appliedclinicaltrialsonline.com

Volume 28 Number 4 April 2019

APPLIED CLINICAL TRIALS

YOUR PEER-REVIEWED GUIDE TO GLOBAL CLINICAL TRIALS MANAGEMENT

1992–2019 ACT 27th Year of Service

STUDY DESIGN

OPTIMIZING RWE GENERATION

REGULATORY

FDA'S 'REAL-WORLD' FRAMEWORK

DATA SHARING

ADOPTING GLOBAL DATA STANDARDS

Emerging Biopharma Deserve TLC from CROs

LISA HENDERSON
Editor-in-Chief

Early this month, I presented on the topic "Why Aren't Lessons Learned from Big Pharma/Big CROs Translating to Smaller Biopharma Relationships?" For years, *Applied Clinical Trials* has been the operational resource for clinical trials for sponsors, CROs, and academia. The transition of the contract research industry from a small cottage to a fully-matured one has seen numerous outsourcing models from the low-touch, very focused and accountable transactional service, to the high-touch, highly involved and complex strategic partnership. Another example has featured CROs taking on internal departments from pharma companies to ensure employees still had jobs, as well as the knowledge transfer relative to that pharma's specific development needs and culture.

It may seem that by this point, sponsors of all shapes and sizes would know how to respond to the challenges relative to working with an outsourcer, no matter what type of relationship is involved. But in discussions I have with smaller biopharma, that does not seem to be the case. Why should CROs be concerned about their service offerings for these emerging players? Because they are not an insignificant part of the drug development landscape.

Consider the following:

- Of all companies sponsoring one or more clinical development programs, 61% now fall outside the ranks of the top 50 largest pharma.
- The percentage of FDA approvals obtained by big pharma in 2018 was 26%, compared to 60% that went to firms that had never before received an approval.

The days of quickly selling a promising compound early on to a large pharma company or launching a full-fledged IPO are dwindling. An Ernst & Young report showed 75 biotech IPOs in 2014 compared to nine in the second half of 2016. These days, innovation hubs in major regions of the country are home to a vast number of companies looking to enter clinical trials, take their compound at least through Phase II, and some are fully intending to stick it through Phase III and even make it to a commercial stage company.

This is all evidence that these emerging biopharma are CROs' latest boom of customers. But, as the new crop of uninitiated companies, they are having similar angst with their CRO. They feel unimportant to large CROs, costs get away quickly, oversight is difficult, projects don't come in on time or budget…the same list that we've seen with outsourcing relationships over the years. The problem is that emerging biopharma companies face constraints that the large, traditional pharma never have. A single clinical trial could be the only thing the company has; basically, its whole survival. They are more financially-constrained and may either make poor decisions or cut corners, not to mention taking advice from new executives or external stakeholders that deviate from the original plan, which then results in out-of-scope charges or change orders.

The future of drug development does not lightly rest on an outsourced expert, it almost solely exists in their confines. It wouldn't be a stretch to ask CROs to be more mindful of their smaller clients and take extra attention in handling their needs.

EDITORIAL OFFICES

485 Route 1 South, Building F, Second Floor, Iselin, NJ 08830 USA
+1 (732) 346-3080 fax: +1 (732) 647-1235,
www.appliedclinicaltrialsonline.com

EDITOR-IN-CHIEF Lisa Henderson, lhenderson@mmhgroup.com

MANAGING EDITOR Michael Christel, mchristel@mmhgroup.com

ASSOCIATE EDITOR Christen Harm, charm@mmhgroup.com

ASSISTANT EDITOR Miranda Schmalfuhs, mschmalfuhs@mmhgroup.com

ART DIRECTOR Dan Ward, Dward@hcl.com

WASHINGTON EDITOR Jill Wechsler
+1 (301) 656-4634 fax: +1 (301) 718-4377

SALES OFFICES

GROUP PUBLISHER Todd Baker

485 Route 1 South, Building F, Second Floor, Iselin, NJ 08830 USA

+1 (732) 346-3002. fax: +1 (732) 647-1235, tbaker@mmhgroup.com

DIRECTOR OF ADVERTISING Wayne K. Blow

UK: +44 1244 629 304 fax: +44 1925 732 798, wblow@mmhgroup.com

NATIONAL SALES MANAGER Bill Campbell
+1 (847) 283-0129 fax: +1 (847) 282-1456, wcampbell@mmhgroup.com

SALES SUPPORT COORDINATOR Kristi Stevenson
+1 (732) 346-3006 fax: +1 (732) 596-0012, kstevenson@mmhgroup.com

ACT CHESTER UK OFFICE: +44 1244 393 100

MARKETING SERVICES

AUDIENCE DEVELOPMENT MANAGER, C.A.S.T. DATA AND LIST INFORMATION
Melissa Stillwell
(218) 740-6831, mstillwell@mmhgroup.com

PERMISSIONS/INTERNATIONAL LICENSING
Jillyn Frommer
+1 (732) 346-3007 fax: +1 (732) 647-1101, jfrommer@mmhgroup.com

REPRINTS Licensing and Reuse of Content: Contact our official partner, Wright's Media, about available usages, license fees, and award seal artwork at Advanstar@wrightsmedia.com for more information. Please note that Wright's Media is the only authorized company that we've partnered with for MultiMedia Healthcare materials.

SUBSCRIPTIONS +1 (888) 527-7008 (toll-free within USA) +1 (218) 740-6477 (outside USA), fulfill@hcl.com

BACK OR CURRENT ISSUES +1 (800) 598-6008, +1 (218) 740-6480 (outside USA)

PRODUCTION OFFICES

PRODUCTION MANAGER Karen Lenzen
131 W. 1st Street, Duluth, MN 55802 USA
+1-248-823-7808 fax: +1 (408) 962-1125

APPLIED CLINICAL TRIALS (Print ISSN: 1064-8542, Digital ISSN: 2150-623X) is published 4 times/year in March, June, Sept & Dec by MultiMedia Healthcare LLC, 325 W 1st Street, STE 300 Duluth, MN 55802. Subscription rates: $70 for 1 year (4 issues), $120 for 2 years (8 issues) in the United States and possessions; $90 for 1 year, $140 for 2 years in Canada and Mexico; all other countries $130 for 1 year, $235 for 2 years. Single copies (prepaid only): $23 in the United States and possessions; $28 in all other countries. Add $6.50 per order for shipping and handling. **Periodicals postage paid** at Duluth, MN 55806 and additional mailing offices. **POSTMASTER:** Please send address changes to APPLIED CLINICAL TRIALS, P.O. Box 6115, Duluth, MN 55806-6115. PUBLICATIONS MAIL AGREEMENT NO. 40612608, Return Undeliverable Canadian Addresses to: IMEX Global Solutions, P. O. Box 25542, London, ON N6C 6B2, CANADA. Canadian G.S.T. number: R-124213133RT001. Printed in the U.S.A. Digital-only editions will publish 6 times/year in Jan/Feb, April, May, July/Aug, Oct, and Nov.

APPLIED CLINICAL TRIALS
VOLUME 28, NUMBER 4

IMAGE SOURCE/GETTY IMAGES; KJPARGETER/SHUTTERSTOCK.COM

FEATURED

WASHINGTON REPORT

FDA SEEKS TO BROADEN CLINICAL TRIAL ELIGIBILITY CRITERIA

To reduce the time and cost involved with conducting clinical research on new biopharmaceuticals, regulators and researchers are looking to broaden outmoded criteria for identifying and enrolling individuals in clinical trials. Sponsors traditionally have shied away from including young patients and those with infectious diseases or comorbidities that might raise safety issues or compromise efficacy results. Their aim is for clinical trials to generate data that will support market approval of new drugs, biologics, and medical products, and to avoid studies that involve children, small patient populations, or patients in poor health that may generate confusing or questionable data.

At the same time, restrictions on clinical trial participation may "slow patient accrual, limit patients' access to clinical trials, and lead to trial results that do not fully represent treatment effects in patients that ultimately will use the drug," explained outgoing FDA Commissioner Scott Gottlieb March 12 in unveiling a series of guidance documents that provide strategies for achieving broader patient inclusion in oncology trials (see: http://bit.ly/2CvAP2j). New policies are needed because long-established and widely-used eligibility criteria become set in stone over time, discouraging new approaches despite changing technology and patient preferences.

The new guidances specifically address clinical studies for new cancer treatments, where a broadening of eligibility criteria would enable more people to participate in oncology trials, a field where patient accrual is difficult, but where the serious nature of disease may lend itself to more flexibility in enrollment criteria. Including children, adolescents, and individuals with infectious diseases, malignancies, and metastases may yield results that are more generalizable and help improve our understanding of a therapy's benefit-risk profile across populations likely to receive the drug in clinical practice, Gottlieb explained.

One guidance finalizes an earlier proposal for including adolescents in adult oncology trials where the age of 18 is the traditional cutoff point. The rationale is that many cancers found in younger patients often behave similarly in adults, and that excluding adolescents from clinical studies may delay their access to potentially effective therapy. The guidance advises on criteria for enrolling these patients in studies, recommendations for dosing and safety monitoring, and relevant ethical considerations to support treatment (see: http://bit.ly/2MxcRHg). A draft guidance similarly addresses how and where it is appropriate for sponsors to include pediatric patients in adult cancer trials to obtain more accurate information more quickly on appropriate dosing and treatment of children of different age levels (see: http://bit.ly/2CwFFN0).

Patients with HIV or hepatitis infections should not be ruled out of participation in clinical trials, says FDA in another draft guidance, especially when cancer treatment may be particularly important for individuals with such chronic conditions. Eligibility may be limited for AIDS patients with particularly low CD4 cell counts or more serious infections, and timing of treatment may be important in certain situations (see: http://bit.ly/2FsfwJ9).

Another guidance addresses criteria for cancer trials to include patients with organ dysfunction (renal, cardiac, hepatic) or previous malignancies. The policy notes that excluding such individuals may skew recruitment to younger patients, which may not be fully representative of the population likely to receive treatment (see: http://bit.ly/2TQadE4).

FDA sees a need to study patients living with cancer who increasingly are diagnosed with brain metastases, in an effort to discourage exclusion of such individuals from clinical research (see: http://bit.ly/2TQadE4).

The four new draft guidances were developed by FDA with input from the American Society of Clinical Oncology and Friends of Cancer Research. These policies aim to ensure that clinical trials are designed to reflect "the diversity of the population that receives drugs in the real world," said Gottlieb, and to help design oncology trials to be "more representative of the patients that may ultimately benefit from novel treatments."

— Jill Wechsler

SHARPLESS ASSUMES REINS AS AGENCY COMMISSIONER

Health and Human Services (HHS) secretary Alex Azar wasted little time in naming a new leader for FDA. Norman (Ned) Sharpless, director of the National Cancer Institute (NCI) at the National Institutes of Health (NIH), become acting FDA commissioner this month as Scott Gottlieb departs the agency.

