenous injection. Tissue level mAb dynamics can also play a role. Greater than 98% of mAbs enter tissues via convection, and several major organs play a role in catabolism and elimination (e.g., skin, muscle, liver, gut). mAb distribution is also dependent on physiological properties, like the rate and extent of tissue extravasation, the rate of distribution within a tissue, the rate and extent of tissue binding, and the rate of elimination. It is important to study the estimated antibody biodistribution coefficients derived from the time-independent relationship between plasma and tissue concentrations of mAbs. Target expression also governs mAb distribution: on-target off tumor, on-target on tumor, non-specific binding and internalization. Distribution in the tumor is based on tumor extravasation, the binding-site barrier and tumor draining lymph node-based clearance. Finally, receptor-mediated transcytosis (RMT) receptors can aid in overcoming the blood–brain barrier (BBB).

Further focus was on several types of analytical platforms to study biodistribution, separated into two approaches: Terminal readout approaches and Continuous readout approaches. Conventional LCMS and LBA PK analysis from tissue homogenates are still the standard methodologies due to their quantitative measurement while multiplexed immunofluorescence (mIF), imaging mass cytometry (IMC), and digital special profiling (DSP) by Nanostring are all examples of methods that fall into the Terminal readout category applicable for obtaining spatial/subcellular details and in depth immunophenotyping. These methods are well established in the industry; IMC and DSP are for quantitative analysis and mIF is only for semi-quantitative analysis. Consensus was that the main challenge with all these methods is the sampling bias. mIF also has limited plex ability and unknown sensitivity, but high throughput. IMC methods have a large selection of metals allowing for the labeling of different doses with different isotopes, but assay development is needed to mitigate the loss of surface antibodies during sample preparation. DSP is less used than IMC possibly due to the fact that it has lower plex options and some samples may have high background noise. However, it is applicable to both proteins and transcripts. Continuous readout approaches include positron emission tomography (PET) imaging and large pore microdialysis. PET imaging is a highly sensitive method that can be used to quantitatively examine overall tissue distribution using antibodies labeled with radioactive isotopes. The isotopes can substitute for naturally occurring atoms, allowing for targeted analysis in antibody and cellular therapy studies. It was mentioned that some disadvantages of this method that may impact PK profiles include low spatial resolution, half-life decay of the isotypes, and the need for specialized equipment (PET cyclotron or generator). Large pore microdialysis involves implanting probes into specific regions of the tissues to collect real-time fluid (typically mucosal fluid and CSF) which provides a direct quantitative measurement in tissue interstitial fluid. However, this method is invasive, requiring a highly technical surgical procedure [105,106].

Case studies were used to illustrate and discuss some of these concepts and their applications at different stages of drug discovery and development. The first discussed biodistribution studies to enable lead selection for a BBB enhanced delivery platform with terminal readout. To accurately measure drug concentration in brain tissue, the development and optimization of tissue processing to isolate and analyze drug concentrations in capillary-depleted brain tissue [107] in combination with in-life perfusion of animals was required. Custom hemoglobin-specific assays in human and cynomolgus monkey whole blood were then used to assess the extent of blood contamination in relevant BBB matrices (CSF, ISF, brain tissue, etc.). Results showed increased accumulation in the brain tissues from the dosed human target expressing transgenic mice with the candidates which had the same Aβ affinity but different affinities to the RMT receptor with differential accumulation in the brain tissues. When studying the lead selection in these mice, they demonstrated different, increased level of accumulation with a combination of effect from different affinities to both Aβ and the RMT receptor. With all this information, the lead was then selected and verified in cynomolgus monkey brain tissue versus serum demonstrating the translatability from mouse model to NHP.

A second case study was discussed to understand mAb therapeutic accumulation in tumors with continuous and terminal readout. Key factors in TMDD are the target expression pattern/density and the biology of the antibody (affinity, avidity, size, half-life). A typical mouse PK/PD oncology study design to assess the impact of the biology of the antibody begins with generating the mAb tools. Some drug formats include Fab, (Fab)2, non-targeting mAbs (negative control), monovalent or divalent mAbs and biparatropic mAbs. Next the mAbs were characterized *in vitro* using SPR to confirm their cell binding and intrinsic properties. Finally, *in vivo* studies in mice occur, resulting in tissue PK samples used to determine "effective" tumor concentrations and concentrations in a non-targeted tissue, tissue IHC to determine spatial distribution of therapeutics relative to blood circulation and longitudinal imaging to determine the spatial extent and mAb penetration in order to select the right format of the therapeutic [108–110].

Table 1. Evaluation of different levels of DoL and aggregation in three types of commonly used PK LBAs.

Platform	Capture		Detection	
	DoL	Aggregation	DoL	Aggregation
Chemiluminescent ELISA	NA	NA	↑ = better sensitivity	↑ = loss in sensitivity, decrease in assay range
MSD-ECL	↑ = loss of sensitivity	No effect	Optimal concentration found	Loss of sensitivity and decrease in assay range
Gyrolab®	↑ = decrease in assay range	No effect	↑ = loss of sensitivity and decrease in assay range	No effect

Overall, there was consensus that biodistribution studies are critical in drug discovery and development for PK/PD based lead selection, understanding MOAs, first-in-human dose justification/efficacious dose projection, differentiation/competitiveness assessments, and for diagnosis and precision medicine. However, it was widely recognized that there are still many challenges and gaps in industry knowledge, e.g., translatability from nonclinical studies to clinical studies during lead selection, best imaging options to assess tumor penetration, data alignment across different methods on effective tissue drug concentration estimation, tumor binding site barrier, the impact of avidity and mAb formats on tumor uptake, and the assessment of tumor target engagement in tissue (PET/tissue digestion). Due to current lack of specific regulatory guidelines for mAb therapeutic distribution, it was suggested that the recent ICH Guideline S12 Nonclinical Biodistribution Considerations for Gene Therapy Products may be consulted [111].

Some of the major bioanalytical challenges were discussed. Tissues were used by the bioanalysts primarily to support preclinical distribution studies and were not typically used in clinical studies due to the challenges of sampling in humans. Tissues are always a challenge for LBA PK assays because representative QC samples cannot be prepared due to the difficulty of "spiking" drug into a tissue. It was noted that some use plasma or serum QCs to monitor performance of tissue assays. Additionally, variability can be observed between samples given the inherent heterogeneity of some tissues (e.g., tumors). Finally, interference from blood contamination and interstitial spaces add complexity to the use of tissues in bioanalysis. In the discovery stage, a more robust and easily accessible imaging platform is needed in order to collect data faster for decision making; the mass cytometer was recommended as a promising technology.

It was recommended that surrogate animal tissues may be used in clinical assay development, and to assess sample stability during homogenization and in tissue lysates. Bridging to human tissues can be performed if sufficient biopsy volume can be collected, such as with bone marrow, skin tissues, etc. The minimum tissue sample amount required for homogenization is recommended to be at least 50 mg. In most cases, partial validation of rare tissue matrices is sufficient if the same assay was fully validated for serum or plasma PK samples [29].

Advanced Approaches in Critical Reagent Selection for PK Assays

LBAs are typically used for PK determinations in support of large molecule biotherapeutic programs. These assays require high-quality, well characterized critical reagents for successful method development. Recommendations for the use of critical reagents in PK LBA were provided in the 2019 and 2020 White Papers in Bioanalysis [25,27], and updated based on new case studies in 2021.

In these assays, antibodies are often utilized as the capture and detection reagents. To facilitate this function, the antibodies may be conjugated with a chemical moiety (tag). This then allows the antibody to either be coated onto the plate as the capture reagent or used for signal generation as the detection reagent. Antibody conjugation is often performed using amine-reactive crosslinking chemistry, where N-Hydroxysuccinimide (NHS ester) on the tag reacts with the primary amines of the antibody. Once labeled, the average conjugate number per antibody is determined; this is referred to as the DoL. The DoL of the tag on the antibody can influence its stability since binding alters the physical and chemical properties of both antibody and conjugate. Additionally, the propensity of unconjugated antibody to aggregate can either be compounded or reduced during labeling and can affect reagent stability over time. Therefore, understanding the effects that different levels of aggregation and DoL may have on different PK LBAs is of vital importance to ensuring high PK LBA performance. The evaluation of different levels of DoL and aggregation in three types of commonly used PK LBAs (MSD-electrochemiluminescent [ECL], Chemiluminescent ELISA, and Gyrolab®) were discussed. The characteristics of each platform examined included an effect on binding, assay sensitivity, and assay range; these are summarized in Table 1. The data demonstrated the

importance of maintaining a consistent DoL and monitoring of aggregation when producing new lots of reagents to ensure assay reproducibility.

It was concluded that when designing LBAs, it is important to consider the assay platform, the capture and/or detection reagents as well as the desired conjugation and associated DoL for the particular platform, as this could impact the desired assay characteristics. This information can allow for adjustments of the DoL to meet assay requirements. Mass spectrometry (MS) is considered the most accurate method to test DoL; HRMS can be used for intact mass analysis and for labeling analysis, therefore this technique can also be a good option, however the use of MS is not mandatory. DOL measuring kits can also be used, provided the same kits are used consistently. However, it was noted that kits may not be sensitive enough to detect specific labeling details, which can manifest as assay performance issues. Kits will provide the overall average DoL, while HRMS will provide information regarding the distribution of labeling (% one tag, two tags etc.). Combined use of both kits and MS can provide contrasting results, so this approach is not recommended. Assay performance is the most important measure of suitability, and the necessary accuracy level should be established based on need.

Well-characterized critical reagents are essential components for the successful validation of high-quality PK and immunogenicity LBA, therefore it is important to control the reagent generation process to minimize bioanalytical variability during the drug development life cycle, avoiding costly problems and delays in the clinic. Reaction conditions such as antibody, label, concentration, time, temperature, pH, and molar ratio can impact reagent quality. Despite the use of similar preparation protocols, results can differ in stability, SNR and low and high concentration controls.

