NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

# Digging Into Dilemmas of Pragmatic Clinical Trials

## WORKSHOP SUMMARY

### Table of Contents


**Executive Summary** ..... **2**

**Introduction** ..... **3**

**Keynote: The Landscape of Healthcare in America: Gaps and Opportunities (That Researchers Need to Hear)** ..... **3**

**Understanding the Needs and Priorities of Health Systems Across Communities** ..... **4**

**Realizing the Potential of Platform Trials** ..... **7**

**What Do We Learn From Null Pragmatic Trials?** ..... **12**

**What Is the Role of ePCTs in the Age of Artificial Intelligence?** ..... **16**

**References** ..... **21**

## Executive Summary

The NIH Pragmatic Trials Collaboratory held a 2-day workshop, "Digging Into Dilemmas of Pragmatic Clinical Trials," focusing on the perspectives of providers, healthcare systems, regulators, and payers that can inform better approaches to research embedded in healthcare delivery. This report summarizes discussions from each workshop session.

Access the complete workshop materials, including recordings.

### Keynote: The Landscape of Healthcare in America: Gaps and Opportunities (That Researchers Need to Hear)

The keynote emphasized the need for patient-centered health systems that offer equitable access to innovation (including in rural areas), and leveraging technology such as AI to improve workforce well-being, care quality, and cost efficiency. Breaking down silos between clinicians and academics was highlighted as essential for progress.

### Panel 1: Understanding the Needs and Priorities of Health Systems Across Communities

Panelists discussed challenges and benefits for underresourced clinics to participate in research, the importance of aligning research with clinic priorities, and the need for capacity building. The panel offered suggestions for researchers and funders to make clinic participation in embedded pragmatic clinical trials (ePCTs) more feasible and mutually beneficial. Building trust, sharing results, and ensuring tangible benefits for clinics were seen as critical for success.

### Panel 2: Realizing the Potential of Platform Trials

Platform trials, which allow multiple therapies to be tested within a single, reusable research infrastructure, were described as a way to accelerate evidence generation for both rare and common diseases. Panelists stressed the importance of simplicity and making thoughtful design choices that will lead to reliable evidence.

### Panel 3: What Do We Learn From Null Pragmatic Trials?

An analysis of NIH Collaboratory Trials revealed that those with more proximal outcomes (ie, early on the intervention's causal path) were more likely to show positive results. Lessons from null trials included the need for thoughtful outcome selection and thorough pilot testing. Examples illustrated that even null trials can yield valuable insights.

### Panel 4: What Is the Role of ePCTs in the Age of Artificial Intelligence?

Despite rapid implementation, many AI technologies have weak evidence to support their clinical use. The panel highlighted the need for appropriate evaluation of AI tools, ethical oversight, and ongoing validation and monitoring. Evaluation approaches should be tailored to each specific AI technology and balance the need for evidence with potential harms from delaying innovation.

## Introduction

The NIH Pragmatic Trials Collaboratory held a virtual workshop on November 4-5, 2025, to share knowledge from the program and explore emerging issues to help advance pragmatic research that benefits providers and recipients of healthcare.

To date, the NIH Collaboratory has launched 35 pragmatic trials that are conducted within healthcare systems at over 1,400 clinical sites. Over the course of these trials, we have learned much about generating evidence capable of driving change in complex and evolving healthcare systems, including the many unique challenges and opportunities of pragmatic research.

In this workshop, a series of panel discussions addressed considerations for research in healthcare systems across a variety of communities; platform trials as an emerging method; implications of pragmatic trials with null findings; and the role of pragmatic trials in the evaluation of AI technologies.

Access the complete workshop materials, including slides and recordings.

## Keynote: The Landscape of Healthcare in America: Gaps and Opportunities (That Researchers Need to Hear)

*Speaker: David W. Zaas, MD, MBA, President, Atrium Health Wake Forest Baptist*

Dr. Zaas leads one of the largest rural healthcare providers in North Carolina. He described his belief that healthcare being patient centered means having access to innovation, research, and excellence. Because of this philosophy, his health system's most rural hospital has the highest enrollment per capita in clinical trials among all their hospitals. Successful research in these settings requires use of technology and tools to break down barriers, building infrastructure (eg, IRB, research coordinators), and, most importantly, a supportive culture with provider buy-in.

> "When we define access to the best quality care, we should include research as part of that."

He described the 3 biggest challenges he sees with the US healthcare system:

1. Making the switch from a sick care system to a system of health.
2. Ensuring health inequities are mitigated—and not worsened—by technology.
3. Not leveraging advances in technology, which can transform healthcare much more rapidly than they do today.

Dr. Zaas advocated for health systems to approach patients and families as customers. Currently, their trust and loyalty are primarily with their physicians, but this relationship could be expanded to their health system. Health systems also need the trust of insurance companies, which involves being patient and having difficult conversations. Scale helps

with influence: when smaller hospitals join larger health systems, they can get to the table with large payers.

His health system sets bold goals, like ensuring that every patient is able to get a specialist appointment within 72 hours regardless of necessity, and offering every patient access to clinical trials regardless of their location. He described a "loved one's" standard whereby everyone should have the same access to care, whether they are an average person or are connected to someone in a leadership position.

Regarding the AI transformation of healthcare, Dr. Zaas expects rapid changes as the technology is embraced clinically and administratively. He argued that AI use in healthcare should do 3 things:

1. Improve the well-being of the workforce (eg, address burnout, turnover).
2. Improve outcomes for quality, safety, and care delivery.
3. Lower expenses so the healthcare industry can grow faster with fewer people.

An example of a potential transformation is making changes to nursing roles so that staff have more time to lead teams and critically evaluate data instead of just collecting data. By using tools like remote patient monitoring, nurses could focus on treating and caring for patients, leading to greater career fulfillment.

Key to the future of healthcare will be breaking down silos so clinicians and academics can work together to solve the same problems. These relationships will allow both groups to align their strengths and differentiate in key areas.

