HHS Public Access

Author manuscript

Clin Trials. Author manuscript; available in PMC 2023 August 01.

Published in final edited form as:

Clin Trials. 2023 August ; 20(4): 416–424. doi:10.1177/17407745231160459.

Pragmatic guidance for embedding pragmatic clinical trials in health plans: Large simple trials aren't so simple

Noelle M Cocoros[1], **Jerry H Gurwitz**[2], **Mark J Cziraky**[3], **Christopher B Granger**[4], **Thomas Harkins**[5], **Kevin Haynes**[6], **Xiaojuan Li**[1], **Lauren Parlett**[3], **John D Seeger**[7], **Sonal Singh**[2], **Cheryl N McMahill-Walraven**[8], **Richard Platt**[1]

[1.]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA

[2.]Meyers Health Care Institute, a joint endeavor of UMass Chan Medical School, Reliant Medical Group, and Fallon Health, Worcester, MA, USA

[3.]HealthCore, Inc, Wilmington, DE, USA

[4.]Duke Clinical Research Institute, Duke University, Durham, NC, USA

[5.]Humana Healthcare Research, Inc., Louisville, KY, USA

[6.]Janssen Research & Development, Titusville, PA, USA

[7.]Optum Epidemiology, Boston, MA, USA

[8.]CVS Health Clinical Trial Services, Blue Bell, PA, USA

Abstract

Background: There are unique opportunities related to the design and conduct of pragmatic trials embedded in health insurance plans, which have longitudinal data on member/patient demographics, dates of coverage, and reimbursed medical care, including prescription drug dispensings, vaccine administrations, behavioral healthcare encounters, and some laboratory results. Such trials can be large and efficient, using these data to identify trial-eligible patients and to ascertain outcomes.

Methods: We use our experience primarily with the NIH Pragmatic Trials Collaboratory's Distributed Research Network, which comprises health plans that participate in the United States Food and Drug Administration's Sentinel System, to describe lessons learned from the conduct and planning of embedded pragmatic trials.

Results: As of April 2022, information is available for research on more than 75 million people with commercial or Medicare Advantage health plans. We describe three studies that have used or plan to use the Network, as well as a single health plan study, from which we glean our lessons learned.

Corresponding Author: Noelle M. Cocoros, 401 Park Drive, Suite 401 East, Boston, MA 02215. Phone: 617-867-4980; Noelle_Cocoros@harvardpilgrim.org.

Conclusions: Studies that are conducted in health plans provide much needed evidence to drive clinically meaningful changes in care. However, there are many unique aspects of these trials that must be considered in the planning, implementation, and analytic phases. The type of trial best suited for studies embedded in health plans will be those that require large sample sizes, simple interventions that could be disseminated through health plans, and where data available to the health plan can be leveraged. These trials hold potential for substantial long-term impact on our ability to generate evidence to improve care and population health.

Keywords

Pragmatic trials; distributed research network; claims data; real-world data

Background

In the United States, embedding clinical trials in health insurance plans (or health plans; organizations that provide healthcare insurance), is a promising approach to generating real-world evidence. Pragmatic clinical trials test an intervention on participants who represent the patients who would receive the new care or treatment if it were adopted in routine practice.[1] When appropriate, pragmatic trials embedded in health plans can be especially pragmatic by leveraging data collected or generated during routine care such as electronic health record and/or administrative health insurance claims data. These data can be used to identify the cohort(s) of interest, support participant contact, or conduct the analyses. Most important, the health plans' ability to use their routine interactions with providers and health plan members allows the possibility of working with external researchers to test selected interventions in a manner that would be implemented in regular practice. This approach can increase efficiency and reduce costs – and observed effects are expected to be generalizable at least to the large segment of the U.S. population that receive medical care covered by health plans. However, these trials come with many important considerations relating to identification of study subjects, choice and measurement of study outcomes, and selection of study settings.[2,3] Others have described the conduct of pragmatic trials embedded in *healthcare delivery systems* (i.e., an organization or group of organizations that provide healthcare) that leverage clinical care settings and electronic health records.[4,5] However, there are unique opportunities related to the design and conduct of pragmatic trials embedded in *health plans*.[6] Both models provide important opportunities to generate evidence, and the optimal model may depend on the intervention of interest.