A systematic approach to developing reaction, purification, and formulation conditions was discussed for three different types of critical reagents, including antibodies of low and high isoelectric points as well as a 3 kDa peptide molecule. A platform method for the preparation of biotinylated or ruthenylated antibodies is not advised. The performance of the conjugation reaction of the antibodies varied by the base antibody and the type of label used. Additionally, the selection of formulation buffers conferring the greatest stability for the labeled antibodies could be predicted by *in silico* modeling and confirmed by analyzing isoelectric shifts generated by the label incorporation. Platform methods were also found not to be appropriate to generate critical reagents using a 3 kDa peptide base molecule. Successful production of biotinylated and ruthenylated peptide reagents required reactions of increased alkalinity, time, and label ratio. Additionally, the purification of these reactions required novel separation techniques prior to formulation in a buffer informed by the properties of the reagents. The development approaches and findings shown in the case studies have been applied to the generation of other conjugated critical reagents.

The selection of a mAb reagent for a specific purpose can be driven by a unique property such as specificity, neutralizing activity, or affinity to the intended drug or target. Guidance documents in the literature provide insight to method refinement to meet the requirements of the LBA [25,27]. Scientists often evaluate a panel of mAb reagents to identify the one that provides the optimal result, however, this orthogonal approach can be time-consuming and an inefficient use of resources. Furthermore, additional challenges are introduced with the employment of solid-phase extraction or depletion steps using acid dissociation or conjugated beads. Therefore, identification and understanding of key attributes such as the kinetic parameters of mAb reagents and how they contribute to LBA performance can provide strategic advantages to improved LBA performance. A retrospective analysis of mAb reagents selected for specific use in LBA was discussed, outlining the process by which they were catalogued by the kinetic properties of apparent K_D, on-rate, and off-rate. The goal was to look for trends that align the function of the reagent with the kinetic data to offer recommendations on how to streamline the selection of a mAb reagent with fit-for-purpose characteristics. Kinetic data obtained from the Octet were cross-referenced with qualitative assessments of anti-id performance for four types of reagents: NAb PC, ADA PC, PK capture reagent, PK detection reagent. The conclusions are presented in Table 2.

When applying these conclusions to a case where a new NAb PC was needed, the recommendations aligned with the findings in Table 2. This confirms that predicting the appropriate kinetic parameters needed for these reagent types will enable bioanalysts to design the suitable screening strategy upfront for low-, mid- or high-affinity antibodies, choose the appropriate panel of clones to move forward with based on kinetic properties and other characterization parameters and reduce the need for repeat campaigns for mAb reagents, thereby saving time, resources and money. A single parameter should not be used to select a reagent, but an integrated approach including k_{on} and/or k_{off} rates, signal to noise ratios, etc. is recommended. Further discussions highlighted that caution should be used when assessing KD values since each technology has its own drawbacks that can affect KD. The experience of the attendees indicated that there were cases where octet pairings did not match with

Table 2. Correlation between reagent kinetic attributes and assay performance.

Reagent	Parameter	Recommendation	Comments
NAb PC	Affinity, K_D	10 pM–5 nM range	mAbs with single-digit pM may be difficult to acid dissociate
	K_{on} (1/Ms)	10^5–10^7	NA
	K_{off} (1/s)	10^3–10^5	mAbs in the 10^6–10^7 did not perform well in LBA
ADA PC	Affinity, K_D	10 pM–1 nM range	• Recommend mid-level affinity PC, in the high pM to mid nM range, if possible • mAbs with single-digit pM may be difficult to acid dissociate
	K_{on} (1/Ms)	10^6–10^7	NA
	K_{off} (1/s)	10^3–10^5	NA
PK capture	Affinity, K_D	1 pM–50 pM range	High affinity Ab
	K_{on} (1/Ms)	10^5–10^7	• No obvious correlation • Possibly due to relatively long LBA incubation times (ELISA, MSD) • 1 Gyrolab® example with faster on-rate (fast contact time) of 10^6
	K_{off} (1/s)	10^5–10^7 (slow)	• Some faster off-rates were acceptable, possibly due to relatively high capture concentrations on the plate which offset the effect of a slow off-rate • A slow off-rate can often be compensated with a higher antibody concentration or a longer incubation time, whereas an antibody with a fast off-rate is washed off quickly
PK detection	Affinity, K_D	1 pM–1 nM range	• No obvious correlation • Most detectors worked well in the range obtained
	K_{on} (1/Ms)	10^5–10^6	• No obvious on-rate correlation for PK assay detector reagent performance • Possibly due to relatively long LBA incubation times, most of these assays are ELISA or MSD, except for Gyrolab® LBA with fast on-rate more desirable (quick contact time) • Although only 1 example available, in general, Gyrolab® assays should use higher affinity antibodies (fast on-rate in particular) due to short contact time
	K_{off} (1/s)	10^4–10^7	• No obvious off-rate correlation for PK assay detector reagent performance, however, off-rate range slow in general (10^4–10^7)

immunoassay results; this can be due to an erroneous K_D value. Therefore, it was concluded that assay format can also play a role, and the intended use of the reagent should inform the affinity.

Regulatory guidelines regarding critical reagents continue to evolve, and increased scrutiny must be placed on critical reagent characterization, evaluation, and selection for use in clinical PK assays. Not all clinical PK assays need to be sensitive (<20 ng/ml). Depending on the planned doses and dosing routes, sensitive clinical PK methods are needed when compounds have limited safety margins and are administered at low doses (e.g., T-cell redirected therapies) or for alternative dosing routes that result in low systemic exposure (e.g., intravitreal administration). A strategy was discussed that included the use of automated high throughput screening, assessment of binding specificity and potency, and a multiple platform approach for the characterization and final selection of the critical reagents that are used in the development of sensitive clinical PK methods needed to support ocular and low dose compounds. Two case studies exploring the use of automation for early critical reagent pair identification were discussed, Octet and ELISA assessment of reagent binding specificity, Biacore for binding potency assessment, and multiple platforms (including SMCxPro, Gyrolab® and MSD) for method development and final critical reagent selection.

In the first case, MSD PK pairing data was used for direct comparison in method development using sensitive platforms to identify 5 out of 8 mouse anti-ids for use as reagents for a humanized IgG1 mAb therapeutic. Once the 5 anti-ids were identified, they were analyzed using BLI analysis of K_{off} to the mAb. In parallel, the specificity of the 5 anti-ids was assessed using a competitive LBA. Three different platforms were used for development of the assays: MSD, Gyrolab®, and SMCxPRO. The results of the binding kinetic analyses suggest that anti-ids with slower K_{off} rates are helpful in the development of more sensitive clinical PK assays. When comparing the platform results, it was observed that the PK pairing relationship was similar across all three platforms, but assay sensitivity was different (SMCxPRO > Gyrolab® > MSD). The best pairs had similar signals in matrix and buffer blanks, which indicated that minimal matrix interference was observed in the assay, a key to passing selectivity in sensitive clinical PK methods. When the manual screening process was compared to the high throughput system used, advantages were seen in the number of plates, pairs and replicates assayed per day (3 plates with 72 pairs in singlets versus 16 plates with 96 pairs in duplicate). The run time for the automated system was slightly longer (4.5 hours versus 3.5 hours) but the data analysis time for the automated system was shorter (20 min per run instead of 1 hr per run for a manual assay). Finally, the analyst performing the manual assay must be active throughout the entire

Table 3. Advantages and Disadvantages for PC Options.

	Advantages	Disadvantages
mAb	• Can obtain a panel of anti-id • Potential to select PCs to assess NAb and binding antibodies • Relatively easy to purify and resupply individual mAbs	• Single mAb as PC will only represent a single epitope with a single affinity • mAb technologies labor intensive
pAb	• Specific for multiple epitopes and potential for a range of affinities • Rabbits typically mount robust antibody responses with a few boosts	• Finite supply; need to pool and store bleeds • Affinity purification may result in loss of low and high antibodies
Antibody display technologies (e.g., phage yeast)	• Can screen for specificity of interest • Can modify affinities by engineering	• Affinities may be low if there is no affinity maturation step implemented • Can exhibit nonspecific binding • Requires antibody engineering expertise
Human purified ADA	• Clinically relevant immunogenic epitopes, binding kinetics and isotypes/subclasses	• Need a large pool of donors • Will not be available for early trials • Requires informed consent, scientific rigor and expertise

Table 4. Possible differences between animal-derived PCs and human purified ADA.

Features	Animal-Derived		Human Purified ADA
	mAb	pAb	
Isotypes	Single IgG subclass	IgG isotypes	IgM, IgG144, IgA, IgE
Affinity	Single affinity	A range of affinities after hyperimmunization	Ranges in affinities which may change with time over the course of the immune response
Epitope	Single	Multiple	Multiple
Epitope specificity	Non-conserved human regions in CDR/FW/Fc, non-native structures	Non-conserved human regions in CDR/FW/Fc, non-native structures	Variable regions (residual mouse or amino acid modifications), non-native structures

run, whereas analyst participation using the high throughout automated method was intermittent, allowing for a more efficient use of human resources.

In the second case discussed, the high throughput automated system was used to test the 441 potential PK pairs quickly and efficiently, allowing for the selection of seven anti-ids for scale-up. SPR was used for the binding kinetic and specificity assessments of the seven anti-ids. This case showed that there is an evolution in the process for the assessment of reagent binding specificity and binding potency when establishing blocking, partial blocking and non-blocking reagents. Furthermore, higher throughput and sensitivity with SPR is advantageous, since SPR can be refined for kinetics and specificity assessments.

It is not practical to characterize all reagent candidates prior to use, so reagent candidate selection based on preliminary tests is recommended. Characterization can then be more efficiently performed on a smaller number of reagents (purified, conjugates) to assess biophysical and functional properties. Important acceptance criteria for bridging a new lot of reagent in an LBA will be defined by assay performance.

Challenges with Positive Control Generation for ADA Assays

During biotherapeutic drug development, assessment of immunogenicity via the measurement of ADA is required to evaluate its potential impact on drug safety and efficacy. Detection of clinical ADA responses relies heavily on the ADA assay method employed, and such methods are validated based on regulatory guidance available [112,113]. An important feature of the ADA assay is the selection of an appropriate ADA positive control (PC). Building on past recommendations [21,25,27], several PC-related topics were discussed at the 2021 Workshop in order to harmonize industry experience.

