



HHS Public Access

Author manuscript

Circ Cardiovasc Qual Outcomes. Author manuscript; available in PMC 2021 June 01.

Published in final edited form as:

Circ Cardiovasc Qual Outcomes. 2020 June ; 13(6): e006292. doi:10.1161/
CIRCOUTCOMES.119.006292.

Computable phenotype implementation for a national, multi-center pragmatic clinical trial: Lessons learned from ADAPTABLE.

Faraz S. Ahmad, MD MS¹, Iben M. Ricket, MPH², Bradley G. Hammill, DrPH^{3,4}, Lisa Eskenazi, MHA⁴, Holly R. Robertson, PhD⁴, Lesley H. Curtis, PhD⁴, Cecilia D. Dobi, MS⁵, Saket Girotra, MD, SM⁶, Kevin Haynes, PharmD, MSCE⁷, Jorge R. Kizer, MD, MSc⁸, Sunil Kripalani, MD, MSc⁹, Mathew T. Roe, MD, MHS^{3,4}, Christianne L. Roumie, MD, MPH⁹, Russ Waitman, PhD¹⁰, W. Schuyler Jones, MD^{3,4}, Mark G. Weiner, MD¹¹

¹Northwestern University Feinberg School of Medicine, Chicago, IL, USA

²Louisiana Public Health Institute, New Orleans, LA, USA

³Duke University School of Medicine, Durham, NC

⁴Duke Clinical Research Institute, Durham, NC, USA

⁵Lewis Katz School of Medicine at Temple University, Philadelphia, PA, USA

⁶University of Iowa Carver College of Medicine, Iowa City, IA & Iowa City Veteran Affairs Medical Center, Iowa City, IA

⁷HealthCore, Willington, DE

⁸Cardiology Section, San Francisco Veterans Affairs Health Care System, and Departments of Medicine, and Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA

⁹Department of Medicine, Vanderbilt University Medical Center, Veterans Health Administration-Tennessee Valley Healthcare System Geriatric Research Education Clinical Center (GRECC), HSR&D Center, Nashville, TN

¹⁰Department of Internal Medicine, University of Kansas Medical Center, Kansas City, KS

¹¹Weil Cornell Medicine, New York Presbyterian-Weill Cornell Campus, New York, NY, USA.

Abstract

Background: Many large-scale cardiovascular clinical trials are plagued with escalating costs and low enrollment. Implementing a computable phenotype, which is a set of executable algorithms, to identify a group of clinical characteristics derivable from electronic health records (EHRs) or administrative claims records, is essential to successful recruitment in large-scale

Corresponding Author: Faraz Ahmad, MD MS. 676 North Saint Clair Street, Suite 600, Chicago, IL 60611. Office: 312-695-4965. Fax: 312-926-8620. faraz.ahmad@northwestern.edu.

*Dr. Ahmad and I.M. Ricket contributed equally to this work as first authors.

Disclosures

The authors report no conflicts.

pragmatic clinical trials. This methods paper provides an overview of the development and implementation of a computable phenotype in ADAPTABLE (Aspirin Dosing: A Patient-Centric Trial Assessing Benefits and Long-Term Effectiveness), a pragmatic, randomized, open-label clinical trial testing the optimal dose of aspirin for secondary prevention of atherosclerotic cardiovascular disease (ASCVD) events.

Methods and Results: A multi-disciplinary team developed and tested the computable phenotype to identify adults 18 years or older with a history of ASCVD without safety concerns around using aspirin and meeting trial eligibility criteria. Using the computable phenotype, investigators identified over 650,000 potentially eligible patients from the 40 participating sites from PCORnet, a network of Clinical Data Research Networks, Patient Powered Research Networks, and Health Plan Research Networks. Leveraging diverse recruitment methods, sites enrolled 15,076 participants from April 2016 to June 2019. During the process of developing and implementing the ADAPTABLE computable phenotype, several key lessons were learned. The accuracy and utility of a computable phenotype are dependent on the quality of the source data, which can be variable even with a common data model. Local validation and modification were required based on site factors, such as recruitment strategies, data quality, and local coding patterns. Sustained collaboration among a diverse team of researchers is needed during computable phenotype development and implementation.

Conclusions: The ADAPTABLE computable phenotype served as efficient method to recruit patients in a multi-site pragmatic clinical trial. This process of development and implementation will be informative for future large-scale, pragmatic clinical trials.

Trial Registration: [Clinicaltrials.gov](https://clinicaltrials.gov); Unique identifier: [NCT02697916](https://clinicaltrials.gov/ct2/show/study/NCT02697916)

Keywords

Electronic Health Record; Computable phenotype; Pragmatic Clinical Trial; Patient Recruitment

Cardiovascular clinical trials are often plagued by escalating costs, challenging regulations, and declining site-level participation.^{1,2} One source of high costs and complexity is the lack of efficient methods for the identification and enrollment of participants. The electronic health record (EHR) represents one tool to facilitate clinical trial recruitment, potentially at a lower cost.³ The EHR is particularly important for pragmatic clinical trials (PCTs), which are studies performed within the context of routine care on diverse populations.⁴

The National Patient-Centered Outcomes Research Network (PCORnet), similar to other data networks,⁵⁻⁸ leverages a standardized common data model with a single, overarching governance process. PCORnet is a national network of 13 Clinical Data Research Networks, two Health Plan Research Networks, and 21 Patient-Powered Research Networks with the goal of enabling investigators to use electronic health data to support clinical research across broad, diverse populations.⁹ With the adoption of a standardized data model across multiple institutional participants, a single, well-written query developed by one institution can run successfully across all institutions to identify large cohorts of potentially eligible patients.

