

# UCSF

## UC San Francisco Previously Published Works

### Title

Achieving diagnostic excellence through prevention and teamwork (ADEPT) study protocol: A multicenter, prospective quality and safety program to improve diagnostic processes in medical inpatients.

### Permalink

<https://escholarship.org/uc/item/3v17247t>

### Journal

Journal of Hospital Medicine, 18(12)

### Authors

Schnipper, Jeffrey

Raffel, Katie

Keniston, Angela

et al.

### Publication Date

2023-12-01

### DOI

10.1002/jhm.13230

Peer reviewed



# HHS Public Access

Author manuscript

*J Hosp Med.* Author manuscript; available in PMC 2024 March 26.

Published in final edited form as:

*J Hosp Med.* 2023 December ; 18(12): 1072–1081. doi:10.1002/jhm.13230.

## Achieving diagnostic excellence through prevention and teamwork (ADEPT) study protocol: A multicenter, prospective quality and safety program to improve diagnostic processes in medical inpatients

Jeffrey L. Schnipper, MD, MPH<sup>1,2</sup>, Katie E. Raffel, MD<sup>3,4</sup>, Angela Keniston, PhD, MSPH<sup>3</sup>, Marisha Burden, MD, MBA<sup>3</sup>, Jeffrey Glasheen, MD<sup>3,4</sup>, Sumant Ranji, MD<sup>5</sup>, Colin Hubbard, PhD<sup>6</sup>, Peter Barish, MD<sup>6</sup>, Molly Kantor, MD<sup>6</sup>, Julia Adler-Milstein, PhD<sup>7</sup>, W. John Boscardin, PhD<sup>8</sup>, James D. Harrison, MPH, PhD<sup>6</sup>, Anuj K. Dalal, MD<sup>1,2</sup>, Tiffany Lee, BA<sup>6</sup>, Andrew Auerbach, MD, MPH<sup>6</sup>

<sup>1</sup>Hospital Medicine Unit, Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, Massachusetts, USA

<sup>2</sup>Harvard Medical School, Boston, Massachusetts, USA

<sup>3</sup>Department of Medicine, Division of Hospital Medicine, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA

<sup>4</sup>Institute for Healthcare Quality, Safety, and Efficiency, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA

<sup>5</sup>Division of Hospital Medicine, Zuckerberg San Francisco General Hospital, San Francisco, California, USA

<sup>6</sup>Department of Medicine, Division of Hospital Medicine, University of California, San Francisco, California, USA

<sup>7</sup>Center for Clinical Informatics and Improvement Research (CLIIR), University of California, San Francisco, California, USA

<sup>8</sup>Department of Medicine and Department of Epidemiology and Biostatistics, University of California, San Francisco, California, USA

### Abstract

---

**Correspondence:** Jeffrey L. Schnipper, MD, MPH, Hospital Medicine Unit, Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, 1620 Tremont St, BC-3, Boston, MA 02120-1613, USA. [jschnipper@bwh.harvard.edu](mailto:jschnipper@bwh.harvard.edu); Twitter: [@drjschnip](https://twitter.com/drjschnip).

#### CONFLICT OF INTEREST STATEMENT

J. L. S. received funding from Synapse Medicine for an investigator-initiated study of their medication decision-support software; and a stipend from the American Society of Health-System Pharmacists for the creation of an online course on medication history-taking.

#### SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**Background:** Few hospitals have built surveillance for diagnostic errors into usual care or used comparative quantitative and qualitative data to understand their diagnostic processes and implement interventions designed to reduce these errors.

**Objectives:** To build surveillance for diagnostic errors into usual care, benchmark diagnostic performance across sites, pilot test interventions, and evaluate the program's impact on diagnostic error rates.

**Methods and Analysis:** Achieving diagnostic excellence through prevention and teamwork (ADEPT) is a multicenter, real-world quality and safety program utilizing interrupted time-series techniques to evaluate outcomes. Study subjects will be a randomly sampled population of medical patients hospitalized at 16 US hospitals who died, were transferred to intensive care, or had a rapid response during the hospitalization. Surveillance for diagnostic errors will occur on 10 events per month per site using a previously established two-person adjudication process. Concurrent reviews of patients who had a qualifying event in the previous week will allow for surveys of clinicians to better understand contributors to diagnostic error, or conversely, examples of diagnostic excellence, which cannot be gleaned from medical record review alone. With guidance from national experts in quality and safety, sites will report and benchmark diagnostic error rates, share lessons regarding underlying causes, and design, implement, and pilot test interventions using both Safety I and Safety II approaches aimed at patients, providers, and health systems. Safety II approaches will focus on cases where diagnostic error did not occur, applying theories of how people and systems are able to succeed under varying conditions. The primary outcome will be the number of diagnostic errors per patient, using segmented multivariable regression to evaluate change in  $y$ -intercept and change in slope after initiation of the program.

**Ethics and Dissemination:** The study has been approved by the University of California, San Francisco Institutional Review Board (IRB), which is serving as the single IRB. Intervention toolkits and study findings will be disseminated through partners including Vizient, The Joint Commission, and Press-Ganey, and through national meetings, scientific journals, and publications aimed at the general public.

## INTRODUCTION

Many factors contribute to diagnostic errors (DEs), but key among them are foundational issues in health care: complex and fragmented care systems, limited time, and the work systems and cultures that impede improvements in diagnostic performance.<sup>1-7</sup> While approaches to identifying inpatient DEs exist, few studies have linked the identification of underlying systemic and structural causes of errors (such as handoff problems, equipment failures, or changes in workload) to existing quality improvement programs. Even fewer have applied theories of how people and systems are able to succeed under variations (i.e., "Safety II")<sup>8,9</sup> or positive deviance methods to characterize optimal diagnostic processes and then use those findings to catalyze health system improvement.