PCs can be mAbs or pAbs and are typically an anti-id antibody or a group of ADAs generated in an animal species that bind to epitopes in the variable region of a biotherapeutic antibody, or an antibody that recognizes an epitope in a protein replacement biotherapeutic, such as a growth factor or cytokine. PCs can also be human purified ADAs or derived using antibody display technologies (e.g., phage, yeast). The PC for ADA assays serves as a surrogate for patient ADA to ensure that assay specificity, sensitivity and drug tolerance meet the study's needs [25]. Table 3 outlines the advantages and disadvantages of the different PC options, while Table 4 outlines the possible differences between animal-derived PCs and human purified ADA.

There are many considerations when selecting a PC. Polyclonal PCs are considered more representative of clinical samples. However, batch-to-batch variation during polyclonal PC generation may impact assay performance. To support long-term clinical development and life-cycle management, mAb PCs would be a preferred choice because of their lot-to-lot consistency. It should be noted that while results for certain parameters such as sensitivity and/or drug tolerance of an immunogenicity assay may vary with different PCs, the assay itself does not change; the results obtained for those validation parameters are solely dependent on the choice of the PC. Use of high affinity antibodies as PCs may result in methods that appear overly sensitive (<10 ng/ml) and do not have the desired drug tolerance while use of a different antibody with lower affinity as the PC may generate adequate sensitivity and drug tolerance. It is important to choose a positive control antibody that is most representative of the potential clinical ADAs rather than one that makes the assay appear acceptable for clinical use. To overcome these challenges, it is usually helpful to test a diverse panel of PCs (mAbs and/or affinity purified, cross absorbed pAbs) with different binding affinities under multiple acid treatment conditions and no acid treatment conditions. Both pAbs and mAbs have been used as PCs during method development and validation. It is recommended to use both mAb and pAb PCs to characterize assay sensitivity and drug tolerance during method validation, and use mAb PCs to monitor assay performance during study support. In biosimilar development, polyclonal PCs derived from immunizing with either the biosimilar or the reference product are both considered adequate given the high analytical similarity between the biosimilar and its reference.

Generation and screening processes for mAb PC selection begin with the immunization of the animal with the drug in order to generate single cell anti-Id secreting clones. If binding to the drug is observed, the cells are further screened for unique binders that specifically bind the variable region of the biotherapeutic antibody through counter-screening against the Fc and irrelevant control. The desired clones are then scaled up and purified by protein A/G. pAb generation also begins by immunizing the animal, followed by checking the titers and pooling the positive samples. The pools are purified chromatographically in steps (protein A/G > affinity > cross absorption against human Fc and/or other animal species). pAbs and/or mAbs are then characterized and selected for use as PCs by determining the binding kinetics and blocking activities (towards target binding) by using SPR or BLI. The broad panel of PCs (and domains, if there are multiple domains) is then tested in the ADA and NAb assays to characterize performing characteristics.

During the generation of drug-related surrogate PCs, the negative depletion step during PC purification is critical to ensure that the antibodies that bind to common backbones are removed and only drug specific PCs are enriched. An improved PC generation process that overcame purification challenges and made the process more efficient in terms of time-saving and PC quality was described whereby a two-step purification for a clinical ADA PC was used. The antiserum from rabbit containing pAb underwent negative depletion followed by affinity purification. Using human IgG mix to replace subclass-specific mAb in the negative depletion step proved to be time saving (shortened by at least ~3 months) and cost efficient (saved large amounts of subclass-specific mAbs), and provided PCs with good quality (specificity and purity).

Reagents such as dye-, biotin- or other small molecule-conjugated antibodies are also essential components of ADA assays. These reagents' unique characteristics directly impact assay performance. Therefore, effective management of the critical reagents is crucial to ensure not only consistent assay performance throughout the life cycle of a method, but also timely progression of projects. For example, biotin–drug conjugates and digoxigenin (DIG)–drug conjugates are often used as the capture and detection reagents, respectively, in ADA ELISA assays. Case studies using these conjugates were discussed to illustrate the various challenges encountered with the critical reagents used in LBA based ADA assays.

The first case study illustrated the importance of having accurate information about the source materials used for reagent production. Several months after the production of a first batch of biotinylated conjugate from an anti-id mAb, a second batch of source material was provided for biotinylation. The requester believed that the second lot of source material was the same as the first lot. However, the conjugate made from the second lot had a significantly lower DoL (3.7) than the first conjugate (5.1). LCMS intact mass analysis of both source materials revealed that these materials were not exactly the same; two different dominant peak groups were detected in separate MW ranges. Investigation indicated that there had been communication gaps in the chain of custody. The first batch of material was from a transient transfection of HEK293 cells while the second batch was produced from a Chinese hamster ovary cell stable cell line.

A second case study was discussed where LCMS proved to be a valuable tool for assay troubleshooting. Use of a new DIG-conjugate lot in an ADA assay resulted in a significant drop in ELISA optical density signal, causing an

assay failure. LCMS was used to assess the conjugation ratios of the old and new lots of DIG-conjugate that were used in the ADA assay; a different incorporation ratio of the new DIG-conjugate was the main reason for the drop of assay signal. Subsequent efforts in preparing another lot of DIG-conjugate that had a comparable incorporation ratio with the original lot restored the assay signal.

Another case study demonstrated the impact of formulation buffer composition on conjugation efficiency. In this study, unusually low conjugation ratios of biotin- and DIG-conjugate reagents was observed during the development of a human ADA assay for Drug A, which had been formulated in a buffer containing arginine succinate. It was determined that, with increasing arginine succinate concentrations, the biotin conjugation efficiency for a different drug (Drug B) having a similar structure as Drug A also decreased, which confirmed the negative effect of arginine succinate in conjugation was not a molecule-specific phenomenon. It was hypothesized that the primary amine group on arginine succinate contributed to low conjugation efficiencies by competing with the primary amine groups on the antibody. According to further tests performed, as the sample volume load was lowered to decrease the concentration of arginine succinate during buffer exchange to phosphate buffered saline prior to biotinylation, a decrease in arginine succinate interference was observed for both Drug A and Drug B, which resulted in higher biotin incorporation ratios for both drugs. By applying this data to real world practice, reducing the sample volume load during the pre-conjugation buffer exchange significantly improved conjugation efficiency for both the biotin and DIG-conjugates.

In a final case study, a significant nonspecific binding signal in an ADA assay was linked to high molecular weight (HMW) aggregate content (3.7 MDa) by size exclusion chromatography-multi angle light scattering (SEC-MALS) analysis. These data indicated even a very low level (0.06%) of HMW aggregates in a conjugate could impact assay performance. Filtration of the conjugate removed the aggregates and reduced the nonspecific binding signal in the assay.

As shown above, biophysical characterization of reagents using key QC methods (e.g., LCMS and SEC-MALS) enables the production of multiple lots of material that have consistent analytical characteristics and reproducible assay performance. In addition, effective reagent management includes streamlined processes and systems for reagent request, production, characterization, and reporting. These infrastructures enable the seamless and timely supplies of high-quality reagents as well as better customer support and resource management. Additional efforts to better understand reagent long-term stability and relationships between assay performance and biophysical characterization data are currently underway for better future reagent management.

As a future consideration, there may be potential utilities for purified ADA from clinical trial samples. Pooled ADA from a large donor set can be used to develop a "quantitative" ADA assay to serve as a common reference standard across labs and to set thresholds for impact on efficacy and PK. Reference standards can improve the understanding of how ADA methods developed with animal PCs detect clinically meaningful ADA and/or identify the relevance of PC and "true" assay drug tolerance. Finally, reference standards can be used to identify and characterize immunogenic epitopes supporting predictive tools used in immunogenicity de-risking.

Critical Reagent Assay Comparability

In trying to maintain the continuous performance of an assay, routine equipment performance assessment, fidelity of assay execution, implementation of automated liquid handling and understanding the interrelationship of each assay component are aspects of an assay that can be controlled to provide faster throughput and fewer assay failures. However, they may not provide any additional control of day-to-day assay variation.

Consistency with the performance of critical reagents forms the foundation of maintaining accurate and precise sample analysis throughout the lifespan of its use. If the production, storage and use of these reagents are not closely controlled, then the assay will generate a wide range of variations or even fail. Variability, within acceptance criteria, can exist within and across lots of critical reagents with patterns of performance sometimes aligned with certain periods of time or particular studies. Criteria outlined by health authorities through their guidance [111–113] provide boundaries for assay acceptance, which can be used to demonstrate control of assay results reflected in lot-to-lot bridging assays throughout the years of assay use.

Although standardized testing results of production and labeling quality are comparable across lots, there may still be a need to alter the concentration of the critical reagents applied to the assay to maintain acceptable performance. Outcomes of small variabilities seen in performance during the lifespan of an assay have not appeared to result in any impactful differences in sample accuracy or precision in determining PK values or immunogenicity status.

Prior recommendations of critical reagent comparability were first provided in the 2018 White Paper in Bioanalysis [21]. These initial recommendations covered the need for SOPs addressing preparation qualification, labeling and storage of critical reagents, characterization strategy and documentation, electronic tracking, and the need to address reagents prone to aggregation. In 2019, additional recommendations were provided [25]. Highlights included addressing the concepts of expiry and retest dates, long-term supply of reagents, and performance assessments recommended for ADA assays. These were reconfirmed in the 2020 White Paper [27], along with lot-to-lot bridging recommendations. The last topic was expanded in 2021 to provide new inputs to update previous recommendations.

