medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Title: **Aggregating data to accelerate personalized therapy in heart failure (ADAPT-HF)**

Running title: **Transparent harmonization of clinical trial data**

Chris Roeder,[1] Carsten Gorg,[1] Atefeh Talebi,[2] Laura Stevens,[3] Denise Scholtens,[4] Laura Rasmussen-Torvik,[4] Laura M Alagna,[4] Sanjiv Shah,[4] Jennifer Hall,[5] Amar Das,[6] Pardeep Jhund,[2] David Kao[1]

1. University of Colorado School of Medicine, Aurora, CO
2. University of Glasgow School of Cardiovascular and Metabolic Health
3. Databricks Inc., San Francisco, CA
4. Northwestern University Feinberg School of Medicine, Chicago, IL
5. American Heart Association, Dallas, TX
6. Guardant, Inc., Palo Alto, CA

Corresponding author:

David Kao

1890 N Revere Ct.

Box F-563 - Room 7020

Aurora, CO 80045

David.Kao@cuanschutz.edu

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

ABSTRACT

Background: Increased public access to data from disparate sources provides opportunities to study and validate predictive and subphenotype models in heterogeneous disease conditions using aggregated individual patient data. Robust, explicit, and transparent harmonization of data elements is critical to ensure interpretability, reproducibility, and generalizability of secondary and retrospective analyses.

Methods & Results: We designed and implemented ADAPT (Aggregating Data to Accelerate Personalized Therapy), a scalable framework using multiple software packages (R, SQL, BigQuery) that enables rapid, explicit harmonization of structured data elements from randomized trials and observational studies using a standard spreadsheet interface. User-specified criteria are applied to primary study data to produce harmonized longitudinal datasets comprised of demographics, medical history, quantitative observations, repeated measures, and clinical outcomes. We demonstrate this functionality using 26 clinical studies found in the National Heart, Lung, and Blood Institute's BioLINCC resource. We illustrate the scalability of ADAPT to the order of billions of datapoints using administrative clinical data in a cloud-computing platform. We also present examples of collaborators using ADAPT for independent harmonization tasks for secondary analyses and democratization of publicly available data.

Conclusion: ADAPT is a disease-agnostic, extensible, and scalable platform to support robust, transparent harmonization of structured research data using interfaces accessible to a variety of researchers regardless of programming ability. It extends FAIR principles beyond research data to also represent harmonization analyses by improving Findability of harmonization decisions, Accessibility of methods to other stakeholders, Interoperability with independent analyses and datasets, and Reusability through efficient implementation in a variety of analysis environments.

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

BACKGROUND

Biomedical research has been revolutionized by the development of high-throughput molecular experimental methods and robust techniques available for analyzing large, complex translational datasets. Analogously, large amounts of data are generated during clinical research and routine care, and there is growing interest in establishing frameworks for harmonizing multiple clinical datasets.[1–6] Despite increasingly robust conceptual frameworks, practical curation and computational methods for leveraging secondary data analysis in clinical research remain comparatively underdeveloped. Aggregation of structured study data has been limited, in part due to uncertainty regarding comparability of data between sources, lack of consensus regarding clinical definitions, and inter-study differences in harmonization decisions. For example, if one study defines a chronic disease like heart failure (HF) based on self-reported history, another uses a set of clinical criteria such as those from the Framingham Heart Study and yet another uses administrative codes, patients identified as having HF in each study may not be truly comparable (**Figure 1**). Recommendations and consensus statements that articulate frameworks for harmonizing disparate data describe the key components in the process of codifying relationships between potentially similar data fields, but there remains no widely used platform for completing these tasks in a discoverable and reusable way. Uncertainty and difficulty in scaling clinical data harmonization have a number of consequences including lost opportunity for discovery, unnecessary effort and resource utilization through redundant harmonization of datasets for different projects, lack of transparency regarding reusability and comparability of different harmonization projects, and uncertainty regarding generalizability of any research findings to specific populations.

