Chapter 14
Pragmatic Trials and New Informatics Methods to Supplement or Replace Phase IV Trials

Eneida Mendonca and Umberto Tachinardi

Introduction

A report developed by the U.S. Department of Health and Human Services in 2014 [1], identified that one of the main challenges to the process of developing new drugs are complex and expensive clinical drug trials. Given the necessity of clinical trials (CT) to approve a new drug, obstacles to trials result in fewer new drugs becoming available. The list of barriers is long and widespread and includes high costs (phase IV CT costs, in particular, are almost the same as the sum of the three preceding phases combined), lengthy processes, recruitment and retention issues; regulatory and administrative barriers, drug-sponsor imposed barriers, and the disconnect between clinical and academic worlds. The report suggests some solutions to the CT problems that include: the use of electronic health record (EHR) systems; simpler enrollment processes; and the wider use of mobile and electronic

E. Mendonca
Department of Biostatistics and Medical Informatics, School of Medicine and Public Health, Department of Industrial and Systems Engineering in the College of Engineering, Clinical/Health Informatics, UW Institute for Clinical and Translational Research, University of Wisconsin Madison, Madison, WI, USA
e-mail: emendonc@regenstrief.org

U. Tachinardi (✉)
Department of Biostatistics and Medical Informatics, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, USA

UW Institute for Clinical and Translational Research, University of Wisconsin-Madison, Madison, WI, USA

UW Health, University of Wisconsin-Madison, Madison, WI, USA
e-mail: utachina@iu.edu

T. Adam, C. Aliferis (eds.), *Personalized and Precision Medicine Informatics*, Health Informatics, https://doi.org/10.1007/978-3-030-18626-5_14

technologies. Informatics tools are replacing traditional processes with significant potential to overcome most of the problems listed above. Automation and reuse of data can reduce costs and time; new technologies (e.g., EHR systems, social media, mobile systems) are helping to improve recruitment and retention rates; and the use of EHRs also helps bridge the research and the clinical sides towards improved clinical trials.

It is clear that CT as previously defined, need to be dramatically improved to respond to the urgent need for more and better drugs [2]. It is even more critical that we improve how we use those drugs [3] and that we identify when we should not use them. That is the role of Phase IV. This chapter discusses enhancements to Phase IV along with new alternatives (that completely change the original definitions and may be seen as full replacements to Phase IVs).

A good example of the new "Phase IV" is the potential to improve drug repurposing. It is a fact that the drug development pipelines are not pumping out enough new drugs [4] to supply the growing need for more and better therapeutical options (This has been called "Eroom's Law", which is the literal and semantic reverse of Moore's Law). Drug repurposing is one of many solutions that can be used to alleviate this problem.

What Are Phase IV Trials, and Why Are They Needed?

Phase IV studies are developed to test the efficacy and safety of drugs *after* they are approved to be marketed by a designated regulatory authority (FDA in the United States). Both characteristics are critically important to the patients that depend on drugs that are efficacious and safe in the short and long term horizons. Randomized Clinical Trials (RCT), work well in efficacy determination, but drug safety assessment may require a different approach. The calculation of sample size is critical in establishing drug safety. Phase III studies usually enroll 1000 to 3000 patients who use the new drug. The probability of identifying a rare adverse event in this small population is low [5]. In fact, defining the right sample size is so critical that the European Medicines Agency (EMA) adopted well-defined guidelines for it: the post-authorization safety studies (PASS) [6]. PASS is designed to identify, characterize or quantify a safety hazard; confirm the safety profile of a medicine; and measure the effectiveness of risk-management measures [7] in healthcare.

While not specific to Phase IV drug safety testing, an interesting aspect of the EMA guidelines is the inclusion of non-interventional [8] alternatives. The *Guideline on good pharmacovigilance practices (GVP)* describes:

> …non-interventional studies to include database research or review of records where all the events of interest have already happened (this may include case-control, cross-sectional, cohort or other study designs making secondary use of data). Non-interventional studies also include those involving primary data collection (e.g., prospective observational studies

and registries in which the data collected derive from routine clinical care), provided that the conditions set out above are met. In these studies, interviews, questionnaires, blood samples and patient follow-up may be performed as part of normal clinical practice.

Pragmatic Clinical Trials

The term pragmatic clinical trial (PCT) was coined nearly 50 years ago to distinguish between clinical trials that were explanatory in orientation (i.e., understanding whether a difference exists between treatments that are specified by strict definitions) and trials that were pragmatic in orientation (i.e., understanding whether a difference exists in treatment as applied in practice) [9]. PCT offers the potential to assess comparative effectiveness in broadly based patient populations receiving care in real-world clinical settings [10].

In August 2018, the website clinicaltrials.gov listed around 500 studies defined as "pragmatic clinical trials", 63 of those were labeled as Phase IV studies. As expected, the majority of those studies were funded at academic centers (total 40), but 18 of them were sponsored by industry. While small, when compared to almost 300,000 total studies in that database, those numbers seem to show a trend, since 60 of those 500 studies were not yet open to enrollment at the time of the query.