Critical reagent management is one important component that should be monitored and controlled. However, day-to-day assay variation is unavoidable. Two case studies were discussed to illustrate this point. The first case was a low ADA consistency control for a screening assay to assess ADA response to a mAb drug across multiple studies. Aliquots of the controls were used within the validated stability and the same lot of reagent was used across all studies. Variation was observed within studies from run to run (plate to plate) but no trend of performance was observed during the 3 years evaluated. Study ADA incidence did not correlate with the variation in the control, although some variation appeared to be associated with the analysts within studies, but all within acceptance criteria. The average normalized value of controls showed similar patterns of performance between screening and specificity runs that appeared to be time dependent in terms of performance. Low variation of average values compared to acceptance criteria reflected upon control of the process. Precision variability was observed, but it was not time-dependent with no trend of increase. The second case was a low ADA positive control for the screening assay used for 8 studies over 4 indications. Aliquots of controls were used within the validated stability. Variation was observed within studies from run to run (plate to plate), and during the 7 years evaluated, no trend relating to performance was observed. Variation was within the validated acceptance criteria, and previous results of variation in the screening assay did not dictate the amount of potential ADA positives, i.e., number of runs in specificity assay per study. Changes in the ADA control lots between studies did not affect the assay performance. There were no changes of lot within a study. Precision variability was observed, but it was not specific to either lot used. Average normalized values varied little over time. Thus, although it is important to monitor critical reagents and assay performance over time, some variation is expected.

Novel Critical Reagent Modalities: "Thinking out of the Box"

Last year, the EU Reference Laboratory for alternatives to animal testing (EURL ECVAM) issued a report that presented specific recommendations on the use of non-animal-derived antibodies [114]. The EURL specifically stated that based on the ESAC opinion on the scientific validity of replacement for animal-derived antibodies, the EURL ECVAM recommends that animals should no longer be used for the development and production of antibodies for research, regulatory, diagnostic, and therapeutic applications. The report goes on to request that provisions of Directive 2010/63/EU should be respected [115], and EU countries should no longer authorize the development and production of antibodies through animal immunization, where robust, legitimate scientific justification is lacking. Although the EU ECVAM recommendations are specifically stated for EU countries, these still may impact global sponsors filing in the EU.

Antibody-based reagents are broadly utilized in a range of analytical applications that are critical for the drug discovery and development processes, such as methods designed for detection of drug in circulation (PK), immune response detection and immunohistochemistry assays designed to further understand the mechanism of action of a therapeutic *in vitro* [116]. Although many of these applications currently rely on access to high-quality, animal-derived monoclonal antibodies, bioanalysts have been attempting to use non-animal-derived critical reagents for many years. Recommendations on the use and management of all types of critical reagents were provided in both the 2017 and 2019 White Papers in Bioanalysis [17,25]. In 2021, the discussions continued, and new experiences in the implementation of LBAs were discussed.

Historically, animal-derived antibodies have been the preferred format for antibody generation since they are less labor intensive than *in vitro* generation and have high affinity and specificity. Animal-derived antibodies have already undergone affinity maturation *in vivo* and do not typically require additional lead optimization. However, there is already strong support for the reduction of animal use in the drug development process, using the principles of the 3R (Replacement, Reduction, Refinement). Animal production of antibodies typically includes the collection of B cells from hyperimmunized animals, directly harvesting immunoglobulin genes for the production of recombinant antibodies or fusing B cells with myeloma cells to form hybridomas. The use of hybridomas is one of the most

traditional methods to generate high-affinity mAbs at an affordable price [117]. Despite their popularity, these types of antibodies can take time to generate, and it is not possible to define the specificity of the reagent during the generation process. Furthermore, the use of animals needs to be justified. Finally, there is the potential for lot-to-lot differences with resupply via a hybridoma. It was suggested that, when possible, animal-derived antibodies should be sequenced so that it can be made recombinantly when more reagent is needed.

Fully synthetic libraries also exist and offer an alternative approach to mAb generation, although additional rounds of affinity maturation may be required. Other protein scaffolds with high binding affinity are available, offering the advantage of diverse material, reduced constraints imposed by antigenicity, and the ability to select for novel additional functions [117]. Using non-animal-derived critical reagents adds important capabilities and benefits to the antibody production toolbox. Some advantages include increased stability, easy optimization/functionalization, cost-effective resupply, and animal-free generation, allowing for compliance to the recommendation by the 2020EURL ECVAM. Novel reagent modalities can include those created by display library-based approaches (phage, yeast, mRNA), affibodies, affimers, virus-like particles (VLPs), peptides and aptamers.

Display-based antibody production begins with antigen immobilization and 233 rounds of selection and panning. The resulting fragments are then subcloned into antibody expression vectors, screened, sequenced, and purified [118]. The advantage with phage display is that specificity is determined on the front-end of the process with the selection of the contributing fragments, as opposed to specificity screening at the back end of *in vivo* generation. This allows for the negative selection on human serum, isotype control mAb or other closely related proteins. Recombinant fragments are also easier and quicker to generate from bacteria than mammalian cell cultures. Finally, the gene sequence is known, allowing for genetic engineering. However, antibody display is limited by the size of the library, a greater number of low affinity binders and the fact that the heavy- and light-chain pairing may not reflect that of the *in vivo* immunoglobulin. Display-based methods require additional affinity maturation and protein engineering steps, potentially extending time required to produce high quality material [119,120].

Among display-based antibody generation, antibody phage display is a versatile, *in vitro* selection technology that can be utilized to discover high-affinity mAbs specific to a wide variety of antigens. Potential concerns with this technology are reduced solubility, viscosity, expression yield, and stability leading to impacted manufacturability and mAb potency. Risks can be associated with self-interaction and polyreactivity, as well as increased content of aliphatic residues in complementarity-determining regions (CDRs) when compared to the non-phage-derived antibodies. Phage display generated antibodies are produced in *Escherichia coli* and therefore lack glycosylation [121–123].

Affibodies are alternative binding proteins with a scaffold based on a Protein A IgG-binding domain and a triple helix bundle 3D structure. They are smaller (6kDa) than mAbs and are stable at extreme pH, elevated temperatures, and in intracellular conditions. Affibodies can be easily conjugated due to a unique C-terminal cysteine. Experience with the use of affibodies has demonstrated that an anti-IgG affibody capture can perform similarly to a mAb. However, it can cross react with monkey, human and mouse Ig, necessitating method optimization, which adds time and resources to the process.

Affimers are alternative binding proteins with two structurally related scaffolds: 1) Stefin A, engineered from human stefin A protein; and 2) Adhiron, synthetically created from a cystatin sequence. They have low molecular weights, which helps to avoid the steric hindrance issues associated with Abs. Suitable for a wide range of targets, they can be easily modified and are capable of withstanding extreme analytical conditions and multiple freeze-thaw cycles. Finally, they are easy to resupply because there is negligible lot-to-lot variability. However, the utility of affimers is limited as the sole company that produces them is focused on their use in diagnostics; custom reagent generation is not currently available.

Peptides are easy to make and inexpensive. When the target is in the extracellular domain of transmembrane proteins, they can be particularly useful as assay reagents [124]. However, peptides can cause variable assay performance often due to suboptimal deposition and orientation on the plate.

Virus-like particles (VLPs) are lipoparticles that self-assemble from over-expression of a viral envelope gene to produce desired cell-membrane targets for use in assays. VLPs can be advantageous not only because they are structurally intact and present an epitope but also because they are native to their environment, removing the need for optimized extraction techniques. In addition, they are lower complexity particles, lacking many of the cytoplasmic and nuclear proteins expressed in cell-based immunizations. They can be used for the production of any class of integral membrane protein, and, if over-expressed, do not cause cellular toxicity. However, generation of VLPs requires co-transfection and recombinant expression, which can lead to potential lot-to-lot variability. There may also be concerns about the reproducibility of large-scale preparations.

Aptamers are synthetic oligonucleotides with 3D conformational shapes that allow binding to target molecules [125]. They may be considered when the antigen is highly toxic or otherwise is not compatible with *in vivo* generation of specific Abs [119]. The process for producing an aptamer, SELEX (Systematic Evolution of Ligands by Exponential Enrichment), begins with screening randomized sequence libraries for target binding. The bound sequences are purified with subsequent PCR amplification to increase specificity. The resulting sequences are then screened, purified, and amplified again to optimize their quality. The major benefits of using aptamers are their faster generation (6–8 weeks), specificity, good affinity, less lot-to-lot variability, stability, and no issues with loss of antibody-producing cell lines. Aptamer-based reagents have not yet provided sufficient selectivity and affinity of binding and have presented significant IP limitations [119]. Furthermore, experience using aptamers is still limited and can be costly; although there are many academic publications about aptamer use, there is limited literature from biopharma.

There has been great advancement of alternative antibody producing technologies in the previous decades. Other technologies, including full-length protein sequencing, single cell emulsion, single B-cell screening, or high-throughput paired sequencing, allow for further reduction of the number of animals required in antibody production versus typical hybridoma approaches. Strategies that combine animal immunization with display technologies and full-length antibody sequencing may provide an appropriate path in developing better reagents and therapeutics. Maintaining access to a broad range of antibody generating platforms is critical. The abrupt transition to non-animal-derived antibodies may lead to unintended cost and technology access implications. There should be a gradual approach when implementing advanced technologies; a staged, long-term roadmap and strategy for a transition to non-animal-derived reagents/therapeutics is desired. Both animal and non-animal-derived critical reagents can be fit for purpose, and the scientific community should have the flexibility to use either reagent type when necessary and justified.

RECOMMENDATIONS

Below is a summary of the recommendations made during the 15th WRIB:

Bioanalytical Challenges for Inhalation & Oral Delivery of Biologics

- Inhaled delivery is perceived to result in higher immunogenicity.
- IgA is more relevant to inhaled delivery than to other delivery. If there is no unwanted clinical outcome associated with various isotype, there is no value to isotyping.
- For vaccine and cell therapies, bioanalysts may want to consider cytotoxic T-cell response.
- ADA determination in lung tissue is not required; the onset of ADA in the lung may be different.
- Tissue samples are difficult to obtain during a clinical study, and PK profiles can't be assessed because repeat sampling is not possible.
- Ultra-sensitive assays (e.g., SIMOA, SMCxPro) require more experience, improved vendor support, and better characterization of reagents to support long-term clinical studies.
- Use the MIST guidance to determine if metabolite quantification is needed. Consideration should be given to metabolite identification, in particular if there is activity detected, providing value in evaluating the metabolite.