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

One commonly cited motivation for analysis of harmonized data from multiple studies, also called individual patient data meta-analysis, is the ability to study small but important patient populations. Robust analysis of these subgroups can be difficult due to challenges in explicit, portable classification criteria and small sample sizes, particularly in clinical trials. Evidence from clinical trials has been treated as equally applicable to all patients meeting a trial's inclusion criteria. In reality, phenotypes are often heterogeneous among patients with chronic conditions, which may result in variable prognoses and treatment responses.[7] For example, guideline-directed medical therapy for HF is recommended based on a single measurement of heart function, the left ventricular ejection fraction (LVEF). However, there are dozens of known HF etiologies that could affect long-term prognosis and treatment response, such as coronary artery disease, toxic exposures, and genetic abnormalities affecting muscle contraction.[8] This diversity of phenotypes means that even evidence-based interventions may have beneficial, neutral, or harmful effects in any given subpopulation.

To overcome limited sample sizes, characterization of disease sub-populations can be accomplished through individual patient-level data harmonization and analysis,[5,9] wherein patients with similar subphenotypes can be aggregated from multiple studies. Utility, credibility, and generalizability of secondary analysis findings from aggregated individual patient data analyses from multiple studies, however, depend on the accuracy and consistency of data harmonization between relevant studies as well as the ability for other investigators and clinicians to examine the harmonization mappings. Historically, reconciling these differences has been accomplished by examining different statistical scripts for each unique analysis. For

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

example, the fields 'sex' = "Male" in Study A, 'gender' = "1" in Study B, and 'male' = "Yes" in Study C could all denote male subjects, but the mapping definitions can only be discerned by reading the primary code (e.g. R, SAS, Stata, Python). Due to the complexity of clinical data, harmonizing the wide range of elements needed for aggregate analysis is often labor-intensive and difficult to replicate. Systematic uncertainty is introduced by the limited ability to identify specific harmonization decisions within programming code and to apply those definitions to an independent dataset or clinical implementation system. This uncertainty compounds the impacts on generalizability that arise from aggregate analyses from multiple clinical studies based on study enrollment criteria. Despite the complexity of aggregating these data and potential impacts on the utility of resulting findings, there is little, if any, documentation of harmonization or variable mapping between studies used for aggregate analysis.

The principles of Findability, Accessibility, Interoperability, and Reusability (FAIR) are frequently used to guide collection, annotation, and storage of biomedical data.[10] We propose that FAIR principles are also applicable to data harmonization mappings and pipelines. We further propose that the construction of standardized frameworks to codify and implement harmonization specifications will facilitate explicit, reusable, and generalizable aggregation of clinical data.

Objective

Develop and demonstrate a scalable, reusable framework that enables rapid, flexible, and transparent harmonization of clinical data to produce aggregate harmonized datasets from disparate sources.

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

METHODS

Requirements

Our framework addresses the following properties of analyses using harmonized data:

(i) Clinical data may originate from different sources in different formats.

(ii) Acquisition of new clinical data or discovery of new knowledge may necessitate revision of concept mapping or phenotype definitions.

(iii) A clinical concept may have different definitions in different contexts (e.g. DM)

(iv) Analyses may require an arbitrarily large number of harmonized data elements (e. g. artificial intelligence/machine learning).

(v) Analyses and software may require data outputs in a variety of formats.

The ADAPT (Aggregating Data to Accelerate Personalized Therapy) framework applies user-specified harmonization criteria to patient-level data from disparate sources to create a single harmonized dataset. Inputs are raw clinical data from the primary study, such as medical history, medications, laboratory values, and imaging findings; domain knowledge from the researcher, such as disease definitions, pathophysiology, study design, and time intervals; and data format requirements for the desired analysis such as normalized, entity-attribute-value (EAV) and repeated-measures tables. Unique harmonization criteria can be created by a researcher for a given project, or definitions from previous analyses can be reused from prior analyses. Criteria may be based on formal disease definitions (e.g. HF with reduced ejection fraction = clinical HF with LVEF < 40%), widely accepted clinical knowledge (e.g. chronic kidney disease stage 3a = estimated glomerular filtration rate = 15-29 ml/min), or they can be customized for specific

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

needs (e.g. baseline B-type natriuretic peptide level > median of all subjects enrolled in study A). Outputs of ADAPT are harmonization specifications in tabular format and aggregated patient-centered datasets that can be formatted for many purposes such as disease-association, decision support, or quality assurance analysis. Herein we describe our framework and apply it to datasets and concepts related to HF as our initial use case.

