

APRIL 2017

Clinical Researcher™

The Authority in Ethical, Responsible Clinical Research

Using EHR Data
Extraction to
Streamline the
Clinical Trial
Process:
See page 2

Technology for Trials:

*Are You Falling Behind,
Keeping Up, or
Getting Ahead?*

INCLUDING the Latest in
Workforce Innovation

Using EHR Data Extraction to Streamline the Clinical Trial Process

PEER REVIEWED

Jennifer Stacey | Maulik D. Mehta, MBA

[DOI: 10.14524/CR-17-0004]

Since 2005, the average time from approval by the U.S. Food and Drug Administration of an Investigational New Drug application to a New Drug Application approval has been 8.1 years. From 2003 to 2013, the cost to develop an approved new drug has more than doubled from more than \$1 billion to nearly \$2.6 billion.¹

Much of the cost and slowness of the overall process is a result of difficulties in recruiting appropriate patient populations. Recent research shows that only 13% of investigative sites exceed their enrollment, and that initial Phase II–IV study timelines are often doubled to reach study enrollment goals.² This has resulted in unnecessary protocol amendments that cause delays and dramatically increase costs of developing new therapies.

The three main players in the clinical trial process—biopharmaceutical firms, contract research organizations (CROs), and healthcare organizations—face obstacles as they navigate through the difficult waters of bringing new drugs to market. For instance:

- **Biopharmaceutical firms** lack real-time data, so site selection is often relationship-driven and susceptible to site failures. Clinical investigators are prone to overestimation of patient availability, which leads to under-enrolled study sites. Overly restrictive eligibility criteria, among other trial characteristics, also make some protocols unfeasible.

Further, protocol amendments pose one of the greatest obstacles to effective clinical trial execution. Amendments are costly, time-consuming solutions to underlying clinical trial issues such as increasingly complex protocol design and difficulty recruiting patients. Nearly two-thirds of protocols require at least one

substantial amendment, and a typical protocol ends up with an average of 2.3 amendments. On average, the cost of a single protocol amendment is \$453,932 and the total cost for sponsors to implement “avoidable” protocol amendments is nearly \$2 billion annually.³

- **CROs** are challenged when inclusion/exclusion criteria are chosen without verifying the impact on availability of a cohort, which can create avoidable amendments. The possibility of underbidding the project also increases their risk. CROs strive for competitive differentiation, but the lack of tools to leverage clinical and health-related data can be a barrier to winning more business. CROs endeavor to help their pharma clients develop more pragmatic operational solutions, but require real-world data for better protocol design and feasibility studies.
- **Healthcare organizations** seek to attract more clinical trials—both to generate additional revenue and to help develop new therapies. Unfortunately, competition is increasing for a shrinking pool of National Institutes of Health (NIH) funding and grant funding rates in general are declining. The number of newly registered NIH-funded trials decreased 24% from 2006 to 2014. At the same time, competition from new research areas has increased.⁴

HS

LEARNING OBJECTIVE

After reading this article, participants should be able to discuss the benefits of utilizing EHR technology in planning and conducting clinical trials.

DISCLOSURES

Jennifer Stacey; Maulik D. Mehta, MBA:
Employee of TriNetX Live™

Increasingly, the answer is to extract real-time patient clinical data residing in healthcare organization electronic health records (EHRs). Leveraging these detailed data allows pharma companies and CROs to identify patients who match exactly the eligibility criteria for the cohort they are seeking.

EHR Data are Key

The traditional clinical trial process is broken. The question is how to utilize technology to optimize the process.

Increasingly, the answer is to extract real-time patient clinical data residing in healthcare organization electronic health records (EHRs). Leveraging these detailed data allows pharma companies and CROs to identify patients who match exactly the eligibility criteria for the cohort they are seeking. EHRs are transactional systems, optimized for capturing and quickly retrieving individual observations about single patients.