This study builds directly on our previously conducted study—utility of predictive systems in diagnostic errors (UPSIDE)<sup>10</sup>—which defined risk factors, underlying causes, and prevalence of DEs among patients admitted to 30 hospitals participating in the Hospital Medicine Reengineering Network (HOMERuN), a collaborative of over 50 academic

hospital medicine programs focused on improving quality of health care delivery.<sup>11</sup> UPSIDE refined reference standard approaches to adjudication of inpatient DEs, defined factors associated with these errors, and created collaborations with our sites and national organizations, providing a powerful opportunity to transform how diagnostic process evaluation programs can be used to improve patient safety in hospitals.

The overall goal of achieving diagnostic excellence through prevention and teamwork (ADEPT) study is to turn our highly successful multicenter network into a DE learning health system that will integrate DE assessments into existing quality and safety programs; provide support and expertise needed to reduce DEs; design, implement, and evaluate pilot interventions compared with usual care; and catalyze scientific, personnel, and infrastructure changes, which will last beyond the duration of this study. Figure 1 demonstrates our conceptual model of DEs, and Figure 2 shows our general framework for the study.

The specific aims of this study are:

*Aim 1:* To implement an enhanced case review infrastructure that can accurately identify DEs and characterize diagnostic processes among patients experiencing inpatient deaths, intensive care unit (ICU) transfers, or rapid-response team calls taking place at participating HOMERuN hospitals.

*Aim 2:* To develop site-level and group-wide benchmarking reports of error rates, diagnostic processes, and diagnostic performance and incorporate them into sites' safety and quality programs.

*Aim 3:* To use Aim 2 infrastructure to identify and pilot Safety I and Safety II interventions.

*Aim 4:* To carry out a comprehensive program evaluation, including analysis of rates of DEs and process faults before and after implementation of our program, and analysis of reach, effectiveness, adoption, implementation, and maintenance (RE-AIM) of Aims 1 and 2 programs and Aim 3 pilot interventions.

**Hypothesis 4a.** That our collaborative program is associated with a reduction in DEs and diagnostic process faults compared with baseline performance.

**Hypothesis 4b.** That our collaborative program is associated with the adoption of programs targeting DEs.

The design of this study is a pragmatic, multicenter quality, and safety program with exploratory outcome evaluation using interrupted time-series techniques.

## METHODS AND ANALYSIS

### Study setting

This study will be carried out at 16 sites in the HOMERuN network, including 10 that participated in the UPSIDE study: six safety-net hospitals, one community-based teaching hospital, and nine traditional academic centers. UPSIDE sites were preferred because of their past experience with our adjudication processes, but it was not a requirement for

participation. Sites were also recruited from the HOMERuN collaborative. All sites have access to support teams and patient populations needed for our research, site leaders with roles in their sites' quality and safety infrastructure, and electronic health record (EHR) systems (Epic or Cerner) which will permit deployment of standard data queries to identify potentially eligible patients, important covariates, and candidate predictors of DE.

### **Patient and public involvement**

The HOMERuN Patient and Family Advisory Council (PFAC) will provide perspectives on key elements of the study. The HOMERuN PFAC includes 10 patient and family caregivers from 10 US hospital systems. PFAC members will be engaged and provide insights during design (identification of Safety II examples, review of interview guides, surveys, and adjudication forms), implementation (consent materials, study operations, review of preliminary findings), and dissemination (interpretation of study results, dissemination of study findings).

### **Participant eligibility criteria**

**Patients**—Adult patients admitted to general medicine services at participating hospitals between September 1, 2022 and June 30, 2026 will be eligible. Using these criteria, sites will use local electronic data queries to identify in-hospital deaths, ICU transfers 48 h after admission, and rapid response team (RRT)<sup>14</sup> calls.

Cases will first be identified retrospectively to gather baseline data for 12 months (i.e., using Research Electronic Data Capture [REDCap] to randomly select 10 events per month per site among eligible patients, as defined above, admitted from September 1, 2022 to August 31, 2023). Starting from September 1, 2023, cases will then be identified within a week of the qualifying event (two alternating with three randomly selected cases per week per site) and proceed concurrently through study completion. We will use block randomization to select cases in both the retrospective and concurrent case identification. Only one event per patient-day can be chosen for review (practically, all events on that day would be adjudicated together). In all cases, reviewers will exclude cases identified in error (e.g., not a medical diagnosis), rapid response or ICU transfer due to policy rather than acuity (e.g., for desensitization to a medication), admission for comfort care only, admission following an out of hospital cardiac arrest, or those with unavailable medical records.

**Clinician subjects**—During the concurrent phase, we will identify clinicians (attending physician; senior resident, physician assistant or nurse practitioner, as applicable; and responsible physician at the time of escalation event) involved with all cases.

### **Aim 1: Identifying DEs**

Adjudications will utilize EHRs as source information. Reviewers will be asked to focus their review on the patient's medical record for that hospitalization, but reviewers may also review records before hospitalization to provide context and review any autopsy reports to confirm the final diagnosis. Results of adjudications will be entered into REDCap tools. Adjudications will generally occur after (and cannot be finalized until) patients have been discharged or died in the hospital.