Health plans have longitudinal data on member/patient demographics, dates of coverage, and reimbursed medical care, including prescription drug dispensings, vaccine administrations, behavioral healthcare encounters, and outpatient laboratory results. Thus, these data, from across health systems, epitomize the type of data routinely collected in patient care that can support assessments of safety and effectiveness of medical interventions, including interventions performed as randomized clinical trials. Embedded pragmatic trials can use these data to identify trial-eligible patients,[7] and to support the intervention, as well as ascertain many outcomes of interest.[8] The claims data can be supplemented by other sources, such as self-reported patient or provider information, medical charts/electronic health records, and additional data captured such as those captured through Healthcare

Effectiveness Data and Information Set. Using an extant network of health plans, such as the NIH Pragmatic Trials Collaboratory Distributed Research Network, can further increase efficiency as well as potential to implement trials in multiple health plans, both to increase generalizability and to enlarge sample sizes. The NIH Collaboratory's Distributed Research Network comprises health plans that participate in the United States Food and Drug Administration's (FDA) Sentinel System.[9–11] In this article, we describe lessons learned from the conduct and planning of multisite pragmatic trials utilizing this network as well as from a single health plan study.

Methods

The NIH Pragmatic Trials Collaboratory Distributed Research Network leverages the Sentinel System's technical infrastructure, which was originally developed for post-marketing medical product surveillance, but was also envisioned as a national resource, including one for pragmatic trials and other research studies.[10] The health plans that participate in Sentinel transform their data into a common data model that harmonizes their data to streamline analyses and research across the health plans.[12] Those partners that are part of the Collaboratory are actively accruing data on ~75 million individuals with commercial or Medicare Advantage. Importantly, beyond efficiency and technical abilities, the health plans in the NIH Collaboratory Distributed Research Network are deeply engaged in, committed to, and experienced in the conduct of research. The projects we describe were all approved by applicable institutional review boards.

IMPACT-AFib - IMplementation of a randomized controlled trial to imProve treatment with oral AntiCoagulanTs in patients with Atrial Fibrillation (Clinicaltrials.gov NCT03259373)

This FDA-funded proof-of-concept trial assessed whether an educational intervention mailed to health plan patients, and their providers, would prompt treatment with an oral anticoagulant among those untreated patients with atrial fibrillation who met clinical guideline recommendations for their use. The study identified patients of five health plans with evidence of atrial fibrillation, additional risk factors for stroke, and no evidence of having been dispensed an oral anticoagulant via the health plans' claims data.[13] Patients identified as eligible through the Sentinel System infrastructure were randomized to the intervention or control arm, where the intervention was a carefully constructed educational mailing to them and the provider who most recently diagnosed them with atrial fibrillation, with the intent to prompt the patients to discuss treatment with their providers; controls received usual care during the initial study period. The primary outcome was the proportion of previously untreated patients dispensed an oral anticoagulant within one year of the intervention. Secondary outcomes included rates of stroke and hospitalized bleeding events. The health plans' data were also used for the analyses. Among more than 47,000 patients randomized and included in the final analysis, there was no statistically significant evidence of an effect of the intervention on the primary or secondary outcomes.[14] We have previously described some of the lessons learned from this trial.[15,16]

D-PRESCRIBE-AD - Developing a PRogram to Educate and Sensitize Caregivers to Reduce the Inappropriate prescription Burden in Elderly with Alzheimer's Disease (Clinicaltrials.gov NCT05147428)

The NIH Collaboratory Distributed Research Network was used to assess feasibility of a trial to reduce the occurrence of prescribing cascades among health plan patients with dementia. Prescribing cascades are defined as situations in which an adverse drug effect is not recognized as such and, as a consequence, instead of discontinuing or reducing the dosage of the agent causing the new symptom, a second medication is added to counter what is perceived to be a new medical problem.[17] One of the most common prescribing cascades described in the literature is antidopaminergic drugs causing symptoms misinterpreted as Parkinson's disease and triggering the dispensing of antiparkinsonian medication. There was particular concern about this phenomenon among persons living with dementia. A feasibility analysis was conducted in two health plans using the Distributed Research Network data. The same computer program was used by both organizations, simplifying the work and ensuring comparability of the analyses in the two organizations. Among 121,538 individuals with Alzheimer's disease and related dementias who were found to be eligible for inclusion in the study cohort, the incidence of the antidopaminergic-antiparkinsonian prescribing cascade was low, obviating the need for an intervention in this population. Planning for such a trial was thus discontinued.[18] This ability to characterize the intended study population, determine the potential available cohort size, and feasibility for the trial is a major benefit of embedding trials in health plans.