ADAPTABLE (Aspirin Dosing: A Patient-Centric Trial Assessing Benefits and Long-Term Effectiveness) is the flagship, pragmatic, PCORI-funded clinical trial evaluating the

effectiveness and safety of 81mg vs 325mg aspirin doses in patients with a history of atherosclerotic cardiovascular disease (ASCVD) in 15,000 patients.¹⁰ Leveraging a computable phenotype was central to the pragmatic nature of ADAPTABLE. A computable phenotype is the product of using an executable set of algorithms to identify specific measurable constructs present in patient records.¹¹ A computable phenotype translates study eligibility criteria into a query, enabling the efficient identification of potentially eligible patients for research.^{9,11–14} The components of a computable phenotype include the presence or absence of diseases, procedures, patterns of medication use, laboratory results, and clinical events.¹⁵

In this methods paper, we describe the development, implementation, and lessons learned from the ADAPTABLE computable phenotype. The lessons were identified through discussions of the ADAPTABLE computable phenotype working group and other ADAPTABLE team members, including clinical investigators, informaticians, data scientists, and health services researchers. Data were also used from an internal ADAPTABLE survey of participating sites administered by the trial coordinating center in 2016.

SUMMARY of ADAPTABLE

PCORnet operates as a distributed research network where each Clinical Data Research Network, Patient Powered Research Network, or Health Plan Research Network retains control of its data, but makes it available to others for querying. Partner institutions harmonize the structure of their data through the use of a common data model to enable this interoperability. The implemented version (v3.1) of the PCORnet Common Data Model at time of computable phenotype development and initial execution included the 15 data tables. The specifications of the PCORnet Common Data Model is available online.¹⁶

The Common Data Model contains a focused set of discrete data elements including demographics, diagnoses, procedures, laboratory data, vital signs, smoking status, medication orders, and dispensing data. With only 15 tables, the Common Data Model includes only a subset of all data available in the EHR or claims-based system.

Computable Phenotype Development

The major eligibility criteria for ADAPTABLE includes a history of ASCVD defined as requiring ischemic heart disease, the presence of one or more “enrichment factors” that elevates the risk for cardiovascular events, and the absence of contraindications to aspirin use, such as prior severe gastrointestinal bleeding or aspirin allergy. After study protocol finalization in October 2015, The Duke Clinical Research Institute, serving as the trial coordinating center, started computable phenotype development and leveraged prior work by The MidSouth Clinical Data Research Network. The MidSouth Clinical Data Research Network developed and validated a coronary heart disease computable phenotype with excellent performance with positive predicted value of 98.5% and sensitivity of 94.6%.¹⁷ The next step of development was the creation of technical specifications based on the ADAPTABLE protocol language.

As shown in Figure 1, concepts that appear in the inclusion and exclusion criteria were generally divided into three categories: (1) concepts directly obtained from Common Data Model fields (e.g., EHR-derived smoking status, systolic blood pressure); (2) concepts requiring generated code lists from coded medical terminologies stored in Common Data Model fields (e.g., medication usage); and (3) concepts not readily available from Common Data Model fields (e.g., prior coronary angiography with 75% stenosis of at least one epicardial coronary vessel). Code lists for the second category were created using a number of different tools, most of which are publicly available,^{18–21} and were derived from encounter diagnosis and procedure codes. Codes from problem lists were not used due to data quality concerns. Codes from structured past medical history are not mapped to a table in the Common Data Model.

This computable phenotype was expected to be a “base phenotype” that individual sites would customize locally. Sites had the option to run the customized computable phenotype on their PCORnet datamart and include the addition of local data not available in the Common Data Model. Most sites adapted the code to run exclusively on local systems instead of their PCORnet datamart. In winter 2015/2016, the computable phenotype was posted on GitHub²² and presented to participating ADAPTABLE sites on multiple conference calls. After this release, each site tested, revised, and validated the computable phenotype locally.

Computable Phenotype Updates

In October 2016, PCORI approved a revised protocol, which included a number of additions to the inclusion/exclusion criteria (Figure 1). Based on previously published work,^{17,23} additional chart review testing the proposed phenotype changes, and extensive steering committee discussions, the definition of ASCVD was expanded from requiring prior myocardial infarction, coronary revascularization, or severe coronary artery disease on coronary angiogram to include patients with a history of clinical ASCVD, defined as a history of chronic ischemic heart disease, coronary artery disease, or atherosclerotic cardiovascular disease, and identified using International Classification of Diseases (ICD) codes. This expansion was a response to sites who indicated that prior coronary angiography results were neither available in the PCORnet Common Data Model or readily available as coded data in their EHR systems. Moreover, for the health system sites, these patients often had tests and procedures at other institutions that were not documented in their EHR systems. This change, along with a broadened list of enrichment factors, were incorporated into the computable phenotype and more than doubled the eligible patient population.

The coordinating center posted updates to the computable phenotype via GitHub throughout the duration of the trial. Sites making changes to their base code were encouraged to post their work to GitHub or share it with other sites and the coordinating center.

Computable phenotype implementation and patient enrollment

All sites leveraged the computable phenotype as the main part of their strategy to identify eligible participants. The first phenotype was run at 19 sites across eight participating Clinical Data Research Networks. The computable phenotype identified 174,785 eligible

patients. The second phenotype was run at 40 sites across nine participating Clinical Data Research Networks and one Health Plan Research Network and it identified 657,215 eligible patients. A total of 15,076 participants were enrolled from April 2016 to June 2019. Figure 2 illustrates the trial recruitment metrics, stratified by the Clinical Data Research Networks and Health Plan Research Network recruitment efforts.