Reviewers will undergo extensive training in the identification of DEs using methods previously established for the UPSIDE study,<sup>10</sup> including a series of webinars on topics, such as defining DEs and episodes of care, how to systematically perform medical record review, using the Safer-Dx and Diagnostic Error Evaluation and Research (DEER) taxonomy tools, how to assess patient harm, and how to avoid hindsight bias. We have created a series of deidentified “gold standard” training cases, some with and some without DE. Gold standard cases will then be completed by all reviewers, and training will continue until  $\kappa$  statistics (compared with the gold standard) exceed 0.7.

A key principle of our adjudication is a two-reviewer requirement, with the conclusions representing the viewpoint of two trained individuals not connected to the case. Consensus will be reached between the two reviewers before final results are recorded; in the event that a case is unclear or consensus cannot be reached, one of the site leads will serve as the third reviewer.

After initial training, every 10th case will be over-read by the site leads at each site (with full access to the medical record) to calculate inter-rater reliability between the two-adjudicator consensus and the site leads. We will also support sites to ensure consistency of reviews through semimonthly webinars for discussion of difficult case reviews and teaching, and creation of online resources, including frequently asked questions on how to handle common situations.

Finally, our case review tools will track the initial adjudication results (particularly whether or not an error took place) after the initial review by each reviewer, whether any disagreements were noted after the initial reading of the case materials, and how these differences were resolved.

**Identification of DEs**—We will use the modified Safer-Dx<sup>10,13</sup> (Figure S1) to prompt reviewers to consider all potential aspects of DE propagation. Reviewers will make a final assessment (using a six-level scale) to determine whether a DE occurred: no evidence, slight evidence, less than 50–50 but close call, more than 50–50 but close call, strong evidence, and virtually certain evidence, with the last three levels considered DEs.

**Understanding of diagnostic failure points**—Reviewers will then complete the modified DEER taxonomy tool (Figure S2)<sup>6,10,15</sup> to identify diagnostic failure points for all patients with a DE. Our modified DEER Taxonomy covers eight categories: presentation/triage; patient history; physical exam; diagnostic test ordering, performance, and interpretation; patient follow-up and monitoring; health care team dynamics; and clinical reasoning/assessment. Lastly, we will use the National Coordinating Council for Medication Error Reporting (NCC-MERP) tool<sup>16</sup> to categorize the degree of harm among cases with errors.

**Definition of Safety II resilience measures**—We will also ask reviewers to identify cases where a patient was at high risk for DE but no error was found and “good catches” where there was no DE despite rare or difficult diagnoses.

**Surveys of clinicians**—The goal of clinician surveys will be to gather information not easily obtained from the medical record, such as impressions of communication effectiveness or clinical uncertainty, to illuminate those aspects of the diagnostic process and inform intervention development. Our surveys (Figure S3) will utilize a modified version of the DEER taxonomy to identify clinicians' impressions of problems with care, as well as items asking about underlying systems factors (e.g., operational procedures, workload, technology), patient factors (e.g., limited English proficiency), and diagnostic resilience activities (e.g., how the case demonstrated diagnostic excellence), as well as open-ended items asking about Safety I and Safety II issues.

In all cases, the SaferDx tool will be completed before clinician survey responses to avoid any bias in DE determination between the two study arms. However, for concurrent cases, the DEER taxonomy will be completed after a review of survey responses to provide a richer understanding of the causes of DE and to inform interventions to improve diagnostic processes.

**Selection of exemplar cases**—We will conduct a deep review of 10 cases adjudicated by experts (five with errors and five without) using a positive deviance approach.<sup>17,18</sup> Cases will be selected to maximize the contrast between those with and without DEs to guard against erroneous conclusions.<sup>19–21</sup> Using qualitative thematic analysis, we will conduct instrumental collective case study analyses to identify themes that emerge from clinician surveys, redacted charts or case review documents, free-text descriptors from our adjudication tools, and clinician interviews.<sup>22</sup> Results from these qualitative analyses will be used to identify possible targets for improvement and better understand Safety II issues.

**Systems analysis of cases**—As in prior studies, we will conduct systems analyses of eight cases of DE (i.e., one for each DEER diagnostic process dimension) at selected sites to provide additional insight into potential targets for Safety I interventions. We will apply tools, such as health care failure modes and effects analysis, cause-and-effect diagrams, and root cause analyses to understand the root causes of reviewed DEs.

## **Aim 2: Benchmarking of diagnostic performance across sites**

Aim 2 has two related goals, the first being the development and dissemination of data reports representing error rates and diagnostic processes, and the second being the incorporation of these reports into local quality programs.

**Feedback report development**—Initial forms of our reports will allow sites to view their data only and will present error prevalence, harms related to errors, diagnostic process faults as represented by DEER taxonomy groupings, and population-attributable fractions indicating the highest priority process faults. In addition, prototypes will include exploratory measures indicating possible Safety II events (e.g., “good catches”) and resilience features identified from exemplar cases and expanded through qualitative efforts. Following iterative refinement of these reports, we will allow sites to see their own performance compared to the overall group's, followed by additional rounds of iterative refinement in response to user feedback.



**Incorporation of reports into existing programs**—Once reports have been validated by working groups, we will then turn to help site leads to incorporate reports into existing quality improvement infrastructure at their institutions. We will ask site leads to identify and engage key stakeholders (such as the chief quality or chief safety officers, residency program directors, and clinical chiefs), as well as important committees where adverse events, deaths, or ICU transfers are reviewed. We will also ask site leads to consider venues or committees where clinical excellence is taught, such as clinical-pathological conferences or resident reports.