Free, Bound, Total, Active, Monoactive, Biactive, & Multiactive PK Assays

- Discussions re: assay format should happen early and are often ongoing
- Different formats may be evaluated during early development, with designs changing over time depending on design limitations, stage of development, reagent availability, timelines, resources, etc.
- The decision about which form to measure is driven by appropriate stakeholders (pharmacokineticists, modelers, bioanalytical scientists); PK assay should complement other available data (e.g., target measurements, ADA).
- Domain-specific PK assays can be developed initially as investigational tools but only one PK assay validated as the primary method for regulated study support.
- Incorporating multiplexing adds additional complexity and while technically it may be achievable, for some species there may be challenges with sensitivity.
- Soluble target interferences should be assessed to evaluate any potential impact from one or all targets in a formed complex, if expected in the study.

Implementing Free/Total PK Assays in Regulated Bioanalysis

- Measurement of total or free is a discussion with appropriate stakeholders. If antigen/target level is much lower than drug level, then free and total are essentially the same except maybe at terminal phase.
- PK/PD team wants to understand how drug is working. If you know what the assay measures, the PK scientist can model based on that. Additional assays may not be relevant or needed.
- Regulators look at the totality of evidence presented along with the rationale. Regulators have accepted free and total – it is based on providing a rationale of what is being tested and a justification for why.
- 2020 White Paper in Bioanalysis recommendation still upheld: free assay is preferred if the target is soluble; however, total assay is usually preferred in early discovery along with orthogonal assays.
- The free form is important for PK/PD and total form is important for safety.
- With free assays, neutralizing Ab can impact analyte detection.
- Dilutions can also impact free assays; higher diluted samples may be tested differently.
- LCMS has proven reliable for measuring total drug. Ultra-sensitivity is not always achievable with this platform.

Multi-Domain Biotherapeutic PK Assays in Regulated Bioanalysis

- One PK assay measuring intact molecules was recommended.
- For tri-specific molecules, the team must decide which target binding domain is most relevant for PK assays.
- The bioanalytical strategy should be adapted on a case-by-case basis depending on the requirements of assay sensitivity and specificity, molecular structure, endogenous target biology, the PK or PK/PD related question that needs to be addressed, the availability of high-quality critical reagents, and the availability of the technical platforms.
- Monitoring single target binding domain should be the main driver for bioanalytical strategies supporting MDB. Understanding target biology and solubility (at different levels systemically) is necessary.
- Ultrasensitive assays for MDBs can be used to improve sensitivity and theoretically reduce variability due to automation.
- In early drug development, commercial recombinant targets can be used as a capture reagent although the assay may not always frankly mimic physiological function.
- Screen cell surface receptors or difficult to express targets in the assay format that would be used; it is preferred to keep targets in the "native" linkage.
- There is a trend in the industry to move to non-animal-derived reagents. Polyclonal Ab reagents are not usually suitable for regulated PK work, but if that is what is available, they can be used as critical reagents as long as they are characterized, and a long-term supply is available.
- Phage and yeast display-derived anti-ids can be used for the PK assays. However, they may have lower specificity and binding affinities.
- Characterization by SPR methods is adequate but other methods capable of characterizing binding can also be considered.

Bioanalytical Challenges to Study the Biodistribution of Biotherapeutics

- Surrogate animal tissues can be used in clinical assay development, as well as for sample stability during homogenization and in tissue lysates. Bridging to human tissues can be done if enough volume of biopsy can be collected, such as bone marrow, skin tissues, etc.
- Partial validation of rare tissue matrices is sufficient based on the FDA BMV guidance document [29] if the same fully validated serum/plasma PK assay is used.
- It was suggested that plasma or serum QCs could be an alternative for tracking tissue assays if properly qualified.
- The minimum tissue sample amount required for homogenization is recommended to be at least 50 mg.
- In the discovery stage, a more robust and easily accessible imaging platform is needed in order to collect data faster for decision making. Mass cytometry is currently the front runner.

Advanced Approaches in Critical Reagents Selection for PK Assays

- A single parameter should not be used to select a reagent, but an integrated approach including k_{on} and/or k_{off} rates, signal to noise ratios, etc. is recommended.

- Caution should be used when assessing K_D values since each technology has its own drawbacks that can affect K_D. Assay format will play a role, and the intended use of the reagent should inform the affinity.
- When designing LBAs, it is important to consider the assay platform as well as the desired DoL for that platform, as this could impact the desired assay characteristics.
- MS is considered the most accurate method to determine DoL; TOF-MS, known for intact mass analysis, is closely related to labeling analysis, therefore HRMS can also be a good option but the use of MS is not mandatory.
- For those without MS capability, other types of methods (e.g., HABA dye assay for biotin quantitation) can also be used, as long as the same types of methods are used consistently. However, kits may not be sensitive enough to see specific labeling details like distribution of label per molecule, which can manifest as assay performance issues.
- The use of both kits and MS can provide variable results, so this approach is not recommended.
- Assay performance is the most important outcome, and the necessary accuracy level should be established based on need.
- Reagent candidate selection based on preliminary tests is recommended. Characterization can then more efficiently be performed on a smaller number of reagents (purified, conjugates) to assess biophysical and functional properties.
- Acceptance criteria for bridging a new lot of reagent in an LBA will be defined by assay performance.

Challenges with Positive Control Generation for ADA Assays

- PC can help to monitor assay performance but may not represent ADAs in patient samples. Thus, PC in general may not likely have clinical relevance.
- Either a pAb or mAb can be used as a PC; using a mAb to determine drug tolerance is acceptable.
- Data with both mAb and pAb PCs, if used, may need to be provided in the submission. Since the PC is a surrogate control, clinical samples may behave differently in terms of sensitivity or drug tolerance.
- In biosimilar development, polyclonal PCs derived from immunizing with biosimilar or reference product are both considered as adequate.
- pAb PCs are considered more representative with respect to clinical samples and preferred to be used in assay characterization. However, reproducibility issues associated with the new batches may impact assay performance.
- mAb PCs can be the preferred choice for monitoring assay performance to support long-term clinical development and life cycle management, because of their lot-to-lot consistency.
- Important to establish systematic and organized reagent production processes to deliver quality reagents in time to support assay development.
- Important to have biophysical characterization capabilities to characterize and provide a new lot of reagent that has consistent biophysical characteristics with the previous lot using key QC methods (e.g., LCMS and SEC-MALS).
- Important to perform reagent troubleshooting promptly when needed using various alternative biophysical characterization methods other than the QC methods (e.g., charge variant analysis, stability analysis).
- Lessons learned from reagent troubleshooting using biophysical characterization:
  - LCMS: In many LBA reagent troubleshooting cases, variation of the label incorporation ratio of the detection reagent (e.g., DIG-conjugate in ELISA) impacts assay performance more significantly than the capture reagent (e.g., biotin conjugate) as the quality of the detection antibody directly affects assay signal.
  - SEC-MALS: Even very low percentage of HMW aggregates (e.g., >1 MDa for Abs) can impact assay performance; outsourcing of reagent production and management to support late-stage projects can be more efficient (e.g., reagent production CROs for large-scale conjugate reagent production and life-cycle management CRO for reagent management/shipping to CROs).

Critical Reagent Assay Comparability

- The consistency of the reagent production, labeling, storage and use are critical to the initial qualification of new reagent lots and to ensure seamless bridging to existing lots. A well-documented process is recommended to ensure proper reagent control.
- Lot-to-lot bridging cannot always be expected to be a replicate of current performance.

- Variability within acceptance criteria is observed throughout the life cycle of an assay.
- Variability of performance of assays over time will exist even when additional control parameters are in place, e.g., automated liquid handling. This variability can exist across and within studies, with no determined singular root cause.
- Assay variability may not impact the final determination of ADA incidence.
- For vaccines, change of reagents or other variability can impact final endpoints.
- Variability and bias are different and can have different impacts.
- It is recommended to look for bias between lots.

Novel Critical Reagent Modalities: "Thinking out of the Box"

- Depending on the timing of reagent generation campaigns, consider using multiple approaches to ensure required reagents are generated on time.
- Phage display is used primarily for anti-id generation.
- Alternatives are typically used due to the inability to find a mAb or pAb suitable as PCs.
- Animal-generated reagents are still generally necessary, but use should be justified and other types of reagents should not be discounted.

Acknowledgements

- US FDA, Europe EMA, UK MHRA, Austria AGES, Norway NoMA, Brazil ANVISA, Health Canada, Japan MHLW and WHO for supporting this workshop.
- Dr Eugene Ciccimaro (BMS), Dr Anna Edmison (Health Canada), Dr Fabio Garofolo (BRI), Dr Swati Gupta (AbbVie), Dr Shannon Harris (HilleVax), Dr Carrie Hendricks (Sanofi), Ms. Sarah Hersey (BMS), Dr Steve Keller (AbbVie), Dr Lina Loo (Vertex), Dr Mark Ma (Alexion), Dr Joel Mathews (Ionis), Dr Meena (Stoke), Dr Manoj Rajadhyaksha (Alexion), Dr Ragu Ramanathan (Vertex), Dr Susan Spitz (Incyte), Dr Dian Su (Mersana), Dr Matt Szapacs (Abbvie), Dr Albert Torri (Regeneron), Dr Jian Wang (Crinetics), Drs Jan Welink (EU EMA), Dr Yuling Wu (AstraZeneca) for chairing the workshop and the White Paper discussions.
- All the workshop attendees and members of the Global Bioanalytical Community who have sent comments and suggestions to the workshop to complete this White Paper.
- Future Science Group as a trusted partner.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