The recruitment strategies include: mailed letters, secure messaging, phone calls, electronic health record-based trial recruitment tools, social media campaigns, and traditional in-clinic recruitment. The details of the recruitment strategies and results will be reported separately. The section below describes the four principal lessons learned from the implementation of the computable phenotype.

LESSONS LEARNED

Choose your data sources based on local data quality and study needs

Ideally, in a multi-site trial in a distributed research data network like ADAPTABLE, a single computable phenotype could be run that would accurately identify eligible participants. However, during the creation of the computable phenotype, the development team recognized that completeness and quality of the data can greatly influence the performance of a computable phenotype,²⁴ and they designed an approach to account for these issues. For example, using a standardized data model like the Common Data Model does not guarantee that all of the data are the same (or “normalized” in informatics terms). The use of a common data model ensures all participating sites will have a local version (or “instance”) of a database with the same set of tables having the same field names, all populated according to a set of rules (or “specifications”). On a quarterly basis, the PCORnet Coordinating Center facilitates data characterization at each site through a series of queries, which are reviewed for each datamart by the Coordinating Center and local team, and include assessments for data completeness, data plausibility, and data model conformance.²⁵ This ensures a baseline level of data quality needed to execute a distributed query like the ADAPTABLE computable phenotype. Despite following a common set of rules and undergoing systematic data curation processes, the content of the Common Data Model often differs among institutions. These differences arise from the makeup of the participating institutions (e.g. academic facilities vs. community-based care centers), clinical practice and coding styles, the interval over which electronic data are available or the relative balance of inpatient and ambulatory data.

The underlying source data, which may come from either EHR systems or administrative/billing systems, can affect content, and the data quality can differ based on the EHR system in use. Many of the clinical criteria in the computable phenotype were extracted from all available codes in the diagnosis and procedure tables in the Common Data Model at each site; however, sites, based on guidance from the PCORnet Coordinating Center and their local source systems, decided whether to load codes from EHR or administrative/billing systems or both. The specific extraction (E), transformation (T) and loading (L) practices used to move data from the source system into the data model can also affect the data content. While the Common Data Model ensures a computable phenotype will run at all participating sites, the results of the computable phenotype “output” in terms of the true

number of patients meeting the desired criteria at all the sites needs to be interpreted with caution.

Local factors contribute to the general quality of the available data, which in turn will affect the accuracy of the computable phenotype. For example, some EHR systems facilitate mapping key structured data elements, such as medications and labs, to the standards used in the Common Data Model and by the computable phenotype. Sites with EHR systems more conducive to accurate mappings, in general, have better data quality and specifically may have more complete data to use in the execution of the computable phenotype. Some sites have more accurate death data through linkage to state or national vital records while others rely exclusively on health system data. These differences led to variations in the ability to exclude deceased patients. Another factor is frequency of updating the data used for the computable phenotype. Sites in PCORnet refresh their datamarts approximately every three months. If the computable phenotype is run directly on the Common Data Model datamart, new data between the time of refresh and the time of computable phenotype execution are not included. This concept, referred to as data latency, has implications for accurately identifying eligible patients for the trial.

Understanding these challenges with EHR data quality and common data models are essential during computable phenotype development. In ADAPTABLE, these challenges were addressed by creating a base algorithm that was then tested and customized locally at the sites by teams of clinical leaders, clinical informaticians, researchers, and data experts. In most cases, site teams elected to customize and validate the computable phenotype on local source systems, in part due to concerns with data latency and need to access data not included in the Common Data Model. These additional data enhanced the identification of eligible participants and facilitated recruitment.

Recognize and accept translation limitations

Ideally, all of the inclusion and exclusion criteria in ADAPTABLE could be precisely mapped to electronic data elements. Unfortunately, the translation from written criteria into valid and actionable code may not be feasible for multiple reasons. Categorizing a disease, which is inherently complex, into codes creates an opportunity for information loss. For example, a condition such as three-vessel artery disease is not reliably identified in a single code or set of codes. Instead, valid and reliable identification would likely be based on several structured and unstructured data elements within the patient record. Moreover, some clinical constructs may require advanced computing techniques to fully extract (e.g. the use of natural language processing to extract left ventricular ejection fraction from unstructured documents).

When identifying patients with a particular disease, the concept of time in EHR data is important and requires attention during computable phenotype development. The time associated with the first diagnosis code reflects the time of entry, and not necessarily time of onset of disease. Diagnosis codes often persist in a patient's record well after the disease or condition has resolved while procedure codes remain fixed to a single time-point in the record. A patient may be admitted with a "suspected" diagnosis of acute myocardial infarction, which is appropriately entered as a diagnosis in the EHR at the time of admission.

However, subsequent inpatient evaluation may have ruled out acute myocardial infarction and established an alternative diagnosis such as pericarditis. Yet, the diagnosis code for acute myocardial infarction may persist in the medical record for an indefinite period of time.

Another challenge can arise from using procedure codes. For example, the PCORnet datamarts predominantly include data after 2010. The first version of the ADAPTABLE computable phenotype used procedural codes—ICD and Current Procedural Terminology (CPT)—to identify patients with a history of coronary artery bypass grafting (CABG) surgery or history of percutaneous intervention (PCI). However, if a patient had a CABG or PCI in 2005, those procedural codes would not necessarily be available in the PCORnet datamarts. A clinician seeing that patient after 2010 would likely only use clinical codes for chronic ischemic heart disease; thus, querying only for procedure codes would miss these patients. The second version of the computable phenotype improved identification of these patients via expanding the definition to include history of chronic ischemic heart disease, coronary artery disease, or atherosclerotic cardiovascular disease.