### **Aim 3: To use Aim 2 infrastructure to identify and pilot Safety I and Safety II interventions**

The overall goal of this aim is to pilot test interventions that reduce DEs by simultaneously focusing on Safety I and Safety II areas of improvement. Table 1 provides examples of possible interventions, but final interventions will be based on lessons learned from Aims 1 and 2 of the study.

We will utilize our collaborative calls to first undertake an environmental scan to understand any current or recent programs seeking to reduce DEs or improve resilience on general medicine services at our sites. Using these data and the data collected as part of Aims 1 and 2, we will undertake intervention development and refinement using the IDEAS framework for innovation and intervention planning.<sup>23,24</sup> Collaborative calls with site leaders will focus on data review, identification of contributing factors for Safety I and II focus areas, development of general blueprints, goals, and objectives for interventions, identification of target populations for each intervention, and development of implementation plans for each (including factors such as leadership engagement, EHR changes, and change management needs). We do not expect all sites to implement all interventions but rather we will allow sites to choose from among a variety of toolkit components best suited to their individual situation. We will aim to have intervention pilots launched by the second half of Year 3 of this project, with plans to continue them through the first 6 months of Year 4. As in our prior studies, we will use a mentored implementation approach to assist sites in the sociotechnical approach to quality improvement.<sup>25</sup> Following best practices for studies of complex interventions,<sup>26</sup> we will allow for modifications of the form of the interventions (e.g., mode of delivery, dose) while maintaining fidelity to the core functions. We will specify allowable adaptations and a description of how planned and unplanned adaptations will be managed, measured, and reported over time.

### **Outcomes**

Our primary outcome will be rates of DEs among our study populations using our adjudication methods, extrapolated to the entire population of medicine inpatients at each site. To do this, we will first measure the proportion of patients with inpatient death, ICU transfer 48 h after admission, and rapid responses among medicine inpatients at each site. Then we will measure proportions with DE among each of these trigger populations. Finally, we will use both sets of data to extrapolate DE rates among the entire population. This allows for an unbiased estimate of the effects of the program even if the number of trigger events is reduced (i.e., by improvements in the diagnostic process).



Secondary outcomes will include DEs contributing to death, permanent patient harm, or requiring life-sustaining treatment using the NCC-MERP criteria (i.e., categories G, H, and I)<sup>16</sup> and number of diagnostic process faults per patient (as determined by the DEER taxonomy during adjudication).

**Evaluation of implementation**—We will use the RE-AIM framework<sup>27</sup> to evaluate the extent to which the implementation of ADEPT programs was successful at each of the study sites.

Measures of *Reach* will include a description of hospitals, teams, or units where our DE measurement methodology is built into usual care (Aim 1), when and where benchmarking data are incorporated into usual care (Aim 2), and where pilot interventions are adopted (Aim 3) compared with those where they are not adopted, and a description of patients who receive patient-level interventions (Aim 3) compared with those who do not. *Effectiveness* will be measured using our primary and secondary outcomes, as above. *Adoption* will be measured as the number and types of audit/feedback, bench-marking, and Safety I and Safety II interventions adopted at each site. *Implementation* will be measured quantitatively, including the proportion of patients in each trigger population who undergo adjudication for DE, the number of surveys administered to clinicians, the number of benchmarking reports produced, and the proportion of patients who receive patient-level interventions. To measure *maintenance*, we will track the above metrics over time (using run charts) and re-evaluate all metrics 6 months after the program is complete.

**Assessment of barriers and facilitators of implementation**—To understand barriers and facilitators of implementation of our program, we will conduct focus groups of a purposive sample of front-line clinicians and hospital leaders in the setting of mentorship pod meetings and collaborative calls during phase 2 of the study. The interviews will be conducted virtually by two coinvestigators using a semistructured interview guide. The interview guide will be informed by the consolidated framework for implementation research (CFIR), which consists of 26 constructs (and 13 additional subconstructs) within five domains: intervention characteristics, outer setting, inner setting, characteristics of individuals, and process. Interviews will be modified based on preliminary findings.

## Analysis plan

**Aims 1, 2, and 3 statistical approach**—These Aims will be evaluated descriptively using the RE-AIM framework (see Aim 4b statistical approach, below).

**Hypothesis 4a statistical approach**—The incidence of DEs (primary outcome) in the entire cohort will be characterized using proportions and 95% confidence intervals. In addition, the incidence of each diagnostic process failure point in the DEER taxonomy (secondary outcome) will be calculated for all patients with DEs. These results will also be rolled up for each category (i.e., presence of any element within a given category). We will follow a similar approach to analyzing measures of Safety II criteria (e.g., “good catches”).

The effect of the program as a whole on primary and secondary outcomes will be assessed using interrupted time series methodology. We will track error rates during the 12 months

before the start of any interventions. We will then track error rates during the 36 months once the program starts. We will measure the change in the  $y$ -intercept (“change in step,” i.e., sudden improvement with the initiation of the program) and change in slope (change in the baseline temporal trend). Significant improvements in one (without significant worsening in the other) will be considered evidence of benefit.

We will use segmented multivariable regression by month to evaluate baseline and changes in  $y$ -intercept and slope over time. We will use logistic regression for dichotomous outcomes (presence or absence of any DE, presence or absence of any severe DE); and Poisson regression for “count” outcomes (number of DEs per patient, number of diagnostic process faults per patient). We will use generalized linear mixed models (e.g., Proc GLIMMIX in SAS) for fixed and random effects. Random effects will include attending physician at admission and study site. Fixed effects for potential confounders (gathered from EHR data, Vizient Data, and adjudication forms) will include: teaching versus nonteaching team, admission by night-float, admission service (medicine vs. other), admission source (emergency department [ED] vs. interhospital transfer vs. other), admission diagnosis category (using clinical classification software for ICD-10 codes), all patients refined diagnosis-related group weight, Elixhauser comorbidity score, hospitalization in the previous 30 days, ED visits in the previous 30 days, number of outpatient visits in the previous 30 days, and demographics (age, sex, race, ethnicity, primary language, median income by zip code, and insurance).