## Understanding the Needs and Priorities of Health Systems Across Communities

*Panelists: Aimee Eden, Rachel Gold, Caleb Holtzer, Ann Sheehy, Sebastian Tong; Moderator: Cherise Harrington*

This panel included investigators who work closely with health systems of a variety of types and a health system serving a rural area. The panelists shared considerations for conducting research in settings where health systems face challenges in providing care, along with strategies for building trust and credibility to better understand the priorities of health systems, providers, and communities.

### Barriers

Panelists first discussed barriers to working with lower-resourced clinics.

#### *Resources*

Panelists described that it is difficult for under-resourced clinics to partner in research because of the lack of resources. Study budgets are so tight that not enough money can go to the clinics, which is problematic when clinics do not have their own research staff. Asking clinics to participate without providing necessary resources is not respectful of the clinic's realities. It is also not enough to simply pay for clinician time because there is also lost revenue when clinician participation in research activities reduces the number of

patients they see. These clinics are under pressure to generate revenue and be financially sustainable.

### *Provider Burnout and Staffing*

If providers are already overwhelmed, it can be difficult to get them to buy into a research project, yet they are often essential to ePCTs. Taking a nurse away from providing care to do research is a hard sell, even if there is a benefit to patients. Clinics may be hesitant to get involved in research if the requirements and day-to-day responsibilities are unclear. In addition, clinic staff might not have research training. It takes years to build this capacity, involving significant training time that also results in lost revenue.

### *Timing*

The timing of research projects can create several challenges for clinics. Studies often wait years for funding. Once funding comes through, investigators want recruitment goals met quickly. Clinics are trying to meet their revenue goals and stay afloat while researchers are asking them to meet aggressive and seemingly arbitrary timelines.

If a grant application is submitted without clinics that have agreed to participate, it can be penalized during review. Yet it is difficult to recruit clinics in the prefunding stage because it can be years before the project begins, if it begins at all. A panelist described that her institution states in its funding applications that the institution does not recruit clinics in advance.

In preparing for a study to begin, clinics cannot open and hire for positions until the contract and funding come through. It can take months to fill a position. Despite this reality, investigators often tell clinics they are behind the recruitment and startup timelines.

### *Sustainability and Capacity Building*

Taking part in research can be exciting and rewarding for community clinics. When staffing an ePCT, it is not advisable to bring someone in who lacks understanding of how the clinic works. Clinics may view their staff receiving training in new skills as a benefit, but there is still tension in how to provide training so that it does not take staff away from clinical care.

Panelists recommended building ongoing relationships with clinics that have made a commitment to research as part of their strategic plan. Ideally, these clinics could receive funding for capacity building, but this is difficult to obtain. One suggestion was for NIH to provide $100,000 per community clinic involved in a study, with this funding going directly to the clinic. Without such funding, clinics with resources are often the only clinics able to participate.

Another suggestion was to offer bridge funding for research staff in clinics. In rural areas, it is best to have local research staff who know their community. Yet hiring and training someone as a research coordinator in a rural area, only to lose them after a study, is a huge cost. Many clinics cannot handle a gap in research funding for such a staff member. Having exchanges with other sites is another option to maintain skilled staff between projects. For example, the Oregon Rural Practice-Based Research Network has a model with a traveling research support person across the state.

### *Alignment of Priorities*

Panelists recommended having a community advisory board and getting provider input on grants. Providers may identify elements that are too burdensome or not relevant. Research teams could also engage regularly with clinic champions across a region. To develop a study that appeals to clinics, patients, and communities, consider the following questions:

- Does the study solve an urgent problem that affects many clinic patients?
- Is there potential benefit to patients?
- Does the study align with health system goals?
- Does the study address something the clinic is struggling with?
- Is the study tied to a care quality measure that will help with clinic reimbursement?
- Is sustainability built in so clinics can keep the program going after the study ends?

## Benefits

Researchers should ensure studies have tangible benefits for participating clinics.

> "When you're recruiting clinics, you can't be asking them to give more than they get."

### *Benefits to Staff*

Potential benefits to clinic staff include:

- Ability to contribute to improving practice
- Exposure to new concepts and evidence
- Training to develop new skills
- Access to tools and solutions that can reduce burden
- Ways to help patients with more challenging conditions (eg, chronic pain)
- Potential to engage in academic paper writing
- Greater job satisfaction

Maximizing such benefits may help researchers when engaging health system administrators who otherwise lack interest in research. Researchers should be mindful to design studies so all participating clinics and patients receive benefit. Stepped-wedge designs or proxy controls, for example, can help maintain rigor without clinics receiving nothing.

### *Bidirectional Relationships*

Panelists emphasized that learning in ePCTs goes both ways: rural settings can teach members of academic medical centers, and vice versa. Recognizing this can lead to repeat clinic participation. When possible, in-person visits are great for meeting people, building trust, and learning about the environment (though travel costs can be a barrier).

> "You have to remember to come to the clinics with humility. They are the ones that are working the long hours taking care of people."

### *Sharing Findings*

Panelists felt that participating clinics should get access to study data and results as quickly as possible, though this can be challenging. Often, when results are published, the study is complete and has no more staff. Contacts on the research team and at participating clinics

may have left, making communication difficult. One option is hosting community events to share findings, including with providers and participants. Other tips included:

- Consider providing interim data with the caveat that it is not peer reviewed.
- Communicate results in a way that is accessible but not patronizing.
- Make the intervention and tools available once the study is over.

### Advice for Funders

Panelists expressed that funders are moving in the right direction with programs like the NIH Collaboratory and requests for applications for ePCTs in rural areas, which may help prevent misunderstanding in application review. They suggested that study section members be aware that ePCTs require some flexibility and recognize the value of mixed methods.

## Realizing the Potential of Platform Trials

*Panelists: Marianne Kearney Chase, Gregory Levin, Martin Landray, Christopher Lindsell, Bijoy Menon; Moderator: Adrian Hernandez*

The FDA defines a platform trial as "a trial designed to evaluate multiple medical products for a disease or condition in an ongoing manner, with medical products entering or leaving the platform."[1] While not all platform trials are pragmatic, there can be benefits to incorporating pragmatic elements. This panel focused on that intersection, discussing a continuous method of evidence generation that uses real-world data and randomization.