For these reasons (and many more), determining and defining incident and prevalent disease is particularly problematic. The ADAPTABLE computable phenotype accounted for these challenges by using previously-tested definitions of diseases, including broad criteria to identify patients with ASCVD, allowing for local validation and customization, and building in confirmation of key inclusion and exclusion criteria during the enrollment process.

Validate and tailor locally

In a multi-site pragmatic clinic trial using a computable phenotype, the study leadership has to provide guidance on validation. In ADAPTABLE, the coordinating center asked each site to develop its own validation strategy of the computable phenotype. Among participating ADAPTABLE sites, nearly everyone validated the computable phenotype. Validation practices varied in terms of the process, data availability, and number of charts reviewed. The Health Plan Research Network validation, for example, focused on optimizing positive predictive value of the computable phenotype because the network only has access to claims data and could not routinely review patient charts. Their implementation of the computable phenotype limited the sample to patients with codes for acute myocardial infarction or revascularization to minimize risks of misclassification. Validation of the Health Plan Research Network computable phenotype through chart review of 185 patients demonstrated a positive predictive value of 90.8%.²⁶

Health system sites developed and executed their own validation plan based on local factors, including previous experience identifying patients with coronary artery disease for trials, site recruitment strategies, and resource constraints. Based on local validation efforts, certain sites further optimized the computable phenotype by including additional criteria on their computable phenotype and leveraging local data beyond that available in the Common Data Model. For example, some sites increased the minimum age requirements to reduce the likelihood of identifying patients with congenital heart disease. Other sites sought to improve the accuracy of clinical ASCVD identification by including additional requirements, such as at least 2 instances of ASCVD diagnostic codes, a history of an aspirin

prescription, or documentation of an ASCVD ICD code by a cardiologist. Due to the heterogeneity in validation methods across sites and the decentralized nature of the Common Data Model, we are unable to describe computable phenotype performance across all study sites.

Providing flexibility for the validation of the computable phenotype locally has benefits. First, the coordinating center does not have direct access to the data. Second, local investigators and data scientists understand their data best and are ideal for designing a validation process that aligns well within their own system. Third, the goal of the phenotype may vary among sites. For example, a site that plans to send out mass mailings to all participants may adapt the algorithm to optimize positive predictive values. In contrast, a site planning for only in-clinic recruitment with chart review by a research assistant may elect to optimize sensitivity. Thus, it would be challenging for the coordinating study team to provide significant oversight over the validation process. However, not having guidelines for the validation process will lead to variation in the extent and quality of validation processes across sites, which may introduce bias into the study sample selection. The extent of the guidelines for validation of the computable phenotype should depend on its complexity, resources dedicated to validation, and ongoing discussions between local investigators and the coordinating study team.

Questions as a safety net

A computable phenotype will never capture all criteria of interest and it will suffer from tradeoffs between sensitivity and specificity. To protect the safety of participants and the integrity of the trial, it is essential to recognize the limitations of the study's computable phenotype within and across sites and understand whether additional screening is needed using mechanisms such as chart review or asking self-reported questions to potential participants either in-person, over the phone, or via a patient portal. If the limitations of the phenotype are identified, they can be used to inform the types of questions patients should be asked prior to enrollment. To this end, during the participant enrollment process using an online portal in ADAPTABLE, each participant is asked a series of questions related to key inclusion and exclusion criteria. For the participating Health Plan Research Network, these questions were essential because their data was exclusively claims-based and lacked information about exclusion factors such as medication allergies.

Within ADAPTABLE, potentially eligible patients were given the opportunity to confirm their eligibility by answering questions in a patient portal (Table 1). During the enrollment period from April 2016 to June 2019, 32,087 patients entered the portal and 18,497 completed the eligibility questions (remaining patients dropped out of the portal prior to answering questions). As shown in Table 1, the computable phenotype did not identify 896 unique patients with self-reported exclusions to the trial, representing a 4.8% failure rate. Failure to identify current use of a contraindicated medication represented over half of the failure rate cases (543/896). An additional 1,342 patients deemed ineligible because they self-reported an inability to change their aspirin dose. Beyond matching into the computable phenotype and answering the portal screening questions, additional eligibility confirmation occurred locally, although these practices varied by site and recruitment strategy.

DISCUSSION

The wide-scale adoption of EHRs and the development of national, electronic data research networks with standardized data models have created the opportunity to revitalize the clinical enterprise infrastructure and optimize the research process. Many challenges experienced by ADAPTABLE investigators mirror those described by other groups using EHR and claims data to develop computable phenotypes and conduct observational epidemiological studies and clinical trials over the last thirty years.^{12,13,27,28}

The ADAPTABLE computable phenotype successfully identified over 650,000 potentially eligible participants for the trial and helped the trial achieve its recruitment goal of 15,000 patients. Although computable phenotypes can enhance the speed and efficiency of recruitment efforts, there are many challenges with their implementation within national data research networks. Through the experience with ADAPTABLE, several key recommendations were identified that can inform the future use of computable phenotypes (Figure 3).

Historically, EHR data have not been collected to support clinical research, but instead are collected for clinical care and billing. As such, it is important for researchers to maintain reasonable expectations about the utility of EHR data quality to support secondary research.^{29–37} Prior to implementation of a computable phenotype, sites may require additional data quality exercises to improve the quality of clinical constructs identified by the computable phenotype. Beyond data quality issues, important clinical elements are not always easily extractable from the EHR. Instead, some data are stored in text fields, which is difficult, if not impossible, to consistently parse. Computable phenotypes may benefit from using a combination of EHR and non-EHR data sources, such as state or national death data or disease registries. Administrative claims data have not historically been utilized to identify eligible research participants. While these systems typically lack clinical granularity, they do capture qualifying clinical events across health systems.