Combing through individual records to find groups of patients is something most forms of EHRs do not support well, if at all. The class of software tools designed to identify patient cohorts relies on data extracted from EHRs and transformed to allow nimble cross-patient searching. The data “liberated” from EHRs frequently represent a subset of all available patient information, are typically limited to observations stored as discrete elements, and are therefore easy to extract.

Cohort identification tools use the extracts of data to provide a first pass at defining patient cohorts that match the criteria of interest. These cohorts are “coarse,” and require additional refinement. Nonetheless, cohort identification tools eliminate the need to “boil the ocean” to find the specific patients required by significantly narrowing the target population to be reviewed, screened, and eventually enrolled into a trial.

A data-based approach reduces overall site attrition and results in fewer sites with more applicable patients. Ultimately, it will decrease the overall cost and accelerate the development of new drug therapies.

Emerging Enabling Technology

Some providers are already using healthcare information technology (IT) solutions to conduct clinical trial design and site feasibility studies. Although many of these data analytic offerings provide access to large patient populations, these solutions (e.g., data aggregators) are typically based on centralized data sets in single institutions.

For example, the Case Comprehensive Cancer Center at Case Western Reserve University in Cleveland, Ohio has developed an automated tool that matches patients with ongoing clinical trials at the

point of care. Using this tool, physicians were able to facilitate patient enrollment in active clinical trials in conjunction with existing clinical workflows.⁵ This ability to find the types of patients that exactly meet trial criteria quickly and easily illustrates the benefits of EHR data extraction technology.

Success in single institutions highlights the power of extracting and leveraging EHR data. The key to industry-wide success, however, is expanding this enabling technology to include larger databases collected from multiple healthcare organizations, broadening the scope of data they make available (e.g., biomarkers, imaging, information “locked” in narrative text of notes and reports, etc.), and increasing adoption of these tools across the spectrum of the biopharma research enterprise.

The nearly universal adoption of EHR technology, maturing standards and interoperability, a desire to use accumulating clinical data to improve care delivery, and growing appreciation that data collaboration is ultimately required to realize its full potential have opened the door to the widespread sharing of EHR data that represents the next step in improving the clinical trial process. Up to now, there hasn’t been a real-time patient data resource available to help develop protocols and recruit patients. Pharma companies have been forced to use epidemiology data, which are often several years old or worse before being published, and may no longer be relevant.

Technology solutions now allow pharma companies and CROs to access EHR data from healthcare organizations globally on a near real-time basis. Advances have made it possible to study patient data securely. Companies can query de-identified, federated databases to research actual patients by reviewing aggregated EHR-based patient records. They can alter eligibility criteria, instantly see the effect on their overall cohort, and learn whether relevant sites have access to sufficient number of eligible patients. They also can identify problems with inclusion/exclusion criteria earlier during protocol development, significantly reducing the cost and delays caused by protocol amendments.

This new technology protects patient privacy by providing de-identified data during research, then allowing re-identification only after a healthcare organization has agreed to participate in a trial. This greatly improves the recruitment phase of the trial process.

Success in single institutions highlights the power of extracting and leveraging EHR data. The key to industry-wide success, however, is expanding this enabling technology to include larger databases collected from multiple healthcare organizations.

The cost to access a user-friendly EHR platform may strain the budgets of many small pharma companies or CROs, but affordable pricing models are becoming available to address this issue.

A Few Words of Caution

However, while EHR data offer many advantages to clinical research, some downsides exist. Extreme diligence is required to shield sensitive protected health information from cyber breaches, some data types may be missing from a given EMR, and coherent, consistent policies and practices for secondary use of EHR data need to be developed worldwide.

Further, the cost to access a user-friendly EHR platform may strain the budgets of many small pharma companies or CROs, but affordable pricing models are becoming available to address this issue. Despite concerns with EHR usage in clinical research today, the advantages of using this “big data” still outweigh these few current drawbacks.