Data sources

We obtained publicly available, deidentified data from two primary sources to develop and test the scalability of ADAPT. Use of these data is therefore not considered human subjects research, but we did conduct this work with review and approval of the Colorado Multi-Institutional Research Board as required by agencies providing the data. A full list of data sets, total subjects in each, and the source data agency are found in the **Supplement**.

BioLINCC

Data for 20 clinical HF trials (n=32,442 subjects) and 5 observational cohorts (n=54,085) were obtained from the National Heart, Lung, and Blood Institute's Biological Specimen and Data Repository Information Coordinating Center (BioLINCC, https://biolincc.nhlbi.nih.gov/home/).[11][11]Primary data were received as *.csv (comma-separated values), *.sas7bdat (SAS) or *.dta (Stata) files. Data dictionaries were obtained from BioLINCC and study sites (e.g. ARIC: https://aric.cscc.unc.edu/) in PDF, word-processing document, and spreadsheet formats.

Administrative hospital discharge data

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Deidentified records for 20 states between 1994 and 2023 totaling 288,341,132 hospital discharges were obtained from individual state public health agencies as available and/or the AHRQ Healthcare Utilization Project Statewide Inpatient Databases.[14] Diagnoses, external injuries, and procedures performed were reported using both International Classification of Disease (ICD) version 9 (up to 2015 quarter 3) and version 10 (2015 quarter 4 and after).

Data management software

Data were aggregated, harmonized and analyzed in multiple platforms including Protégé-OWL 3.3.1 (Stanford University, Palo Alto, CA, https://protege.stanford.edu/),[12,13] RStudio (version 2025.09.1, Posit Software, Boston, MA, https://posit.co/products/open-source/rstudio), the R statistical package (version 4.5.0, R Foundation for Statistical Computing, Vienna, Austria, https://www.r-project.org/), Stata Statistical Software (Release 18, StataCorp LLC), MySQL (version 8.0.27, Oracle Corporation. https://www.mysql.com/), PostgreSQL (version 9.6, PostgreSQL Global Development Group, https://www.postgresql.org), and Google BigQuery (Google LLC, https://cloud.google.com/bigquery).

ADAPT Framework

To harmonize data, domain-specific knowledge asserting functional equivalency between data elements must first be encoded by the user. Four primary types of information are required: 1) primary clinical data such as tests, measurements, physical examination or patient self-reported history; 2) harmonization criteria (e.g., a patient has anemia if their hemoglobin level is < 13 g/dL for men and < 12 g/dL for women); 3) domain knowledge regarding concepts, phenotypes, or disease processes relevant to the current project (e.g. studying the impact of anemia in HF);

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

and 4) the analysis being performed and data format needed to execute that analysis (e.g. survival analysis vs. random forest prediction vs. growth curve analysis).

Figure 1 shows a schematic of relationships between subjects, data, and harmonized phenotype concepts as linked by a set of simple relationships. The key differentiator of ADAPT is the prioritization of rapid, transparent and scalable harmonization. Although not exhaustive, we have found a wide variety of analyses can be accomplished using a combination of the following archetypal methods: 1) "identity" or "exact value" (harmonized value = original value), 2) "exact match" (harmonized value = x if original value = y), 3) "ordinal" (harmonized value = x if any of a set of specified values are present), and 4) "scalar" (harmonized = x if original value is "greater than," "less than," or "between" user-specified values). Ordinal variables may be handled either using exact string matching or using a scalar comparison (e.g. '> 12^{th} grade education'), so we elected not to create a distinct method for ordinal variables. Specific harmonization criteria may represent accepted clinical definitions (e.g., patients with a body mass index > 30 kg/m2, or a waist circumference ≥ 94 cm in men or ≥ 80 cm in women indicates obesity), or they may be analysis- or dataset-specific (e.g., lab value greater or less than the median of a specified dataset). Specification of harmonization criteria or rules is accomplished in a standard spreadsheet format, eliminating the requirement for programming knowledge and allowing use of convenience tools like copy/paste. Importantly, these assertions can be reused between multiple datasets, making subsequent harmonization tasks more efficient and, where desired, standardized.