As learned in ADAPTABLE, a computable phenotype is not a static program. Instead it should be prepared to respond to changes. Throughout the course of the recruitment period, the phenotype should respond to changes in source data or alterations to the study protocol. If the quality of source data changes, the phenotype should be prepared to respond, in kind. In addition, validation efforts along with recruitment efforts should inform changes to subsequent computable phenotype iterations. For example, if through validation efforts, a diagnosis code contributes significantly to false positive rates then removing the code might be necessary. Alternatively, if the computable phenotype fails to identify sufficient patients for recruitment, trial leadership should consider altering the phenotype to broaden the eligible pool.

Implementing a high-quality computable phenotype requires collaboration between the study coordinating center, the data teams, and study sites. Within each study site, adaptation and implementation of the computable phenotype requires close partnership between clinical leaders and data/informatics experts. Their work requires further guidance by their respective compliance policies. Each team member brings a set of different expertise, from

understanding clinical workflows, designing appropriate data and IT architecture, addressing data quality issues along with technical expertise in writing, testing, and customizing computable phenotypes. Supporting a diverse group requires adequate resources for personnel and meetings.

A better understanding of electronic data quality, along with developing methodologies to account for data quality issues, all are important areas for future development and research. Another key area of research is the development of methodology to combine structured and unstructured data elements into a portable computable phenotype across multiple sites.^{38–40} Developing the infrastructure to integrate health system and claims data across large, diverse populations and the methods to develop a single computable phenotype that runs across both types of data will help advance the capacity to conduct large, efficient pragmatic clinical trials. Lastly, improvement in EHR usability and workflow to better capture data, either at the point-of-care by clinicians or with natural language processing pipelines, could improve clinical care as well as increase the efficiency of pragmatic clinical trials. This type of work involves a diverse team, including human factors experts.

This report has important limitations. The ADAPTABLE trial is an ongoing PCORnet demonstration study, and these lessons reflect the initial experience of sites participating in this trial. Similarly, the computable phenotype used in ADAPTABLE was designed specifically for the PCORnet Common Data Model and may not be easily extensible to other data models or analytical programs other than SAS. Due to the heterogeneity in validation efforts along with the decentralized nature of the PCORnet, we are unable to report computable phenotype performance characteristics. However, validation from the Health Plan Research Network and prior work identifying patients with coronary heart disease demonstrated a positive predictive value of over 90%.^{17,26} Finally, the results presented herein are descriptive reflections of a diverse team of researchers working on the ADAPTABLE computable phenotype. While the results are a product of a diverse authorship group (in terms of site location and position within the study), they are descriptive and serve to provide broad recommendations. Future research on the ADAPTABLE computable phenotype will focus on quantitative analysis to provide more specific recommendations.

Conclusion

To our knowledge, this is the first report on the process of development of an EHR-based computable phenotype and the lessons learned during its implementation in a large, national US cardiovascular clinical trial. The results of the investigation include broad lessons learned from experience deploying and implementing a computable phenotype that may be broadly generalizable to other clinical studies. In conclusion, the most important lesson learned in the process of computable phenotype implementation is that challenges must be met with a degree of flexibility, or adaptability, which enables changes based on the data, resources, sites, and patients.

Acknowledgments

We thank the ADAPTABLE study leadership team including Drs. Richard Platt, Adrian Hernandez, Robert Harrington, and Russell Rothman. In addition, we thank the PCORnet coordinating center and the PCORnet Distributed Research Network Operations.

The base code for the ADAPTABLE computable phenotype is located on GitHub here: <https://github.com/ADAPTABLETRIAL/PHENOTYPE>

Funding/Support

ADAPTABLE is funded by a PCORI (Patient Centered Outcomes Research Institute) grant (Contract Number: ASP-150227079) and is the first pragmatic clinical trial conducted by PCORnet, the National Patient-Centered Clinical Research Network. PCORnet is an initiative funded by the PCORI, and aims to improve the nation's capacity to conduct clinical research by creating a large, highly representative network that directly involves patients in the development and execution of research. All statements, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of PCORI, its Board of Governors or Methodology Committee. Dr. Ahmad is in part funded by grant number K12HS026385 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality. The funders had no role in this project, the decision to publish, or preparation of the manuscript.