1. Savoie N, Booth BP, Bradley T *et al.* 2008 White Paper: The 2nd calibration and validation group workshop on recent issues in good laboratory practice bioanalysis. *Bioanalysis* 1(1), 19–30 (2009).
2. Savoie N, Garofolo F, van Amsterdam P *et al.* 2009 White Paper on recent issues in regulated bioanalysis from the 3rd Calibration and Validation Group Workshop. *Bioanalysis* 2(1), 53–68 (2010).
3. Savoie N, Garofolo F, van Amsterdam P *et al.* 2010 White Paper on recent issues in regulated bioanalysis and global harmonization of bioanalytical guidance. *Bioanalysis* 2(12), 1945–1960 (2010).
4. Garofolo F, Rocci M, Dumont I *et al.* 2011 White Paper on recent issues in bioanalysis and regulatory findings from audits and inspections. *Bioanalysis* 3(18), 2081–2096 (2011).
5. DeSilva B, Garofolo F, Rocci M *et al.* 2012 White Paper on recent issues in bioanalysis and alignment of multiple guidelines. *Bioanalysis* 4(18), 2213–2226 (2012).
6. Stevenson L, Rocci M, Garofolo F *et al.* 2013 White Paper on recent issues in bioanalysis: "hybrid" – the best of LBA & LC/MS. *Bioanalysis* 5(23), 2903–2918 (2013).
7. Fluhler E, Hayes R, Garofolo F *et al.* 2014 White Paper on recent issues in bioanalysis: a full immersion in bioanalysis (Part 1 – small molecules by LCMS). *Bioanalysis* 6(22), 3039–3049 (2014).
8. Dufield D, Neubert H, Garofolo F *et al.* 2014 White Paper on recent issues in bioanalysis: a full immersion in bioanalysis (Part 2 – hybrid LBA/LCMS, ELN & regulatory agencies' input). *Bioanalysis* 6(23), 3237–3249 (2014).
9. Stevenson L, Amaravadi L, Myler H *et al.* 2014 White Paper on recent issues in bioanalysis: a full immersion in bioanalysis (part 3 – LBA and immunogenicity). *Bioanalysis* 6(24), 3355–3368 (2014).

10. Welink J, Fluhler E, Hughes N *et al.* 2015 White Paper on recent issues in bioanalysis: focus on new technologies and biomarkers (part 1 – small molecules by LCMS). *Bioanalysis* 7(22), 2913–2925 (2015).
11. Ackermann B, Neubert H, Hughes N *et al.* 2015 White Paper on recent issues in bioanalysis: focus on new technologies and biomarkers (part 2 – hybrid LBA/LCMS and input from regulatory agencies). *Bioanalysis* 7(23), 3019–3034 (2015).
12. Amaravadi L, Song A, Myler H *et al.* 2015 White Paper on recent issues in bioanalysis: focus on new technologies and biomarkers (part 3 – LBA, biomarkers and immunogenicity). *Bioanalysis* 7(24), 3107–3124 (2015).
13. Yang E, Welink J, Cape S *et al.* 2016 White Paper on recent issues in bioanalysis: focus on biomarker assay validation (BAV) (part 1 – small molecules, peptides and small molecule biomarkers by LCMS). *Bioanalysis* 8(22), 2363–2378 (2016).
14. Song A, Lee A, Garofolo F *et al.* 2016 White Paper on recent issues in bioanalysis: focus on biomarker assay validation (BAV): (part 2 – Hybrid LBA/LCMS and input from regulatory agencies). *Bioanalysis* 8(23), 2457–2474 (2016).
15. Richards S, Amaravadi L, Pillutla R *et al.* 2016 White Paper on recent issues in bioanalysis: focus on biomarker assay validation (BAV): (part 3 – LBA, biomarkers and immunogenicity). *Bioanalysis* 8(23), 2475–2496 (2016).
16. Welink J, Yang E, Hughes N *et al.* 2017 White Paper on recent issues in bioanalysis: aren't BMV guidance/guidelines 'Scientific'? (Part 1 – LCMS: small molecules, peptides and small molecule biomarkers). *Bioanalysis* 9(22), (2017).
17. Neubert H, Song A, Lee A *et al.* 2017 White Paper on recent issues in bioanalysis: rise of hybrid LBA/LCMS immunogenicity assays (part 2: hybrid LBA/LCMS biotherapeutics, biomarkers & immunogenicity assays and regulatory agencies' inputs). *Bioanalysis* 9(23), 1895–1912 (2017).
18. Gupta S, Richards S, Amaravadi L *et al.* 2017 White Paper on recent issues in bioanalysis: a global perspective on immunogenicity guidelines & biomarker assay performance (part 3 – LBA: immunogenicity, biomarkers and PK assays). *Bioanalysis* 9(24), (2017).
19. Welink J, Xu Y, Yang E *et al.* 2018 White Paper on recent issues in bioanalysis: 'a global bioanalytical community perspective on last decade of incurred samples reanalysis (ISR)' (Part 1 – small molecule regulated bioanalysis, small molecule biomarkers, peptides & oligonucleotide bioanalysis). *Bioanalysis* 10(22), (2018).
20. Neubert H, Olah T, Lee A *et al.* 2018 White Paper on recent issues in bioanalysis: focus on immunogenicity assays by hybrid LBA/LCMS and regulatory feedback (Part 2 – PK, PD & ADA assays by hybrid LBA/LCMS & regulatory agencies' inputs on bioanalysis, biomarkers and immunogenicity). *Bioanalysis* 10(23), (2018).
21. Stevenson L, Richards S, Pillutla R *et al.* 2018 White Paper on recent issues in bioanalysis: focus on flow cytometry, gene therapy, cut points and key clarifications on BAV (Part 3 – LBA/cell-based assays: immunogenicity, biomarkers and PK assays). *Bioanalysis* 10(24), (2018).
22. Garofolo W, Savoie N. The Decennial Index of the White Papers in Bioanalysis: 'A decade of recommendations (2007–2016)'. *Bioanalysis* 9(21), 1681–1704 (2017).
23. Fandozzi C, Evans C, Wilson A *et al.* 2019 White Paper on Recent Issues in Bioanalysis: chromatographic assays (part 1 – innovation in small molecules and oligonucleotides & mass spectrometric method development strategies for large molecule bioanalysis). *Bioanalysis* 11(22), (2019).
24. Booth B, Stevenson L, Pillutla R *et al.* 2019 White Paper On Recent Issues in Bioanalysis: FDA BMV guidance, ICH M10 BMV guideline and regulatory inputs (Part 2 – recommendations on 2018 FDA BMV guidance, 2019 ICH M10 BMV draft guideline and regulatory agencies' input on bioanalysis, biomarkers and immunogenicity). *Bioanalysis* 11(23), (2019).
25. Piccoli S, Mehta D, Vitaliti A *et al.* 2019 White Paper on Recent Issues in Bioanalysis: FDA immunogenicity guidance, gene therapy, critical reagents, biomarkers and flow cytometry validation (part 3 – recommendations on 2019 FDA immunogenicity guidance, gene therapy bioanalytical challenges, strategies for critical reagent management, biomarker assay validation, flow cytometry validation & CLSI H62). *Bioanalysis* 11(24), (2019).
26. Neubert H, Alley SC, Lee A *et al.* 2020 White Paper on Recent Issues in Bioanalysis: BMV of hybrid assays, acoustic MS, HRMS, data integrity, endogenous compounds, microsampling and microbiome (part 1 – recommendations on industry/regulators consensus on BMV of biotherapeutics by LCMS, advanced application in hybrid assays, regulatory challenges in mass spec, innovation in small molecules, peptides and oligos). *Bioanalysis* 13(4), (2021).
27. Spitz S, Zhang Y, Fischer S *et al.* 2020 White Paper on Recent Issues in Bioanalysis: BAV guidance, CLSI H62, biotherapeutics stability, parallelism testing, CyTOF and regulatory feedback (part 2a – recommendations on biotherapeutics stability, PK lba regulated bioanalysis, biomarkers assays, cytometry validation & innovation part 2B – regulatory agencies' inputs on bioanalysis, biomarkers, immunogenicity, gene & cell therapy and vaccine). *Bioanalysis* 13(5), (2021).
28. Corsaro B, Yang T, Murphy R *et al.* 2020 White Paper on Recent Issues in Bioanalysis: Vaccine assay validation, qPCR assay validation, QC for CAR-T flow cytometry, NAb assay harmonization and elispot validation (part 3 – recommendations on immunogenicity assay strategies, NAb assays, biosimilars and FDA/EMA immunogenicity guidance/guideline, gene & cell therapy and vaccine assays). *Bioanalysis* 13(6), (2021).
29. US FDA, Center for Drug Evaluation and Research, Center for Veterinary Medicine. Guidance for Industry, Bioanalytical Method Validation. Rockville, MD, USA (2018).
30. Kalluri R, LeBleu VS. The biology, function, and biomedical applications of exosomes. *Science* 367(6478), eaau6977 (2020).