### Platform Trial in a Rare Disease: HEALEY ALS

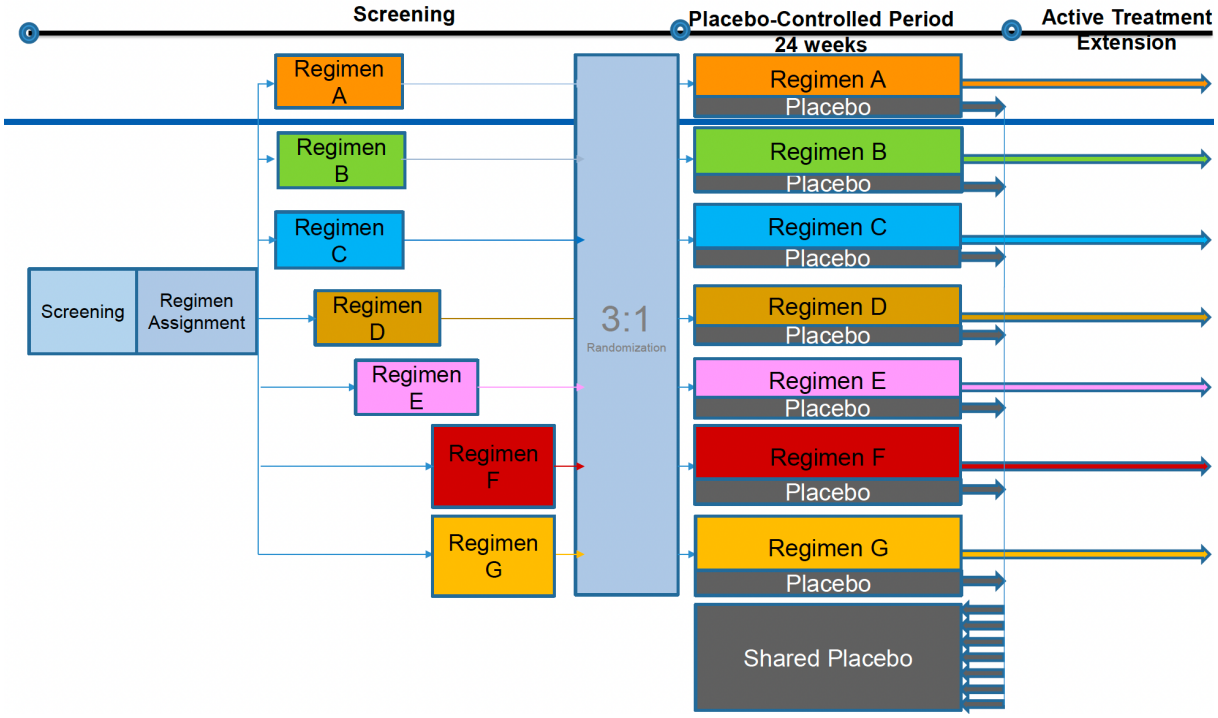
HEALEY ALS is a platform trial studying treatments for amyotrophic lateral sclerosis (ALS), a rare and fatal disease. It was designed to meet the needs of the ALS community by rapidly moving forward drugs with potential, reducing cost, and increasing the amount of people who could receive an active drug rather than placebo.

The research team kept patients at the center while working with industry partners, ALS foundations, and an ALS site network (NEALS) to develop the platform trial. They also talked to the FDA to work through questions. The infrastructure involves a master protocol, a single IRB, and central governance. In its first 5 years, HEALEY ALS tested 7 drugs with more than 1,300 patients. It has 88 sites and nearly 1,000 staff contributing.

Patients are screened at the master protocol level with minimal inclusion and exclusion criteria, then assigned randomly to a treatment regimen (Figure 1). Patients and their physicians know the regimen arm but not whether the patient is receiving an active drug or placebo. The randomization is at a 3:1 ratio due to pooled placebo data.

So far, 2 of the drugs showed promise and are proceeding to phase 3 studies. Recently, the team used clinician and statistician input to revise the master protocol, increasing statistical power, streamlining operations, and making the trial more patient friendly. The researchers share anonymized placebo group data and biospecimens via the PRO-ACT database, making them available for public use. They have also published about their operational efficiencies and shared knowledge on platform trials.[2,3]

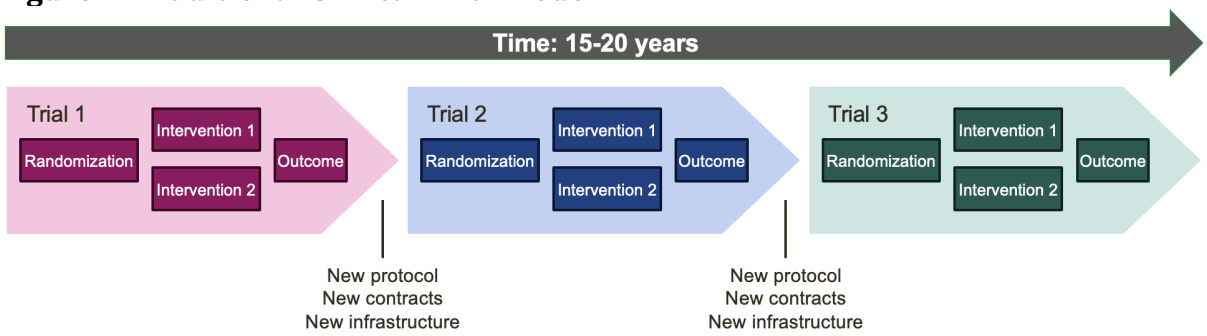
**Figure 1. HEALEY ALS Trial Design**




## Platform Trial in a Common Disease: ACT GLOBAL

ACT GLOBAL is a platform trial addressing a health problem at scale with potentially global impact. Stroke is one of the largest global health burdens. To detect modest treatment effects requires large trials enrolling tens of thousands of patients, which is virtually impossible to do in academic medicine using current funding models. Furthermore, the traditional model of a single clinical trial with a single investigator, repeated in sequence, can be costly and time consuming (Figure 2).