We will conduct a limited number of subgroup analyses, including by sex, age, income, and race/ethnicity to evaluate any disparities in DE rates and whether the intervention is more or less effective in certain populations (for the latter analyses, we will use interaction terms, e.g., subgroup  $\times$  change in slope, to evaluate the statistical significance of any effect modification). We also plan secondary analyses to determine whether rates of DEs change between phases 1 and 2 (when interventions will be implemented—see Figure S4 timeline).

Throughout we will use weighting to produce population-based estimates reflecting the differential sampling percentages at each site, particularly if we notice differential detection of cases because of differences in RRT, ICU transfer, or inpatient death data availability.

**Hypothesis 4b statistical approach**—We will use the RE-AIM framework<sup>27</sup> to evaluate the extent to which the implementation of ADEPT programs was successful at each of the study sites. Reach, adoption, and implementation measures will be analyzed using simple statistics to define rates of adoption over time within sites, between sites, and compared to each site’s starting place (defined by our environmental scans administered at the beginning of our study).

To measure maintenance, we will track the reach, adoption, and implementation metrics over time (using run charts) and re-evaluate all metrics in the 4 months after the program is complete (phase 3—see Figure S4 timeline).

**Qualitative analyses of information collected during case reviews (Aim 1)**—We will conduct a thematic analysis.<sup>28</sup> Inductive and deductive coding will be conducted by

co-investigators.<sup>29</sup> A preliminary codebook will be developed with the goals of identifying the case characteristics, external and internal contexts, and characteristics of the individuals involved. The team will iteratively meet and discuss their coding with the goal of reconciling the proposed codebook until a final codebook is agreed upon. The final codebook will be applied to all transcripts. Qualitative data analysis will be managed in Dedoose software (sociocultural research consultants). From the coded data, we will assign themes that capture the major concepts from interviews.<sup>30</sup>

**Qualitative analysis of collaborative calls (Aims 2 and 4)**—Using thematic analysis,<sup>28</sup> we will analyze moderated small group discussions held during collaborative webinars. We will use rapid qualitative methods to conduct a mixed inductive–deductive thematic analysis at the semantic level. Templated summaries and matrix analysis will be used to analyze the data.<sup>30,31</sup>

During each small group discussion, the moderators will take field notes and observations to supplement the audio recordings. The audio recordings and field notes will be used to support Aim 2 collaborative efforts as well as for Aim 4 program evaluations.

**Qualitative analysis of implementation (Aim 4)**—To qualitatively analyze barriers and facilitators of implementation, we will conduct clinician and leader focus groups in Year 4 using a standard semistructured focus group guide. We will use rapid qualitative methods, as described above. We will use a combination of inductive and deductive approaches, starting with a priori themes based on updated CFIR constructs<sup>32</sup> (deductive), and then use an inductive approach to draw on emerging themes from the interviews. Codes will be revised by two coinvestigators from the core 4 team based on analysis of initial transcripts. Coding of themes will be done by two independent coders, then codes will be reviewed together until an agreement is reached. We will stop analyses when we reach “saturation,” that is, no new insights are being gleaned, likely 6–12 from each stakeholder group (attendings, residents/advanced practice providers, leadership).

### Sample size

We anticipate 1920 total cases in the preintervention period (phase 0) and 5280 total cases in the postintervention period of the study (phases 1 and 2). All computations below are based on achieving 80% power and a significance level ( $\alpha$ ) of .05.

**Primary outcome: DE**—Table S1 provides the absolute and relative (%) minimally detectable differences in the proportion of patients with at least one diagnostic error for the pre- and postintervention sample sizes as described above, using various baseline DE proportions. This range of baseline DE proportions is based upon results from the UPSIDE study, in which patients who died had a prevalence of errors of 21.7% and patients who were transferred to the ICU had an error rate of 28%. Assuming a combined diagnostic error rate of 24% in the preintervention period (boldfaced row in Table S1), we will achieve 80% power to detect an absolute reduction of 3.2% (15.5% relative rate reduction) in DE rates.

Based upon data from UPSIDE on the incidence of death and ICU transfer in general medicine patients, as well as population-level data on the incidence of RRT calls,<sup>6</sup> we

estimate that the preintervention rate of DEs in patients undergoing one of these events, weighted for their prevalence in the general medicine population, to be 2.67%. If the minimally detectable difference in DE rates is a 15.5% relative reduction, as per our analysis above, the minimally detectable absolute difference in the weighted rate of DEs is 0.41%, that is, from 2.67% to 2.26%.

**Secondary outcome: DE that contributed to severe patient harm or death**—In the UPSIDE cohort, 78.7% episodes of care in which DE occurred were associated with harm to the patient. If we assume a low rate of diagnostic error in the ADEPT cohort (21%), which provides a conservative estimate of power to detect a change in this secondary outcome, the minimally detectable absolute reduction in the rate of harm in patients with DE is between 6.3% and 7.4% (7.5%–10.0% relative reduction).