References

1. Eapen ZJ, Vavalle JP, Granger CB, Harrington RA, Peterson ED, Califf RM. Rescuing clinical trials in the United States and beyond: a call for action. *Am Heart J*. 2013;165(6):837–847. [PubMed: 23708153]
2. Jones WS, Roe MT, Antman EM, Pletcher MJ, Harrington RA, Rothman RL, Oetgen WJ, Rao SV, Krucoff MW, Curtis LH, Hernandez AF and Masoudi FA. The Changing Landscape of Randomized Clinical Trials in Cardiovascular Disease. *J Am Coll Cardiol*. 2016;68:1898–1907. [PubMed: 27765193]
3. Mc Cord KA, Ewald H, Ladanie A, Briel M, Speich B, Bucher HC, Hemkens LG, initiative RCDfR and the Making Randomized Trials More Affordable G. Current use and costs of electronic health records for clinical trial research: a descriptive study. *CMAJ open*. 2019;7:E23–E32.
4. Weinfurt KP, Hernandez AF, Coronado GD, DeBar LL, Dember LM, Green BB, Heagerty PJ, Huang SS, James KT, Jarvik JG, Larson EB, Mor V, Platt R, Rosenthal GE, Septimus EJ, Simon GE, Staman KL, Sugarman J, Vazquez M, Zatzick D and Curtis LH. Pragmatic clinical trials embedded in healthcare systems: generalizable lessons from the NIH Collaboratory. *BMC Med Res Methodol*. 2017;17:144. [PubMed: 28923013]
5. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, Suchard MA, Park RW, Wong IC, Rijnbeek PR, van der Lei J, Pratt N, Noren GN, Li YC, Stang PE, Madigan D and Ryan PB. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Stud Health Technol Inform*. 2015;216:574–8. [PubMed: 26262116]
6. Psaty BM, Breckenridge AM. Mini-Sentinel and regulatory science--big data rendered fit and functional. *N Engl J Med*. 2014;370(23):2165–2167. [PubMed: 24897081]
7. Forrow S, Champion DM, Herrinton LJ, Nair VP, Robb MA, Wilson M and Platt R. The organizational structure and governing principles of the Food and Drug Administration's Mini-Sentinel pilot program. *Pharmacoepidemiol Drug Saf*. 2012;21 Suppl 1:12–7. [PubMed: 22262588]
8. Platt R, Carnahan RM, Brown JS, Chrischilles E, Curtis LH, Hennessy S, Nelson JC, Racoosin JA, Robb M, Schneeweiss S, Toh S and Weiner MG. The U.S. Food and Drug Administration's Mini-Sentinel program: status and direction. *Pharmacoepidemiol Drug Saf*. 2012;21 Suppl 1:1–8.
9. Richesson R, Smerek M and Cameron CB. A Framework to Support the Sharing and Re-Use of Computable Phenotype Definitions Across Health Care Delivery and Clinical Research Applications. *EGEMs (Wash DC)*. 2016;4.
10. Hernandez AF, Fleurence RL and Rothman RL. The ADAPTABLE Trial and PCORnet: Shining Light on a New Research Paradigm. *Ann Intern Med*. 2015;163:635–6. [PubMed: 26301537]

11. Richesson RL, Hammond WE, Nahm M, Wixted D, Simon GE, Robinson JG, Bauck AE, Cifelli D, Smerek MM, Dickerson J, Laws RL, Madigan RA, Rusincovitch SA, Kluchar C and Califf RM. Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory. *J Am Med Inform Assoc.* 2013;20:e226–31. [PubMed: 23956018]
12. Richesson RL, Sun J, Pathak J, Kho AN, Denny JC. Clinical phenotyping in selected national networks: demonstrating the need for high-throughput, portable, and computational methods. *Artif Intell Med.* 2016;71:57–61. [PubMed: 27506131]
13. Pathak J, Kho AN, Denny JC. Electronic health records-driven phenotyping: challenges, recent advances, and perspectives. *J Am Med Inform Assoc.* 2013;20(e2):e206–211. [PubMed: 24302669]
14. Boxwala AA, Peleg M, Tu S, Ogunyemi O, Zeng QT, Wang D, Patel VL, Greenes RA and Shortliffe EH. GLIF3: a representation format for sharable computer-interpretable clinical practice guidelines. *J Biomedical Inform.* 2004;37:147–61.
15. Richesson RL, Smerek M, Rusincovitch S, Zozus MN, Chaudhuri PS, Hammond WE, Califf RM, Simon G, Green Beverly, Kahn M, Laws R. *Electronic Health Records-Based Phenotyping In: Rethinking Clinical Trials: A Living Textbook of Pragmatic Clinical Trials.* Bethesda, MD: NIH Healthcare Systems Research Collaboratory; 2014.
16. PCORnet. Common Data Model (CDM) Specification, Version 5.1. https://pcornet.org/wp-content/uploads/2019/09/PCORnet-Common-DataModel-v51-2019_09_12.pdf. Accessed April 2, 2020
17. Roumie CL, Patel NJ, Munoz D, Bachmann J, Stahl A, Case R, Leak C, Rothman R and Kripalani S. Design and outcomes of the Patient Centered Outcomes Research Institute coronary heart disease cohort study. *Contemp Clin Trials Commun.* 2018;10:42–49. [PubMed: 29696157]
18. RxNav. <https://rxnav.nlm.nih.gov>. Accessed April 2, 2020.
19. National Drug Code Directory <http://www.accessdata.fda.gov/scripts/cder/ndc>. Accessed April 2, 2020.
20. 2016 ICD-10-PCS and GEMs. <https://www.cms.gov/Medicare/Coding/ICD10/2016-ICD-10-PCS-and-GEMs.html> Accessed April 2, 2020.
21. Healthcare Cost and Utilization Project Clinical Classification Software (CCS) for ICD-9-CM. <https://www.hcup-us.ahrq.gov/toolsoftware/ccs/ccs.jsp> Accessed April 2, 2020.
22. ADAPTABLETRIAL/PHENOTYPE. <https://github.com/ADAPTABLETRIAL/PHENOTYPE> Accessed April 2, 2020.
23. Roumie CR, Shirley-Rice J, Kripalani S. MidSouth CDRN - Coronary Heart Disease Algorithm. Vanderbilt University PheKB; 2014 Available from: <https://phekb.org/phenotype/234>. Accessed April 2, 2020.
24. Kahn MG, Raebel MA, Glanz JM, Riedlinger K, Steiner JF. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. *Med Care.* 2012;50 Suppl:S21–29. [PubMed: 22692254]
25. Qualls LG, Phillips TA, Hammill BG, Topping J, Louzao DM, Brown JS, Curtis LH and Marsolo K. Evaluating Foundational Data Quality in the National Patient-Centered Clinical Research Network (PCORnet®). *EGEMS (Wash DC).* 2018;6:3. [PubMed: 29881761]
26. Fishman E, Barron J, Dinh J, Jones WS, Marshall A, Merkh R, Robertson H and Haynes K. Validation of a claims-based algorithm identifying eligible study subjects in the ADAPTABLE pragmatic clinical trial. *Contemp Clin Trials Commun.* 2018;12:154–160. [PubMed: 30480162]
27. Weng C, Tu SW, Sim I, Richesson R. Formal representation of eligibility criteria: a literature review. *Journal of biomedical informatics.* 2010;43(3):451–467. [PubMed: 20034594]
28. Richesson RL, Green BB, Laws R, Puro J, Kahn MG, Bauck A, Smerek M, Van Eaton EG, Zozus M, Hammond WE, Stephens KA and Simon GE. Pragmatic (trial) informatics: a perspective from the NIH Health Care Systems Research Collaboratory. *J Am Med Inform Assoc.* 2017;24:996–1001. [PubMed: 28340241]
29. Kopcke F, Trinczek B, Majeed RW, Schreiweis B, Wenk J, Leusch T, Ganslandt T, Ohmann C, Bergh B, Rohrig R, Dugas M and Prokosch HU. Evaluation of data completeness in the electronic health record for the purpose of patient recruitment into clinical trials: a retrospective analysis of element presence. *BMC Med Inform Decis Mak.* 2013;13:37. [PubMed: 23514203]