The appointment won praise on all sides, as Sharpless is a respected cancer researcher and well-acquainted with FDA operations and challenges. Before becoming head of NCI in October 2017, Sharpless ran the Lineberger Comprehensive Cancer Center at the University of North Carolina. As a physician-scientist at that institute, he worked with industry on drug development programs and co-founded early-stage biotech companies involved with developing new cancer drugs and diagnostic blood tests.

Gottlieb praised the choice, as did the cancer research community and biopharmaceutical companies. Ellen Sigal, founder of Friends of Cancer Research, noted Sharpless' expertise as an oncologist in treating leukemia, as well as research on the role the cell cycle plays in cancer and aging. She predicted that Sharpless will continue FDA policy "that best benefits patients."

While Sharpless will work to advance many of the initiatives backed by Gottlieb, particularly the campaign to reduce teen use of e-cigarettes and excessive opioid use, as acting commissioner, new initiatives will be put on hold pending the nomination and confirmation of a permanent commissioner for the agency, which may take some time.

— Jill Wechsler

LEARN MORE ABOUT

RBM CASE STUDIES:

How Artificial Intelligence and Machine Learning with Advanced Analytics are Driving New Levels of Patient Safety and Study Quality

On-demand webinar

Aired March 28, 2019

View now for free!

www.appliedclinicaltrialsonline.com/act_p/quality

PRESENTERS:

Dr. Rajneesh Patil
Senior Director, Head – Process Design & Analytics, Centralized Monitoring Services, Clinical Operations, IQVIA

Dr. Jonas Renstroem
Associate Director, Centralized Monitoring Process Design & Analytics, IQVIA

MODERATOR:

Lisa Henderson
Editorial Director, *Applied Clinical Trials*

Risk-based Monitoring (RBM) continues to transform clinical monitoring, and many leading pharma firms are already realizing its benefits. Learn from RBM experts as they share insights from years of executing RBM trials.

- Highlight the evolution of advanced analytics and how application of contemporary statistical science can strengthen risk-based strategies
- Share learnings from practical models using predictive and advanced analytics that help reduce white noise/false positives in signal detection deployed for RBM
- Provide insights on where machine learning, and artificial intelligence models have potential application in clinical monitoring to address risks based on study design and patient population

Register now to discover the latest capabilities that are changing how clinical trials are being delivered – with better, more timely patient insights, enhancing patient safety and improving overall study quality.

Key take-aways:

- Use a global network of sites for targeted recruitment of NHV and Learn the impact RBM can have on study quality, patient safety and cost
- Understand how innovative technologies and methodologies are driving new levels of study quality
- Hear how artificial intelligence, machine learning, predictive and advanced analytics are improving patient safety and overall study quality and operational efficiencies

Presented by:

Applied Clinical Trials

Sponsored by:

IMS Health & Quintiles are now IQVIA™

CONTACT US
iqvia.com

For technical questions about this webinar, please contact Kristen Moore at KMoore@mmhgroup.com

EU REPORT

CHALLENGES IN CONVEYING TRANSLATIONAL RESEARCH MESSAGE

Europe's outstanding achievements in basic life sciences research are not getting through to patients as useful treatments and diagnostics—and the urgency of filling those gaps is not getting through to European policymakers, says yet another new coalition jockeying to make its voice heard in the crowded space of strategic lobbying in Brussels.

A joint statement from Europe's drug industry and Europe's medical societies calls for a "bold vision" to ensure adequate funding and coordination for translational research. Issued on the eve of the European Union's summit meeting in Brussels on March 21 (where the agenda was dominated yet again by Brexit woes), it makes a plea for recognition from a world where, it says, it is not understood.

Laboriously, the coalition explains the role and importance of translation research, and highlights the challenges it faces. "This type of research is costly and risky, involving many steps of collaboration among all stakeholders. Funding activities are scarce," it says. And, it goes on, "instruments aiming to strengthen the various steps in the process of the translational research ecosystem are often fragmented and lack tangible follow through the full medicine's lifecycle."

In language so convoluted that it risks intensifying rather than clarifying confusion in the outside world it aims to impress, it urges EU policymakers "to work together with the research community, industry, and patient representatives to ensure that these directions are implemented in inter-sector collaborations to progress synergies within and across projects and initiatives."

Something better in terms of communication might have been expected from the combined intellects of the drug industry's European Federation of Pharmaceutical Industries and Associations and the 400,000 researchers and health professionals in the BioMed Alliance—the two partners in this coalition. So far, the top item in their plan, as revealed by their statement, consists of "increasing visibility on the role and needs of translational research to better inform policymakers, academia, patients, and the general public."

With no irony, BioMed Alliance presents itself in its statement as conveying "the views of its members in a cohesive and comprehensible form to policymakers, professionals, and the public at large."

As it happens, they may be in luck. One of the items on the agenda of EU leaders at the summit was developing an "integrated approach" to promote growth in the EU through "an assertive industrial policy allowing the EU to remain an industrial powerhouse." The leaders are scheduled to request their officials "to present, by March 2020, a long-term vision for the EU's industrial future, with concrete measures to implement it."

And in another stroke of potential good luck, a breakthrough on research strategy occurred just days before the summit, when agreement was reached among EU legislators on the outlines of their next long-term research program. This will put more than $100 billion behind European researchers over the next seven years, with a hefty slice of that earmarked for life sciences and health topics. And the entire program is geared to "societal challenges and industrial competitiveness" and "market-creating innovation" will provide a one-stop shop to help bring promising and breakthrough technologies from lab to market application, and assist in scale-up of ideas. Parallel streams are to be set up for early stages of research and for development and market deployment.

If the translational research community can't talk their way into benefiting from some of that, they will have no one to blame but themselves.

— Peter O'Donnell

JOINT HTA TALK DOESN'T STOP

The European Union discussions on collaborative health technology assessment (HTA) have been so intense over the last year or so that it might be thought that Brussels is where all the HTA action is. Until now, anyway, because since the collapse in December of Austria's gallant bid to move the discussions forward on the bid for European collaboration, Brussels has gone very quiet on the subject. Romania, which took over the EU presidency from Austria in January, has very low ambitions for advancing the talks among member states. And meanwhile, Soledad Cabezón, the Spanish socialist who piloted the complex legislation through the European Parliament, has announced that she is standing down as an MEP at the upcoming elections in May.

But there is plenty going on elsewhere. An international consortium tasked with finding a common definition of HTA has just published its suggestion, and is seeking comments on it until the end of April. This is a rather longer definition than the EU is proposing in its legislation: "multidisciplinary," "explicit and scientifically robust methods," assessing value "at different points in the lifecycle," "comparative," "systematic," "transparent," "multiple stakeholders" are all in there. So too are "clinical effectiveness and safety, costs and economic implications, wider implications for the patient and caregivers, and any ethical, social, cultural, or legal issues, as well as organizational and environmental aspects." The whole thing can be seen here: https://bit.ly/2FNxtse and comments can be sent to HTAdefinition@ihe.ca by April 30.

The first HTA pilots are in progress within the Nordic cooperation between the authorities in Norway, Finland, and Sweden. This mechanism, initiated a year ago, claims that it will speed up evaluations and generate uniform reports while making applications smoother for companies. Roche has volunteered its Tecentriq product for one of these pilots. In Amsterdam, where the European Medicines Agency (EMA) has now officially opened up for business there, the European technical collaboration on HTA, EUnetHTA, held its annual conference earlier this month.

— Peter O'Donnell

REGULATORY

FDA MODERNIZES CONTROLLED CORRESPONDENCE INTERACTIONS WITH INDUSTRY

Portal provides a centralized location for inquiries—a welcomed upgrade from traditional email submission

When generic drug manufacturers and related industry submit written inquiries to the FDA requesting information on a specific element of a generic drug development or certain postapproval submission requirements, the inquiries are referred to as "controlled correspondence." Controlled correspondence are labelled as standard or complex, and until October 2018, these inquiries could only be submitted by emailing the Office of Generic Drugs, as outlined in the *Controlled Correspondence Related to Generic Drug Development* draft guidance (see: http://bit.ly/2OmPvp7).

Submission of controlled correspondence via the email method involved data entered manually by FDA staff. As part of FDA's ongoing effort to modernize interactions with industry, FDA launched the Center for Drug Evaluation and Research (CDER) Direct NextGen Collaboration Portal (see: http://bit.ly/2WnoeGd). Effective October 2018, CDER completed enhancements to the portal so that controlled correspondence can now be received, managed, and tracked online. The portal allows for more efficient processing of the controlled correspondence and provides "one-stop-shop" access to a prospective applicant's controlled correspondence request history.

Handling controlled correspondence through the portal allows greater consistency, faster triage, simplified acknowledgement, and easier tracking. Fully integrated with CDER's internal work management systems, the portal's expandable cloud capability will also accommodate future growth. After reviewing a request for information via controlled correspondence, FDA sends the requestor a response that can be accessed through the portal. This central location for submissions and related communications offers a chronological view of all communications and documents.

The Portal also streamlines FDA's review of controlled correspondence by providing real-time status updates and notifications about controlled correspondence submissions to FDA personnel. Streamlining the review process for the thousands of controlled correspondence inquiries received each year saves time and aids FDA and industry in developing generic drugs that meet FDA's Generic Drug User Fee Amendments obligations.

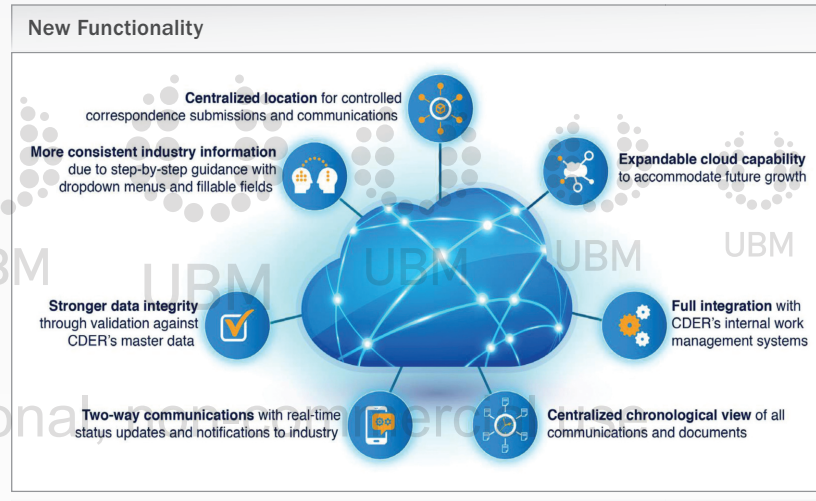

Figure 1. Steps to improved accuracy with built-in validation.

FDA personnel note more accurate information from industry because submitted data is more consistent. This improved accuracy is due to step-by-step guidance and a built-in validation against CDER master data, which provides stronger data integrity. Figure 1 above illustrates these benefits.

Industry has welcomed the addition of controlled correspondence to CDER's Direct NextGen Collaboration Portal. In the first two months after its release, 272 company representatives signed up for the new functionality and FDA received 689 controlled correspondence. The agency is receiving about a dozen controlled correspondence daily through the portal. Those in industry who have used the portal's controlled correspondence functionality describe it as more convenient and a better format for submissions compared to the process of sending requests via email.