Whereas clinical trials are widely-accepted designs to establish the presence or absence of Rx efficacy as well as toxicity, they are often too rigid and with too short horizons. As a result, the efficacy and toxicity of approved drugs is not entirely known. Pragmatic trials take advantage of secondary use of EHR and other types of data (e.g., tumor registries and claims) to determine longer-term effects and personalized responses to treatments. Recent initiatives like PCORnet [11] are designed to share and exchange data across institutional boundaries to enable pragmatic trials by augmenting the sample size for all populations of interest.

The Role of EHR as a Phase IV Tool

Clinical Trials are designed to be highly controlled processes that define "how" and "what" data are collected and organized. Various mechanisms are usually in place to ensure data completeness and accuracy. Statistical methods are often used to analyze data, sometimes pre-analysis of data will require cleansing and or semantical harmonization (making sure all codes have one explicit and reproducible meaning).

Data in EHR systems are usually not at the same level of quality and standardization. For instance, both structured and unstructured data may be used for the same information. Therefore, mining data in EHR systems is a complex task. Nonetheless, the value of repurposing the wealth of EHR data available is high. This secondary

use provides faster and cheaper ways to obtain data from patients. Different than "explanatory" trials that measure efficacy, "pragmatic clinical trials", designed to test "effectiveness" [12], match the goals of CT Phase IV.

EHR systems were primarily designed to support financial, clinical and administrative functions, while collecting data to support those processes. However, along their evolution, EHR databases became a valuable resource for other uses, such as quality and clinical research. Now virtually all data used in Phase IV studies (laboratory results, chief complaints, ER admissions, prescriptions) are being collected in modern EHR systems.

There are many strategies to assess drug safety, including active surveillance (pharmacovigilance), intensive monitoring schemes and registries. Active surveillance are continuous processes used to identify adverse events by tracking prospective findings for a group of patients that are using drugs of interest. Intensive monitoring is a system to collect data in specific areas of the healthcare system (e.g., ICUs, ERs). The selected sites may provide information, such as data from specific patient subgroups that would not be available in a passive spontaneous reporting system. Registries are systems that curate and organize data for specific populations, conditions or outcomes.

Is There a Difference Between Phase IV Clinical Trials and Drug Re-purposing?

Phase IV seeks to define if the drug is effective for the approved uses and if new adverse events develop in the long term that were not identified in the previous clinical phases (II and III). The combination of EHR data and PCT approaches can be used to either find new adverse effects; prove or disprove drug effectiveness for the approved uses; or identify new benefits of the drug that were not initially tested or approved, but are capable of yielding important benefits in areas where drugs do not exist or are still being tested [13].

Drug repurposing (or repositioning) is the process of expanding the use of currently available drugs to other indications than the approved ones. A well-known example is sildenafil (Viagra) which was initially developed to treat angina [14]. Many companies and academic centers are working in this area due to the reduced costs (when compared to brand new drugs) and the lack of new compounds past preclinical phases. The same process that results in drug repurposing can also be used for the prediction of adverse events of known or novel drugs [15].

In reality, in both Phase IV and drug repurposing, the challenge is quite similar: to establish the association of a drug with an outcome (positive or negative) and find out if that relationship is causal or not. If PCT is the chosen method, the technical approach for both should include a precise, computable and reproducible definition of markers (phenotypes) based on data that is already being collected.

Identifying and Acting on Adverse Drug Events (ADE)

Identifying adverse events is a critical step for any system aimed to provide data for conventional or pragmatic clinical trials. An analogy can be made with trigger tools designed to support Learning Health System models. Those are resources developed to help with a standardized identification of an adverse event (in this case any negative outcome determined by some healthcare action). The Institute for Healthcare Improvement Global Trigger Tool (GTT) has become one of the most widely used trigger tools for detecting harm in hospitalized patients [16].

A GTT "trigger" is a medical record-based "hint" (such as the use of the antidote naloxone) that "triggers" the search of the medical record to determine whether an adverse event (such as a clinical overdose of an opiate, as opposed to a therapeutic use in response to non-prescribed opiate use) might have occurred [17]. Similar triggers can be developed to track the use of individual or associate drugs, thus mimicking part of what Phase IVs are designed to do.

Once identified an ADE candidate needs to be processed to properly identify the cause of the problem (misuse, prescription error, drug adverse event), then appropriate action should follow: suspension/change of therapy, internal and external reporting (e.g., FDA, EMA).