**Secondary outcome: Number of DEER taxonomy process faults present**—In the UPSIDE cohort, the mean number of process faults was 6.8 (SD: 3.7) in patients with a DE. With our proposed sample size, we achieve 80% power to detect an absolute reduction of 0.62 (9.1% relative reduction).

Timeline (Figure S4).

## ETHICS AND DISSEMINATION

University of California, San Francisco (UCSF) is serving as the central Institutional Review Board (IRB) for the study and has approved the study protocol. The other participating sites either ceded to the central IRB or determined that the study was exempt from IRB approval at the local level. The central IRB approved a waiver of patient consent on the grounds that this is a minimal-risk study, obtaining consent would be a potentially poor use of limited research funds, and that requiring patient consent would greatly limit the generalizability of the research. Personal health information will be retained at the local hospital level in a secure cloud-based location, but only deidentified data (using study ID numbers) will be shared with UCSF for analysis. PHI will be exchanged between each site and Vizient but no more than is currently done for operational purposes; Vizient will only return deidentified data back to study sites using study ID numbers. Investigators at UCSF (A. D. A., C. H., W. J. B.) will have access to the final (deidentified) study data set; other investigators will have access to aggregated data. Given the minimal-risk nature of this study, a data monitoring committee is not needed. The co-principal investigators and project manager will audit the conduct of the study on a monthly basis; they will also communicate any important protocol modifications to the central IRB, Agency for Healthcare Research and Quality, and update [clinicaltrials.gov](https://clinicaltrials.gov) as needed.

We will make our study methodologies, data collection tools, analytic approaches, interventions, and approach to implementation of the intervention available for widespread distribution, in conjunction with our dissemination partners.

Authorship eligibility will follow the guidelines of the Editors. We do not plan to use professional writers. Access to the full protocol and statistical code will be made available upon reasonable request.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

The work was supported by AHRQ, grant number R18HS029366. **Roles and responsibilities of the funder:** AHRQ has had and will have no role in the study design, collection, management, analysis, and interpretation of data; writing of this manuscript or future manuscripts; and the decision to submit this manuscript and future manuscripts for publication. **Other roles and responsibilities:** Consultants provide specific content expertise in diagnostic reasoning (Gurpreet Dhaliwal, Sanjay Saint, Shoshana Herzig, and Janice Kwan) and regarding intervention development (Gopi Astik, Andrew Olson, Robert Trowbridge). The Advisory Board provides guidance and expertise to ensure the successful completion of all study aims. Board members include David Bates, Tejal Gandhi, Kaveh Shojania, Urmimala Sarkar, Neil Powe, Vineet Arora, and Gordon Schiff. The HOMERuN Patient–Family Advisory Council provides patient and family stakeholder input into the development of all patient-facing interventions, outcomes to be evaluated, interpretation of all results, and all dissemination activities, including messaging of the study’s results to the public. Members include Martha Carnie, Beverly Rogers, Patricia Evans, Georgiann Ziegler, Melissa Wurst, Cathy Hanson, D’Anna Holmes, Jim Banta, and Gina Symczak.

## REFERENCES

1. Ely JW, Kaldjian LC, D’Alessandro DM. Diagnostic errors in primary care: lessons learned. *J Am Board Fam Med.* 2012;25(1):87–97. [PubMed: 22218629]
2. Groszkruger D Diagnostic error: untapped potential for improving patient safety? *J Healthc Risk Manag.* 2014;34(1):38–43. [PubMed: 25070255]
3. McCarthy M Diagnostic error remains a pervasive, underappreciated problem, US report says. *BMJ.* 2015;351:h5064. [PubMed: 26396160]
4. Norman GR, Eva KW. Diagnostic error and clinical reasoning. *Med Educ.* 2010;44(1):94–100. [PubMed: 20078760]
5. Scarpello J Diagnostic error: the Achilles’ heel of patient safety? *Clin Med.* 2011;11(4):310–311.
6. Schiff GD. Diagnostic error in medicine: analysis of 583 physician-reported errors. *Arch Intern Med.* 2009;169(20):1881–1887. [PubMed: 19901140]
7. Singh H, Giardina TD, Forjuoh SN, et al. Electronic health record-based surveillance of diagnostic errors in primary care. *BMJ Qual Saf.* 2012;21(2):93–100.
8. Carson-Stevens A, Donaldson L, Sheikh A. The rise of patient Safety-II: should we give up hope on Safety-I and extracting value from patient safety incidents? Comment on “False Dawns and New Horizons in Patient Safety Research and Practice”. *Int J Health Policy Manag.* 2018;7(7):667–670. [PubMed: 29996588]
9. Turenne CP, Gautier L, Degroote S, Guillard E, Chabrol F, Ridde V. Conceptual analysis of health systems resilience: a scoping review. *Soc Sci Med.* 2019;232:168–180. [PubMed: 31100697]
10. Dalal AK, Schnipper JL, Raffel K, Ranji S, Lee T, Auerbach A. Identifying and classifying diagnostic errors in acute care across hospitals: early lessons from the utility of predictive systems in diagnostic errors (UPSIDE) study. *J Hosp Med.* Published online May 21, 2023. doi:10.1002/jhm.13136
11. Auerbach AD, Patel MS, Metlay JP, et al. The hospital medicine reengineering network (HOMERuN): a learning organization focused on improving hospital care. *Acad Med.* 2014;89(3):415–420. [PubMed: 24448050]
12. National Academies of Sciences Engineering, and Medicine. *Improving Diagnosis in Health Care.* The National Academies Press; 2015.