**Figure 2. Traditional Clinical Trial Model**



In contrast, platform trials create a reusable infrastructure with several efficiencies:

- **Governance:** Master ethics frameworks, master contract, reusable IRB packages
- **Data:** Harmonized electronic data and metadata allowing quality at scale
- **Fiscal:** Economy of scale, shared budgets with reduced site burden, faster activation
- **Logistical:** Common monitoring, training, processes, e-consent, e–patient-reported outcomes

Setting up the ACT GLOBAL infrastructure involved thinking about what could be common, what needed to change per therapy, and what was outside the possibilities of a master protocol. An adaptive design is key to continuous learning. Each new trial in the platform has its own investigator and team, which strengthens the shared network. Sites also gain capabilities and can move faster.

> "When disease burden is large and evidence must evolve continuously, platform trials are not optional—they're essential."

Enrollment started in 2024. ACT GLOBAL is active on 3 continents with plans to expand to 6 continents and 30 to 40 countries in the next year. The goal is to have approximately 400 hospitals in 6 to 8 regional nodes that are quasi-autonomous and run their own trials. So far, ACT GLOBAL has 6 trials enrolling nearly 1,000 patients per month, with another 6 to 7 trials in the pipeline.

In summary, platform trials require systems-based thinking with intentional design. ACT GLOBAL is a framework for common-disease research that addresses pain points in contracts, ethics, data, and funding to make global science faster, fairer, and more inclusive.

### Principles of Platform Trials

Clinical trials are fundamentally simple: ask an important question, answer it reliably, and look after the participants. These principles are critical to any clinical trial, including pragmatic trials, platform trials, and the combination of the two.

ISIS 2 was a large simple trial conducted in the 1980s that had a great impact on the prescribing of aspirin and thrombolytics.[4] It enrolled approximately 17,000 people and had a simple outcome (death). The protocol was 10 pages, and it stated that extra work for providers must be absolutely minimal. In today's money, the trial cost about $10 million. However, if the trial were conducted today, it would likely cost 100 times that amount. This raises the question of why trials are so much more expensive today with all the modern systems available.

RECOVERY is another example of a streamlined, productive large randomized trial. Early in the COVID-19 pandemic, it was set up within 9 days. It had an important outcome (mortality) and simple eligibility criteria, follow-up (a 1-page case report form), and consent. It leveraged linkages to routine NHS data sets, though non-NHS sites also successfully participated. Researchers changed the treatments and randomization structure over time, studying 15 treatments in total. This trial showed that "streamlining and quality are not opposed."[5] The first results were achieved in 99 days, with 4 hours to

change practice in the UK and 4 weeks globally, resulting in 1 million lives saved in the first year.

Principles of good clinical trial design are applicable regardless of intervention or setting:[6,7]

- Design to produce scientifically sound answers to relevant questions
- Respect the rights and well-being of participants
- Be collaborative and transparent
- Be feasible for the context
- Manage quality effectively and efficiently

Overall, trials do not have to be complicated. They should be tailored to the available infrastructure and not wasteful of staff or participants' time, supplies, or other resources. Strengths and safeguards in existing systems should not be duplicated or altered without careful justification. The closer trial processes are to routine practice, the more efficiently and effectively they are likely to be executed, and with fewer mistakes.

> "If you want to think about how to do platform trials in the future, think about the principles and learn from the past."

While platform trials are not suitable for every circumstance, they provide a practical and efficient way to answer multiple questions. Clinical research is not a matter of copying and pasting the previous protocol. Rather, researchers should think fundamentally about the principles that will result in good evidence while looking after the participants, and doing this as efficiently as possible by focusing on the issues that really matter to trial quality.

### FDA Perspectives on Platform Trials

The FDA released a Master Protocols Draft Guidance in 2023.[1] Platform trials can offer substantial operational and statistical efficiencies while providing reliable evidence on safety and effectiveness. For example, use of a shared control arm can reduce the sample size needed (by 25% or more) to evaluate multiple products, compared to individual trials.

However, there are limitations to consider, such as the use of concurrent vs nonconcurrent controls. The FDA recommends use of a concurrent control that only includes patients randomized at the same time as those who received the medical product. This approach preserves the integrity of the randomized comparison for the primary analysis. If a researcher proposes to use a nonconcurrent control, the FDA would need a strong justification due to the potential for temporal drifts (eg, changes in standard of care).

Blinding can also create complexities. In some settings, open-label comparisons may be reasonable, but others need blinding. Blinding becomes more challenging as the number of products studied increases because they may have different routes of administration. Researchers should consider the pros and cons of complete vs partial blinding.

Statisticians may find platform trials challenging. Recommendations were to use a straightforward comparison and work with regulators to make design choices that will lead

to reliable evidence. There can also be complexities in coordinating the trial in terms of reporting, regulatory submissions, and maintaining trial integrity.

## When and How to Use Platform Trials

### *Rationale*

The key to platform trials is using them when the patients and clinical context are similar. The diseases and therapies studied might change, but the operations remain mostly consistent. For example, the RECOVERY trial moved from COVID-19 to influenza. Despite this shift, the trial leverages the same types of patients, providers, and outcomes. The ability to study questions that evolve over time makes these trials extremely useful.

Platform trials also offer opportunities within a disease with many possible treatments. Efficiencies arise from studying multiple treatments at once, and patients find it desirable that, if their initial treatment is not working, there are other options within the trial.

Some drug companies are calling for platform trials because they allow drugs to be studied for secondary indications, such as rare diseases. Panelists supported having more platform trials because they can answer meaningful questions efficiently with the caveats that they are not a panacea and need to be done thoughtfully.

> "The platform provides an opportunity for something that mirrors practice, which is if the treatment doesn't work, let's try another one."

### *Design Considerations*

When designing platform trials, they need to be fit for purpose. The main goal is to develop a master protocol with as many common elements as makes sense for the context. This approach balances efficiency with validity.

As new therapies come onto the platform, optimize each appendix to leverage as much of the master protocol as possible. It is not feasible to know at the start of a platform trial all the treatments that may be added, so appendices allow the platform to accommodate treatments that push the master protocol slightly.

An umbrella trial is a type of platform trial that does concurrent trials for a single disease or condition. It can have great operational, statistical, regulatory, and administrative efficiencies that exceed those of other platform trials.