Informatics Methods in Phase IV

The Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS) is a database that contains adverse events reports, medication error reports, and product quality complaints resulting in adverse events that were submitted to the FDA [18]. FAERS adopts MedDRA (Medical Dictionary for Regulatory Activities), used in pharmacovigilance processes in the US, Europe and some eastern countries. MedDRA provides a single standardized international medical terminology which can be used for regulatory communication and evaluation of data pertaining to medicinal products for human use [19]. The database is a good resource for post marketing drug information, but investigators have shown that the resource is not sufficient because of challenges related to underreporting [20, 21] and patterns of missing drug exposures [22, 23]. Data mining with effective analytical frameworks and large-scale medical data is a potentially powerful method to discover and monitor ADEs [24]. Pharmacovigilance studies have explored the examination of ADEs using a diverse number of data sources, including scientific literature, online publicly available databases, social media, and EHRs [20, 25].

Studies have also shown that just a small fraction of ADEs recorded in EHRs are reported to federal databases and authorities, making EHRs an important source of ADE information.

Existing observational studies have mainly relied on structured EHR data to obtain ADE information; however, ADEs are often buried in the EHR narratives and not recorded in structured data. A number of studies using EHRs have focused on using structured/coded information and drug ontologies [26, 27], showing somewhat limited results [24, 28, 29]. Furthermore, information contained in clinical notes are not likely to be presented in structured form in other parts of the record (e.g., signs, symptoms, severity of findings, disease progress).

Most studies that attempt to discover ADE information from narratives use a combination of natural language processing (NLP) methods and machine learning. Most recently, approaches integrating multi-faceted, heterogeneous data sources have become more common [25]. Two recent reviews of the literature on the use of NLP methods for pharmaco-vigilance and medication safety show a growing number of algorithms for automated detection of associations between medications and adverse events [20, 25].

Luo and colleagues [25] categorized the findings based on the characteristics of the NLP components and their complexity. Methods evaluated included basic keyword and trigger phrase search, algorithms exploring the syntactic and semantic patterns of drugs and adverse events, methods extending existing biomedical NLP systems and methods using existing or custom-built ontologies. The study also identified recent trends in EHR-based pharmacovigilance, such as the increased adoption of statistical analysis and machine learning, integration of temporal resolution, and the use of multiple data sources. Wong and colleagues [20] illustrated the fundamentals of NLP and discussed the application of these methods on medication safety in different data sources (e.g., EHRs, scientific literature, Internet-based data, and reporting systems). Both reviews demonstrate that it is important to consider the different approaches, as some of them are context and task-dependent. Combined approaches (hybrid) involving computational (statistical and machine learning) and linguistic methods may yield better results.

Despite the growing number of NLP, machine learning, and statistical methods for adverse event detection in EHR systems, several challenges remain. One example is the data fragmentation caused by movement of patients between multiple EHR systems. This is a big problem when longitudinality is required. Techniques designed to combine EHR data from multiple institutions while still protecting privacy are becoming increasingly available [30].

Data exchanges provide a powerful means to rapidily and significantly expand cohorts. Whether the data comes from research [31] or directly from EHR systems [32], the intent is to expand the cohorts faster than traditional methods. Larger cohorts increase the probability of identifying outliers (i.e. rare conditions), but also confirm key trends and patterns. Initiatives that make secondary use of data require additional measures to protect privacy and confidentiality. Several automated de-identification methods are available [33], helping promote safer data sharing.

Polypharmacy is the use of multiple medications [34], and one of the most understudied aspects of adverse event detection using EHR data. With the aging of the population and the increased number of chronic diseases, it is expected that a substantial percent of the population take more than one medication. In a national

population-based study, Qato and collaborators found that 36% of older US adults were regularly using 5 or more medications or supplements and 15% were potentially at risk for a major drug-drug interaction [35]. Despite that, polypharmacy has not been the focus of the scientific community [25], with most studies assessing the adverse events based on a single drug. The "new Phase IV" paradigm presents a good opportunity to fix this problem, since EHR and pharmacy systems can more naturally identify associations of drugs, versus the specific targeted drugs monitored by traditional RCTs.

Data Integration and Analytical Tools

Phase IV studies (traditional or pragmatic) depend on collaboration from multiple sites. With EHR systems being added to the research protocols, data harmonization and integration becomes central to the process. Different sites may adopt different EHR systems, and even when the same EHR systems are used, the data may be represented in different ways at different levels of granularity.

Several initiatives have focused in data harmonization processes (i.e. use of Common Data Models). Others have focused on efforts on shared resources and community-wide tools to promote analytical solutions to the use of electronic health records.

Common data models standardize the representation of healthcare data from diverse sources in a consistent way. The goal is to facilitate the mapping of clinical observation to standard vocabularies and, consequently, improve how these data can be reused for research purposes and shared across institutions. This chapter does not intend to give a comprehensive view of common data models, but it is worth mentioning some examples.