Mapping Disparate Data to Enable Collaboration

A core element of cohort identification based on federated databases of EHR data is the mapping of disparate clinical data coding standards to a common terminology for ease of use and seamless research collaboration. This eliminates the need for healthcare organizations, pharma companies, and CROs to struggle with translating coding language from multiple systems and organizations.

Clinical data captured by EHRs and extracted for cohort identification is typically coded, meaning that individual data elements are assigned codes from relevant controlled terminology, or coding systems like ICD-10-CM, ICD-10-PCS, and CPT. Some data elements, while coded, are used under different standards at different organizations (i.e., providers of medication standards include Walters Kluwer’s Medi-Span, Cerner’s Multum, First DataBank, and others).

To provide interoperability, disparately coded data must be mapped to a unified set of standards. The mapping process can be costly, since current standards are at different stages of maturity and have varying levels of support and relevant tooling for mapping. A typical mapping exercise requires extensive manual review by terminology experts to ensure high quality. In addition, every mapping is dynamic, in that the effort requires ongoing maintenance due to changes in both the underlying source data and the target standard terminology.

In short, harmonizing data to a unified set of standard terminologies is a necessary step in enabling the functions of cohort identification tools and is a key feature of the new technology.

Using EHR Data to Avoid Costly Amendments

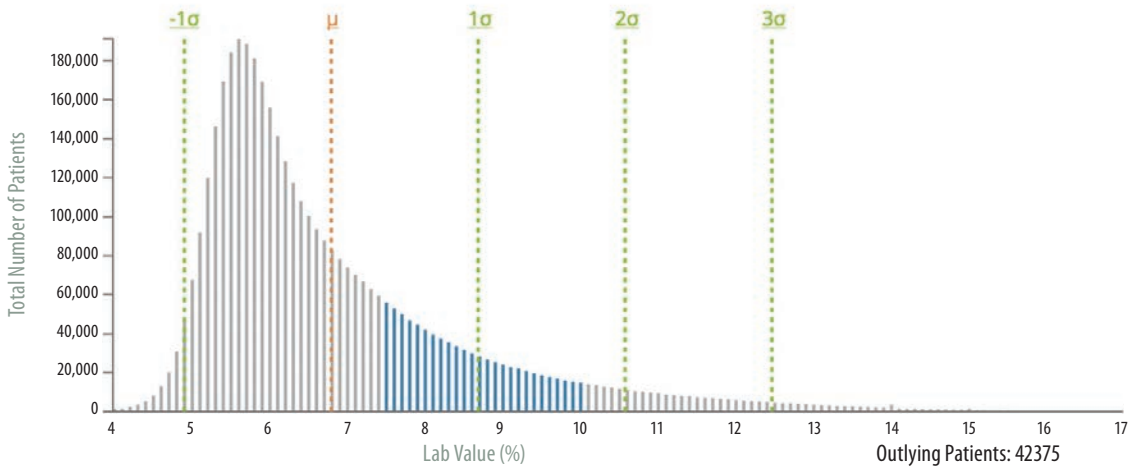
Some organizations have already begun using federated EHR data from multiple healthcare organizations to develop their protocols and recruit patients, and early results are encouraging. Planners, investigators, protocol writers, and strategy teams have been able to move recruitment planning upstream to align with the clinical design process. This has helped to ensure trial feasibility and reduce the number of preventable clinical trial amendments.

ICON, a CRO based in Ireland, was able to leverage EHR data from a global research network to support a bid defense for a European pharmaceutical company. The firm had been initially dropped from consideration, but was later able to become a viable contender because of its use of real-time EHR data.

At the bid defense, ICON presented an HbA1c sensitivity analysis, as the client was contemplating changing the lower range of its protocol from 7.5 to 6.5. Using the cohort identification technology, ICON was able to quickly run the analysis at both 6.5 and 7.5, and found that the difference in the number of matching patients was only 30 for that specific cohort (see Figure 1). Since the cohort already had more than 8,000 matching patients, ICON recommended that the client keep the study entry criterion at 7.5. Another CRO had advised changing the criteria, but the client was hesitant, as its entire program had been based on the 7.5 criterion. The client was pleased that ICON had been able to quickly provide real data from real patients to justify keeping the original higher threshold.