If statistical efficiency is the primary driver of a master protocol, ensure it can be achieved. Consider how to borrow information and infrastructure from other study arms. If the statistical efficiencies are small and it will require a lot of operational effort to accommodate a therapy, then use of the platform trial may not be worthwhile.

## What Do We Learn From Null Pragmatic Trials?

*Panelists: Corita Grudzen, Michael Ho, Vincent Mor, Miguel Vazquez; Presenter: Kevin Weinfurt; Moderator: Josephine Briggs*

### Introduction

Null trials—defined as trials in which the intervention does not meet criteria for statistical significance for the primary outcome—are common in all types of clinical investigation. Often, null findings lead to the conclusion that the intervention lacks promise, and it is abandoned. The impetus for this session was the surprising observation that a number of the interventions tested in null NIH Collaboratory Trials still looked promising to investigators and in some cases were still being implemented by health system partners. The goal of this session was to explore factors in the pragmatic trial approach that may lead to null outcomes and clarify what can be learned from null trials.

### Characteristics of Positive vs Null NIH Collaboratory Trials

As of this session, 16 NIH Collaboratory Trials, all performed in partnership with healthcare systems, had been completed. All studied interventions were viewed by peer review panels as promising and were supported by some efficacy data. Seven trials were positive, that is they met criteria for statistical significance for their primary endpoint (7 "positive trials"), and nine did not (9 "null trials"). The characteristics of trial design, intervention, and outcomes/results were compared.

#### *Design*

All of the positive trials and all but 3 of the null trials were effectiveness trials (as opposed to hybrid effectiveness-implementation trials). There were a variety of study designs (ie, cluster, stepped-wedge, and individual randomization) distributed across the 2 groups with no clear patterns. Reaching the needed sample size is a common issue with null trials in general, but only 2 of the null trials in this analysis had that problem.

#### *Intervention*

Intervention types (ie, operational/educational, comparing therapies, or a combination) were well distributed, with slightly more null trials having operational/educational interventions (5 null trials vs 2 positive trials). Clinician-targeted interventions were also slightly more prevalent in the null trials (4 null trials vs 1 positive trial), though there was a good mix of intervention targets (ie, patient, clinician, or both) across groups. From data available on intervention complexity (ie, low, moderate, or high), no differences in distribution were noted. Several positive trials had high-complexity interventions.

Several null trials were not completed because of implementation challenges. Methods to monitor implementation of the intervention are very valuable in the pragmatic trial setting and can lead to well-reasoned study termination. For example, dialysis seemed like an optimal setting for ePCTs, but two NIH Collaboratory Trials of dialysis interventions (TiME and HiLo) were terminated largely because adequate separation of the trial groups could not be achieved.[8,9]

### *Outcomes/Result*

The one feature that appeared differentially distributed between positive and null trials was the position of the trial's primary outcome on the causal path. Members of the NIH Collaboratory Coordinating Center team categorized outcomes as proximal or distal in nature (without prior knowledge of the trial outcome), with proximal outcomes being those tied closely to the intervention's mechanism of action and distal outcomes being later effects. For example, in a trial of advance care planning, a proximal outcome might be completing an advance directive, whereas a distal outcome might be hospitalization.

All of the positive trials had primary outcomes that were judged to be proximal, whereas several of the null trials had primary outcomes categorized as distal.

## Lessons From Four Effectively implemented Null Trials

The session focused on a discussion by the lead investigators from four effectively implemented NIH Collaboratory Trials that had null findings (Table).

**Table. Four Null NIH Collaboratory Trials**

| Trial | Population | Intervention | Design | Primary Outcome |
|---|---|---|---|---|
| **ICD-Pieces[10]9/7/26 5:43:00 AM** | Concurrent diabetes, chronic kidney disease, hypertension | Collaborative primary care program | Cluster randomization | All-cause hospitalization at 1 year |
| **Nudge[11]** | Patients with chronic cardiovascular conditions | Text messages and chat bot | Individual randomization | Medication adherence at 12 months |
| **PRIM-ER[12]9/7/26 5:43:00 AM** | Adults older than 65 years | Multicomponent palliative care education | Stepped-wedge, cluster randomization | Hospital admission |
| **PROVEN[13]** | Nursing home residents | Advance care planning video | Cluster randomization | Hospital transfer at 12 months |

Overall, all 4 panelists felt the tested interventions still had promise. Participants found it noteworthy that in some cases the interventions were still considered promising by their partnering healthcare system.

The panel discussion focused on which aspects of the pragmatic design might have constrained the ability to observe an effect, reflecting on factors that influenced the results and what they might have done differently.

### *Pragmatic Design*

The pragmatic approach requires implementing the intervention with no interruption of ongoing care and without much additional staffing. Panelists noted that this impacts fidelity and adherence monitoring. Examples included:

- ICD-Pieces operated in 141 primary care facilities, and uneven adherence was felt to impact the failure of the of the intervention to change hospitalization rate.
- Nudge provided patients with text reminders to fill prescriptions but had no tool to measure impact of the messages.
- In PROVEN, the 2 partnering nursing home corporations enrolled all the nursing homes they owned, regardless of whether the nursing home administrators viewed the study as a priority. After the null primary analysis, PROVEN applied a per-protocol analysis, which suggested greater intervention fidelity might have modified the outcome. This analysis is not considered definitive since it loses the rigor of random assignment.

A positive outcome for these trials generally required behavior change—of patients or providers—and all panelists agreed that behavior change is hard. Overall, investigators acknowledged that a proximal measure assessing behavior of the patient or the health system staff would have been valuable. It is challenging to do this pragmatically, but critical.

### *Outcome Measures*

The pragmatic design creates important limitations on possible outcome measures. In these multisite trials, choice of outcome measures was limited to those that could be captured consistently across all sites. Panelists noted that most outcome measures that can be extracted from the electronic health record (EHR) are either quite proximal or very distal, and they are often not patient centered.

> "If trying to change a behavior that we know affects a health outcome, we can power the study for the behavior change and not the health outcome."