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

The process of transforming raw data into many harmonized values can be broken down into the three primary steps: 1) collation of raw data from disparate sources into a single table of primary or raw data points with contextual annotation, such as the study, source table, table, field, and units for each datum; 2) application of scalar comparisons, string-matching, or assertions of equivalence (i.e. identity) comparing individual patient data and user-defined criteria (e.g., obesity = body mass index > 30 kg/m^2) to coalesce source data into common clinical concepts and values, and 3) formatting and export of harmonized results for analysis (e.g. descriptive statistics, hypothesis testing, machine learning in R, Stata, or SAS). Using a denormalized data structure, transparent harmonization specifications, and a modular pipeline, ADAPT enables more rapid ingestion of new datasets than standard methods as well as simplified reuse of harmonization definitions and analytic methods from prior projects on data acquired from disparate sources.

Implementation

The flow of data through ADAPT is shown in **Figure 3**. Primary data from sources of interest are then transformed and assembled into a single denormalized data repository using either a performance-optimized custom schema or an established Common Data Model such as that of the Observational Medical Outcomes Partnership. In the resulting repository, each line represents a single raw observation about one patient as recorded in the primary data source, e.g. Framingham subject 2345 (patient) -> Field *AB12* (sex) -> 1 (male). A unique index is assigned to each primary datapoint in order to maintain the provenance of harmonized values. Harmonized datapoints are produced by applying community-defined or user-specified criteria to the denormalized primary data using code that applies one of the harmonization method

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

archetypes described above: *exact matching* (e.g. study field demo.gender = 'M' -> concept:Sex = value:Male), *scalar comparisons* (study field demo.age < 18 -> concept:Adult = 'Yes'), or *identity* (e.g. study field demo.agey = 18 -> concept:age_yrs = 18). We implemented the ability to use phenotype-contingent criteria (e.g. race- and sex-specific ranges) if desired.

Application of harmonization specifications is performed using conditional joins or table merges based on user specifications. Example code blocks using R, SQL, and BigQuery are provided in **Table 1**. All versions of the ADAPT pipeline utilize the same simple tabular format to specify harmonization criteria through standard spreadsheet applications and as a harmonization criteria repository that may be readily searched and explored for existing harmonization rules (**Table 2**). This table can be manipulated easily by domain experts even if they lack coding expertise. The resulting file is used directly to execute the harmonization code, eliminating the possibility of discordance between the harmonized data produced and the repository of definitions. The product is a second denormalized EAV repository where each line represents one harmonized observation about one subject: e.g., Framingham subject FHS2345 (patient) -> common concept = "Sex" -> harmonized value = "Male" using datapoint F87654. Comprehensive BioLINCC harmonization mappings for BioLINCC data as well as harmonization code for the R statistical package and Google BigQuery are available at https://github.com/dkao42/ADAPT-HF.

When a harmonized value is created, it is annotated with the index of the primary datum used to create it (e.g., Subject FHS2345 from Clinical Trial A's concept:Sex has the harmonized value "Male" based on datapoint F87654 (Clinical Trial A, Subject 2345, form *a_demo*, field gender = 1). This index allows datum metadata such as the source table, field, or timepoint, to be

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

retrieved. An example of the harmonization criteria table and application to compiled raw data is provided in **Figure 4**. Using this approach, researchers can choose to treat measurements using different methods as equivalent (e.g. left ventricular ejection fraction via echocardiography can be compared to same measurement via magnetic resonance imaging) and represent this equivalence explicitly in harmonization criteria while retaining the ability to audit the source data in the event of unexpected findings.

The final result is a patient-centered graph-like structure for each subject that is a greatly expanded version of **Figure 1**. These graphs contain all available primary data relating to an individual, harmonized observations created from primary clinical data according to the user-specified phenotype definitions, and explicit representation of the data point(s) used to generate each harmonized value. Queries from any standard statistical package can be used to produce specific output formats for specific data analysis packages and algorithms. We have used this approach to conduct analyses ranging from simple logistic regression to unsupervised machine learning.