From those learnings, the team modified the trial design and is implementing the D-PRESCRIBE-AD trial which is currently in progress.[19] Two health plans in the Distributed Research Network are being used to address whether a patient/caregiver-centered mailed educational intervention can modify the use of potentially inappropriate medications among adults > 50 years with Alzheimer's disease and related dementias. The target drug classes are sedative-hypnotics, strong anticholinergics, and antipsychotics. The trial is randomized on the individual-level and has three arms: combined patient/caregiver and prescribing provider educational intervention, provider-only educational intervention, and usual care. The health plans' data were used for cohort identification and randomization of >14,000 subjects. The health plans' usual mechanisms for contacting providers and their members were used to deliver the intervention. Health plan baseline and follow-up data will be used for the analysis as well. The primary outcome is cessation of the identified inappropriate drug of concern after six months of follow-up, after a three-month blackout. Secondary outcomes include dose reduction of the potentially inappropriate medication, healthcare utilization measures, and inpatient mortality. Institutional review boards waived a requirement for informed consent by providers or members because the intervention was deemed to be consistent with the health plans' usual quality improvement activities.

Co-CARE-AD – Collaborative Care Coordination Program for Alzheimer's Disease and Related Dementias (Clinicaltrials.gov NCT05281744)

This ongoing randomized trial, implemented in a single health plan, is aimed at evaluating the effectiveness of a collaborative care coordination program for health plan members with dementia living at home and to their caregivers. Programs that use multidisciplinary teams

to provide care coordination and support can benefit patients and their caregivers across all disease stages. Eligible members are identified through claims data with an oversampling of the racial and ethnic minority members to increase their representation. The trial engages primary care providers who can remove their patients and their caregivers from the pool eligible for contact, followed by outreach to patients and their caregivers via a mailed introductory brochure and subsequent telephone calls by social workers.

ACHIEVE - A Controlled trial to improve use of High IntEnsity statins for Vascular protEction

The ACHIEVE trial is planned to be a multi-health plan, cluster randomized trial designed to assess the impact of a pharmacist-led provider- and patient-directed intervention to improve evidence-based statin initiation among individuals with atherosclerotic cardiovascular disease, particularly in the Black population. A low proportion of patients with atherosclerotic cardiovascular disease are on high-intensity statins in commercially insured populations, thus there is an important need to develop effective interventions.[20] Eligible patients and their providers will be identified through claims data, and clinicians will be asked permission for the pharmacist to contact their patients. The pharmacists will work with the both the patient and the clinician to increase the prescription of, and adherence to, appropriate statin therapy. A combination of claims data and pharmacist- and patient-reported data will be used for the analysis. The trial will target enrollment of at least 50% Black members, with race ascertainment based on patient self report. To ensure that an adequate sample of Black patients is offered an opportunity to participate, the investigators will use health plan data about race when this available, and will supplement this information as needed by oversampling in census areas with higher proportions of Black residents. Note that while bias in cluster randomized trials can be a concern,[21] reliance on health plan data to identify eligible members should help avoid the bias that might arise from decisions made by individuals who are aware of treatment assignment.

Results

Key lessons learned are shown in Table 1, along with the major advantages and disadvantages of this approach and key ethical issues.

Planning

Pragmatic clinical trials embedded in health plans need committed, internal staff employed with the skills to navigate, champion, and oversee the work in their organizations and ensure the necessary data are available. These individuals play an essential role in coordination across health plans to ensure that the interventions are implemented consistently in the different organizations. While some of the lessons from the IMPACT-AFib trial have already been described,[15,16] some notable ones are worth repeating. For example, staff from health plans must be involved throughout the trial, including the pre-planning phase. This allows the study team to understand health plan concerns, ensure the trial protocol is feasible and aligns with each organization’s mission, culture, communication style to patients and providers, and processes. One simple outcome of early engagement is understanding the time needed for health plans’ clinical and operational leadership to review and approve all

materials, such as protocols and mailings of educational materials to patients and providers. Gaining approval for identical communications and processes across organizations requires both special attention and additional time, compared to implementation in a single health plan.

Even with engagement from health plan staff, it can be difficult to identify existing or planned programs that could directly affect the proposed trial intervention. Learning about such programs – including those that may conflict with or modify the uptake or impact of the trial intervention – can be challenging given the size and complexity of health plans.