In another case, ICON was able to help a U.S. client determine triglyceride parameters to use as an inclusion criterion for a large cardiovascular trial being planned. In this situation, ICON, again using cohort identification technology, was able to show the full distribution of triglyceride lab results across a large representative population. It then adjusted the upper range so the client could see the effect on the patient population that still met the target cohort size (see Figure 2).

FIGURE 1: HEMOGLOBIN A1c SENSITIVITY ANALYSIS



A comparison of 6.5–10% vs. 7.5–10% HbA1c lab results. While the lab criteria alone yield a larger number of patients for the broader 6.5–10% range, the overall effect of modifying the lower range from 7.5% to 6.5% was negligible (less than 0.4% difference) when applied to the full study inclusion and exclusion requirements.

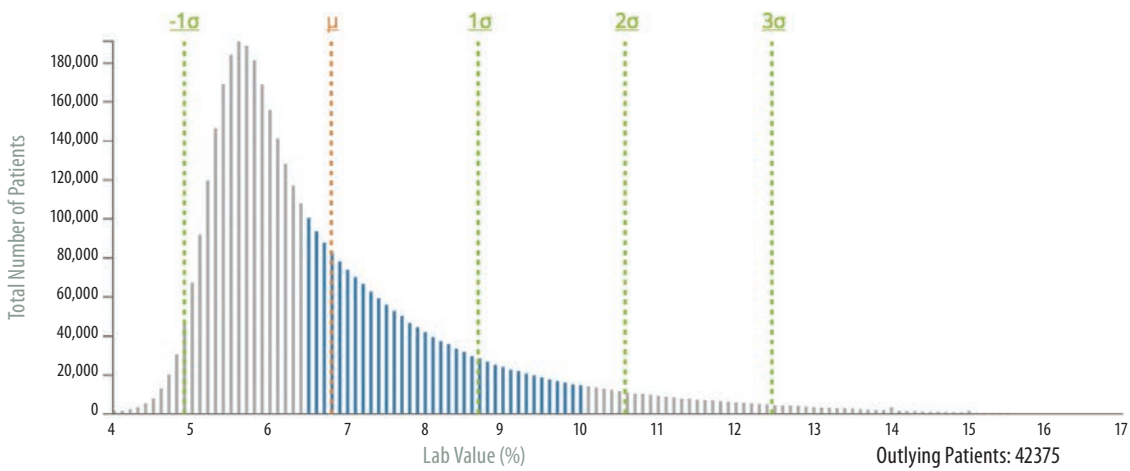
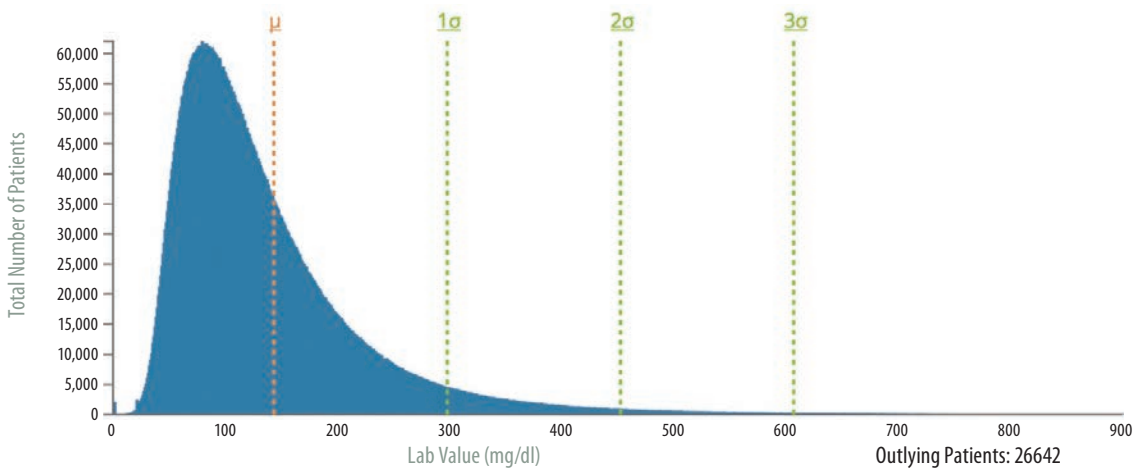