Overall, FDA received more than 2,500 controlled correspondence inquiries in 2018, with 100% of them email until October and then 99% via the portal since October. With its addition, life sciences industry submissions are more consistent and simpler to acknowledge and triage, and controlled correspondence are easier to track. In just a few steps, industry initiates and submits controlled correspondence to a convenient online location. In short, the portal provides a win-win solution for FDA and industry and opens up opportunities to further improve the regulatory communication process.

— Saundra Middleton, FDA's Office of Generic Drugs, Office of Regulatory Operations; Daniil Graborov, Office of Business Informatics, Division of Regulatory Review and Drug Safety Services and Solutions

Additional contributors to this article include Shannon Hill, Christina Marshall, Marissa McNall, and Kelly Miller.

A Q&A

Protecting Sponsors Against Bias and Variability

Mark Opler, PhD, MPH
Chief Research Officer
WCG, MedAvante-ProPhase

Placebo response is growing and contributes to the risk of trial failure.

Clinical trials succeed or fail based on the ability of the primary endpoint to differentiate study drug from control conditions. In the case of placebo-controlled studies, the levels of random error, sources of noise, variability introduced by patient or investigator factors, and placebo response rates can have a profound influence on the outcome. Design and execution teams can take several steps to reduce these risks, improve signal-to-noise ratios, and mitigate the impact of placebo response. *Applied Clinical Trials* recently spoke with Mark Opler, PhD, MPH, chief research officer of WCG, MedAvante-ProPhase, to learn how these approaches need to be incorporated into standard practice to reverse prevailing trends going forward for certain therapeutic areas and conditions.

Applied Clinical Trials: What's the difference between positive, negative, and failed trials?

Opler: A *positive trial* is what we all strive for in clinical research: the experimental treatment (e.g., the drug, the device) is clearly and unequivocally better than the control (e.g., placebo). A *negative trial* is the regrettable, but sometimes inevitable, consequence of research in which the experimental treatment is not better than the control. And, a *failed trial*—where the outcome cannot be interpreted—is probably the worst possible outcome because we've spent a lot of money and time, we've exposed patients to an experimental treatment, and we've come no closer to the answer than when we started.

Applied Clinical Trials: Is placebo response really a problem for clinical research?

Opler: Yes, definitely. The placebo response is probably the leading cause of failed trials. Placebo response and high placebo response occur when patients in a placebo-controlled study respond well to what is essentially no active treatment. The sugar pill produces the same or better outcome than the experimental treatment. It's a very serious problem in clinical research. Those of us who have been studying it for many years have realized that this problem is actually growing. The placebo response was once negative or absent in certain therapeutic areas. Now, we're seeing it routinely sometimes outstripping effect sizes from the treatments we're studying.

Applied Clinical Trials: Why does placebo response occur?

Opler: The most probable cause of placebo response is therapeutic expectation (i.e., the

expectation of improvement). We, as an industry, have not adequately addressed it in our clinical research work. The average patient that comes into a study needs to be very carefully educated about their role and about the use of placebos. We want patients to get better and they may come in expecting to get better when they enter a clinical trial, particularly if they don't fully appreciate the difference between clinical research and medical care.

Applied Clinical Trials: What should sponsors and study teams know about measurement reliability?

Opler: Another contributor to failed trials is the lack of reliability of measures. For instance, if a thermometer is used incorrectly, we get the wrong result. Measurement reliability is about ensuring that, from visit to visit, from patient to patient, and from site to site, we have reliability in our approach to evaluating the primary outcome. Whether that primary outcome is driven by a thermometer, a clinical interview, or a specific examination procedure, we can reduce the risk of failed trials and increase the likelihood of trial success by paying appropriate attention to reliability.

> "The most probable cause of placebo response is therapeutic expectation (i.e., the expectation of improvement). We, as an industry, have not adequately addressed it in our clinical research work."

Applied Clinical Trials: What role do you think technology plays in all of this?

Opler: Like anything else that we do in clinical research, technology is omnipresent. In our efforts to combat placebo response and improve measurement reliability, technology can play a very important role. Whether you are using electronic forms for clinical outcome assessments or technology to evaluate the level of noise in data over time, consider every technological aspect of the program being conducted and ask, "Is this contributing to study success? Is it improving the reliability of measurement? And is it getting me closer to my ultimate goal, a positive trial?"

Applied Clinical Trials: What are your top three recommendations to sponsors?

Opler: For sponsors in the process of planning studies, I would urge them to do three things. First, think about study design. There are aspects of study design that can contribute to lower placebo response and higher success, whether that's the number of arms in the trial or the selection of outcomes and endpoints All of these can contribute in subtle, and not so subtle ways, to a positive study.

Second, make sure that for almost every therapeutic area, sponsors and study teams have a strategy to mitigate the risk of high placebo response. This is clearly recognized in some therapeutic areas, but we have yet to build a meaningful awareness in others.

Third, the sponsor should be aware of what is being done to ensure measurement reliability, to ensure that methods and procedures are in place to make sure the most vital data—the primary endpoint—in the study is being protected from noise and from error. Those are the top three recommendations to anyone, almost regardless of disease therapeutic area or stage of development.

Q&A

REGENERATIVE MEDICINE COMPANIES SUCCESSFULLY ADVANCE STEM CELL THERAPIES

Stem cell therapy development is a new area where several companies are emerging with promising results utilizing stem cell therapies. Some companies have failed, whereas others, such as Athersys, are succeeding. Ahead, Gil Van Bokkelen, PhD, CEO of Athersys, discusses his perspective on the recent advancements in stem cell therapies and their multiple application on a variety of therapeutic areas. In addition to joining the Athersys family, Van Bokkelen serves as the Chairman of the Board of Governors for the National Center for Regenerative Medicine. He also served as the Chairman of the Alliance for Regenerative Medicine from 2010 through 2012, and he served *ex officio* from 2013 to 2014.

Van Bokkelen has also served on several other boards, including the Biotechnology Industry Organization's board of directors. He has extensive leadership experience and a strong scientific and financial knowledge of the biotechnology industry.

Moe Alsumidaie: What led you to work in the field of stem cell research and regenerative medicine?

Gil Van Bokkelen: Athersys didn't start with a focus on the field of cell therapy and regenerative medicine. Regenerative medicine wasn't even a field yet. We had developed some powerful genomics technologies that led to multiple partnerships with major pharma companies. These technologies attracted the attention of a group that had done some pioneering work in the stem cell field—they had discovered a novel cell type that had some exciting properties.

This connection led to a collaboration to explore the unique characteristics of these cells, which can promote healing in multiple ways, exhibit robust growth properties, and can also be administered like Type O blood, without requiring tissue matching or immune suppression. These characteristics mean that the technology could be scaled robustly.

However, we didn't know yet whether or where the technology might be relevant therapeutically. So, we collaborated with independent labs and research teams to help us figure out potential therapeutic areas of relevance for our technology. Since then, we've discovered that MultiStem has profound significance in areas of critical care medicine such as stroke and, as we just recently announced, acute respiratory distress syndrome (ARDS), and other areas we have been working in with leading independent research teams.

MA: What results came out of the Phase II MultiStem study for stroke?

GVB: The key finding that we observed was that if we treated stroke patients with Intravenous (IV) administration of MultiStem within 36 hours or less, there was evidence of robust recovery and that patients continued to improve over time. One of the main things we clinically evaluated was something called Excellent Outcome, which is essentially the proportion of patients that achieved full recovery. In Excellent Outcome, patients have to demonstrate an excellent or normal score in each of three different clinical assessments: the NIH Stroke Scale (NIHSS), the Modified Rankin Scale, and the Barthel Index for Activities of Daily Living (ADL), which indicates if the patient can function without assistance. In practical terms, this means the patients recovered to the level of independence and quality of life they enjoyed before the stroke.

We also saw significant evidence of improvement among other clinical parameters, which eventually led to a series of critical regulatory designations from the FDA, including both fast track and regenerative medicine advanced therapy (RMAT), as well as similar designations from international regulators.

Another observation from the trial was the importance of developing and validating a simple, easy-to-prepare version of the product—something that we refer to as a genuine "off the shelf" product. Essentially, MultiStem consists of cells that are kept in frozen form in a vial until needed, and then the product is thawed and administered to the patient using a simple and straightforward process.

Gil Van Bokkelen

MA: Before we dive into the science, how does your technology differ from competitors that are targeting stroke with stem cells?

GVB: Our fundamental approach relies on treating patients within the first 36 hours after a stroke has occurred, which we have shown is a critical window following a stroke. MultiStem is administered intravenously, using a very simple and straightforward approach. We don't need to alter the cells that comprise MultiStem genetically, and we have developed proprietary methods and technologies to manufacturing the product in a scalable manner. The key features are that the cells stimulate tissue repair and healing through multiple, distinct mechanisms of action, have robust scalability, and are easy to prepare and deliver using a simple intravenous administration. Like a traditional biologic, the cells are then cleared from the body over time.

SanBio, our competitor, has been focused on using cell therapy for the treatment of patients with chronic stroke damage, using genetically modified bone marrow-derived cells that are then surgically implanted into the brain, six months to seven years after a stroke has occurred. They are injecting these modified cells right into the brain hoping that they can help stimulate recovery. While they saw some promising signs of recovery in a Phase II study on traumatic brain injury patients, they followed that up with

Q&A

some disappointing results in patients that had suffered chronic stroke damage.

However, it's important to recognize that there are several fundamental differences in these two approaches, from a technological perspective and a timing perspective. First, we are using a different cell type that has different therapeutic properties. Second, we administer our product intravenously, without the need for surgery. Third, and perhaps most importantly, we're intervening within 36 hours after the stroke has occurred, whereas SanBio was focused on trying to help correct damage that had happened months to years beforehand, which is a lot more challenging.

Our data shows that if we administer MultiStem within the relevant time frame, meaning within that 36-hour window, we can blunt or neutralize the hyperinflammatory cascade that causes a lot of the long-term damage, and promote better recovery. Essentially, MultiStem is stopping the counter-productive hyper-inflammatory cascade before it starts, while also stimulating fundamental reparative mechanisms. Our clinical data shows that this can improve things pretty dramatically. While the long-standing stroke dogma is that there is no additional improvement after 90 days, we saw that the patients who received MultiStem within the 36-hour window continued to improve through the one year clinical follow-up, and many of them experienced a full recovery.

MA: Are there any side effects to the hypo-inflammatory profile that would impact the recovery?

GVB: We've seen very consistent tolerability and safety profile in the clinical studies that we have run. We conducted many studies preclinically before we ever got to the clinical stage, and many of those were performed with the guidance and input of the FDA, and so we were confident that it was a well-characterized and safe product.

MA: Mechanistically, how does MultiStem work?