Data in the NIH Collaboratory Distributed Research Network can be crucial in the feasibility or pilot stages of a trial. In the prescribing cascade example above, the team was able to make a “no-go” decision based on appreciation from real-world data that this prescribing cascade was not a major contributor to inappropriate prescribing in this population. In addition, specifications related to cohort and outcome definitions can be explored prior to trial start. When planning a pragmatic clinical trial embedded in health plans, timing of the work is key. Health plan enrollment and disenrollment are most likely to occur at the start of the calendar year. Therefore, researchers should identify eligible patients early in the year, but after the membership changes have stabilized. Similarly, attrition and loss to follow up due to disenrollment or death should be anticipated. Plans can provide estimates of expected annual turnover to inform power calculations. Turnover often varies by employment status, age, and the presence of chronic diseases.[22]

Many health insurers serve different populations via various plans (commercial, Medicare Advantage, managed Medicaid), and therefore, researchers should be aware of the particular population served by a health plan and resulting impacts on bias and generalizability. Importantly, the inclusion of multiple health plans in the projects discussed has enabled the researchers involved to assess heterogeneity in results. Whether patients can be included in trials may also depend on the specific plan they are in and health plans’ obligations under their contracts with purchasers. For example, for commercial insurance, the purchaser is typically an employer; for Medicare, the purchaser is the Centers for Medicare and Medicaid Services. Thus, different restrictions apply and different populations may be eligible for inclusion. For example, including Medicare patients in a pragmatic trial can come with additional challenges; in IMPACT-AFib a letter of support from the Centers for Medicare and Medicaid Services facilitated plan participation. Inclusion of Medicaid members of health plans typically requires state by state approval.

Identification of eligible patients via claims-based algorithms can result in misclassification (e.g., patients incorrectly identified as having the condition(s) of interest). It is in the best interest of the study itself and the health plans in general to reduce the inclusion of “false positives” in the study cohort or to introduce a screening step to confirm eligibility. We recommend use of validated algorithms with known performance characteristics.[23,24]

An early assessment is required about how to conduct certain parts of the study, such as cohort identification and analysis. If the health plans’ data have been transformed to a common data model and are updated and curated regularly, as is the case with the NIH

Collaboratory Distributed Research Network, then it is possible to distribute the same analytic code to each health plan, which then executes it against its local data in that common model. This approach makes more efficient use of programmer effort and ensures that the analyses at each site are identical, versus a shared protocol model where each health plan/site develops de novo programming. Which option to choose, or a hybrid of the two, will depend on the details of the trial.

Saving data

Regardless of what organization develops the programming, there must be careful and deliberate thought about when and what data to save. While historical data are typically available, important changes occur over time, with new information coming in sometimes years after a calendar period. Changes can include addition of episodes of care, new diagnoses, and changes in eligibility. There often is no record of these changes in the health plan data warehouses which is one of the challenges associated with repurposing data for research. To ensure the data of interest for a trial are available for future use, key information must be saved in real time. For example, in D-PRESCRIBE-AD, multiple cohorts are saved, including all those potentially eligible patients based on diagnosis and prescription criteria, all potentially eligible after restricting to one patient per provider (to avoid contamination across arms), and the final cohort randomized to each of the three study arms. By saving multiple cohorts' data, all claims history for relevant patients is available, if needed.

"Fresh" data

Prospective clinical trials based in health plans typically cannot rely exclusively on "settled" claims data for identifying and randomizing the eligible patients. Health plans conducting pragmatic trials must ensure they only approach patients who meet the eligibility criteria, including being actively enrolled, alive, and still having evidence of being on – or not on – the medication(s) of interest to the study, as applicable. In the data used routinely for the Sentinel System, claims transformed into the common data model are generally lagged by several months to avoid incomplete data. If fresher data are desired for outpatient or inpatient encounters, it is important to understand that the time to data completeness varies by data element, care setting, and by health plan. It can take 6 months or more for complete or near-complete inpatient data. Therefore, in general, fresh data must be generated – and reviewed to assess and estimate completeness based on prior data. Enrollment data can be obtained with an approximate one-week lag and outpatient pharmacy dispensing data with an approximate two-week lag. In addition, data capturing deaths outside of the hospital setting are notoriously delayed in many data sources in the U.S.[25] Therefore, trials conducted in a source such as the Distributed Research Network will likely need to focus on in-hospital death instead of mortality overall. While pragmatic trials are expected to need fresh data at the time of assessing patient eligibility and randomization, more complete data will be appropriate for ascertainment of outcomes or else outcome misclassification will need to be addressed (i.e., lags will need to be built in). Understanding these details is key to developing an appropriate protocol and analytic plan.