PCORNet [11], the National Patient-Centered Clinical Research Network, is an initiative of the Patient-Centered Outcomes Research Institute (PCORI). The goal of PCORNet is to facilitate clinical research by facilitating the sharing of electronic health records across institutions. The PCORNet CDM [36] is a platform that enables rapid responses to research-related questions. The CDM is based on the FDA Sentinel Initiative Common Data Model [37]. It leverages the use of standard terminologies and coding systems such as ICD, SNOMED, CPT, and LOINC among others. PCORNet also provide a platform that allows simple creation, operation, and governance of distributed health data networks, called PopMedNet. PopMedNet allows for distributed querying, customizable workflows, and auditing and search capabilities, while enables the enforcement of varying governance policies.

The Observational Health Data Sciences and Informatics (OHDSI) [38, 39], an international collaborative initiative whose goal is to create and apply open-source data analytic solutions to a large network of health databases to improve human health and wellbeing. OHDSI was initially based on the Observational Medical Outcomes Partnership (OMOP) [40], which generated the OMOP CDM. In addition

to EHR data, the OMOP CDM supports administrative claims data. ODSHI also provides tools for tools for data quality and characterization (ACHILLES), database exploration, standardized vocabulary browsing, cohort definition, and population-level analysis (ATLAS).

Another example is the Accrual to Clinical Trials (ACT) network, a federation of academic research institutions. The goals of this network are somewhat similar to the ones described above. ACT aims to facilitate cohort discovery by determining recruitment feasibility and patient identification. ACT uses the i2b2 tool's multi-site Shared Health Research Information Network (SHRINE), and the i2b2 CDM.

Despite these initiatives, data harmonization and data sharing are still major challenges in the design and implementation of Phase IV trials.

Data Challenges in Pragmatic Clinical Trials

Conventional Phase IV studies adopt several mechanisms to ensure that the data is complete, accurate and standardized. When EHR systems are used as the source, as opposed to traditional data collection tools like Case Report Forms (CRF), data quality becomes an important issue. There are informatics techniques that can help improve the quality: data harmonization, use of standard coding systems, data linkage and NLP are part of the informatics toolbox.

Data harmonization methods are used when the data comes from a variety of sources that originally used different definitions for the same concepts. The data harmonization process equalizes the granularity of the definition (e.g., reducing sex concepts to two genders M/F) at the coarsest common level of granularity. Standardized coding systems help data to be shared more easily. Examples of those coding standards include ICD, SNOMED, LOINC, RxNORM, and MedDRA. Data linkage is important when there are multiple sources containing part of the necessary data, a common identifier (usually name, date of birth or identity document number) is used as the link anchor. When a patient, for instance, has his or her data in multiple EHR systems those methods need to be used. NLP can produce codes out of unstructured data (i.e., plain text). The ability to extract codes from text can overcome some deficiencies like missing structured data (e.g., a behavioral condition) or confirm the accuracy of certain structured codes (e.g., an ICD code entered for billing).

Most EHR systems have the data available in two different databases. The first is the database used primarily by the system to support transactions using the user-interface in real-time, called the transactional database. The transactional database is optimized for performance, referential integrity and multiple users simultaneously editing the same information. Transactional databases are not good for analytics like machine learning, where intense querying occurs at very high frequencies. Consequently, EHR systems usually have a secondary database for analytics work. The Clinical Data Warehouses (CDW) are databases designed to respond to complex queries, and not to perform changes in the data (edits, insertions or deletions).

The CDW usually has a temporal lag with the transactional database, usually lagging around 24 h behind.

Those differences are important in designing a solution like a continuous Phase IV, where some processing can use past data, but others need to be computed in real-time. EHR systems are consolidators of data of several sub-systems (i.e., labs). Interfaces allow data to be transferred from the ancillary systems to the EHR. A popular interfacing standard is HL7 (Health Level Seven), where data is streamed from the source system (i.e., lab system) and received by the target system (i.e., EHR). Some solutions to track adverse events actually tap directly in that data stream to get the results faster (closer to realtime). Based on HL7, FHIR (Fast Healthcare Interoperability Resources) can also help applications (like adverse event detectors) request data from EHR systems quickly, process it and return an action (i.e., decision support) if applicable.

Linking Patient Data Across Multiple EHR Systems

There are two basic ways that EHR systems help with Phase IV trials: as a resource to the conventional Clinical Trials Management Systems (CTMS) or by replacing the CTMS with a pragmatic Phase IV solely using EHR systems data. The first model (Fig. 14.1—left box) assumes that a conventional Phase IV protocol will be in use, subjects will be enrolled given a defined criterion, subjects will consent and