GVB: In contrast to a traditional drug that is designed to do one specific thing, MultiStem is essentially a living drug that does multiple things to help recovery and healing. For example, a few years ago we discovered that shortly after a stroke has occurred, your brain starts telling the immune system that it needs help by sending signals directly to the spleen. The immune cells in the spleen then become activated and inflammatory, leave the spleen and enter the circulatory system, and then head to the brain where they create a hostile, inflammatory environment, which kills off a lot of additional brain tissue that could be saved. All of this happens within the first couple of days following the stroke. Also, these activated immune cells lay down a boundary of scar tissue that contains a specific substance called glycosaminoglycans or GAGs for short, which is essentially the biological equivalent of a brick wall. When neurons encounter inflammation and GAGs they physically pull back from that area in a process referred to as "neuronal die-back." The GAGs act like a boundary that prevents the neurons from ever reentering the region—impeding or preventing patient recovery and healing.

> Our fundamental approach relies on treating patients within the first 36 hours after a stroke has occurred, which we have shown is a critical window following a stroke.

Usually, when immune cells leave the spleen and travel toward the region of damage to exert their hyper-inflammatory cascade, you end up with a lot of collateral damage and an empty spleen. What we predicted, and saw in our Phase II clinical trial, was that if we stopped the hyper-inflammatory cascade from happening, it's like keeping all the horses in the barn, so that they are ready to defend the body against opportunistic infections or other problems. Moreover, importantly, patients that have a stroke, particularly if they are elderly, are very susceptible to secondary infection and other immune-related complications. We predicted that we should see less complications when we treated patients with MultiStem, and that's precisely what we saw.

So, there are lots of reasons to be excited about the Phase II data and our ongoing Phase III and registrational trials. Our partner in Japan is running a pivotal study, which gives us a lot of confidence and optimism that we might very well be on the cusp of changing stroke care as we know it.

MA: You just recently announced results from another challenging area, treating patients with acute respiratory distress syndrome (ARDS). What did that study show?

GVB: ARDS is a severe condition where patients experience severe inflammation and fluid build-up in their lungs. These patients aren't getting enough oxygen, and so they have to be placed on a ventilator and are cared for in the ICU. The clinical trial results from our randomized, double-blind placebo-controlled study showed that when patients with ARDS were treated with MultiStem, there was a meaningful reduction in mortality, and patients also showed increased days off the ventilator and a higher number of ICU-free days than patients being treated under the standard of care. The data also suggests that the treatment effect was more significant in the more severely ill patients.

People in the ARDS field have found these results to be very exciting. Our partner in Japan, Healios, is already launching an ARDS clinical trial there under the new accelerated regulatory framework for regenerative medicine therapies, so we are in a really good position to advance science.

— Moe Alsumidaie, MBA, MSF, is a thought leader and expert in the application of business analytics toward clinical trials, and Editorial Advisory Board member for and regular contributor to Applied Clinical Trials

Optimizing Study Design in Real-World Evidence Generation

Rob Sambrook

With outcomes from "real life" a critical compliment to clinical trial data, the importance of involving an epidemiologist at study inception is explored.

STUDY DESIGN

Real-world evidence (RWE) is needed in addition to clinical trial data to understand drug effectiveness in a real-life setting and to profile patient populations in terms of their clinical characteristics and drug utilization. Studies designed to quantify the incidence and prevalence of health conditions and/or population attributes such as risk factors are, in the scheme of clinical research, relatively straightforward studies to conduct. Clearly, they are not in the same league as randomized controlled trials (RCTs) in terms of their cost, operational complexity, timelines, or degree of risk. Consequently, many pharmaceutical sponsors assume that they can pursue these studies without input from an epidemiologist.

This assumption can be risky, as epidemiologists understand how to minimize bias and ensure external validity—necessary elements of even these "simple" studies.

Let's look at a few of the most basic pharmacoepidemiological study designs and the associated risks of bias.

Cross-sectional studies: The mainstay of incidence/prevalence studies

Cross-sectional surveys (CSS) select participants based on the inclusion/exclusion criteria defined in the study protocol to profile a cohort of individuals at a single point in time. That is to say, there is no baseline established, no outcome measured, and no follow-up data collected. Rather, certain populations characteristics are identified that merely describe the cohort under study at a snapshot in time.

CSS designs inherently lack an ability to examine the temporality of events, e.g., cause and effect.

Cross-sectional surveys are among the simplest and least expensive studies to conduct, and indeed, are very low on the hierarchy of evidence (See Figure 1 on facing page).

Medical chart reviews: Adding a longitudinal perspective

Medical chart reviews (MCRs), in which data are drawn from patient's medical charts (either manually in the case of physical charts or through a database of electronic medical records[EMRs]) can either be used to generate cross-sectional or longitudinal data. When these studies are conducted longitudinally, they may explore causality, providing insights beyond incidence and prevalence. At the same time, it must be acknowledged that medical charts are rarely complete for sponsors' research purposes, given that they were, of course, created for a different purpose.[1]

Hybrid studies: An innovative blend of CSS and MCR designs

Hybrid CSS-MCR studies are a way of overcoming the limitations of one study design by adding another, complementary, study. Hybrid CSS-MCR studies can overcome the limitation of CSS studies with respect to being able to look at causality and outcomes since the patient history recorded in the medical chart can be used as the control, prior to exposure to a variable. And hybrid studies can fill in gaps in the medical record by gathering additional details from patients, including on outcomes and adverse events.

Hybrid CSS-MCR studies are conducted in the same way as individual CSS and MCR studies, but taken together, they allow for a much more comprehensive research investigation.

The caveats

Simple though they may be, CSS, MCR, and hybrid studies can produce robust findings that are generalizable to the population in the real world that will be using the treatment—in other words, they can have external valid-

ity. However, great care must be taken in the study design, and specifically the sampling strategy, to ensure generalizability.

Below are some of the pitfalls that need to be guarded against:

- **Non-response bias**

Cross-sectional surveys are frequent offenders of non-response bias, given the fact that patients who participate in research are different from those who do not. Admittedly, it is very difficult to eliminate all bias in CSS studies, but the goal must be to minimize it.

- **Selection bias**

Selection bias can occur when the sampling is not well considered, such as in studies that deploy a convenience sampling strategy, rather than a probability sampling strategy (or make use of a representative data source). The difference between the two is marked. Probability sampling is achieved when every study participant has an equal chance of being selected for inclusion in the study. This ensures that the study population is representative of the target population since the selection occurs in a random fashion. Understanding representativeness is especially relevant for incidence and prevalence studies and requires an evaluation of how similar the sample population is to that of the population of interest by comparing certain characteristics (often demographic).[2]

Conversely, a non-probability sampling strategy does not enable the study population to have an equal probability of selection into the study, otherwise called non-random selection. CSS studies frequently rely on panels, clinician referrals, patient groups, traditional advertising, proprietary databases, and social media to generate evidence and in so doing are using non-probability sampling. Even if participants are selected from these sources at random, the question always comes back to whether or not the source was representative of the population of interest. Including a random selection process as part of a flawed sampling strategy will not ensure the generalizability of the study estimates.

> Epidemiologists know how to minimize (or avoid) the issues of selection bias, sampling bias, and underpowered studies.

- **Sample size**

In order to test a study hypothesis, the study has to have sufficient "power." In other words, the sample size has to be large enough to ensure that the study estimates are reliable. Conducting a study that is purely descriptive without considering the study size may result in findings that cannot exclude the role that chance may have played. Having a sufficient study size is important to providing confidence that the findings are not owing to sampling bias.

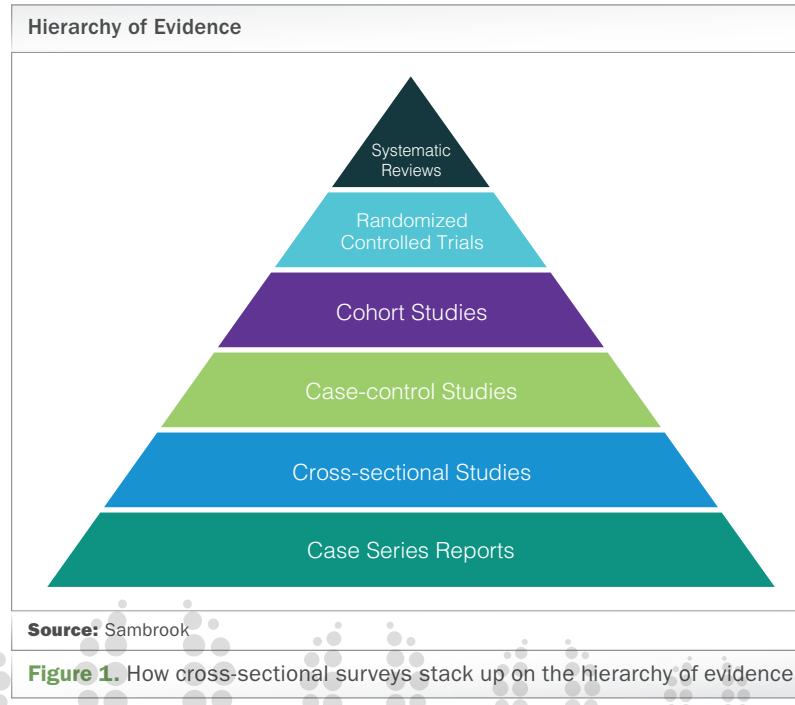

Figure 1. How cross-sectional surveys stack up on the hierarchy of evidence.

- **Confounding**

At study inception, potential confounders need to be identified so that data pertaining to these covariates can be collected at baseline, allowing the statistical modelling approach to control for them. If confounding is not assessed and addressed during the design and analyses stage, any resulting evidence of an association between two variables may be diluted or inflated due to the confounding. The range of the data collected in medical chart and EMR databases may sometimes make it challenging to control for all confounders.

The Value of an epidemiologist

As specialists in study design, epidemiologists know how to minimize (or avoid) the issues of selection bias, sampling bias, and underpowered studies. They will propose a study design and methodology for identifying, recruiting, and selecting participants and deliver robust generalizable data that provide insights into the target population.

In working on a CSS, MCR, or hybrid study, an epidemiologist will:

- Consider the sampling strategy up front as part of the study design.
- Explore real-world data sources and consider those that are fit for the research question.
- Identify a data source that enables the generalizability of the sample population.
- Determine the most appropriate sampling strategy.

SAMPLING PATIENTS AT THE PHYSICIAN LEVEL

Sampling patients based on study inclusion/exclusion criteria at the general practitioner level assumes that all patients would have an equal chance of visiting their primary care physician and therein an equal chance in participating in a research study. This is faulty reasoning, since:

- Certain patients may be more likely to frequent their GP (frequent visitors tend to be older, female, have more psychological distress, or physical disease).[1]
- Those with chronic conditions and multimorbidity use more medical resources.[2]
- Patients may visit their GP in a non-random fashion for prescription refills.[3]

Implementing a recruitment ceiling for sites can only ensure the representativeness of sites themselves, not of the patients visiting those sites.

[1]Vested P, Christensen MB. Frequent attenders in general practice care: a literature review with special reference to methodological considerations. *Public Health*. 2005;119(2):118-37.