Data quality

Even with a fully distributed approach that uses shared cohort identification and analytic programs, there is still substantial work with the data to be done within each health plan. Ongoing data quality checks are necessary throughout the trial to ensure the data are of sufficient quality for the intended use. The data in the NIH Collaboratory Distributed Research Network are regularly updated and go through extensive, routine quality checks as part of the Sentinel System process, but supplementary data such as fresh enrollment data are sometimes added. We have implemented limited versions of the Sentinel routine data quality program to run on the fresh enrollment and outpatient pharmacy data used for IMPACT-AFib and D-PRESCRIBE-AD. This step can take days to weeks depending on whether any issues are found that require investigation. We have found it imperative to build in additional data quality checks at various points in the trial to ensure the health plans are coding and extracting data in the same way, and that programs perform as intended (e.g., the health plans will conduct checks of cohort eligibility prior to mailings for the D-PRESCRIBE-AD trial).

Race and ethnicity data are frequently missing from claims-based data sources, and the feasibility of capturing these data should be assessed early in the process. Even if the data are captured there may be differences in these data depending on the source (self-report vs imputed data). Ensuring a study population is diverse is a challenge in many settings and with many trial approaches. In the ACHIEVE study, which is currently in proposal phase, we anticipate the need to capture self-reported race from participants. Planning work has shown which potential health plans have a membership that is diverse enough to justify participation. Additionally, it will be possible to focus recruitment in census areas with a diverse mix of residents.

Data management

To avoid introducing bias, health plan sites must handle all their arms the same way, even if the control arm is assigned to usual care. Assessments for all aspects of inclusion and eligibility are still needed in the control arm, such as identifying patients/providers with a complete address and those on a ‘do not contact’ list. These actions must be performed at the same time for both groups, to avoid introduction of error resulting from changes over time. While this may seem like an obvious requirement, the logistical complexities involved in pragmatic trials mandate attention to these details. For example, in D-PRESCRIBE-AD, sites conducted all eligibility criteria checks prior to randomization via the distributed program.

Data sharing

Health plan data contain proprietary information about patients, formularies, care practices, and business processes.[26] Thus, data sharing plans must be discussed early, allowing time for the execution of data sharing agreements as necessary between the health plan partners and the trial coordinating and/or data centers. Certain data are business sensitive and are not shareable (e.g., data on patientship within small geographical areas). There should be discussions around what level of sharing is most appropriate to perform the analysis. Some studies can be completed using aggregate data, some studies use distributed regression and

other “hands off” approaches to performing individual-level analyses, whereas other studies may require deidentified datasets. With IMPACT-AFib, no individual-level data were shared outside of the health plans. Instead, a distributed approach was used for all aspects of the work (i.e., a single program was run by all five plans to conduct the analysis within their site; site-specific results on the aggregate level were then used for meta-analysis[14]). For the analysis of D-PRESCRIBE-AD, plans will provide deidentified individual-level data to be combined for analysis and data use agreements will be required.

Provider considerations

For trials that include contact with a healthcare provider, there are numerous considerations when working with health plan data that inform the feasibility of the work. In particular, identifying the provider in claims data can be challenging because a provider can be an individual who provides a diagnosis, treatment, or prescription – or a provider can be a facility or other healthcare company like ambulance company, laboratory provider, or durable medical equipment company. The development of a hierarchy may be helpful because many “providers” may be attached to a single patient and the apportionment of clinical responsibility may be unclear. If a provider associated with the most recent diagnosis for example is not an individual, it may be necessary to consider previous diagnoses or identify a primary care provider. What is feasible will depend on the health plan and choosing the right provider will depend on the nature of the trial and the research question. Of note, providers associated with prescribed medications are more likely to be individuals than facilities or clinics.

Because health plans may have difficulty identifying *provider groups* in their data, advance discussions are necessary about how to identify the required target provider (facility, prescriber, diagnosing physician). Providers care for patients in more than one health plan, and this may lead to a provider or prescriber being assigned to the intervention group in one health plan and to the comparison group in another. In our experience, administrative barriers have not allowed trial personnel to create a master provider index that includes all providers in each participating health plan. To avoid cross-contamination, one approach can be to limit participants in a geographic region to a single health plan (e.g., each plan is assigned a mutually exclusive set of states, metropolitan statistical areas, or some other level of geography).