Example 1: Secondary analysis of clinical HF studies

We have used ADAPT to produce harmonized datasets comprised of all clinical randomized HF trials in BioLINCC as well as the ARIC, CHS, Framingham (Original, Offspring, Offspring Spouse, Generation III, Omni 1-2), JHS, MESA observational cohorts (**Table 3**). Taken together, we harmonized > 7519 study variables to 692 harmonized concepts from both study groups (>85,000 subjects, > 140 million raw data points, > 30 million harmonized data points using ~17,800 individual harmonization statements. This comprised a wide range of variables

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

including demographics, comorbidities, medications, physical findings, test results, survey instruments, and clinical outcomes. Harmonized baseline characteristics are provided in **Table S1**. We implemented and tested ADAPT in multiple software environments to produce different degrees of computational performance, user accessibility, expressivity, and scalability. Listed roughly in order of maximal expressivity to maximal scalability, we used the R statistical package, local and cloud-based Structured Query Language (SQL) relational databases (MySQL, PostgreSQL), and the cloud-based Google BigQuery. Code for implementation of ADAPT in these systems is available from the authors on request. We have demonstrated the ability to implement ADAPT in web-based workbook environments including the American Heart Association's Precision Medicine Platform and the University of Colorado Health Data Compass Eureka environment. The latter enables access to deidentified EHR data for our health system using the Observational Medical Outcomes Partnership (OMOP) Common Data Model.

Example 2: Administrative hospital discharge data

After establishing key workflow components and testing harmonization pipeline integrity on BioLINCC data, we migrated the same data schema, workflow, and code to Google BigQuery to assess scalability. We used ADAPT to harmonize and analyze statewide administrative hospital discharge from data (**Table 4**, **Supplement Table S2**) from 20 states. Slight modifications were incorporated to accommodate unique features of administrative hospital discharge records (e.g. diagnosis present on admission, designation of primary diagnosis or procedure). Harmonization specifications were derived using a combination of state-specific documentation, standardized case report forms, and International Classification of Disease (ICD) version 9 and 10 codes.[15,16] To illustrate adaptability to independently derived harmonization schema, we also harmonized

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

diagnosis data to phecodes, [17,18] which are curated, standardized sets of ICD codes used in phenome-wide association analysis (PheWAS). There were total of ~9.0 billion demographic, administrative, and outcomes-oriented primary data points ~2.1 billion diagnosis ICD codes, ~461.7 million procedure ICD codes, and ~45 million external injury codes in the primary raw dataset. We then used > 15,800 demographic/administrative/outcomes criteria, > 7,900 diagnosis codes (ICD-9 and -10), and > 84,500 phecode specifications to produce a harmonized dataset comprising 8.1 billion administrative data points, 1.1 billion diagnoses, 109 million procedures, 45.1 million external injuries, and 1.2 billion phecodes (total ~10.5 billion datapoints).

Clinical research productivity (**Supplemental Table S3a-b**)

Using these data, we have published 16 peer-reviewed manuscripts (46 distinct coauthors and presented 49 abstracts at national and international conferences (55 distinct coauthors). We have found these data to be particularly helpful for trainees. Among 27 young investigators using these data, 23 presented abstracts at national or international conferences and 13 produced at least 1 peer-reviewed publication. Topics have ranged from association between coffee intake and heart failure risk to impact of climate and air quality on heart failure, and in-hospital outcomes of necrotizing fasciitis. We have also used ADAPT to create interactive R Shiny-based worksheets for self-directed exploration and cluster analysis of HF clinical trial data, a project which was chosen as a Selected Solution for the NHLBI Big Data Analysis Heart Failure Challenge. [19]

Portability

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Two groups of collaborators have successfully used ADAPT to harmonize clinical trial data independently from the development team. First, HF researchers at the University of Glasgow have harmonized data from industry-sponsored clinical trials, which we have merged with BioLINCC data. Second, we collaborated with the AMP HF/HeartShare network to create[20] a large harmonized dataset for use by independent researchers comprised of all publicly available HF trials from BioLINCC HF trials and several observational cohorts. This helped achieve an important part of the network's mission to democratize publicly available clinical trial data. Harmonized data are now published on BioData Catalyst (https://www.nhlbi.nih.gov/science/biodata-catalyst).