FIGURE 2: TRIGLYCERIDE LAB



A view of the number of patients according to their most recent triglyceride lab results. The EHR platform allowed the user to enter various value ranges to determine the best-suited triglyceride parameters for inclusion criteria for an upcoming study.

| | Min | Max | Median | Mean(μ) | Std. Dev(σ) |
|---------------|-----|----------|--------|---------------|----------------------|
| Network Stats | 0.0 | 20,957.0 | 113.00 | 143.37 | 154.44 |

Greater than or equal to

 mg/dl

Less than or equal to

 mg/dl

Most Recent Result

Ever

FIGURE 3: ORIGINAL VS. EXPANDED AMENDED DISEASE CRITERIA

14,715 Patients

Population 5–17 years/any gender (7,127,901)

| | |
|----|---|
| OR | M08 Juvenile arthritis |
| OR | L93 Lupus erythematosus |
| OR | M32 Systemic lupus erythematosus |
| OR | M33 Dermatopolymyositis (juvenile dermatomyositis) |
| OR | M34 Systemic sclerosis (scleroderma) |
| OR | M35.0 Sicca syndrome (Sjogren) |
| OR | M35.1 Other overlap syndromes |
| OR | M30 Polyarteritis nodosa and related conditions |
| OR | M31.3 Wegener's granulomatosis |
| OR | M31.5 Giant cell arteritis with polymyalgia rheumatica |
| OR | M31.6 Other giant cell arteritis |
| OR | M31.7 Microscopic polyangiitis |
| OR | M31.4 Aortic arch syndrome (Takayasu) |
| OR | M30.1 Polyarteritis with lung involvement (Churg-Strauss) |
| OR | D89.1 Cryoglobulinemia |
| OR | L95.9 Vasculitis limited to the skin, unspecified |
| OR | M35.2 Beçhet's disease (Beçhet's syndrome) |
| OR | K50 Crohn's diseases (regional enteritis) |
| OR | K51 Ulcerative colitis |

53,351 Patients

Population 5–17 years/any gender (7,127,901)

| | |
|----|---|
| OR | M08 Juvenile arthritis |
| OR | L93 Lupus erythematosus |
| OR | M32 Systemic lupus erythematosus |
| OR | M33 Dermatopolymyositis (juvenile dermatomyositis) |
| OR | M34 Systemic sclerosis (scleroderma) |
| OR | M35.0 Sicca syndrome (Sjogren) |
| OR | M35.1 Other overlap syndromes |
| OR | M30 Polyarteritis nodosa and related conditions |
| OR | M31.3 Wegener's granulomatosis |
| OR | M31.5 Giant cell arteritis with polymyalgia rheumatica |
| OR | M31.6 Other giant cell arteritis |
| OR | M31.7 Microscopic polyangiitis |
| OR | M31.4 Aortic arch syndrome (Takayasu) |
| OR | M30.1 Polyarteritis with lung involvement (Churg-Strauss) |
| OR | D89.1 Cryoglobulinemia |
| OR | L95.9 Vasculitis limited to the skin, unspecified |
| OR | M35.2 Beçhet's disease (Beçhet's syndrome) |
| OR | K50 Crohn's diseases (regional enteritis) |
| OR | K51 Ulcerative colitis |
| OR | G80 Cerebral palsy |
| OR | G71.0 Muscular dystrophy (Duchenne) |
| OR | E08-E13 Diabetes mellitus |
| OR | K90.0 Celiac disease |
| OR | E84 Cystic fibrosis |

EHR technology allows for patient cohort size comparisons based on the addition of several diseases. In this example, an amendment that opened enrollment to patients with cerebral palsy, Duchenne muscular dystrophy, diabetes, celiac disease, or cystic fibrosis in addition to all other prior allowed diseases expanded the number of eligible patients by 72% within the 5–17 year age range based on target indications alone.