Implementation

Engagement of providers and patients.—Implementing an intervention in a health plan often involves engaging with the provider and/or the patient. Providers can be engaged to screen potential subjects or to allow the use of their name in contacting patients, as is planned for the ACHIEVE study. In Co-CARE-AD, providers can opt out patients and/or their caregivers of the study on their behalf. Engagement with patients is intensely scrutinized by health plans, who frequently require contact materials to go through legal, privacy, communication, and clinical committee reviews and thus may limit the frequency and method of repeated patient contact.

Modalities for connecting with patients include postal mailings, emails, websites, text messages, and/or phone calls. Our direct experience with some of these methods has met with limited success. For example, for the IMPACT-AFib trial, there was a relatively high rate of undeliverable mail particularly for providers and some variation by plan [7% (range 3-11%) for providers, 2% (range <1-4%) for patients]. For this trial we also developed a study website targeted to patients; a mailer to the patients directed them to the website for in depth information about atrial fibrillation and the need for treatment. There was extremely low traffic to the study website. For Co-CARE-AD, we anticipated challenges with contacting patients by phone by sending a letter in advance.

Analysis

There is typically a lag of several weeks between randomization and actualizing the intervention with patients or providers due to the logistics of working with vendors who administer the intervention, as applicable, and working with a list of hundreds or thousands of individuals. In the time between randomization and actual intervention date, people may become ineligible by health plan disenrollment, death, or changing treatment status. Therefore, investigators may need to do a modified intent-to-treat analysis where people who became ineligible between randomization and intervention date are excluded from the analysis.

Well before the time of analysis, the study team will need to determine how the final analysis will be conducted. Will each health plan generate plan-specific results for subsequent pooling or meta-analysis – or will the plans generate deidentified, limited datasets that can be combined with other health plans to create a pooled analytic dataset? Both scientific and practical considerations influence these decisions. The most straightforward scientific approach may be to create and analyze a pooled dataset, but distributed methods may be needed if the health plan is unable to provide an individual-level dataset containing all required data elements for privacy or intellectual property protection.

Whatever the final data to be generated for the analysis, just as with the cohort identification step, it is possible to distribute the same analytic code to each health plan for execution of parts of all of the analysis. This approach makes more efficient use of programmer effort and ensures that the analyses at each site are identical, versus a shared protocol model where each health plan/site develops de novo programming.

Conclusions

Studies that use data from health plans, such as those conducted by the NIH Pragmatic Trials Collaboratory Distributed Research Network, can provide much needed evidence to drive clinically meaningful changes in care. These trials can be used to assess whether an intervention is effective in different geographical locations and populations – and in multiple complex organizations. The kinds of pragmatic clinical trials that are the best candidates for embedding in health plans are studies that test an intervention that is suitable for a health plan to implement and that require a large sample size, and where the data already available to the health plan can be leveraged for one or more aspects of the study. While

such studies are logistically challenging, the possible efficiency for both conducting trials and implementing the results provides the potential for substantial long-term impact.

Acknowledgements

The authors thank Karen Staman, MS, of CHB Wordsmith, for her writing and editorial assistance; funding for this support was from the NIH Collaboratory. We also thank Audrey Wolfe and Juliane Reynolds for administrative support. The program is supported by the National Institutes of Health (NIH) through cooperative agreement U24AT009676 from the Office of Strategic Coordination within the Office of the NIH Director. It is also supported by the NIH through the NIH HEAL Initiative under award number U24AT010961. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or its HEAL Initiative.

Declaration of conflicting interests

Lauren Parlett is an employee of Anthem and has received research support from Sanofi for a single study. All other authors declare that they have no conflict of interest.

Data Availability

Not applicable. We summarize lessons from trials planned or conducted.