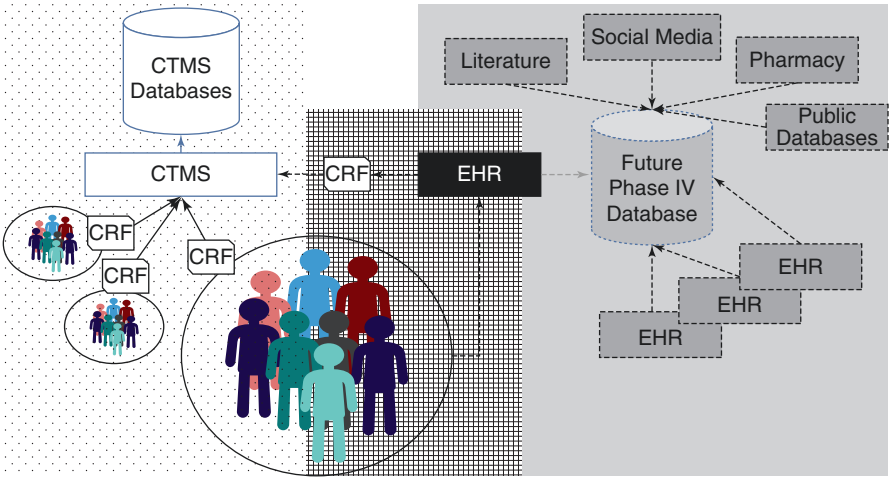

Fig. 14.1 Traditional Phase IV studies (left) use Clinical Trials Management Systems (CTMS) to manage participants, protocols and study teams; to be a source of record for study data and documents; and to produce reports. Those systems can be interfaced with EHR systems (checkered rectangle in the middle) reducing the need for human transcripts. The right box shows a hypothetical scheme where EHR systems and other data sources combined form a "future platform" for Phase IV studies (pragmatic approach)

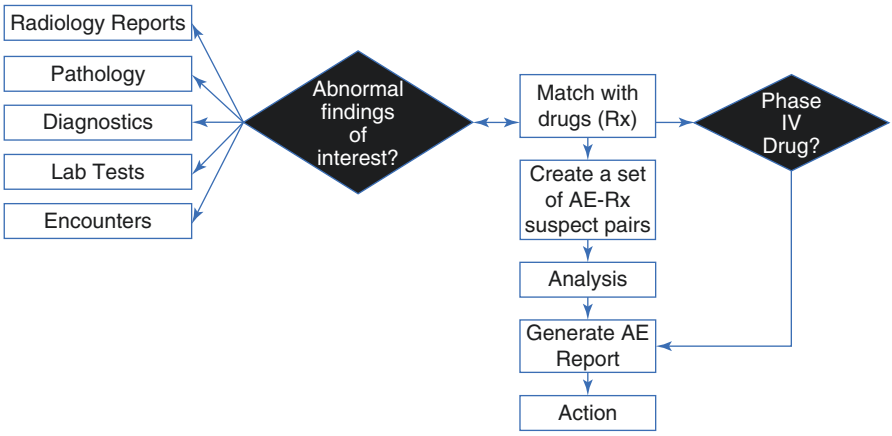

Fig. 14.2 EHR data can be used to identify triggers (potential AEs) out of different data sources: pathology reports, lab results, ICD codes, ER visits (encounters), radiology reports and others. If a pragmatic Phase IV study for a certain drug is in place, the ocurence of a finding of interest will trigger the generation of an AE report. If the system is being used as a pharmacovigilance tool, the pair EFI-Drug will be analyzed as a potential unpredicted AE

data will be collected using Case Report Forms (CRF). The EHR system of record for a given participant will be interfaced with the CTMS system, and the data will be transferred electronically from the EHR (Fig. 14.2).

A possible replacement to Phase IV as traditionally defined (Fig. 14.1—right box), this new platform would also connect to other data sources (e.g., publications, social media and pharmacy databases). The envisioned new Phase IV databases would be used for heavy analytics (i.e. machine learning).

Phase IV and Precision Medicine

The future of Phase IV CTs is one that relies heavily on data collected via the EHR. PCTs are one step forward, but the availability of both historical and real-time data enables a continuous and individualized Phase IV. The use of machine learning and AI can help "learn" the patterns of normality for a given person (or population) and detect changes or anomalies in that pattern. The same technologies can help differentiate the "good changes" (drug efficacy) from the "bad changes" (adverse events). Learning is always a more intense computational effort, but once the patterns are established, the detection of those occurrences can be performed in real-time using clinical decision support (CDS) tools. The detected changes (triggers) then become part of a registry of adverse events or positive outcomes, which in essence would be the basis of an endless Phase IV. The detected anomalies can also immediately become an actionable event (CDS) specific to one particular person (Fig. 14.3).