DISCUSSION

Using a variety of widely-used free and low-cost software platforms, we created a flexible, scalable, and transparent system for encoding, applying, and retaining harmonization criteria for complex integration of clinical research data. In order to make the platform useful to a wide range of investigators and analytic needs, we chose to support multiple harmonization definitions for any given concept rather than limiting mappings to singular definitions of complex traits such as HF. We focused instead on efficient capture and representation of harmonization criteria based on the categories of harmonization (e.g. original value, exact matching, scalar comparison). ADAPT can support consensus disease definitions as easily as project-specific criteria, even within a single project. The explicit transparency achieved by using a familiar spreadsheet format enables investigators to easily find and compare harmonization criteria between projects and to identify easily the raw datum used for each harmonized datapoint. Maintenance of the harmonization repository is simplified by using the same physical files for

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

encoding, querying, execution, and sharing. ADAPT can be deployed locally or in the cloud, and we have found this architecture to be scalable up to tens of billions of datapoints and thousands of harmonization rules by using high performance cloud-based data stores. We have found ADAPT to be a broadly effective tool for our work in a wide range of research applications across a several types of datasets, and it has been used by other research groups for similar tasks. We believe that systems such as ADAPT will be pivotal to realizing the research potential of the ever-increasing number of clinical and translational datasets.

Rapid accumulation of clinical and research datasets with a focus on data sharing democratization has prompted increasing discussion of strategies for harmonizing structured data.[1,2,4–6,21] Recommendations often include standardized data collection templates and terminologies, common data elements and data models, and validation of harmonized data. While these recommendations are central to improving data interoperability, comparability, and efficiency of data sharing, practical guidance and infrastructure regarding implementation of these principles are lacking. ADAPT represents a concrete step towards implementing those community recommendations, in part by focusing on the technical workflows to integrate data from disparate sources rather than curation of specific biomedical concepts.

An important decision in design of a practical harmonization framework is whether to use a reductionist philosophy, wherein the goal is to establish a singular definition for each concept to be used by all investigators. Such an approach provides high confidence in comparability of data from multiple sources. However, it can be difficult to establish and confidently apply formal, specific definitions to a large number of datapoints in diverse contexts from different sources,

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

restricting the amount of data that may be aggregated and the purposes for which they may be used. Alternatively, the harmonization framework may focus on maximizing expressivity in defining harmonization terms without requiring a ground-truth definition for each concept. In practice, this approach is used frequently in both research and clinical care, albeit often in an ad hoc manner. For example, the left ventricular ejection fraction (LVEF) is a critical measurement in HF research and can be measured by echocardiogram, computed tomography, or magnetic resonance imaging, each with different strengths and weaknesses. Precision and accuracy vary between modalities, and the reported LVEF is usually not numerically identical between modalities. However, classification by LVEF ranges of $< 40\%$ (HF with reduced ejection fraction or HFrEF), 40-49% (HF with mildly reduced ejection fraction or HFmrEF) and $\geq 50\%$ (HF with preserved ejection fraction or HFpEF) are often sufficient for an analysis. In such cases, the advantages of comparing less precise LVEFs over a larger population may be larger than the impact of inter-modality variation. However, other scenarios which focus on quantitative correlation between LVEF and, for example, exposure, may require higher precision. These sorts of decisions are quite common in clinical medicine, but are often implicit in statistical analysis. With ADAPT, it is easy to implement explicitly the decision to use imaging modalities interchangeably if appropriate for a given analysis or to be more restrictive for others. Furthermore, it is easy for a subsequent user (e.g. researcher or clinician) to ascertain and evaluate this choice for their needs.