30. Weiner JP, Fowles JB, Chan KS. New paradigms for measuring clinical performance using electronic health records. *Int J Qual Health Care*. 2012;24(3):200–205. [PubMed: 22490301]
31. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc*. 2013;20(1):144–151. [PubMed: 22733976]
32. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. *J Biomed Inform*. 2013;46(5):830–836. [PubMed: 23820016]
33. Weiner MG, Embi PJ. Toward reuse of clinical data for research and quality improvement: the end of the beginning? *Ann Intern Med*. 2009;151(5):359–360. [PubMed: 19638404]
34. Dugas M, Lange M, Muller-Tidow C, Kirchhof P, Prokosch HU. Routine data from hospital information systems can support patient recruitment for clinical studies. *Clin Trials*. 2010;7(2):183–189. [PubMed: 20338903]
35. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV, Lehmann HP, Hripcsak G, Hartzog TH, Cimino JJ and Saltz JH. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Care*. 2013;51:S30–7. [PubMed: 23774517]
36. Ahmad FS, Chan C, Rosenman MB, Post WS, Fort DG, Greenland P, Liu KJ, Kho AN and Allen NB. Validity of Cardiovascular Data From Electronic Sources: The Multi-Ethnic Study of Atherosclerosis and HealthLNK. *Circulation*. 2017;136:1207–1216. [PubMed: 28687707]
37. Capurro D, Yetisgen M, van Eaton E, Black R, Tarczy-Hornoch P. Availability of structured and unstructured clinical data for comparative effectiveness research and quality improvement: a multisite assessment. *EGEMS (Wash DC)*. 2014;2(1):1079. [PubMed: 25848594]
38. Sharma H, Mao C, Zhang Y, Vatani H, Yao L, Zhong Y, Rasmussen L, Jiang G, Pathak J and Luo Y. Developing a portable natural language processing based phenotyping system. *BMC Med Inform Decis Mak*. 2019;19:78. [PubMed: 30943974]
39. Kang T, Zhang S, Tang Y, Hrubby GW, Rusanov A, Elhadad N and Weng C. EliIE: An open-source information extraction system for clinical trial eligibility criteria. *J Am Med Inform Assoc*. 2017;24:1062–1071. [PubMed: 28379377]
40. Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. *J Am Med Inform Assoc*. 2013;20(1):117–121. [PubMed: 22955496]