[2]Van den Bussche H, Kaduszkiewicz H, Schäffer, I et al. Overutilization of ambulatory medical care in the elderly German population? – An empirical study based on national insurance claims data and a review of foreign studies. *BMC Health Services Research*. 2016;16:129

[3]Jepson M, Salisbury C, Ridd MJ, et al. The "On in a Million" study: creating a database of UK primary care consultations. *The British Journal of General Practice*. 2017;67(658):e345-e51.

- Work with biostatisticians to calculate the appropriate sample size.
- Ensure that any study limitations are minimized and clearly explained.

Unfortunately, if a study design was poorly conceived and allowed for bias, the flaw may not become apparent until after the data are collected and the analysis is underway. At that point, there's very little that can be done to salvage it. Sampling bias, for example, may be discovered when the prevalence measured by the study is significantly higher or lower than what is expected (based on findings from other studies). To prevent undermining the study integrity in this way, it is advisable to involve an epidemiologist at the study inception or at the very least to have the study protocol critiqued by an epidemiologist. With their very specialized focus on study design, epidemiologists can ensure the research design is fit for the research objectives and ensure robust and reliable downstream study findings.

Rob Sambrook *is Divisional Principal, ICON;*
email: Robert.sambrook@iconplc.com

References

1. Gearing RE, Mian IA, Barger J et al. A Methodology for Conducting Retrospective Chart Review Research in Child and Adolescent Psychiatry. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 2006; 15(3):126-34.
2. Safdar, N, Abbo LM, Knoblock MJ et al. Research Methods in Healthcare Epidemiology; Survey and Qualitative Research. *Infection Control & Hospital Epidemiology,* 2016; 37(11):1272-77.

The Framework for FDA's Real-World Evidence Program

Beth Schurman

A look at the scope of the agency's draft framework for evaluating the use of RWE to support new drug approvals and the implications for sponsors.

At the end of 2018, the FDA released a draft framework for a new Real-World Evidence (RWE) Program (see https://bit.ly/2E5ZM68). The agency proposes to use the RWE Program to guide generation of data in support of approval for new indications or to help support post-approval study requirements. It is proposed to be used by both the FDA's drug and biologic review programs. This article highlights the scope of this program, its opportunities and gaps, and the implications for the pharmaceutical industry.

Definitions

To understand the scope of this framework, it is important to understand the definition of real-world data (RWD) and real-world evidence (RWE) as defined by the FDA in the current draft framework:

Real-world data (RWD) refers to patient data that is collected through a variety of sources such as electronic health records (EHRs), medical claims and billing data, data from product and disease registries, and patient-generated data (includes in-home-use settings and mobile devices).

Real-world evidence (RWE) is clinical evidence obtained from the analysis of RWD that provides information about usage, risks, and benefits of a medical product derived from sources other than traditional randomized clinical trials.

Scope of the RWE Program framework

The FDA's RWE Program will assess inclusion of RWD to support approval of new indications and to help support or satisfy post-approval study requirements. The following may be included:

- Addition or modification of an indication—change in dose, dose regimen, route of administration
- Addition of a new population
- Addition of comparative effectiveness or safety data
- Any other post-approval requirements

RWD can be used to supplement traditional clinical trial data and improve the efficiency of these trials. Trials within the RWE Program scope can be one of two types:

Hybrid: Traditional clinical trial supplemented with RWD from EHR, medical claims, or pharmacy databases.

Pragmatic: Contains some elements which closely resemble routine clinical practice.

The hybrid trial approach may include the use of RWD as a control arm in a traditional clinical trial. Observational studies are anticipated within the pragmatic approach. The FDA intends to use an existing guidance, *Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data*, to evaluate epidemiologic studies as applicable for hypothesis generating pragmatic studies.

Current challenges of the proposed framework for the pharma industry

The new draft framework leaves some significant gaps with regard to a defined approach that pharmaceutical manufacturers should take in their RWE trial design for regulatory approval. The guidelines do not yet completely define the RWE trial design approaches that are approvable, nor the regulatory submission process. The FDA was actively soliciting feedback up until Feb. 5 that will be used in its final RWE guidance.

The agency has stated it will review on a case-by-case basis and will use three criteria for evaluation.

1. If the RWD selected is fit for use to generate data for product effectiveness decisions.
2. If the study design used to generate RWE can provide

robust scientific evidence to help address the regulatory question.
3. If the execution of the studies generating RWE meets FDA regulatory requirements.

The last criterium forms a catch-22 as the current FDA regulatory requirements regarding RWE do not specifically outline use in regulatory approval. Several FDA regulations and guidance documents are available for the use of electronic data, eHealth records in clinical investigations, electronic informed consents, and Q&As that could potentially be used for the FDA RWE Program. These include guidances such as *Use of Electronic Source Data in Clinical Investigations* and *Use of Electronic Health Records in Clinical Investigations*, among others. Per the draft guidance: "FDA has gained considerable experience assessing electronic health care data (e.g., EHRs, medical claims data, registries) through experience with the Sentinel System and other data systems." The FDA will be using this experience to guide its perspective on data submissions within the RWE Program.

Pharmaceutical manufacturers must seek guidance from the FDA for each trial design proposed to ensure regulatory processes are met. With this draft framework publication, a number of areas are yet to be defined, according to the FDA. The agency has stated the intention to update and design guidance on how to include RWD to design both randomized and non-randomized hybrid clinical trials as well as observational studies. To achieve this, the FDA directly outlines in the draft guidance the factors that would need to be considered, such as:

- Types of interventions or therapeutic areas for which routine clinical data could be used.
- The reliability and quality of data collected from the routine clinical care settings.
- If outcomes are rare, the number of patients need to be tested.
- How to account for variations in routine clinical practice
- The extent to which randomization can and should be included.

With the eight-week solicitation period for suggestions and feedback ended in early February, a mature form of the framework is expected in the future. The FDA states that the RWE Program "…will involve the establishment of demonstration projects, engagement with stakeholders intended to promote shared learning and consistency in applying the framework, and the development of guidance documents to assist sponsors interested in using RWE to support their work." The RWE framework will also include data standards for collection, analysis, and submission of RWD. The need for RWE data may spur the need to develop a common data model (CDM) in order to work with RWD across different sources. The CDM would unify terminology, vocabularies, and coding schemes. It will be incumbent on pharmaceutical manufacturers and clinicians to remain actively involved in the development of the final FDA RWE guidance.

The FDA intends to review gaps in both the RWE framework and in RWE data systems, as well as the interoperability, and provide strategies to address them. The following summarizes gaps currently identified by the FDA that may need to be addressed while framing guidance for the RWE Program.

1. Lack of enough data showing concordance between the results of traditional clinical trials vs. observational studies.
2. Lack of reporting requirements for various observational studies.
3. Lack of data-recording standards and interoperability across various healthcare data-collection systems.
4. Challenges in integrating data for the same patient across various data sources.
5. Challenges in collecting data from electronic patient reported outcomes (ePROs), wearables, biosensors, and mobile app.

Interestingly, the RWE framework also opens the door for possible inclusion of data derived from outside the US. The continued evaluation and publication of data and systems gaps with strategic recommendations by the FDA will have to address how to support these issues and how to address HIPAA privacy rules within RWE.

> Pharmaceutical manufacturers must seek guidance from the FDA for each trial design proposed to ensure regulatory processes are met.

Further implications for the industry

Pharma will have to consider the most efficient way to approach RWE generation for regulatory submission, given the case-by-case approach from the FDA. Additionally, companies will need to address the agency's stated concern that real-world results would be cherry picked for positive and beneficial data. Some of the questions that the pharma industry should consider with regard to RWE evaluation include:

1. Does the pharma company's current organizational structure (decision ownership and stakeholder involvement) align to efficient execution of RWE generation for regulatory approval?
2. Have sufficient resources been allocated within the organization to handle RWE generation for regulatory approvals?
3. What is the decision-making process to evaluate trade-offs between RWE vs. traditional trials to receive follow-on indication approvals for products?
4. How cost effective would it be to choose the RWE program vs. traditional clinical trials?
5. Will the FDA's acceptance of RWE open the door for more partnerships between pharma and academic centers and how this will be handled?
6. Is the company able to effectively choose between different types of RWD and patient reported outcomes?
7. What discussions should be held with the FDA and when should they be held for RWE trial design and submissions—especially given the current case-by-case approach to review and approval?

Conclusion

The RWE Program opens opportunities for more resource efficient approaches to follow-on indications. Increased definition by the FDA is needed to guide resource investment and trial design approaches to ensure efficient investment. Pharmaceutical manufacturers will likely need to evolve their internal structure and systems, including decision models related to RWE trial design and regulatory interaction to accommodate this new route of approval. We look forward to reviewing and understanding the updated guidance expected from the FDA in the future.

Beth Schurman *is Partner at Herspiegel Consulting*

References

1. Framework for FDA's Real-World Evidence Program: https://www.fda.gov/downloads/ScienceResearch/SpecialTopics/RealWorldEvidence/UCM627769.pdf
2. Statement from Scott Gottlieb, MD, on the FDA's new strategic framework: https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm627760.htm
3. FDA public docket collecting RWE draft guidance feedback: https://www.federalregister.gov/documents/2018/12/07/2018-26546/framework-for-a-real-world-evidence-program-availability
4. Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data: https://www.fda.gov/downloads/drugs/guidances/ucm243537.pdf

Global Standardization of Clinical Research Data

Barbara Jauregui, Lynn D. Hudson, Lauren B. Becnel, Eileen Navarro Almario, Ronald Fitzmartin, Frank Pétavy, Nathalie Seigneuret, James K. Malone, Fang Liz Zhou, Jose Galvez, Tammy Jackson, Nicole Harmon, Rebecca D. Kush

Outlining the latest government, industry, and public health efforts to promote increased adoption of common standards in data collection and sharing.

DATA SHARING

Clinical research is only as effective as its ability to have an impact on health. This impact comes when researchers find breakthroughs, discover new diagnostics or treatments, and identify critical pathways that lead to curing diseases. To maximize their utility, clinical research data should be traceable, accessible, interoperable, reproducible, and of good quality, allowing study findings to be imparted and shared in a clear and understandable way.[1] Unfortunately, today, clinical research data are often collected in a variety of formats, leading to difficulties to effectively share and compare the data under the terms allowed by study participants' consent. This disconnect creates an evidence gap that slows scientific advances, which can result in ineffective and even harmful treatments and diagnostics that continue to be employed in clinical practice.[2]

A significant issue that arises when working with research data is the inability to validate and reproduce findings to demonstrate that the experimental result is in fact true. A survey of over 1,500 researchers conducted by *Nature* in 2016 found that more than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments.[3] This effect is commonly caused by divergence from the protocol and the inability to retrace steps in the process.[4] The landmark article by John Ioannidis in 2005, titled "Why Most Published Research Findings Are False" states: "The greater the flexibility in designs, definitions, outcomes, and analytical modes in a scientific field, the less likely the research findings are to be true."[5]

While irreproducibility of research results in the field of genetics is encouraging greater transparency in methods and materials, along with the analytic codes that underlie the conclusions, this does not appear to be the case for clinical trials. There are also efforts to leverage big data, which may provide information on trends, signals, or hypotheses to be tested further, but generally do not provide results of sufficient adequacy to support regulatory submissions.