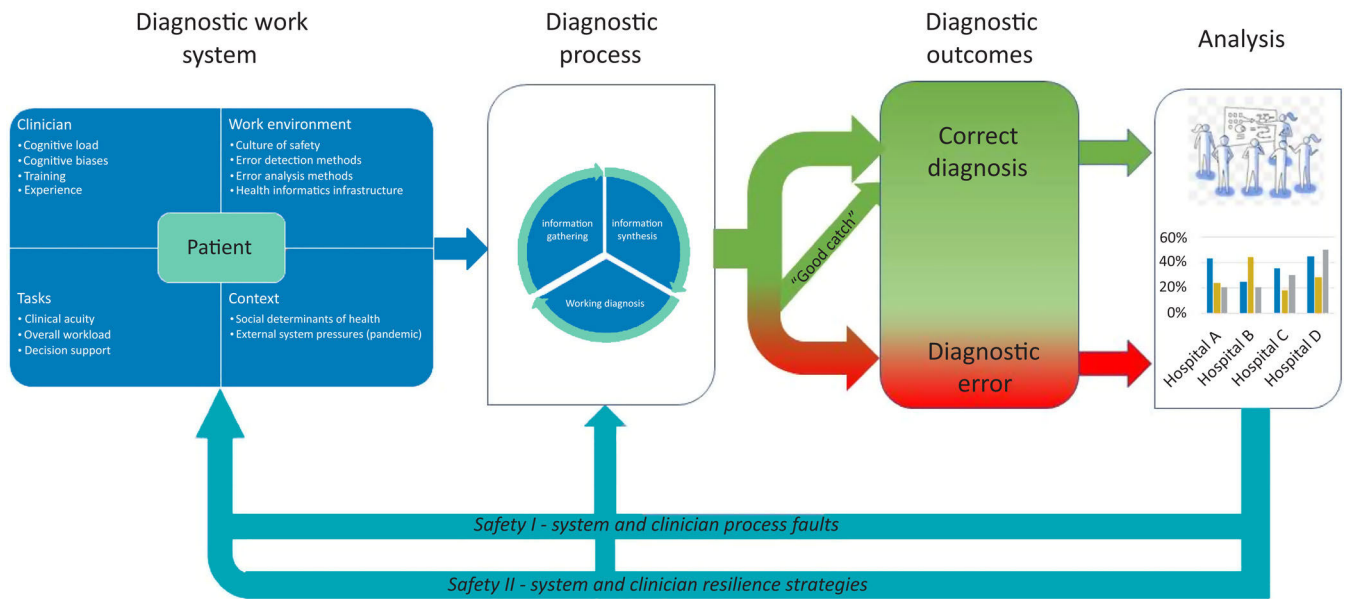
13. Singh H, Khanna A, Spitzmueller C, Meyer AND. Recommendations for using the revised safer Dx instrument to help measure and improve diagnostic safety. *Diagnosis*. 2019;6(4):315–323. [PubMed: 31287795]
14. Institute for Healthcare Improvement. Rapid response teams. 2023. Accessed August 16, 2023. <https://www.ihl.org/Topics/RapidResponseTeams/Pages/default.aspx>
15. Schiff GD. Diagnosis and diagnostic errors: time for a new paradigm. *BMJ Qual Saf*. 2014;23(1):1–3.
16. National Coordinating Council for Medication Error Reporting (NCCMERP). NCC MERP types of medication errors. 2001. Accessed August 16, 2023. <https://nccmerp.org/types-medication-errors>
17. Bradley EH, Curry LA, Ramanadhan S, Rowe L, Nembhard IM, Krumholz HM. Research in action: using positive deviance to improve quality of health care. *Implement Sci*. 2009;4:25. [PubMed: 19426507]
18. Rose AJ, McCullough MB. A practical guide to using the positive deviance method in health services research. *Health Serv Res*. 2017;52(3):1207–1222. [PubMed: 27349472]
19. Palinkas LA. Qualitative and mixed methods in mental health services and implementation research. *J Clin Child Adolesc Psychol*. 2014;43(6):851–861. [PubMed: 25350675]
20. Palinkas LA, Aarons GA, Horwitz S, Chamberlain P, Hurlburt M, Landsverk J. Mixed method designs in implementation research. *Adm Policy Ment Health Ment Health Serv*. 2011;38(1):44–53.
21. Palinkas LA, Mendon SJ, Hamilton AB. Innovations in mixed methods evaluations. *Annu Rev Public Health*. 2019;40:423–442. [PubMed: 30633710]
22. Crowe S, Cresswell K, Robertson A, Huby G, Avery A, Sheikh A. The case study approach. *BMC Med Res Methodol*. 2011;11:100. [PubMed: 21707982]
23. Robertson M, Henning R, Warren N, et al. The intervention design and analysis scorecard: a planning tool for participatory design of integrated health and safety interventions in the workplace. *J Occup Environ Med*. 2013;55(suppl 12):S86–S88. [PubMed: 24284761]
24. Robertson MM, Henning RA, Warren N, et al. Participatory design of integrated safety and health interventions in the workplace: a case study using the intervention design and analysis scorecard (IDEAS) tool. *Int J Hum Factors Ergon*. 2015;3(3–4):303–326. [PubMed: 33898018]
25. Li J, Hinami K, Hansen LO, Maynard G, Budnitz T, Williams MV. The physician mentored implementation model: a promising quality improvement framework for health care change. *Acad Med*. 2015;90(3):303–310. [PubMed: 25354069]
26. Patient-Centered Outcomes Research Institute. PCORI methodology standards: studies of complex interventions. 2020. Accessed September 5, 2020. <https://www.pcori.org/research-results/about-our-research/research-methodology/pcori-methodology-standards#Complex>
27. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health*. 1999;89(9):1322–1327. [PubMed: 10474547]
28. Kiger ME, Varpio L. Thematic analysis of qualitative data: AMEE guide no. 131. *Med Teach*. 2020;42(8):846–854. [PubMed: 32356468]
29. Schrier M *Qualitative Content Analysis in Practice*. Sage Publications Inc; 2012.
30. Hamilton A *Qualitative methods in rapid turn-around health services research*. 2013. Accessed May 1, 2021. [https://www.hsrdr.research.va.gov/for\\_researchers/cyber\\_seminars/archives/780-notes.pdf](https://www.hsrdr.research.va.gov/for_researchers/cyber_seminars/archives/780-notes.pdf)
31. Keniston A, McBeth L, Astik G, et al. Practical applications of rapid qualitative analysis for operations, quality improvement, and research in dynamically changing hospital environments. *Jt Comm J Qual Patient Saf*. 2023;49(2):98–104. [PubMed: 36585315]
32. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated consolidated framework for implementation research based on user feedback. *Implement Sci*. 2022;17(1):75. [PubMed: 36309746]