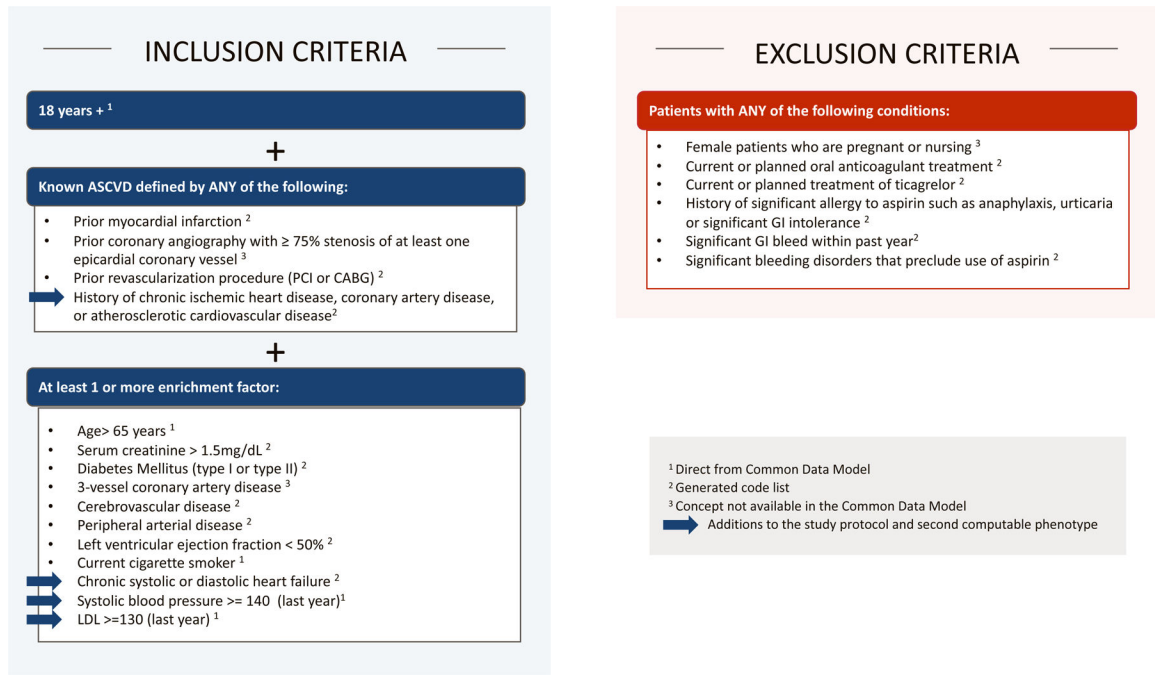


Figure 1. ADAPTABLE Inclusion and Exclusion Criteria Versions 1 and 2.

The inclusion and exclusion for the trial at the time of initial computable phenotype development included data elements that were directly mapped to the common data model, identified via sets of code lists, or not available in the common data model. The new criteria added in Version 2 are annotated with blue arrows. PCI = percutaneous coronary intervention; CABG = coronary artery bypass grafting; GI = gastrointestinal.

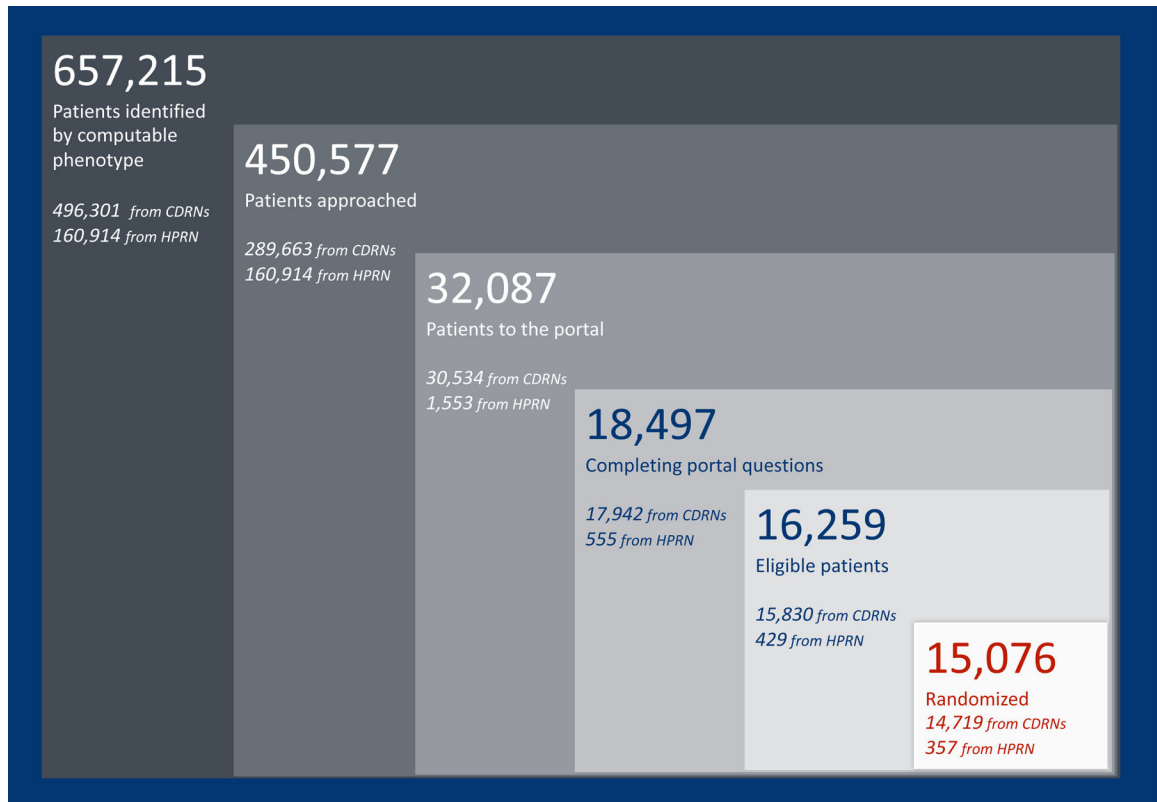


Figure 2. ADAPTABLE recruitment metrics.

The computable phenotype identified a large pool of potentially eligible participants at both clinical data research networks and health plan research networks. CDRN = clinical data research network; HPRN = health plan research network.

KEY RECOMMENDATIONS

✓ 1

Have reasonable expectations about the accuracy and quality of EHR data

✓ 2

Perform extensive computable phenotype testing and be prepared for modifications throughout the study

✓ 3

Create a diverse research team with clinical leaders and data/informatics experts and involve them throughout the study, including protocol development

Figure 3.
Key recommendations from ADAPTABLE computable phenotype implementation.

Table 1.

Eligibility questions answered by potential participants in online portal during enrollment.

Portal Question	Affirmative Patient Response [†]
1. Are you allergic to aspirin?	158 (0.85%)
2. Have you had a severe bleeding problem in the past with aspirin?	282 (1.5%)
3. Are you currently taking any of these anticoagulant medications [*] ?	543 (2.9%)
Total, unique non-eligible [‡]	896 (4.8%)

^{*}Brilinta® (Ticagrelor), Coumadin® (Warfarin), Eliquis® (Apixaban), Pradaxa® (Dabigatran), Savaysa® (Edoxaban), Xarelto® (Rivaroxaban)

[†]Denominator=18,497 patients entering the portal and responding to eligibility questions

[‡]Potential participants may be represented in more than one category.