Most panelists reflected in hindsight that they would change the primary outcome measure. Three of the four studies used hospitalization as a primary outcome. Reducing hospitalization rate is, in general, a shared interest of the public and health payers, and sometimes (although not always), a patient-centered outcome. Nevertheless, it was problematic as the *primary* trial outcome for a variety of reasons.

- In ICD-Pieces, reducing hospitalization was a very distal measure of the effectiveness of improving guideline-directed care for patients with chronic kidney disease. A longer time interval might have led to a different result.
- In PRIM-ER, in the emergency care setting, there are alternative measures important to the health systems, including patient satisfaction with communication and bed type (eg, regular, intensive care unit, observation).
- PROVEN used reducing hospital transfers from nursing homes as the outcome measure. It is a complex outcome since nursing homes have a financial incentive to not send people to the hospital unnecessarily, but a variety of local factors beyond patient preference are often decisive.

All agreed there are now better tools available to measure intermediate or surrogate outcomes that align with guideline-directed care, such as hypertension or glucose control.

Null findings may also occur when an incorrect causal pathway is used during the design phase in selecting the primary outcome.

Employing a mixed-methods approach could also help researchers understand provider and patient experiences with an intervention, providing useful information even in the case of a null primary outcome. However, panelists noted that qualitative findings may not be seen as impactful by grant reviewers and journal editors.

Panelists also challenged the need to achieve 95% confidence in pragmatic trials where the interventions are low risk and already have efficacy evidence. Standards for trials were developed for a regulatory context but do not necessarily reflect how decisions are made in the real world. The threshold for continuing ePCT interventions is different than for drugs. For example, a modest increase in provider satisfaction may be sufficient when there is high provider burden. Panelists noted methods to maintain rigor while viewing results in a more complex way (for example, Bayesian statistics or a decision analysis concept that can be set in advance).

### *Health System Factors*

A common theme was the need to align pragmatic trial outcomes with health system priorities. Panelists noted that for partnering health systems, the cost of the intervention and providers' satisfaction with it drove decisions about ultimate implementation, not statistical significance. If an intervention was low cost and providers found it useful, it might continue to be used, especially if de-implementation would require resources. Several of the interventions that had null findings were maintained or implemented more widely. For example, some nursing homes continue to use the advance care planning videos from PROVEN. PRIM-ER has had some of its intervention components offered nationally.

Health systems want rapid answers. Ideas for speeding evidence generation included:

- Maintaining bidirectional partnerships with health systems
- Platform trials
- Building and sustaining research infrastructure
- Adaptive trial designs
- Choosing more proximal outcomes
- Proposing a program of research with follow-up studies that depend on the outcomes of an early study

### *Pilot Testing*

Panelists offered thoughts on aspects for pilot testing. Moving from efficacy to effectiveness testing is difficult because ePCTs involve making changes in healthcare systems. It is important to demonstrate that the intervention can be delivered and to identify core intervention functions that must be maintained across sites (while allowing some tailoring). Panelists also recommended piloting workflows, measurement of intervention fidelity, and ways to motivate change in providers. ICD-Pieces revealed the need to pilot at all sites, because piloting at a single site did not uncover issues later experienced at other sites. However, extensive piloting may be contradictory to the need to complete trials faster.

### *Advice for Funders*

Large-scale effectiveness testing of promising interventions is important. Randomization provides rigor for which there is no easy substitute, and pragmatic approaches are necessary for effective scaling. Nevertheless, the challenges need to be recognized. Panelist advice for funders included:

- Recognize that moving from efficacy to effectiveness trials requires careful pilot studies.
- Be more aggressive in descoping trials that do not meet milestone measures of adherence and enrollment.
- Offer more resources and 2 years for the pilot phase.

## What Is the Role of ePCTs in the Age of Artificial Intelligence?

*Panelists: Adrian Hernandez, Michael Ho, Rohan Khera, Edward Melnick, Jeremy Sugarman, Kenneth Wiley; Moderator: Emily O'Brien*

The panel discussed the state of AI in healthcare and the role of ePCTs in evaluating AI tools. The speed of AI development is outpacing rigorous evaluation, so new tools are being implemented before they are properly tested. AI use in healthcare has the potential to affect millions of lives, so "deployment before deliberation" was likened to the Wild West. However, there is an opportunity for innovation and evidence to evolve together (Figure 3).

**Figure 3. Chat GPT Depicts AI in Healthcare**




Prompt: Please create an image of the 'Wild West' of AI in healthcare as it stands in November 2025




Prompt: Please create an image of a utopia where innovation and evidence evolve together, with ePCTs appropriately used to evaluate new AI tools

### Overview of AI in Healthcare

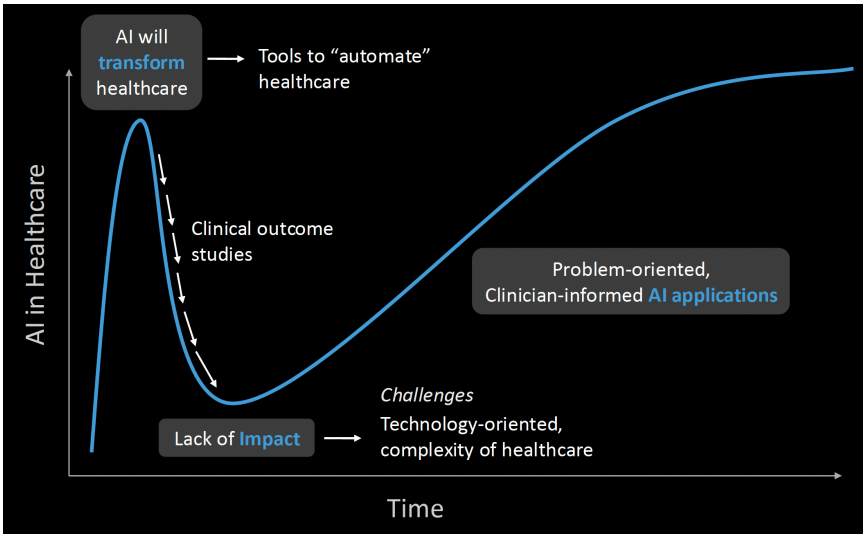
Early development of AI in healthcare was approached with the mindset that the technology could automate healthcare tasks; however, such implementations did not live up to the promise due to the complexity of healthcare. More recently, clinician-informed AI development has focused on building tools to solve challenges (Figure 4). AI has shown an ability to augment diagnosis, enhance outcome prediction, and improve care delivery. It is also empowering new biological discoveries with its ability to analyze large-scale data sets.