Regulated clinical resesearch has become increasingly global, particularly for areas such as rare diseases for which there is a small population of patients spread throughout the world. Efforts to streamline regulatory submissions for new product approvals have encouraged the development, largely through the International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH),[6] to standardize and harmonize the structure of these submissions as eCommon technical documents (eCTD). Such standards are useful not only for sponsors who wish to submit in multiple regions simultaneously, but also for regulators to facilitate reviews. ICH has also provided guidelines for global research on protocols, terminologies, and statistical analyses.

Hidden data

Currently, an estimated 85% of research studies do not translate to a meaningful clinical discovery.[7] The causes for this low level of translation of promising research into meaningful insights and interventions for human health are multiple. One of many examples is the discovery of the relationship between infant sleeping position and sudden infant death syndrome (SIDS). Had it been possible to aggregate and systemically analyze all the evidence available by the year 1970, over 60,000 infant deaths worldwide could have been prevented.[8] Differences in protocols among studies, small sample sizes, numbers of patients,

and families involved per study, and differences in comparisons between SIDS and unaffected infants were among the factors that may have contributed to the delayed recognition of infant positioning on their back while sleeping as a protective factor against SIDS. This is one of many cases where critical health findings were present, but hidden in the data.

Regulatory validation of clinical trial findings involves stringent requirements to ensure that regulators can adequately evaluate the safety and efficacy of the medicinal product. Within the flexibilities afforded by the U.S. Federal Food, Drug, and Cosmetic Act, at least two adequate and well-controlled studies, each convincing on its own, are generally needed to establish effectiveness; similar recommendation was given by the European Medicines Agency (EMA).[9,10,11] Review of trial data includes the "validation" needed to establish that the results have clinical meaning and that the findings are not due to chance alone. Furthermore, the need to provide adequate directions for the use of a drug in relevant subgroups requires an assessment of aggregated data from multiple trials. This regulatory review is facilitated by the use of standards for protocol information, outcome definitions, data terminology, and formats.

Adoption of common standards in research becomes pertinent to the regulatory process as data from early discovery is translated into clinical benefit (e.g., biomarker discovery, mechanistic studies, etc). The terminology standards used in regulatory submissions and healthcare can be similarly adopted in clinical research trials to facilitate this seamless integration of data.[12,13]

How to ensure meaningful exchange of information

Interoperability is "the ability of different information technology systems and software applications to communicate, exchange data, and use the information that has been exchanged."[14] "Semantic interoperability" refers not only to the exchange of information, but also the exchange of meaning such that the recipient of the information can readily understand and interpret the information accurately in the manner intended by the data generator and/or sender.

Recently, FAIR has been cited as an acronym for four requirements that should be provided for a data publishing environment for machines and humans, to support appropriate aspects of data sharing.[15] These FAIR facets are:

- Data should be Findable
- Data should be Accessible
- Data should be Interoperable
- Data should be Reusable

One key to ensuring semantic interoperability and adherence to the FAIR principles or facets is for parties to use the same data standards and terminologies or ontologies. Clearly, the more parties who agree on the data standards and terminologies, the better. This is the rationale behind consensus-building for a robust standards development process.

To maximize the real-world impact of any research study, the data must be collected and analyzed in a common format. Standardization helps build efficient and interoperable research data networks capable of producing high-quality and more reliable data that can support healthcare decisions, detect safety and other signals, and be utilized to generate new hypotheses and new knowledge. It also streamlines research activities by allowing data to be accrued more efficiently, and makes it possible to consolidate digital data available from different sources to support further research and healthcare decisions.[16]

Standardization ultimately leads to more organized evidence, which can be better understood by audiences possessing limited scientific literacy—and increase the ability of researchers and lay people to comprehend and share important findings.

Data standards allow research teams to explicitly name and define the different elements and aspects of their studies. By using standard terms, researchers can precisely describe, manage, and share their data, allowing external research teams to understand what the researchers did, how they did it, how to interpret the results, and accurately reproduce these results in future studies. It also lets researchers perform queries across diverse datasets, which allows for data from different research studies to be consolidated into larger datasets for analysis. In addition to supporting collaboration among researchers, standardization ultimately leads to more organized evidence, which can be better understood by audiences possessing limited scientific literacy. This organization can increase the ability of researchers and lay people to comprehend and share important findings.

There are several clinical research standards in use globally today, which cover the different stages of clinical research. These include those from the Clinical Data Interchange Standards Consortium (CDISC) for clinical and translational research;[17] controlled terminology published through the National Cancer Institute's (NCI) Enterprise Vocabulary Services; MedDRA (Medical Dictionary for Regulatory Activities) for medical history in clinical trials and for adverse events reporting,[18] Health Level Seven (HL7) for structured product labels and ECG waveforms, the International Standards Organization (ISO) for the identification of medicinal products (IDMP); LOINC (Logical Observation Identifiers Names and Codes) for clinical laboratory tests and observations;[19] and the ICH, as previously mentioned. There are also standards that exemplify collaboration among standards development organizations (SDOs) and other entities. For example, the Biomedical Research Integrated Domain Group (BRIDG) Model is a CDISC, HL7, and ISO standard, with NCI and FDA as key stakeholders.[20]

Over the past two decades, CDISC, a global non-profit organization that develops data standards through a volunteer-driven, consensus-based process, has developed a global, open-access suite of clinical

and translational research data standards. These standards support the entire research lifecycle (including preclinical research) from structured protocol information through data collection, exchange, tabulation, analysis, and reporting.[21] Standards specific to certain therapeutic areas have been developed collaboratively through the Coalition for Accelerating Standards and Therapies (CFAST), which has included the Critical Path Institute, CDISC, FDA, NCI, and TransCelerate BioPharma, along with medical experts and patient groups working in these therapeutic areas (TAs). Regulators from Europe and Japan have also contributed to the development of these TA standards.

TA user guides specify how to use these standards to structure the data for research on a given disease or treatment, broadening the circle of collaboration with patient representative groups, research investigators, and public-private partnerships. FDA has published specifications for these TAs in their Study Data Technical Conformance Guide.[22] Working with data in a common format with controlled terminology makes it easier, faster, and more efficient for pharmaceutical companies, CROs, academic organizations, regulators, and other government entities to collaborate on projects.[23] These standards are utilized for both regulated and some non-regulated trials, including interventional and observational studies, nutrition, public health, epidemiology, medical device, and outcomes research. They have even been applied to data from studies on healthy birth, growth, and development.

Data from traditional pharmaceutical, academic, public health, and the healthcare enterprises vary in their level of standardization. This interdependent research continuum highlights the need for standards that translate across the evidence divide.[7] Implementing standards from protocol through analysis stages can enhance the quality and efficiency of clinical research processes and facilitate traceability, particularly when the standards are implemented from the start. Many research teams have made impactful discoveries with the application of data standards in later stages of the research process, but not without significant data transformation effort at the end of the process. For instance, a research team recently conducted a meta-analysis of chemotherapy in head and neck cancer (MACH-NC) by contacting and requesting individual patient data from several published studies. Analyzing the combined data, which included patient and tumor characteristics, dates of failure and death, treatment details, and toxicities, the researchers demonstrated the superiority of concurrent chemotherapy in the treatment of certain cancers, validating the results of the published studies.[24] Their work could have been simplified and enhanced substantially had the different datasets been standardized from the beginning of each individual study.

Standardization allows a significantly faster and less costly avenue for generating evidence and performing robust analyses, by providing the data and processes employed in a common, predictable, and explicit format. A recent research project exploited open-access clinical trial data standardized using CDISC to answer important questions in prostate cancer to save time and reduce costs of the initiative.[25] Data standards also provide great potential for semi-automation of the evidence generation process[26] and for saving substantial human resources and time in the start-up of a clinical trial.[21] If data collection standards are employed from the beginning, study start-up times can be reduced by 70% to 90%, since standard case report forms, edit checks, and validation documentation already exist and many can be reused from trial to trial. Study teams can then focus on protocol-specific additions to the standards, which results in cost savings, faster delivery of results, and higher quality data.[27]

Data standards also facilitate community engagement, data sharing, and transparency. An open-data, crowdsourced project from Project Data Sphere identified predictors for survival in castration-resistant metastatic prostate cancer through prognostic models that used CDISC-standardized data from the comparator arms of four Phase III clinical trials and enabled 50 independent teams.[25] These teams developed a comprehensive set of benchmarked models that uncovered key prognostic variables and novel interactions between them. All method predictions and code from this initiative are available for public use, increasing transparency and facilitating collaboration. Project Data Sphere participants noted that the data provided in a known standard format were easier to interpret and more useful than those that were submitted in proprietary formats.

Responses to epidemics and global public health emergencies, such as outbreaks like Ebola and the Zika virus, realize significant benefit from standards by ensuring that decisions are based on the best available evidence. The earlier treatments can be evaluated, the faster outbreaks can be contained. In 2015, the World Health Organization (WHO) conducted a consultation on research data sharing during public health emergencies. A background briefing for this exercise mentioned multiple opportunities for improvement with regard to data sharing, including the "need to build databases where all data are entered in a uniform way, which can be populated when outbreaks occur and are available worldwide."[28] This solution requires that data standards be available prior to outbreaks. WHO convened a diverse group of stakeholders to discuss the development of global norms and standards for more rapid and transparent data sharing during public health emergencies.[29] Common research data standards have now been collaboratively developed for Ebola,[30] malaria,[31] and influenza,[32] all of which can be leveraged for responding to new outbreaks.[33,34]

Improving drug regulation

Data standards in regulatory submissions supporting new product applications have enabled efficient review through automated validation of data quality. A suite of tools and services for clinical and nonclinical standardized data support high level analysis early in the review process.[35,36] Transparency of the regulatory review processes is enhanced through engagement in the process of standards development and the availability of publicly-shared standard analyses scripts.[37,38] The incorporation of patient-reported outcome (PRO) measures along with the TA standards could draw an even broader set of stakeholders into the process. These standards are freely available and could be adopted to enable the same transformation in all supported clinical research.

Downstream standard development efforts built on standardized data include harmonized research protocol templates and outcomes adapted for TAs. These efforts bring us closer to the possibility of even greater efficiency with master protocols for use in clinical trial networks. FDA and ICH developed a common protocol template concur-

rently with another such development effort by TransCelerate. These templates have now been harmonized and published as one.[39] They are now being "technology-enabled" based upon protocol standards developed previously and incorporated into the BRIDG Model. This common protocol template has already proven to be quite useful in a) ensuring that endpoints to be collected are aligned with protocol objectives; and b) information from the protocol can be re-used across multiple downstream documents such as the statistical analysis plan, the clinical study report, and the product label. These efforts have now led to a new protocol project with ICH.