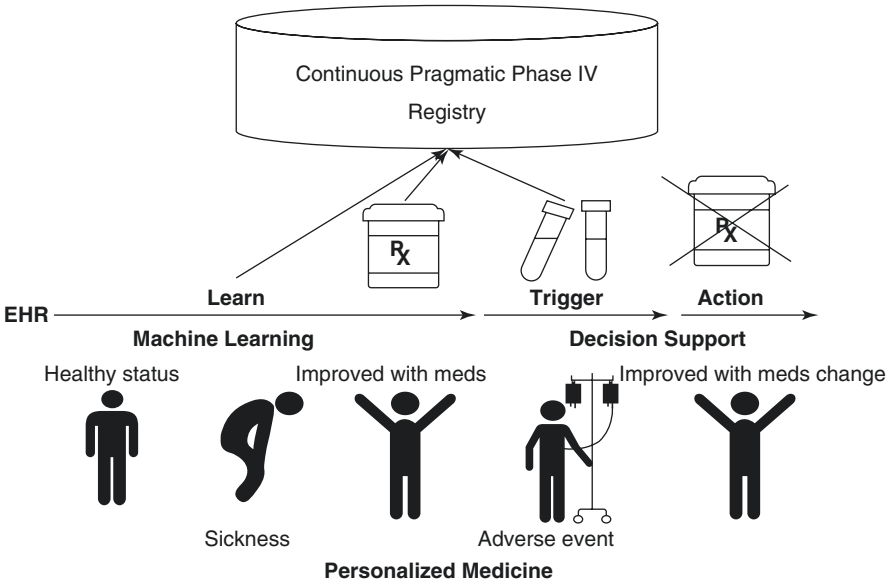

Fig. 14.3 The EHR system as a foundation to a continuous pragmatic Phase IV registry and a key tool in support of precision medicine. Using a person's EHR (or a large number of people's EHRs) as a training set, machine learning methods can define a "normal pattern" and identify when something does not fit the pattern. In this schematic example, the development of an adverse event after a new drug (to treat the sickness) was introduced. That trigger would define an action based on decision support logic implemented in the EHR system's clinical decision support system module. The detection, trigger, decision support sequence is typical of a personalized medicine approach. The finding for that individual, on the other hand, provides insights that, if repeated for other similar cases, can be used to produce generalizations like showing that the drug is not safe when this particular set of findings is present

Solutions like this one would definitely provide early signals in cases like the Vioxx (rofecoxib), given the amount of evidence that was available but not connected to make a compelling case towards a revision of the drug safety [41].

Final Remarks

The pressing and growing need for new more effcective, safer and cheaper drugs is forcing the clinical trials industry toward radical innovation. Phase IV clinical trials, for instance, are transitioning from their original design into an agile and more efficient platform to track drug efficacy and adverse events. Those innovations also support more precise and targeted use of available drugs (i.e., precision medicine and re-purposing), an active way to improve efficacy and safety.

The ubiquity of EHR systems is a key factor driving this transformation. Not only are EHR systems helping improve data collection for traditional Phase IVs,

but more importantly they are showing that a real-time, continuous, efficient solution can completely replace the old model. The current EHR systems still need to be improved in terms of data quality, use of standards, data sharing and data integration. Academic research institutions are developing solutions to overcome some of the current limitations of the EHR systems (e.g. developing standards for phenotyping, augmenting data with NLP and machine learning, integrating data with other sources and establishing data sharing networks). EHR vendors are adding new features (most of which developed by research groups) to new releases of their systems. But the backlog is enormous, and at the current rate of progress, it will take a long time to have all necessary advancements implemented in practice. Some of those changes may even require a complete reengineering of the current systems, since they were not designed to acquire and process bigger volumes of multi-dimensional data (as required in this case).

Certainly the "new Phase IV" will take advantage of more variety (more patients, more conditions, more findings), more data elements, and larger sample sizes for longer periods of time. Those new characteristics impose the need for novel tools and methods. In the current era of big data there are plenty of options for new computing (e.g., cloud computing) and analytics (e.g., deep learning) technologies to support those challenges.

Genomics data are slowly starting to be incorporated into EHR systems [42]. Since EHR systems were not designed to properly incorporate unstructured data like genomics, most institutions are adopting external solutions to provide that function. The addition of those new types of data can potentially transform how cohorts are defined for all clinical phases of clinical trials, including potentially the "N-of-1" model [43]. A "continuous EHR-based Phase IV" combined with a pharmacogenomics component can be truly transformational. The boundaries between the traditional CT phases would be less clear, and may even disappear.

References

1. Sertkaya A, Birkenbach A, Berlind A, Eyraud J. Examination of clinical trial costs and barriers for drug development. Washington, DC: US Department of Health and Human Services; 2014.
2. Bennani YL. Drug discovery in the next decade: innovation needed ASAP. Drug Discov Today [Internet]. 2011;16:779–92. https://www.sciencedirect.com/science/article/pii/S1359644611001826?via%3Dihub.
3. Vallance P. Developing an open relationship with the drug industry. Lancet [Internet]. 2005;366:1062–4. https://www.sciencedirect.com/science/article/pii/S0140673605668353?via%3Dihub.
4. Scannell JW, Blanckley A, Boldon H, Warrington B. Diagnosing the decline in pharmaceutical R & D efficiency. Nat Rev Drug Discov [Internet]. 2012;11:191–200. http://www.ncbi.nlm.nih.gov/pubmed/22378269.
5. Cesana BM, Biganzoli EM. Phase IV Studies: some insights, clarifications, and issues. Curr Clin Pharmacol [Internet]. 2018;13:14–20. http://www.eurekaselect.com/161232/article.