Large language models yield many new capabilities at scale:

- Disease subtyping
- Assessing quality of healthcare
- Automating care processes
- Creating orders based on conversations during care encounters
- Predicting adverse outcomes
- Building just-in-time adaptive interventions using EHR data before a care encounter
- Overcoming EHR interoperability issues across sites by parsing data structure

One study showed that if a primary care physician were to follow all guidelines for providing care for their patients, it would take them 30 hours per day.[14] There is a persistent gap between what providers could be doing for patients vs what they are doing. Even in well-resourced settings, there is a hunger for more support.

**Figure 4. AI in Healthcare**




There is both anxiety and excitement for all the ways AI can improve care. For example, researchers are working on the use of AI tools to identify good transplant candidates. While such tools could be beneficial, they are not completely predictable because they go beyond an algorithm to actual learning. As an AI tool learns, it may make mistakes. It is possible these mistakes would be made at scale.

Despite rapid advances, many AI technologies have weak evidence to support their clinical use. Particularly, prospective randomized studies in diverse populations are lacking. Ongoing conversations among experts are needed to enhance methodological rigor for new AI technologies. Working with the clinical community is also necessary to develop processes for incorporating algorithm-enabled care appropriately in clinical practice.

### Case Study: ADAPT Trial

The ADAPT trial is a follow-up to EMBED, an NIH Collaboratory Trial that studied user-centered tools for initiating buprenorphine in the emergency department. ADAPT is testing a refined set of tools in a larger population. The trial follows a multiphase optimization strategy. The goal is to incrementally refine the intervention based on data collected to understand the effects of the tools and how people use them. In a rapid randomized testing phase, researchers will try small changes to see what works (A/B testing). Finally, they will compare the optimized intervention with the original version.

ADAPT, which is randomizing at the patient level, needed a way to identify patients with suspected opioid use disorder even when their medical record is incomplete. The team developed a model that uses an AI machine learning analysis to look for predictive factors (including nonlinear trends and associations) of opioid use disorder. The model runs nightly and flags patients identified with point-of-care alerts in Epic. They trained and validated the model, performing checks with diagnosis codes and manual chart review.

### Case Study: Chat 4 Heart Health

Chat 4 Heart Health is a follow-up of the Nudge study, both NIH Collaboratory Trials. Nudge tested variations of a text message intervention to improve refill adherence for patients taking cardiovascular medications.[15] The hypothesis was that behavioral nudges via text message with or without a chatbot would perform better than generic text messages and usual care. The chatbot was designed to identify reasons why patients were not refilling their medications and to provide fixed responses for potential solutions based on a response library developed by the research team. This intervention was developed before AI technology. At 3 months, there was a small improvement in refill adherence for all intervention groups compared to usual care, but the difference was not significant at 12 months (the primary outcome).[11]

In a post hoc analysis, the research team found that the subset of patients who responded to the messages had much higher adherence. This led to development of the Chat 4 Heart Health trial, which uses AI as part of the text messaging intervention to engage patients and improve adherence. In the future, the ability to have read receipts could help researchers better understand whether patients are viewing the messages.

### Testing of AI Tools in Healthcare

AI could have many benefits in healthcare, such as allowing providers to engage with their patients as opposed to typing into a computer during a clinical interaction. Use of AI in healthcare can be examined according to whether it improves one of the following:

- Patient experience of care
- Population health
- Value of care (ie, reduced cost of care per capita)
- Work life of healthcare providers

People are exposed to A/B testing in their everyday digital experiences, and the same techniques could be applied in healthcare. Testing short-term outcomes of AI tools in many cases can be relatively low risk. However, at the population health level, deployment and testing becomes more complex and potentially problematic. An example is a policy implemented in South Korea to increase detection of thyroid cancer, which led to many asymptomatic people getting routinely screened. Ultrasound enabled micro-detection of small cancers, resulting in more procedures performed. However, mortality rates did not change. Similarly, overly sensitive measurement by AI tools could cause adverse effects. This illustrates the need to evaluate long-term outcomes, not just short-term benefits.

Conversely, patients and healthcare systems will not tolerate waiting 5 to 10 years to see if an AI tool improves outcomes. There is an ethical balance of deploying a tool without evidence vs delaying innovation that could help people while waiting for proof. Panelists discussed that there is no perfect threshold for every use case, so the healthcare field must approach these issues with humility. In the history of medicine, there have been many times when a doctor came up with an idea, thought it was great, introduced it into practice, even had short-term evidence that it worked, but evidence-based trials often showed that most things we thought worked actually did not. One idea is adaptive licensing in which tools receive conditional approval pending real-world evaluation.

> "We have treated AI all in one bucket, but that's a problem. AI warrants careful attention for each kind of AI."

Another challenge has been the categorization of AI as one concept. There is a need to stratify AI tools into much more granular categories. Some tools, like transcription, will have a very low bar for deployment, whereas diagnostic use may take years to test.

There can be unintended consequences of implementing new technologies, so the technologies must be validated and checked regularly to make sure they work and continue to work. AI should be viewed as a way to enhance human-centered care. As a result of these new tools, the skills and tasks of clinicians will likely change in the next several years.

The return on investment for AI technologies has not yet become evident to many companies. Technology development tends to be siloed, so panelists suggested that the healthcare field not just encourage individual innovation but think collectively.

### Ethical Issues

Clinical uses of AI offer great promise, but there are known and unknown pitfalls that raise an array of ethical issues. In exploring these issues, each type of AI warrants careful attention. Some of the potential negatives of AI use in healthcare include:

- Algorithmic biases
- "Hallucinations"
- Model drift in which the training set used is no longer applicable because the environment changed
- Worsened inequalities
- Dehumanization of care

> "We need to be transparent about the uses of AI across our healthcare systems."