Exchange of "computable biomedical knowledge" (CBK) is also being studied in academia for providing results of research back to practice as in the final portion of a learning health cycle.[40] The Learning Health Community[41] has an initiative called Essential Standards to Enable Learning (ESTEL),[42] which has published a white paper regarding a framework for LHS standards. These LHS-related efforts do not encourage the development of new standards, rather leveraging those that already exist and building upon them. The NIH has also recently invested funds in a Center for Data to Health (CD2H) to encourage adoption of standards across NIH Clinical and Translational Science Awards (CTSAs) as one goal.[43] Another area ripe for standards adoption is electronic health records (EHRs), which will be better leveraged for research purposes when data can readily be shared in a standard format. FDA has issued recent guidance in this regard.[44]

For research studies intended for regulatory review, concerted efforts have been made to create global guidelines and standards for developing new therapies. The ICH developed guidelines for good clinical practices and formats for new product submissions to regulators for review in Europe, the U.S., and Japan. One key data standard output of ICH was MedDRA, which consisted of a rich and highly specific standardized medical terminology, created to facilitate sharing of regulatory information internationally for medical products used by humans. Global data standards for regulated clinical research were collaboratively developed to complement the ICH work, for example, the clinical trial registry (CTR) standard,[45] which can be used to register clinical trials in the NIH/NLM ct.gov, the WHO International Clinical Trial Registry Platform (ICTRP),[46] and the EMA's EudraCT.[47] The European Innovative Medicines Initiative (IMI) also encouraged the use of standards for the research studies they fund by offering a "standards starter pack" as a reference.[36]

Improving policymaking through research standards

Governmental authorities, international public health sponsors and advocates, biomedical research consortia, professional medical societies, and advisory committees charged with recommending ways to improve the efficacy and safety of medicines and other health technologies have promoted data sharing as a way to improve research. At the time of the this writing, the NIH is drafting guidelines to foster the development of scientific evidence with explicit, transparent, and consistently reported methods allowing: 1) decisions to be traced to the underlying evidence; 2) additional analyses of the dataset that may be required for decision-making; 3) new knowledge and insights to be gained through the analysis of pooled data; and 4) routine updating of systematic reviews across studies as new evidence becomes available.[48] The U.S.'s 21st Century Cures Act[49] encourages FDA to develop ways to leverage real-world data (e.g. from EHRs and mobile devices) to augment clinical trial data and specifically referenced CDISC as a standards setting body. The Patient-Centered Outcomes Research Institute (PCORI)[50] has funded, through its Trust Fund, a cross-agency project led by FDA to facilitate the use of real-world data through the harmonization of common data models (CDM) that have been adopted by various research networks, including PCORNet, ODHSI/OMOP, and Sentinel. The "Cures" legislation did not, however, mandate use of standards for federally-funded academic clinical trials.

More generally, funding agencies also have established data-sharing policies, though few require the use of data standards over the course of conducting the funded research. While trials that meet criteria for submission to electronic clinical trial registries will need some degree of protocol description or adverse event standardization, aggregation and secondary use of full datasets is inhibited due to the absence of a requirement that funded researchers utilize standards. As long as federal-funding agencies do not have similar mandates or guidelines for standards as do regulatory agencies, sharing of data between or among agencies is hindered.

> Responses to epidemics and global public health emergencies, such as outbreaks like Ebola and the Zika virus, realize significant benefit from standards by ensuring that decisions are based on the best available evidence.

Some funding agencies have taken another approach—to standardize data from researchers to common structures and semantics. The U.S. National Institute of Allergy and Infectious Disease (NIAID) has created a data warehouse that utilizes CDISC's data-collection and aggregation standards to model and standardize their funded clinical trial data from diverse sources;[51,52] also, NIAID is funding the development of a TA standard and implementing CDISC standards for global research studies. Similarly, NIAID's ImmPort database,[53] which aggregates information from diverse translational or clinical immunology studies, uses CDISC to structure data extracts to support secondary use.[54] These platforms maximize the NIAID investment in research by providing sources of data that share common meaning. Their data can be readily utilized for meta-analyses with similar regulated trials, as the FDA requires use of CDISC standards for submissions, but adoption and use of a common standard within academic federal funding agencies' systems is not yet common globally. Thus, policymakers have the opportunity to multiply the value of federally-funded and regulated trials by not only making provision for data sharing, but also by requiring global clinical research standards.

Contribute to research data standardization efforts

Getting from where we currently operate to a place where standardized research data around the world can truly talk to each other is a great challenge and an immense opportunity. We have a collective responsibility to contribute to this effort; global stakeholders have different roles to play. Researchers and sponsors alike should become aware that the initial training and time required to implement data standards is more than worth the effort, since standards simplify the regulatory submission process, while enabling the data to be repurposed within and outside their research teams. Furthermore, regulatory agencies could continue increasing the amount of information—publicly or via controlled access—from regulatory submissions, following the example of EMA, to allow examination from different parties and enable the wider scientific community to conduct research and answer more questions using the increasingly available data. Coupled with the use of standardized data, it should eventually lead to higher quality submissions and regulatory reviews.[55]

National and international health policymakers have the responsibility to demand a broader evidence base to support their decisions and recommendations, as well as a more rigorous approach for evidence synthesis presented to them or developed by their teams. As FDA and Japan's Pharmaceuticals and Medical Devices Agency (PMDA) have done, national entities, such as the 27 different institutes and centers that comprise the NIH in the US, should avoid unnecessary duplication of efforts and coordinate around existing robust standards that are maintained by global standards development organizations. There are several examples of global standards used within NIH. The National Human Genome Research Institute (NHGRI) relies heavily on the use of international standards to annotate genetic and phenomic data. Without the use of standards such as the Gene Ontology (GO) and the Human Phenotype Ontology (HPO), scientists would not be able to directly compare scientific results. Furthermore, as new discoveries are made, these same scientists contribute back to the ontologies to maintain the standards. Another example of NIH involvement with standards bodies is the Genetic and Rare Diseases Information Center (GARD), which relies heavily on SNOMED, ICD, and Orphanet to find and share resources.

National policymakers should form a team of technical experts to evaluate the best avenues for implementing data standards, adopting and encouraging the use of existing international standards whenever possible, to pave the way for global data exchange. International policymakers, in turn, should promote the adoption of global data standards as means of accelerating and enhancing collaborations among international partners for greater global impact of research. International policymakers are also responsible for providing technical support to countries in the progressive implementation of research data standards, so countries can make more informed national decisions and contribute to the global pool of standardized data. Entities that are part of the healthcare system should continue efforts to bridge the gap between clinical practice and research while implementing data standardization as well.

Imagine a world in which research data can be shared and aggregated seamlessly such that the power of that data can be maximized to accelerate collaborative learning and streamline the path to new therapies. We have an ethical imperative to adopt and leverage robust global data standards that will improve the way research is conducted to benefit all patients.

***Barbara Jauregui**, is an International Consultant, Pan American Health Organization/World Health Organization; **Lynn D. Hudson** is Chief Science Officer, Critical Path Institute (C-Path); **Lauren B. Becnel** is Senior Director, RWDnA & Data Strategy—Oncology Client Partner, Pfizer; **Eileen Navarro Almario** is Associate Director for Clinical Affairs in the Office of Computational Science, OTS, CDER, FDA; **Ronald Fitzmartin** is Data Standards Staff, Office of the Director, CBER, FDA; **Frank Pétavy** is Head of Biostatistics and Methodology Support, EMA; **Nathalie Seigneuret** is Senior Scientific Project Manager, Innovative Medicines Initiative; **James K. Malone** is Senior Medical Director, Eli Lilly and Company; **Fang Liz Zhou** is Director, Global Medical Evidence Generation; **Jose Galvez** is Chief, Office of Biomedical Translational Research Informatics (BTRIS), NIH; **Tammy Jackson** is Senior Director, Clinical Innovation, PPD Inc.; **Nicole Harmon** is Chief of Staff, Clinical Data Interchange Standards Consortium (CDISC); **Rebecca D. Kush** is President, Catalysis Research; Scientific Innovation Officer, Elligo Health Research, and Fellow, Translational Research Center for Medical Innovation, Foundation for Biomedical Research and Innovation, Kobe, Japan*

Authors include representatives of PAHO/WHO, Critical Path Institute, FDA (CBER and CDER), EMA, NIH, IMI, ACRO, CDISC, Lilly (TransCelerate), Sanofi (DataSphere), Pfizer, and Elligo Health Research.

References

1. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data*. 2016 Mar 15;3:160018.

2. Kush R, Goldman M. Fostering responsible data sharing through standards. *N Engl J Med*. 2014 Jun 05;370(23):2163-5.

3. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature*. 2016.

4. Weissgerber TL, Garovic VD, Winham SJ, Milic NM, Prager EM. Transparent reporting for reproducible science. *J Neurosci Res*. [Editorial]. 2016 Oct;94(10):859-64.

5. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005 Aug;2(8):e124.

6. International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH). Homepage. [28 November 2018]; Available from: https://www.ich.org/home.html.

7. Ioannidis JP. How to make more published research true. *PLoS Med*. [Research Support, Non-U.S. Gov't]. 2014 Oct;11(10):e1001747.

8. Gilbert R, Salanti G, Harden M, See S. Infant sleeping position and the sudden infant death syndrome: systematic review of observational studies and historical review of recommendations from 1940 to 2002. *Int J Epidemiol*. [Meta-Analysis Review]. 2005 Aug;34(4):874-87.

9. Government USF. Drug Amendments Act of 1962, Public Law 87-781, 76 STAT 780. 1962.

10. Drug USDoHaHSFa. Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products. 1998: Available from: https://www.fda.gov/ucm/groups/fdagov-public/@fdagov-drugs-gen/documents/document/ucm072008.pdf.

11. The European Agency for the Evaluation of Medicinal Products. COMMITTEE FOR PROPRIETARY MEDICINAL PRODUCTS (CPMP) - POINTS TO CONSIDER ON APPLICATION WITH 1. META-ANALYSES; 2. ONE PIVOTAL STUDY. London2001.

12. Institute NC. Common Terminology Criteria for Adverse Events v.4.032010: Available from: https://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_

13. Shamsuzzaman M, Patel, T., Navarro Almario, E., Wu, C., Tesfaldet, B., Fleg, J., Csako, G., et al. Abstract 18043: Identifying Predictors for All-Cause Mortality in Diabetic Patients in the ACCORD Trial Using Random Survival Forests Circulation. *Circulation*. 2017;136(Suppl 1).

14. HIMSS. Definition of Interoperability. [28 November 2018]; Available from: https://www.himss.org/sites/himssorg/files/FileDownloads/HIMSS%20Interoperability%20Definition%20FINAL.pdf.

15. FORCE11. GUIDING PRINCIPLES FOR FINDABLE, ACCESSIBLE, INTEROPERABLE AND RE-USABLE DATA PUBLISHING VERSION B1.0. [28 November 2018]; Available from: https://www.force11.org/fairprinciples.