6. Kiri VA. A pathway to improved prospective observational post-authorization safety studies. Drug Saf [Internet]. 2012;35:711–24. http://link.springer.com/10.1007/BF03261968.
7. Post-authorisation safety studies (PASS)|European Medicines Agency [Internet]. https://www.ema.europa.eu/human-regulatory/post-authorisation/pharmacovigilance/post-authorisation-safety-studies-pass. Cited 9 Sept 2018.
8. European Medicines Agency, Heads of Medicines Agencies. Guideline on good pharmacovigilance practices (GVP)Module VIII – Post-authorisation safety studies (Rev 3) [Internet]. 2017. https://www.ema.europa.eu/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-gvp-module-viii-post-authorisation-safety-studies-rev-3_en.pdf.
9. Rosenthal GE. The role of pragmatic clinical trials in the evolution of learning health systems. Trans Am Clin Climatol Assoc [Internet]. 2014;125:204–16. discussion 217-8. http://www.ncbi.nlm.nih.gov/pubmed/25125735.
10. Oster G, Sullivan SD, Dalal MR, Kazemi MR, Rojeski M, Wysham CH, et al. Achieve control: a pragmatic clinical trial of insulin glargine 300 U/mL versus other basal insulins in insulin-naïve patients with type 2 diabetes. Postgrad Med [Internet]. 2016;128:731–9. http://www.ncbi.nlm.nih.gov/pubmed/27690710.
11. PCORnet The National Patient-Centered Clinical Research Network. PCORnet, the National Patient-Centered Clinical Research Network - PCORnet [Internet]. https://pcornet.org/.
12. Roland M, Torgerson DJ. What are pragmatic trials? BMJ [Internet]. 1998;316:285. http://www.ncbi.nlm.nih.gov/pubmed/9472515.
13. McCabe B, Liberante F, Mills KI. Repurposing medicinal compounds for blood cancer treatment. Ann Hematol [Internet]. 2015;94:1267–76. http://www.ncbi.nlm.nih.gov/pubmed/26048243.
14. Hernandez JJ, Pryszlak M, Smith L, Yanchus C, Kurji N, Shahani VM, et al. Giving drugs a second chance: overcoming regulatory and financial hurdles in repurposing approved drugs as cancer therapeutics. Front Oncol [Internet]. 2017;7:273. http://www.ncbi.nlm.nih.gov/pubmed/29184849.
15. Deftereos SN, Andronis C, Friedla EJ, Persidis A, Persidis A. Drug repurposing and adverse event prediction using high-throughput literature analysis. Wiley Interdiscip Rev Syst Biol Med [Internet]. 2011;3:323–34. http://www.ncbi.nlm.nih.gov/pubmed/21416632.
16. AHRQ Agency for Healthcare Research & Quality. AHRQ Agency for Healthcare Research and Quality [Internet]. Rockville, MD: AHRQ Agency for Healthcare Research & Quality. https://www.ahrq.gov/.
17. Stockwell DC, Bisarya H, Classen DC, Kirkendall ES, Landrigan CP, Lemon V, et al. A trigger tool to detect harm in pediatric inpatient settings. Pediatr Int. 2015;135:1036–42. http://www.ncbi.nlm.nih.gov/pubmed/25986015.
18. Center for Drug Evaluation and Research- US Food and Drug Administration. Drug approvals and databases - FDA adverse event reporting system (FAERS) [Internet]. Silver Spring, MD: Center for Drug Evaluation and Research; 2017. https://www.fda.gov/drugs/informationondrugs/ucm135151.htm.
19. MedDra Medical Dictionary for Regulatory Activities. Vision for MedDRA [Internet]. VA: McLean. https://www.meddra.org/about-meddra/vision.
20. Wong A, Plasek JM, Montecalvo SP, Zhou L. Natural language processing and its implications for the future of medication safety: a narrative review of recent advances and challenges. Pharmacother J Hum Pharmacol Drug Ther [Internet]. 2018;38:822–41. https://doi.org/10.1002/phar.2151.
21. Wang X, Hripcsak G, Markatou M, Friedman C. Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: a feasibility study. J Am Med Informatics Assoc [Internet]. 2009;16:328–37. https://doi.org/10.1197/jamia.M3028.
22. Munkhdalai T, Liu F, Yu H. Clinical relation extraction toward drug safety surveillance using electronic health record narratives: classical learning versus deep learning. JMIR Public Heal Surveill [Internet]. 2018;4:e29. http://www.ncbi.nlm.nih.gov/pubmed/29695376.