Ethical issues relevant to AI use include privacy, consent, authorization, and potential harms to individuals, groups, or society (Figure 5).[16] There can also be hazards for healthcare organizations, such as operational or reputational risk when employing AI.

There are different ethical considerations for research than for clinical care. "Digital ethics" frameworks are premised on substantive principles and procedural values. There is also a need to examine how these core principles relate to subsidiary principles (Figure 6).[17]

**Figure 5. Potential Harms of AI Technologies**




Reprinted from: National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework. 2023. https://doi.org/10.6028/NIST.AI.100-1

Char and colleagues outlined a range of ethical considerations across the translational cascade (ie, development, implementation, and evaluation).[18] The National Academies' Code of Conduct for Health and Medicine[19] also provides overarching goals to keep in mind when introducing AI tools:

- Advance humanity
- Ensure equity
- Engage impacted individuals
- Improve workforce well-being
- Monitor performance
- Innovate and learn

When new AI tools are implemented, appropriate governance structures and monitoring are necessary. Adopting tools rapidly without evidence of safety and efficacy can be problematic. It is important to maintain clinician accountability and the ability to accept or overrule AI tools' actions. AI tools exist to help but not replace the human. Reports of how ethical issues emerge and are managed will be helpful to help ensure that the rights, interests, and welfare of patients are protected.

**Figure 6. Digital Ethics Principles**




From Becker SJ et al. AI & Soc. 2023;38:2629–2639. Used under CC BY 4.0.

Some institutions have developed a structured committee approach in which every technology undergoes an evaluation of performance and risk-benefit ratio. There is also an ethical assessment of how the technology changes the care of patients and whether there will be harm to a certain segment of the population.

## References


1. U.S. Food and Drug Administration. Master Protocols for Drug and Biological Product Development: Guidance for Industry (Draft Guidance). Published online December 2023. Accessed November 25, 2025. https://www.fda.gov/media/174976/download
2. Paganoni S, Harkey B, Giacomelli E, Cudkowicz M, HEALEY ALS Platform Trial Study Group. Lessons from the HEALEY adaptive platform trial in amyotrophic lateral sclerosis. *Nat Aging*. 2024;4(11):1512-1515. doi:10.1038/s43587-024-00740-x
3. Harkey BA, Distefano S, Pagliaro JA, et al. Operational development and launch of an adaptive platform trial in amyotrophic lateral sclerosis: processes and learnings from the first four regimens of the HEALEY ALS platform trial. *Muscle Nerve*. 2025;72(2):294-305. doi:10.1002/mus.28442
4. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet*. 1988;2(8607):349-360.
5. Califf RM, Cavazzoni P, Woodcock J. Benefits of streamlined point-of-care trial designs: lessons learned from the UK RECOVERY Study. *JAMA Intern Med*. 2022;182(12):1243-1244. doi:10.1001/jamainternmed.2022.4810
6. World Health Organization. Guidance for Best Practices for Clinical Trials. Published online 2024. https://www.who.int/publications/i/item/9789240097711
7. Good Clinical Trials Collaborative. Guidance for Good Randomized Clinical Trials. Published online November 2023. Accessed November 25, 2025. https://www.goodtrials.org/
8. Edmonston D, Isakova T, Dember LM, et al. Higher versus lower phosphate targets for patients undergoing in-center hemodialysis: a randomized controlled trial. *J Am Soc Nephrol JASN*. Published online July 10, 2025. doi:10.1681/ASN.0000000765
9. Dember LM, Lacson E, Brunelli SM, et al. The TiME trial: a fully embedded, cluster-randomized, pragmatic trial of hemodialysis session duration. *J Am Soc Nephrol JASN*. 2019;30(5):890-903. doi:10.1681/ASN.2018090945
10. Vazquez MA, Oliver G, Amarasingham R, et al. Pragmatic trial of hospitalization rate in chronic kidney disease. *N Engl J Med*. 2024;390(13):1196-1206. doi:10.1056/NEJMoa2311708
11. Ho PM, Glorioso TJ, Allen LA, et al. Personalized patient data and behavioral nudges to improve adherence to chronic cardiovascular medications: a randomized pragmatic trial. *JAMA*. 2025;333(1):49-59. doi:10.1001/jama.2024.21739
12. Grudzen CR, Siman N, Cuthel AM, et al. Palliative care initiated in the emergency department: a cluster randomized clinical trial. *JAMA*. 2025;333(7):599-608. doi:10.1001/jama.2024.23696
13. Mitchell SL, Volandes AE, Gutman R, et al. Advance care planning video intervention among long-stay nursing home residents: a pragmatic cluster randomized clinical trial. *JAMA Intern Med*. 2020;180(8):1070-1078. doi:10.1001/jamainternmed.2020.2366
14. Porter J, Boyd C, Skandari MR, Laiteerapong N. Revisiting the time needed to provide adult primary care. *J Gen Intern Med*. 2023;38(1):147-155. doi:10.1007/s11606-022-07707-x
15. Glasgow RE, Knoepke CE, Magid D, et al. The NUDGE trial pragmatic trial to enhance cardiovascular medication adherence: study protocol for a randomized controlled trial. *Trials*. 2021;22(1):528. doi:10.1186/s13063-021-05453-9
16. Tabassi E. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. National Institute of Standards and Technology (U.S.); 2023:NIST AI 100-1. doi:10.6028/NIST.AI.100-1

17. Becker SJ, Nemat AT, Lucas S, Heinitz RM, Klevesath M, Charton JE. A Code of Digital Ethics: laying the foundation for digital ethics in a science and technology company. *AI Soc*. 2023;38(6):2629-2639. doi:10.1007/s00146-021-01376-w
18. Char DS, Abràmoff MD, Feudtner C. Identifying ethical considerations for machine learning healthcare applications. *Am J Bioeth AJOB*. 2020;20(11):7-17. doi:10.1080/15265161.2020.1819469
19. National Academy of Medicine. *An Artificial Intelligence Code of Conduct for Health and Medicine: Essential Guidance for Aligned Action*. (Adams L, Fontaine E, Matheny M, Krishnan S, eds.). National Academies Press; 2025:29087. doi:10.17226/29087

The complete workshop materials, including slides and recordings, are available on the Living Textbook.