Some organizations have already begun using federated EHR data from multiple healthcare organizations to develop their protocols and recruit patients, and early results are encouraging.

72% increase in eligible patient population

FIGURE 4: ORIGINAL VS. EXPANDED AMENDED FRACTURE CRITERIA

1,018 Patients



33,896 Patients

| | | | |
|-----------------------------------|--------|---|---|
| 532.0 Fracture of lumbar vertebra | 45,596 | ⊞ | ⊗ |
|-----------------------------------|--------|---|---|

| Event 2 – All of the terms in this event occurred between today and 24 months ago | | | |
|---|--|----------|-----|
| OR | 532.0 Fracture of lumbar vertebra | 45,596 | ⊞ ⊗ |
| OR | 542.2 Fracture of upper end of humerus | 70,808 | ⊞ ⊗ |
| OR | 542.3 Fracture of shaft of humerus | 44,674 | ⊞ ⊗ |
| OR | 542.4 Fracture of lower end of humerus | 67,734 | ⊞ ⊗ |
| OR | 552 Fracture of forearm | 213,988 | ⊞ ⊗ |
| OR | 572 Fracture of femur | 92,469 | ⊞ ⊗ |
| OR | 582.1 Fracture of upper end of tibia | ⚠ 44,436 | ⊞ ⊗ |
| OR | 582.2 Fracture of shaft of tibia | 65,717 | ⊞ ⊗ |
| OR | 582.3 Fracture of lower end of tibia | ⚠ 28,856 | ⊞ ⊗ |
| OR | 582.4 Fracture of shaft of fibula | 58,482 | ⊞ ⊗ |

The pharmaceutical company implemented another amendment to the same trial in order to expand the size of the potential patient population. In looking at only the fracture criteria, the initial inclusion criteria of vertebral fracture only identified 1,018 patients in the specified 5–17 age group, but it was augmented to 33,896 patients when all types of long bone fractures were later added—even with the requirement that they must occur in the prior two years. Our EHR source was able to show that the expansion of fracture criteria alone allowed a 97% increase in potential patients.

FIGURE 5: ORIGINAL VS. FULLY AMENDED PROTOCOL CRITERIA ANALYSIS

| | | |
|-----------------------|--|-----------------------|
| 20,513,595 – 18 sites | Network | 20,513,595 – 18 sites |
| 3,823,331 – 18 sites | Population – Ages 5–17 | 3,823,331 – 18 sites |
| 8,565 – 17 sites | Inclusion Criteria Diseases (plus all other remaining criteria) | 48,772 – 17 sites |
| 2 – 1 site | Allowed Fracture Types | 615 – 15 sites |
| 0 – 0 sites | Glucocorticoid Requirement | 38 – 7 sites |

Using analytical tools available in our EHR source allows for comparison of the complete original (pre-amendment) protocol vs. the most current protocol (after several amendments). While the size of the EHR network and patients within the 5–17 age range remain the same in the top portion of both funnels, the impact of specific parameters—diseases allowed, types of fractures allowed, medication requirement—can be clearly noted in the resulting patient numbers. At the same time, the effect of all protocol criteria together can be considered. Here we are able to show that there were no potential patients who met the non-amended protocol criteria, while the fully amended protocol was broadened enough to identify 38 patients.

“Being able to use cohort identification technology based on EHR data provides us with the objective data and analytics on real patients to help our clients make decisions that matter,” said Otis Johnson, PhD, MPA, vice president for feasibility and clinical informatics at ICON.