23. Begaud B, Moride Y, Tubert-Bitter P, Chaslerie A, Haramburu F. False-positives in spontaneous reporting: should we worry about them? Br J Clin Pharmacol [Internet]. 1994;38:401–4. http://www.ncbi.nlm.nih.gov/pubmed/7893579.
24. Zhan C, Roughead E, Liu L, Pratt N, Li J. A data-driven method to detect adverse drug events from prescription data. J Biomed Inform [Internet]. 2018;85:10–20. https://www.sciencedirect.com/science/article/pii/S1532046418301394?via%3Dihub.
25. Luo Y, Thompson WK, Herr TM, Zeng Z, Berendsen MA, Jonnalagadda SR, et al. Natural language processing for EHR-based pharmacovigilance: a structured review. Drug Saf [Internet]. 2017;40:1075–89. http://www.ncbi.nlm.nih.gov/pubmed/28643174.
26. Iqbal E, Mallah R, Rhodes D, Wu H, Romero A, Chang N, et al. ADEPt, a semantically-enriched pipeline for extracting adverse drug events from free-text electronic health records. PLoS One [Internet]. 2017;12:e0187121. https://doi.org/10.1371/journal.pone.0187121.
27. Combi C, Zorzi M, Pozzani G, Moretti U, Arzenton E. From narrative descriptions to MedDRA: automagically encoding adverse drug reactions. J Biomed Inform [Internet]. 2018;84:184–99. https://www.sciencedirect.com/science/article/pii/S1532046418301278?via%3Dihub.
28. Nadkarni PM. Drug safety surveillance using de-identified EMR and claims data: issues and challenges. J Am Med Informatics Assoc [Internet]. 2010;17:671–4. https://doi.org/10.1136/jamia.2010.008607.
29. Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, et al. 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured. Health Aff [Internet]. 2011;30:581–9. https://doi.org/10.1377/hlthaff.2011.0190.
30. Kho AN, Cashy JP, Jackson KL, Pah AR, Goel S, Boehnke J, et al. Design and implementation of a privacy preserving electronic health record linkage tool in Chicago. J Am Med Informatics Assoc [Internet]. 2015;22:1072–80. https://doi.org/10.1093/jamia/ocv038.
31. Ohmann C, Banzi R, Canham S, Battaglia S, Matei M, Ariyo C, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. BMJ Open [Internet]. 2017;7:e018647. http://www.ncbi.nlm.nih.gov/pubmed/29247106.
32. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. J Am Med Informatics Assoc [Internet]. 2014;21:578–82. https://doi.org/10.1136/amiajnl-2014-002747.
33. Kayaalp M. Modes of de-identification. AMIA Annu Symp Proc [Internet]. 2017;2017:1044–50. http://www.ncbi.nlm.nih.gov/pubmed/29854172.
34. Masnoon N, Shakib S, Kalisch-Ellett L, Caughey GE. What is polypharmacy? A systematic review of definitions. BMC Geriatr [Internet]. 2017;17:230. http://bmcgeriatr.biomedcentral.com/articles/10.1186/s12877-017-0621-2.
35. Qato DM, Wilder J, Schumm LP, Gillet V, Alexander GC. Changes in prescription and over-the-counter medication and dietary supplement use among older adults in the United States, 2005 vs 2011. JAMA Intern Med [Internet]. 2016;176:473–82. http://www.ncbi.nlm.nih.gov/pubmed/26998708.
36. PCORnet The National Patient-Centered Clinical Research Network. PCORnet Common Data Model (CDM) - PCORnet [Internet]. 2018. https://pcornet.org/pcornet-common-data-model/.
37. Sentinel Coordinating Center. Sentinel Initiative [Internet]. 2018. https://www.sentinelinitiative.org/.
38. OHDSI Observational Health Data Sciences and Informatics. OHDSI – Observational Health Data Sciences and Informatics [Internet]. 2018. https://www.ohdsi.org/.
39. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform. 2015;216:574.
40. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc [Internet]. 2012;19:54–60. http://www.ncbi.nlm.nih.gov/pubmed/22037893.
41. Krumholz HM, Ross JS, Presler AH, Egilman DS. What have we learnt from Vioxx? BMJ [Internet]. 2007;334:120–3. http://www.ncbi.nlm.nih.gov/pubmed/17235089.

42. Ohno-Machado L, Kim J, Gabriel RA, Kuo GM, Hogarth MA. Genomics and electronic health record systems. Hum Mol Genet [Internet]. 2018;27:R48–55. https://academic.oup.com/hmg/article/27/R1/R48/4975618.
43. Silvestris N, Ciliberto G, De Paoli P, Apolone G, Lavitrano ML, Pierotti MA, et al. Liquid dynamic medicine and N-of-1 clinical trials: a change of perspective in oncology research. J Exp Clin Cancer Res [Internet]. 2017;36:128. http://jeccr.biomedcentral.com/articles/10.1186/s13046-017-0598-x.