31. Kawakami K, Fujita Y, Kato T *et al.* Diagnostic potential of serum extracellular vesicles expressing prostate-specific membrane antigen in urologic malignancies. *Sci. Rep.* 11(1), 15000 (2021).
32. Masyuk TV, Radtke BN, Stroope AJ *et al.* Pasireotide is more effective than octreotide in reducing hepatorenal cystogenesis in rodents with polycystic kidney and liver diseases. *Hepatology* 58(1), (2013).
33. Doyle LM, Wang MZ. Overview of extracellular vesicles, their origin, composition, purpose, and methods for exosome isolation and analysis. *Cells* 8(7), 727 (2019).
34. Tiwari S, Kumar V, Randhawa S, Verma SK. Preparation and characterization of extracellular vesicles. *AJRI* 85(2), e13367 (2021).
35. Welsh JA, van der Pol E, Bettin BA *et al.* Towards defining reference materials for measuring extracellular vesicle refractive index, epitope abundance, size and concentration. *J. Extracell. Vesicle.* 9, 1816641 (2020).
36. Margolis E, Brown G, Partin A *et al.* Predicting high-grade prostate cancer at initial biopsy: clinical performance of the ExoDx (EPI) Prostate *Intelliscore* test in three independent prospective studies. *Prostate Cancer Prostatic Dis* https://doi.org/10.1038/s41391-021-00456-8 (2021).
37. Veerman RE, Akpinar GG, Eldh M, Gabrielsson S. Immune cell-derived extracellular vesicles – functions and therapeutic applications. *Trends Mol. Med.* 25(5), 382–394 (2019).
38. Lina B, Rosa S, Eduardo F, Iván P, Teresa P. Role of platelet-derived microvesicles as crosstalk mediators in atherothrombosis and future pharmacology targets: a link between inflammation, atherosclerosis, and thrombosis. *Front. Pharmacol.* 7, 293 (2016).
39. Dignat-George F, Boulanger CM. The many faces of endothelial microparticles. *Arterioscler. Thromb. Vasc. Biol.* 31(1), 27–33 (2011).
40. Zaldivia MTK, McFadyen JD, Lim B, Wang X, Peter K. Platelet-derived microvesicles in cardiovascular diseases. *Front. Cardiovasc. Med.* 4, 74 (2017).
41. Combes V, Simon AC, Grau GE *et al. In vitro* generation of endothelial microparticles and possible prothrombotic activity in patients with lupus anticoagulant. *J. Clin. Invest.* 104(1), 93–102 (1999).
42. Poncelet P, Robert S, Bailly N *et al.* Tips and tricks for flow cytometry-based analysis and counting of microparticles. *Transfus. Apheresis Sci.* 53, 110–126 (2015).
43. European Medicines Agency. Committee for Medicinal Products for Human Use (CHMP). Guideline on bioanalytical method validation. London, UK, EMEA/CHMP/EWP/192217/2009 (2011).
44. Viswanathan CT, Bansal S, Booth B *et al.* Quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays. *Pharm. Res.* 24(10), 1962–1973 (2007).
45. Piccoli SP, Sauer JM. Points to Consider Document: Scientific and Regulatory Considerations for the Analytical Validation of Assays Used in the Qualification of Biomarkers in Biological Matrices Critical Path Institute (C-Path) (2019). https://c-path.org/wp-content/uploads/2019/06/EvidConsid-WhitePaper-AnalyticalSectionV20190621.pdf
46. King LE. Parallelism experiments in biomarker ligand-binding assays to assess immunological similarity. *Bioanalysis* 8(23), 2387–2391 (2016).
47. Barra A, Schulz D, Aucouturier P, Preud'homme J-L. Measurement of anti-*Haemophilus influenzae* type b capsular polysaccharide antibodies by ELISA. *J. Immunol. Methods* 115(1), 111–117 (1988).
48. Padoan A, Bonfante F, Pagliari M *et al.* Analytical and clinical performances of five immunoassays for the detection of SARS-CoV-2 antibodies in comparison with neutralization activity. *EBioMedicine* 62, 103101 (2020).
49. Lu C-H, Kalmar B, Malaspina A, Greensmith L, Petzold A. A method to solubilise protein aggregates for immunoassay quantification which overcomes the neurofilament "hook" effect. *J. Neurosci. Methods* 195(2), 143–150 (2011).
50. Ohtsu Y, Tanaka S, Igarashi H *et al.* Analytical method validation for biomarkers as a drug development tool: points to consider. *Bioanalysis* 13(18), 1379–1389 (2021).
51. Chen J, O'Gorman MT, James LP *et al.* Pharmacokinetics of lorlatinib after single and multiple dosing in patients with anaplastic lymphoma kinase (ALK)-positive non-small cell lung cancer: results from a global phase I/II study. *Clin. Pharmacokinet.* 60, 1313–1324 (2021).
52. US FDA Facts: Biomarkers and Surrogate Endpoints. https://www.fda.gov/about-fda/innovation-fda/fda-facts-biomarkers-and-surrogate-endpoints
53. Table of Surrogate Endpoints That Were the Basis of Drug Approval or Licensure. https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure
54. Kakkis ED, O'Donovan M, Cox G *et al.* Recommendations for the development of rare disease drugs using the accelerated approval pathway and for qualifying biomarkers as primary endpoints. *Orphanet. J. Rare Dis.* 10, 16 (2015).
55. Mabrouk O, Chen S, Edwards AL *et al.* Quantitative measurements of LRRK2 in human cerebrospinal fluid demonstrates increased levels in G2019S patients. *Front. Neurosci.* 14, 526 (2020).
56. Personalized Medicine Coalition. Personalized medicine at the FDA: the scope and significance of progress in 2020. 11 February 2021. https://www.personalizedmedicinecoalition.org/Userfiles/PMC-Corporate/file/PM_at_FDA_The_Scope_Significance_of_Progress_in_2020.pdf

57. S.2209 – Verifying Accurate Leading-edge IVCT Development Act of 2021. https://www.congress.gov/bill/117th-congress/senate-bill/2209/text
58. US FDA, Center for Devices and Radiological Health. Draft Guidance for Industry: Framework for Regulatory Oversight of Laboratory Developed Tests (LDTs). Rockville, MD, USA (2014).
59. US FDA. Developing and labeling *in vitro* companion diagnostic devices for a specific group of oncology therapeutic products. April 2020. https://www.fda.gov/media/120340/download
60. Pauling L, Robinson AB, Teranishi R, Cary P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proc. Natl Acad. Sci. USA* 68(10), 2374–2376 (1971).
61. Nardi-Agmon I, Abud-Hawa M, Liran O *et al.* Exhaled breath analysis for monitoring response to treatment in advanced lung cancer. *J. Thorac. Oncol.* 11(6), 827–837 (2016).
62. Horváth I, Barnes PJ, Loukides S *et al.* A European Respiratory Society technical standard: exhaled biomarkers in lung disease. *Eur. Respir. J.* 49(4), 1600965 (2017).
63. Ibrahim W, Cordell RL, Wilde MJ *et al.* Diagnosis of COVID-19 by exhaled breath analysis using gas chromatography-mass spectrometry. *ERJ Open Res.* 7(3), 00139–2021 (2021).
64. Markar SR, Wiggins T, Antonowicz S *et al.* Assessment of a noninvasive exhaled breath test for the diagnosis of oesophagogastric cancer. *JAMA Oncol.* 4(7), 970–976 (2018).
65. "Call me a heretic" Arnold M. 2 November 2020. https://www.linkedin.com/pulse/call-me-heretic-mark-arnold/?trackingId=gfAGRvpy8dCLITZOIZQU1A%3D%3D
66. ICH M10, Draft Bioanalytical Method Validation (2019). www.ich.org/fileadmin/Public Web Site/ICH Products/Guidelines/Multidis ciplinary/M10/M10EWG Step2 DraftGuideline 2019 0226.pdf
67. Lee JW, Devanarayan V, Barrett YC *et al.* Fit-for-purpose method development and validation for successful biomarker measurement. *Pharm. Res.* 23(2), 312–328 (2006).
68. Amur S, Frueh FW, Lesko LJ, Huang SM. Integration and use of biomarkers in drug development, regulation and clinical practice: a US regulatory perspective. *Biomark. Med.* 2(3), 305–311 (2008).
69. Piccoli SP, Garofolo F. Biomarker assay validation. *Bioanalysis* 10(12), 889–891 (2018).
70. Keller S, Purushothama S. Biomarker assays: is it time for a standalone regulatory guidance? *Bioanalysis* 10(23), 1893–1895 (2018).
71. Lafzi A, Moutinho C, Picelli S, Heyn H. Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies. *Nat. Protoc.* 13(12), 2742–2757 (2018).
72. Lee JW, Kelley M, King LE *et al.* Bioanalytical approaches to quantify "total" and "free" therapeutic antibodies and their targets: technical challenges and PK/PD applications over the course of drug development. *AAPS J.* 13(1), 99–110 (2011).
73. Schmielau J, Finn OJ. Activated granulocytes and granulocyte-derived hydrogen peroxide are the underlying mechanism of suppression of t-cell function in advanced cancer patients. *Cancer Res.* 61(12), 4756–4760 (2001).
74. Olson WC, Smolkin ME, Farris EM. Roby shipping blood to a central laboratory in multicenter clinical trials: effect of ambient temperature on specimen temperature, and effects of temperature on mononuclear cell yield, viability and immunologic function. *J. Translat. Med.* 9, 26 (2011).
75. McKenna KC, Beatty KM, Vicetti Miguel R, Bilonick RA. Delayed processing of blood increases the frequency of activated $CD11b^+$ $CD15^+$ granulocytes which inhibit T cell function. *J. Immunol. Methods* 341(1–2), 68–75 (2009).
76. CLSI. Validation of Assays Performed by Flow Cytometry. 1st ed. CLSI guideline H62. Clinical and Laboratory Standards Institute (2021). https://clsi.org/standards/products/hematology/documents/h62/
77. British Pharmacopoeia Guidance on the application of flow cytometry for the cell and gene therapy community. (2021). https://www.pharmacopoeia.com/bp-consultations/flow-cytometry
78. Illingworth A, Marinov I, Sutherland DR, Wagner-Ballon O, DelVecchio L. ICCS/ESCCA consensus guidelines to detect GPI-deficient cells in paroxysmal nocturnal hemoglobinuria (PNH) and related disorders part 3 – data analysis, reporting and case studies. *Cytometry B. Clin. Cytom.* 94(1), 49–66 (2018).
79. Medina A, Puig N, Flores-Montero J *et al.* Comparison of next-generation sequencing (NGS) and next-generation flow (NGF) for minimal residual disease (MRD) assessment in multiple myeloma. *Blood Cancer J.* 10, 108 (2020).
80. Sommer U, Eck S, Marszalek L *et al.* High-sensitivity flow cytometric assays: considerations for design control and sensitivity validation for identification of rare events. *Cytometry B. Clin. Cytom.* 100(1), 42–51 (2021).
81. CLSI EP17-A2, Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures, 2nd Edition 2020. https://clsi.org/standards/products/method-evaluation/documents/ep17/
82. O'Hara DM, Xu Y, Liang Z, Reddy MP, Wu DY, Litwin V. Recommendations for the validation of flow cytometric testing during drug development: II assays. *J. Immunol. Methods* 363, 120–34 (2011).
83. Wood B, Jevremovic D, Béné MC, Yan M, Jacobs P, Litwin V; ICSH/ICCS Working Group. Validation of cell-based fluorescence assays: practice guidelines from the ICSH and ICCS - part V - assay performance criteria. *Cytometry B. Clin. Cytom.* 84(5), 315–23 (2013).