In an example of the technology’s ability to drive in-depth portfolio planning, a leading pharma company was able to leverage a multisite federated EHR database to evaluate a long-standing inclusion screening criterion that was perceived to be hampering recruiting efforts. Using data extraction to research a larger population of quantitative data, the company was able to see from side-by-side comparisons with and without the criterion how it changed the eligible patient number. The company then removed the criterion from the protocol template, which improved the potential patient pool and recruitment efficiency to potentially avoid costly amendments.

Another global healthcare company using the traditional site and patient selection process ended up with five amendments over an eight-year period and enrolled a total of 23 patients. The initial protocol wasn’t able to enroll a single patient. The study manager felt this would be the case, but had no tangible data at that time to dispute key opinion leaders who insisted there would be patients.

A retrospective analysis revealed how each amendment expanded the potential patient pool and delivered a collective assessment to the updated eligibility criteria overall. A final assessment that took all existing criteria and the five amendments into consideration and drew on EHR data from multiple healthcare organizations yielded 38 potential subjects. A similar analysis of the original protocol found zero patients—the same findings of the actual study before any amendments were considered.

As of the writing of this article, the trial has 23 patients enrolled, supporting the findings in the analysis and demonstrating the viability of EHR data analytics in “stress testing” a protocol for feasibility from conception to avoid costly amendments upstream (see Figures 3–5).

EHR Data–Based Results Match Epidemiologic Findings

The value of EHR-based studies has furthermore been validated in terms of ability to reproduce epidemiologic findings published in medical literature. EHR-based data extraction can provide a proactive method of producing accurately defined patient populations. This allows healthcare organizations, biopharma companies, and CROs to make better, more timely decisions.

Conclusion

Developing new therapies and getting them to market is cumbersome, time consuming, and costly. Flawed protocol design based on anecdote or opinions often fail to find the right patients for trials. Site selection based on art instead of real-world data is fraught with risks of trials closing due to failure to accrue patients. Cohort identification technology based on EHR data provides a better way.

The industry now has a treasure trove of real-time, relevant information in the form of EHR data being collected from nearly every healthcare organization. The key is getting to that information and leveraging it to make better upfront decisions and streamline the clinical trial process.

Along with the emergence of a culture of data sharing that improves availability of data for research, advances in data interoperability and maturing technologies for federated databases and cloud and data analytics are now allowing healthcare organizations, pharma companies, and CROs to tap into a vast wealth of data. As use of these collaborative networks increases, EHR data will soon become the key building block on which the industry can build a more effective, efficient process to bring new therapies to market faster. Eventually, that will lead to better clinical outcomes, which represent everyone’s ultimate goal.

References

1. Getz K. 2015. The cost of clinical trial delays. <https://www.ctti-clinicaltrials.org/files/pgct-session5.1-getz.pdf>
2. Bachheimer JF. 2016. Adaptive patient recruitment for 21st century clinical research. *App Clin Trials*. www.appliedclinicaltrials.com/adaptive-patient-recruitment-21st-century-clinical-research?pageID=3
3. Getz K. 2011. Protocol amendments: a costly solution. *App Clin Trials*. www.appliedclinicaltrials.com/protocol-amendments-costly-solution?id=&sk=&date=%0A%09%09&pageID=2
4. Desmon S. 2015. Industry-financed clinical trials on the rise as number of NIH-funded trials falls. HUB (Johns Hopkins University). <http://hub.jhu.edu/2015/12/15/industry-funded-clinical-trials-may-threaten-objectivity/>
5. Sahoo SS, Tao S, Parchman A, Luo Z, Cui L, Mergler P, Lanese R, Barnholtz-Sloan JS, Meropol NJ, Zhang G-Q. 2014. Trial Prospector: matching patients with cancer research studies using an automated and scalable approach. *Cancer Infts.*



Jennifer Stacey (jennifer.stacey@trinetx.com) is director of clinical operations at TriNetX in Cambridge, Mass.



Maulik D. Mehta (maulik.mehta@trinetx.com) is senior vice president of TriNetX in Cambridge, Mass.