16. Califf RM, Robb MA, Bindman AB, Briggs JP, Collins FS, Conway PH, et al. Transforming Evidence Generation to Support Health and Health Care Decisions. *N Engl J Med*. 2016 Dec 15;375(24):2395-400.

17. CDISC. Clinical Data Interchange Standards Consortium (CDISC) Website Homepage. 2017 [December 1, 2017]; Available from: https://www.cdisc.org/.

18. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). Drug Saf. 1999 Feb;20(2):109-17.

19. Institute R. Logical Observation Identifiers Names and Codes (LOINC) Website Homepage. 2017 [December 1, 2017]; Available from: https://loinc.org.

20. BRIDG. Biomedical Research Integrated Domain Group (BRIDG) Model [28 November 2018]; Available from: https://bridgmodel.nci.nih.gov.

21. Rozwell C KR, Helton E. CDISC Standards: Enabling Reuse without Rework. *Applied Clinical Trials*. 2006.

22. Center for Biologics Evaluation and Research (CBER). STUDY DATA TECHNICAL CONFORMANCE GUIDE. [28 November 2018]; Available from: https://www.fda.gov/downloads/forindustry/datastandards/studydatastandards/ucm384744.pdf.

23. Rozwell C KR, Helton E. Saving time and money. *Applied Clinical Trials*. 2007;16(6):70-4.

24. Eisbruch A. Meta-analyses and Systematic Reviews: Can They Solve Contentious Clinical Questions? *Cancer J*. 2017 Mar/Apr;23(2):84-5.

25. Guinney J, Wang T, Laajala TD, Winner KK, Bare JC, Neto EC, et al. Prediction of overall survival for patients with metastatic castration-resistant prostate cancer: development of a prognostic model through a crowdsourced challenge with open clinical trial data. *Lancet Oncol*. 2017 Jan;18(1):132-42.

26. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med*. 2010 Sep 21;7(9):e1000326.

27. CDISC. Business Case for CDISC Standards. 2014: Available from: https://www.cdisc.org/2014-business-case-use-cdisc-standards.

28. Goldacre B HS, Mahtani KR, Heneghan C. WHO consultation on Data and Results Sharing During Public Health Emergencies. Oxford: Centre for Evidence-Based Medicine. 2015.

29. World Health Organization. An R&D Blueprint for Action to Prevent Epidemics. Geneva: World Health Organization. 2016.

30. CDISC. CDISC Ebola Therapeutic Area User Guide. 2016: Available from: https://www.cdisc.org/standards/therapeutic-areas/ebola.

31. CDISC. CDISC Malaria Therapeutic Area User Guide v1.0. 2017: Available from: https://www.cdisc.org/standards/therapeutic-areas/malaria.

32. CDISC. CDISC Influenza Therapeutic Area User Guide v1.1. 2017: Available from: https://www.cdisc.org/standards/therapeutic-areas/influenza.

33. IDDO IDDO. New CDISC data standardaids development of therapies for Ebola virus. IDDO; 2017.

34. WWARN WARN. Launch of the first malaria therapeutic area data standard and new Case Record Form (CRF). WWARN; 2017.

35. Sakushima K, Sakaguchi, H., Chen, T.J. Electronic Study Data submission for New Drug Application in Japan. 2017: Available from: www.phusewiki.org/docs/2017_CSS_US/PP15_Final.pdf.

36. Law D, Koussis, P., Allard, C., Sviglin, H., Ho, J., Li, J., Kropp, T., et al. JumpStarting Review: Highlights. FDA/PhUSE US Computational Science Conference Proceedings: FDA/PhUSE US CSS Symposium 2017, March 14-17 [serial on the Internet]. 2017: Available from: www.phusewiki.org/docs/CSS2015Presentations/PP21FINAL.pdf.

37. Nilsson MaB, N. Standard Analyses and Displays for Common Data in Clinical Trials - The Journey Continues! FDA/PhUSE US Computational Science Conference Proceedings: FDA/PhUSE US CSS Symposium 2017, March 19-21 [serial on the Internet]. 2017: Available from: www.phusewiki.org/docs/2017_CSS_US/PP14_Final.pdf.

38. Tu H, and Soukup, M. Reducing Duplicated Effort on Standard Analyses through Collaboration - Roadmap for Standard Analyses and Code Sharing Working Group. FDA/PhUSE US Computational Science Conference Proceedings: FDA/PhUSE US CSS Symposium 2017, March 19-21 [serial on the Internet]. 2017: Available from: www.phusewiki.org/docs/2017_CSS_US/PP18_Final.pdf.

39. TransCelerate. TransCelerate BioPharma Common Protocol Template. [28 November 2018]; Available from: https://transceleratebiopharmainc.com/assets/common-protocol-template.

40. Medical School University of Michigan. Mobilizing Computable Biomedical Knowledge. [28 November 2018]; Available from: https://medicine.umich.edu/dept/lhs/service-outreach/mobilizing-computab.

41. Community LH. Homepage. [28 November 2018]; Available from: http://www.learninghealth.org/.

42. Community LH. Community Initiatives. [28 November 2018]; Available from: http://www.learninghealth.org/sample-initiative-1.

43. (CTSA) CaTSAP. Homepage. [28 November 2018]; Available from: https://ctsa.ncats.nih.gov/cd2h/.

44. Food and Drug Administration. Use of Electronic Health Record Data in Clinical Investigations Available from: https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM501068.pdf.

45. CDISC. Clinical Trial Registry XML (CTR-XML). [12 December 2018]; Available from: https://www.cdisc.org/standards/data-exchange/ctr-xml.

46. World Health Organization. WHO International Clinical Trials Registry Platform Available from: https://www.who.int/ictrp/en/.

47. EudraCT. Homepage. [12 December 2018]; Available from: https://eudract.ema.europa.eu/.

48. Valkenhoef G, Tervonen T, Brock B, Hillege H. Deficiencies in the transfer and availability of clinical trials evidence: a review of existing systems and standards. BMC Med Inform Decis Mak. [Research Support, Non-U.S. Gov't Review]. 2012 Sep 04;12:95.

49. U.S. Food and Drug Administration. 21st Century Cures Act and FDA [12 December 2018]; Available from: https://www.fda.gov/RegulatoryInformation/LawsEnforcedbyFDA/SignificantAmendmentstotheFDCAct/21stCenturyCuresAct/default.htm.

50. Patient-Centered Outcomes Research Institute. Homepage. [28 November 2018]; Available from: www.pcori.org.

51. Singh J, Kandaswamy, H., Otoo, J.D., Gumne, P., Duvenhage, M., Whalen, C.J., and Tartovsky, M. CDASH and SDTM for Clinical Trials and Reporting. Proceedings of the 2017 National Institutes of Health Research Festival [serial on the Internet]. 2017: Available from: https://researchfestival.nih.gov/2017/posters/cdash-and-sdtm-clinical-tr

52. Newell KO, Xiao, J., Duvenhage, M., Singh, J., Gumne, P., Kandaswamy, H., Whalen, C., et al. Using a data warehouse model based on CDISC SDTM for centralized data and safety reporting for clinical trials. National Institutes of Health. Research Festival [serial on the Internet]. 2016: Available from: https://researchfestival.nih.gov/2016/posters/using-data-warehouse-model.

53. Bhattacharya S, Andorf S, Gomes L, Dunn P, Schaefer H, Pontius J, et al. ImmPort: disseminating data to the public for the future of immunology. Immunol Res. 2014 May;58(2-3):234-9.

54. Shankar RD, Bhattacharya S, Jujjavarapu C, Andorf S, Wiser JA, Butte AJ. RImmPort: an R/Bioconductor package that enables ready-for-analysis immunology research data. Bioinformatics. 2017 Apr 1;33(7):1101-3.

55. Davis AL, Miller JD. The European Medicines Agency and Publication of Clinical Study Reports: A Challenge for the US FDA. JAMA. 2017 Mar 07;317(9):905-6.

Non-Data Scientists: The Evolving Role of Clinical Data Management

The "citizen data scientist" is the person with no official data scientist training who uses the latest tools and technologies to handle data wrangling duties, analyze data, and create reports and models.

Derek Lawrence
Operational Service Lead, Clinical Data Management, Rho, Inc.

The world of data management today is practically synonymous with electronic data capture (EDC). Data management staff spend most of their time not directly programming systems but maintaining the system itself—completing forms or reconciling queries. As data is changing—and it is, with massive increases in available data sources—we must consider what traditional data management activities look like on studies designed using only external data sources. As trials move to more agile means of data capture, including electronic medical record (EMR) and electronic health record (EHR) systems, biosensors, wearables, bring your own device trials, and more, EDC expertise risks becoming outdated.

A recent Impact Report from Tufts Center for the Study of Drug Development (CSDD) surveyed sponsors and CROs about data management and found a decrease in the prevalence of EDC as a primary point of data capture for clinical trials. The decrease isn't stark—yet—but there is a downward trend. Data management is starting to show strain by trying to combine traditional methods and approaches with more recent advances in tools, data diversity, volume, and velocity.

The same report showed that now there is an average of six apps used to support each study, and that newer sources of data—EMR and EHR systems, biosensors and wearables, to name a few—are predicted to rise two- or three-fold in the next three years. These types of devices send massive amounts of data that are too vast and complex to be distilled into a simple row-by-column spreadsheet. Incredibly complicated data requires a multidimensional data structure—that cannot be reviewed manually—and needs to be addressed differently than standard EDC.

All signs point to a shift in how data management is changing and adapting to new challenges and new tools. Companies are looking for specialized data scientists, instead of just data management. These positions have a heavier technical bent and are prepared to take over more complicated tasks. However, finding trained data scientists prepared to tackle these complex, multidimensional data structures, is not always easy.

Hence, there's been increased attention on a newer industry term, the "citizen data scientist." Our industry hasn't agreed on an official definition, but in short, the citizen data scientist is the person with no official data scientist training who uses the latest tools and technologies to handle data wrangling duties, analyze data, and create reports and models.

The below percentages reflect a break down of what data scientists spend the most time doing:

- Building training sets (3%)
- Cleaning and organizing data (60%)
- Collecting data sets (19%)
- Mining data for patterns (9%)
- Refining algorithms (4%)
- Other (5%)

The tools and technology available today make it possible for non-data scientists, also known as citizen data scientists, to do some of the same work as data scientists. With the right tools and technology, a non-data scientist can manage cleaning and organizing data, leaving the trained data scientists to the more complicated work of building training sets, mining data for patterns, and refining algorithms. Some of these data visualization tool examples include SAS Visual Analytics, D3.js, Tableau, or even potentially home-grown systems, if you have in-house resources (like RhoVer). Additionally, data science/analytics online courses, such as DataCamp and Coursera, can be a good resource for potential citizen data scientists.

As data management experts, we must resist the urge to wait and see which new tool and technology will clear the next path for our staff. We need what is referred to as "adaptagility"—fluid, differentiated, unorthodox thinking that will help build a new model for data while replacing the old.