References

1. Ford I and Norrie J. Pragmatic trials. N Engl J Med 2016; 375: 454–463. [PubMed: 27518663]
2. Mc Cord KA, Al-Shahi Salman R, Treweek S, et al. Routinely collected data for randomized trials: promises, barriers, and implications. Trials 2018; 19: 29. [PubMed: 29325575]
3. Bynum JPW, Dorr DA, Lima J, et al. Using Healthcare Data in Embedded Pragmatic Clinical Trials among People Living with Dementia and Their Caregivers: STATE OF THE ART. J Am Geriatr Soc; 68 Suppl 2: S49–S54. [PubMed: 32589274]
4. Weinfurt KP, Hernandez AF, Coronado GD, et al. Pragmatic clinical trials embedded in healthcare systems: generalizable lessons from the NIH Collaboratory. BMC Med Res Methodol 2017; 17: 144. [PubMed: 28923013]
5. Richesson RL, Marsolo KS, Douthit BJ, et al. Enhancing the use of EHR systems for pragmatic embedded research: lessons from the NIH Health Care Systems Research Collaboratory. J Am Med Inform Assoc 2021; 28: 2626–2640. [PubMed: 34597383]
6. Choudhry NK. Randomized, Controlled Trials in Health Insurance Systems. N Engl J Med 2017; 377: 957–964. [PubMed: 28877014]
7. Shi Q, Shambhu S, Marshall A, et al. Role of health plan administrative claims data in participant recruitment for pragmatic clinical trials: An Aspirin Dosing: A Patient-centric Trial Assessing Benefits and Long-term Effectiveness (ADAPTABLE) example. Clin Trials 2020; 17: 212–222. [PubMed: 32009464]
8. Ma Q, Chung H, Shambhu S, et al. Administrative claims data to support pragmatic clinical trial outcome ascertainment on cardiovascular health. Clin Trials 2019; 16: 419–430. [PubMed: 31081367]
9. Behrman RE, Benner JS, Brown JS, et al. Developing the Sentinel System--a national resource for evidence development. N Engl J Med 2011; 364: 498–9. [PubMed: 21226658]
10. Platt R, Brown JS, Robb M, et al. The FDA Sentinel Initiative - An Evolving National Resource. N Engl J Med 2018; 379: 2091–2093. [PubMed: 30485777]
11. Robb MA, Racoosin JA, Sherman RE, et al. The US Food and Drug Administration's Sentinel Initiative: expanding the horizons of medical product safety. Pharmacoepidemiol Drug Saf 2012; 21 Suppl 1: 9–11. [PubMed: 22262587]
12. Curtis LH, Weiner MG, Boudreau DM, et al. Design considerations, architecture, and use of the Mini-Sentinel distributed data system: USE OF THE MINI-SENTINEL DISTRIBUTED DATABASE. Pharmacoepidemiol Drug Saf 2012; 21: 23–31. [PubMed: 22262590]

13. Al-Khatib SM, Pokorney SD, Al-Khalidi HR, et al. Underuse of oral anticoagulants in privately insured patients with atrial fibrillation: A population being targeted by the IMplementation of a randomized controlled trial to imProve treatment with oral AntiCoagulanTs in patients with Atrial Fibrillation (IMPACT-AFib). Am Heart J 2020; 229: 110–117. [PubMed: 32949986]
14. Pokorney SD, Cocoros N, Al-Khalidi HR, et al. Effect of Mailing Educational Material to Patients With Atrial Fibrillation and Their Clinicians on Use of Oral Anticoagulants: A Randomized Clinical Trial. JAMA Netw Open 2022; 5: e2214321. [PubMed: 35639381]
15. Cocoros NM, Pokorney SD, Haynes K, et al. FDA-Catalyst—Using FDA's Sentinel Initiative for large-scale pragmatic randomized trials: Approach and lessons learned during the planning phase of the first trial: Clin Trials 2019; 16: 90–97. [PubMed: 30445835]
16. Garcia CJ, Haynes K, Pokorney SD, et al. Practical challenges in the conduct of pragmatic trials embedded in health plans: Lessons of IMPACT-AFib, an FDA-Catalyst trial. Clin Trials 2020; 17: 360–367. [PubMed: 32589056]
17. Rochon PA, Gurwitz JH. The prescribing cascade revisited. The Lancet 2017; 389: 1778–1780.
18. Singh S, Cocoros NM, Haynes K, et al. Antidopaminergic-Antiparkinsonian Medication Prescribing Cascade in Persons with Alzheimer's Disease. J Am Geriatr Soc 2021; 69: 1328–1333. [PubMed: 33432578]
19. Gurwitz J. D-PRESCRIBE-AD (The Developing a PRogram to Educate and Sensitize Caregivers to Reduce the Inappropriate Prescription Burden in Elderly With Alzheimer's Disease Study). Clinical Trial Registration NCT05147428, clinicaltrials.gov, https://clinicaltrials.gov/ct2/show/NCT05147428 (8 December 2021, accessed 2 February 2022).
20. Nelson AJ, Haynes K, Shambhu S, et al. High-Intensity Statin Use Among Patients With Atherosclerosis in the U.S. J Am Coll Cardiol 2022; 79: 1802–1813. [PubMed: 35512860]
21. Eldridge S, Kerry S, Torgerson DJ. Bias in identifying and recruiting participants in cluster randomised trials: what can be done? BMJ 2009; 339: b4006–b4006. [PubMed: 19819928]
22. Fang H, Frean M, Sylwestrzak G, et al. Trends in Disenrollment and Reenrollment Within US Commercial Health Insurance Plans, 2006-2018. JAMA Netw Open 2022; 5: e220320. [PubMed: 35201308]
23. Fishman E, Barron J, Dinh J, et al. 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]
24. Cocoros NM, Arlett P, Dreyer NA, et al. The Certainty Framework for Assessing Real-World Data in Studies of Medical Product Safety and Effectiveness. Clin Pharmacol Ther 2021; 109: 1189–1196. [PubMed: 32911562]
25. Haynes K. Mortality: The final outcome. Pharmacoepidemiol Drug Saf 2019; 28: 570–571. [PubMed: 30701612]
26. Simon GE, Coronado G, DeBar LL, et al. Data Sharing and Embedded Research. Ann Intern Med 2017; 167: 668. [PubMed: 28973353]
27. Sabin JE, Cocoros NM, Garcia CJ, et al. Bystander Ethics and Good Samaritanism: A Paradox for Learning Health Organizations. Hastings Cent Rep 2019; 49: 18–26.