## STRENGTHS AND LIMITATIONS OF THIS STUDY

### Strengths and weaknesses:

- Multicenter, real-world design increases generalizability.
- Rigorous two-person adjudication process allows for expert review while maximizing the reliability and validity of diagnostic error assessment.
- Use of existing safety and quality systems to share data describing error rates and underlying causes promotes adoption and sustainability.
- Design allows sites to choose and customize interventions to their setting, maximizing efficacy.
- Toolkit will provide lessons learned for widespread implementation.
- Variability and flexibility in intervention development and implementation will require further studies to definitively demonstrate their benefit.





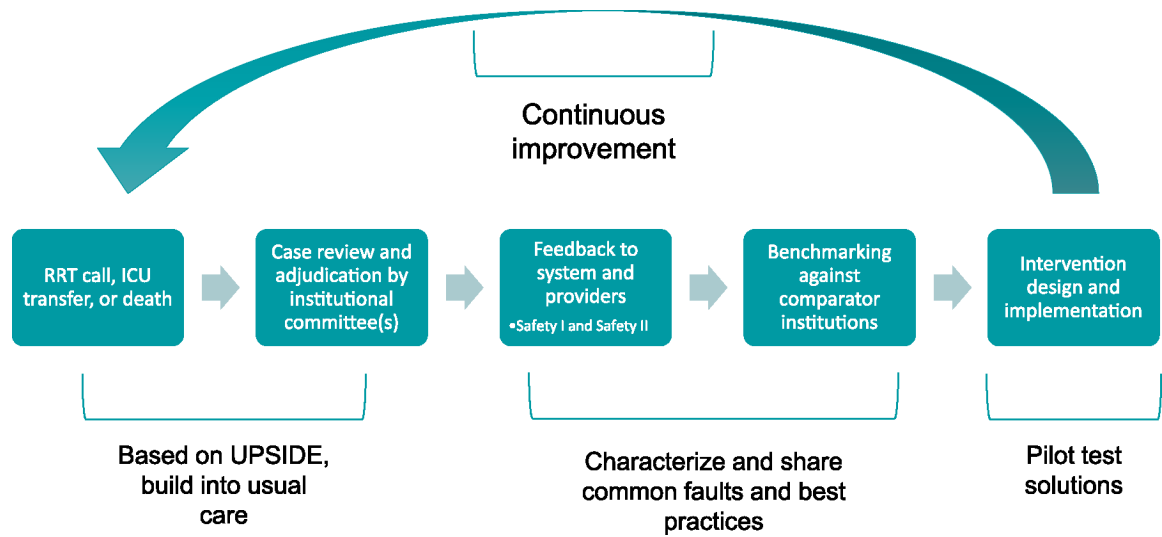
**FIGURE 1.** Conceptual model of diagnostic processes, based on previous work by the National Academies of Sciences, Engineering, and Medicine,<sup>12</sup> and the SaferDx framework.<sup>13</sup>

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**FIGURE 2.**

Framework of the ADEPT study for diagnostic improvement. ADEPT, achieving diagnostic excellence through prevention and teamwork; ICU, intensive care unit; RRT, rapid response team; UPSIDE, utility of predictive systems in diagnostic errors.

**TABLE 1**

Intervention schema with hypothetical examples.

	<b>Focus of intervention</b>	<b>Safety I</b>	<b>Safety II</b>
Patient		<ul style="list-style-type: none"> <li>• Patient access to medical records and documented diagnoses</li> </ul>	<ul style="list-style-type: none"> <li>• Tools for communicating about uncertainty</li> <li>• Surveys to ask patients about whether they understand their diagnosis</li> </ul>
Clinician		<ul style="list-style-type: none"> <li>• Individual feedback regarding DE cases</li> <li>• Alerts for patients who are not improving as expected</li> </ul>	<ul style="list-style-type: none"> <li>• Curricula on “habits of highly effective diagnosticians”</li> <li>• Diagnostic checklists or “time outs”</li> </ul>
System		<ul style="list-style-type: none"> <li>• Increase capacity to read or perform diagnostic procedures off-hours</li> <li>• Decision support to identify incorrect test selection</li> </ul>	<ul style="list-style-type: none"> <li>• Diagnostic peer-consult service to provide expert overreads</li> <li>• More clinical backup when work thresholds are exceeded</li> </ul>

Abbreviation: DE, diagnostic error.