84. Jamin C, Le Lann L, Alvarez-Errico *et al.* Multi-center harmonization of flow cytometers in the context of the European "PRECISESADS" project. *Autoimmun. Rev.* 15(11), 1038–1045 (2016).

85. Bruni D, Angell HK, Galon J. The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. *Nat. Rev. Cancer* 20(11), 662–680 (2020).

86. Battle for High Dimensional Flow Fluorescence vs Metal CYTO2020 Poster. https://learning.isac-net.org/cytovirtual2020-posters

87. Maecker HT, McCoy JP Jr, Amos M ; FOCIS Human Immunophenotyping Consortium. A model for harmonizing flow cytometry in clinical trials. *Nat. Immunol.* 11(11), 975–978 (2010).

88. Cheung M, Campbell JJ, Whitby L, Thomas RJ, Braybrook J, Petzing J. Current trends in flow cytometry automated data analysis software. *Cytometry A.* 99(10), 1007–1021 (2021).

89. Zhang T, Warden AR, Li Y, Ding X. Progress and applications of mass cytometry in sketching immune landscapes. *Clin. Transl. Med.* 10, 1–25 (2020).

90. Estevam J, Krutzik P, Tuig JV, McAuliffe WJ, Smithson G. Development and validation of a high-parameter mass cytometry workflow. *Cytometry B. Clin. Cytom.* 100(1), 92–102 (2021).

91. Alsina M, Shah N, Raje NS *et al.* Updated results from the phase I CRB-402 study of anti-BCMA CAR-T cell therapy bb21217 in patients with relapsed and refractory multiple myeloma: correlation of expansion and duration of response with T cell phenotypes. Program: oral and poster abstracts. Type: Oral. Session: 653. Myeloma/amyloidosis: therapy, excluding transplantation: car t therapies for myeloma: novel approaches and longer-term follow up data hematology disease topics & pathways: multiple myeloma, biological, diseases, therapies, CAR-Ts, plasma cell disorders, lymphoid malignancies, clinically relevant. https://ash.confex.com/ash/2020/webprogram/Paper140410.html

92. Finney OC, Yeri A, Mao P *et al.* Molecular and phenotypic profiling of drug product and post-infusion samples from CRB-402, an ongoing: phase I clinical study of bb21217 a BCMA-directed CAR T cell therapy program: oral and poster abstracts session: 653. Myeloma/amyloidosis: therapy, excluding transplantation: Poster I hematology disease topics & pathways: biological, therapies, immunotherapy. https://ash.confex.com/ash/2020/webprogram/Paper142426.html

93. Rybakowska P, Van Gassen S, Quintelier K, Saeys Y, Alarcón-Riquelme ME, Marañón C. Data processing workflow for large-scale immune monitoring studies by mass cytometry. *Comput. Struct. Biotechnol. J.* 19, 3160–3175 (2021).

94. Kelley RK *et al.* Safety, efficacy, and pharmacodynamics of tremelimumab plus durvalumab for patients with unresectable hepatocellular carcinoma: randomized expansion of a phase I or II study. *J. Clin. Oncol.* 39(27), 2991–3001 (2021).

95. Zappasodi R, Budhu S, Hellmann MD *et al.* Non-conventional inhibitory $CD4^{+}Foxp3^{-}PD\text{-}1^{hi}$ T cells as a biomarker of immune checkpoint blockade activity. *Cancer Cell* 33(6), 1017–1032.e7 (2018); Erratum in: *Cancer Cell* 34(4), 691 (2018).

96. Junker F, Gulati P, Wessels U *et al.* A human receptor occupancy assay to measure anti-PD-1 binding in patients with prior anti-PD-1. *Cytometry* 99, 832–843 (2021).

97. Gennemark P, Walter K, Clemmensen N *et al.* An oral antisense oligonucleotide for PCSK9 inhibition. *Sci. Transl. Med.* 13(593), eabe9117 (2021).

98. Hall P, Vahle JL, Colman K. Inspiration and exasperation: the challenges of inhaled biologics. *Toxicol. Pathol.* 49(2), 232–234 (2021).

99. US FDA. Safety testing of drug metabolites. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/safety-testing-drug-metabolites

100. Dshaies RJ. Multispecific drugs herald a new era of biopharmaceutical innovation. *Nature* 580, 329–338 (2020).

101. Wang Y, Yang S. Multispecific drugs: the fourth wave of biopharmaceutical innovation. *Sig. Transduct. Target Ther.* 5, 86 (2020).

102. US FDA, Center for Drug Evaluation and Research. Draft Guidance for Industry: Bispecific Antibody Development Programs. MD, USA (2019). https: //www.fda.gov/regulatory-information/search-fda-guidance-documents/bispecific-antibody-development-programs-guidance-industry

103. Staack RF, Jordan G, Dahl U, Heinrich J. Free analyte QC concept: a novel approach to prove correct quantification of free therapeutic protein drug/biomarker concentrations. *Bioanalysis* 6(4), 485–496 (2014).

104. Jordan G, Onami I, Heinrich J, Staack RF. Evaluation of the potential use of hybrid LC-MS/MS for active drug quantification applying the 'free analyte QC concept'. *Bioanalysis* 9(21), 1705–1717 (2017).

105. Jadhav SB, Khaowroongrueng V, Fueth M *et al.* Tissue distribution of a therapeutic monoclonal antibody determined by large pore microdialysis. *J. Pharm. Sci.* 106(9), 2853–2859 (2017).

106. Chang HY, Morrow K, Bonacquisti E *et al.* Antibody pharmacokinetics in rat brain determined using microdialysis. *mAbs* 10(6), 843–853 (2018).

107. oulay AC, Saubaméa B, Declèves X, Cohen-Salmon M. Purification of mouse brain vessels. *J. Vis. Exp.* 105, e53208 (2015).

108. Yamaguchi A, Achmad A, Hanaoka H *et al.* Immuno-PET imaging for non-invasive assessment of cetuximab accumulation in non-small cell lung cancer. *BMC Cancer* 19, 1000 (2019).

109. Luo Y, Ronning F, Wakeham N *et al.*QCP in the low carrier density limit. *Proc. Natl Acad. Sci. USA* 112(44), 13520–13524 (2015).

110. Schneider DW, Heitner T, Alicke B *et al. In vivo* biodistribution, PET imaging, and tumor accumulation of ^{86}Y- and ^{111}In-antimindin/RG-1, engineered antibody fragments in LNCaP tumor-bearing nude mice. *J. Nucl. Med.* 50(3), 435–443 (2009).
111. ICH Guideline S12 Nonclinical Biodistribution Considerations for Gene Therapy Products. https://www.ema.europa.eu/en/ich-guideline-s12-nonclinical-biodistribution-considerations-gene-therapy-products-step-2b
112. US FDA, Center for Drug Evaluation and Research. Center for Biologics Evaluation and Research. Guidance for Industry: Immunogenicity Testing of Therapeutic Proteins Products – Developing and Validating Assays for Anti-Drug Antibody Detection. MD, USA (2019).
113. European Medicines Agency. Committee for Medicinal Products for Human Use (CHMP). Guideline on Immunogenicity assessment of therapeutic proteins. London, UK, EMEA/CHMP/BMWP/14327/2006 Rev 1 (2017).
114. Viegas Barroso JF, Halder ME, Whelan M. EURL ECVAM Recommendation on Non-Animal-Derived Antibodies, EUR 30185 EN, Publications Office of the European Union, Luxembourg, ISBN: 978-92-76-18347-1, doi:10.2760/091625, JRC120199 (2020).
115. Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. http://data.europa.eu/eli/dir/2010/63/2019-06-26
116. Chackalamannil S, Rotella D, Ward SE. *Comprehensive Medicinal Chemistry III*. Elsevier (2017).
117. Gray A, Bradbury ARM, Knappik A *et al.* Animal-free alternatives and the antibody iceberg. *Nat. Biotechnol.* 38, 1234–1239 (2020).
118. Wenzel EV, Roth KDR, Russo G *et al.* Antibody phage display: antibody selection in solution using biotinylated antigens. *Methods Mol. Biol.* 2070, 143–155 (2020).
119. Lu RM, Hwang YC, Liu IJ *et al.*Development of therapeutic antibodies for the treatment of diseases. *J. Biomed. Sci.* 27(1), 1 (2020).
120. Kaleli NE, Karadag M, Kalyoncu S. Phage display derived therapeutic antibodies have enriched aliphatic content: insights for developability issues. *Prot.: Struct. Funct. Bioinformat.* 87(7), 607–618 (2019).
121. Alfaleh MA, Alsaab HO, Mahmoud AB *et al.* Phage display derived monoclonal antibodies: from bench to bedside. *Front. Immunol.* 11, 1986 (2020).
122. Spencer S, Bethea D, Raju TS, Giles-Komar J, Feng Y. Solubility evaluation of murine hybridoma antibodies. *mAbs* 4(3), 319–325 (2012).
123. Jain T, Sun T, Durand S *et al.* Biophysical features of the clinical mAb landscape. *Proc. Natl Acad. Sci. USA* 114(5), 944–949 (2017).
124. Heyduk E, Hickey R, Pozzi N, Heyduk T. Peptide ligand-based ELISA reagents for antibody detection. *Anal. Biochem.* 559, 55–61 (2018).
125. Gupta V, Lassman ME, McAvoy T *et al.* An evaluation of an aptamer for use as an affinity reagent with MS: PCSK9 as an example protein. *Bioanalysis* 8(15), 1557–1564 (2016).
126. Companion Diagnostics. https://www.fda.gov/medical-devices/in-vitro-diagnostics/companion-diagnostics
127. Sommer U, Eck S, Marszalek L *et al.* High-sensitivity flow cytometric assays: considerations for design control and analytical validation for identification of rare events. *Cytometry B. Clin. Cytom.* 100(1), 42–51 (2021).
128. Borghardt JM, Kloft C, Sharma A. Inhaled therapy in respiratory disease: the complex interplay of pulmonary kinetic processes. *Can. Respir. J.* 2018, 2732017 (2018).
129. Liang W, Pan HW, Vllasaliu D, Lam JKW. Pulmonary delivery of biological drugs. *Pharmaceutics* 12(11), 1025 (2020).