Table 1.

Key Considerations for Conducting Pragmatic Trials in Health Plans

Overall advantages	The NIH Distributed Research Network and its indivdual partners have well-established relationships amongst one another and substantial experience with conducting pragmatic research. For studies that require large sample sizes, this is an efficient approach to identification of eligible participants and analysis of results when the available data are applicable. Suitable for simple interventions that could be disseminated through health plans. Data do not need to be shared outside of the participating health plans. It is possible to leverage claims data for aspects of study planning and implementation and combine it with data collected specifically for the study.
Overall disadvantages	There are important logistical challenges that require careful planning. Conducting such studies in multiple health plans is only applicable for a select set of study questions (i.e., those that require large numbers and can rely on healthcare data to identify eligible individuals, facilitate patient/provider contact, and/or analyze the study outcomes).
Planning stage considerations	
Engagement	Staff from health plans must be involved throughout the trial, including pre-planning phases.
Pre-trial decision making	Feasibility analyses or a pilot phase should be conducted to ensure adequate sample size and to develop details of the protocol and analysis (e.g., feasibility work for a prescribing cascades trial revealed small numbers and led to the development of D-PRESCRIBE-AD). Determine how major steps in the trial will be conducted (e.g., if all sites have data in a common data model, will a single computer program be used for cohort identification; will individual-level data be needed for analysis).
Intervention	If the trial intervention includes contact with a healthcare provider, there are numerous considerations that require careful thought given the nature of claims data. All of the trials discussed have or will target providers for some level of intervention. The health plans must utilize their source data to identify the correct providers.
Data	Plans related to data needs, management, and quality checking should be discussed early. Anticipate the need for “fresh” data and understand to what degree the data lag (for IMPACT-AFib and D-PRESCRIBE-AD, fresh enrollment and dispensing data were necessary at the time of cohort identification).
Timing and attrition	Timing of the work is key. There are substantial changes in health plan enrollment and disenrollment particularly at the start and end of the calendar year.
Ethical issues	Ethical considerations for embedded trials are generally typical of clinical trials. A waiver of consent may be an opportunity, depending on the study question (this was done for IMPACT-AFib and D-PRESCRIBE-AD). Concerns related to the inclusion of a ‘routine care’ arm in IMPACT-AFib have been previously described.[27]
Implementation stage considerations	
Engagement	There are various modalities for connecting with patients and providers to consider depending on the study question. High levels of engagement may be difficult to achieve. Health plans provide intense scrutiny when patients will be contacted and any materials will likely go through legal, privacy, communication, and clinical committee reviews.
Analysis	There is typically a lag of several weeks between randomization and actualizing the intervention with patients or providers, wherein patients may become ineligible. Investigators must address this in the analysis.