We elected to develop a flexible framework because a) it can still be used to implement fundamental or consensus concept definitions if desired and b) it can be adapted for different communities and requirements to harmonize their data. Importantly, we believe that this strategy

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

is only possible with a very high degree of transparency into harmonization decisions, a quality that is often lost with manual or project-specific harmonization, for example in data collection or statistical code. Full transparency allows investigators to a) evaluate the credibility of prior decisions, b) assess comparability of harmonized datasets and generalizability of related findings, and c) determine the relevance of existing libraries of criteria to new projects and whether the prior content may be reused. Flexibility in harmonization parameters is important to maximize utility for research, but we believe a robust harmonization framework must also support integration with community-oriented resources such as standardized definitions and controlled terminologies. After using our framework in a number of settings, we have concluded that the FAIR principles can be applied to harmonization processes themselves in addition to primary data and their metadata. Using FAIR principles in harmonization projects would facilitate collaboration, support robust validation, and increase the reusability of any given set of harmonization specifications. We propose that FAIR principles may be applied to data harmonization as follows using ADAPT as an example:

Findability

Users may wish to identify harmonization criteria by the clinical concept used (e.g. LVEF), the primary data element (e.g. lvef_baseline), the study (e.g. Framingham Heart Study Original Cohort) or the harmonized value (e.g. "< 40%"). ADAPT captures user-designated harmonization criteria in a generic tabular format (CSV), which can be explored using any spreadsheet application and all statistical packages (**Figure 3**). Because the statistical code contains no criteria directly but rather extracts and applies the criteria from the primary harmonization dictionary, reviewing primary statistical code to recreate harmonization decisions

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

is rendered unnecessary. Use of a familiar user interface enables non-coding subject matter experts to participate directly in the harmonization process and makes available features such as copying/pasting, searching, sorting and filtering to manipulate the research criteria. Differences between concept definitions (e.g. WHO vs. ADA definition of diabetes vs. self-reported history) can be compared given that the harmonization criteria are clearly and explicitly stated. Additionally, use of the same file for both creating criteria, accessing them subsequently AND applying criteria to the data of interest reduces the need for maintenance and dramatically reduces the possibility of unexpected discordant harmonization criteria.

Accessibility

The highly readable and portable harmonization dictionary format used by ADAPT improves researchers' ability to identify and reuse harmonization schemata when trying to validate findings or reuse definitions in independent datasets. Large data aggregation projects have historically required considerable technical proficiency in statistical programming, which frequently limits the participation of clinical or translational subject matter experts who lack those skills. ADAPT allows clear and unambiguous examination of these harmonization decisions. Consequently, investigators with the most domain knowledge often have the least access to creating and reviewing mapping criteria. Because it is unnecessary to review the statistical code the simplicity and familiarity of the spreadsheet interface can enable investigators and subject matter experts to select appropriate harmonization criteria from prior projects or if none exist, encode new criteria without statistical programming ability. Criteria can also be revised and directly applied to data by the same investigator iteratively (e.g. quantile sizes).

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

These advantages therefore enable investigators who could not previously directly interact with the harmonization process to evaluate existing rules or specify new ones fully if they wish.

Interoperability

ADAPT is built upon common functionalities present in a variety of analytic and programming environments. The harmonization criteria components are stored in comma separated value text format, making the core content usable in any computing environment. It is operating-system agnostic and was developed entirely using free software, although it can also be implemented in commercial environments (e.g. Stata, BigQuery). It can be run on local hardware, on-premises servers, or in cloud-based resources. This provides access to ADAPT in most settings upon translation of the core algorithmic routines which we already provide for R, Stata, SQL, and BigQuery. This enables robust reuse and reapplication to independent datasets as well as the ability to critically evaluate published research and the definitions utilized in any given analysis that uses ADAPT. Finally, the ability to collate data sources without limit within a given harmonization specification allows easy comparison of fields from the data source, and transformation across many studies. The ability to understand the details and specificity of definitions can greatly improve (or reduce) confidence in veracity and generalizability of findings. Furthermore, the ability to implement ADAPT quickly in a new environment simplifies harmonization and sharing where raw data is subject to access restrictions.

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Reusability

Reusability is facilitated by the streamlined, clinical domain-agnostic structure which can be applied to new electronic clinical data requiring minimal programming background for new harmonization projects. Reliance on a simple tab-delimited format to transfer harmonization specification reduces the time required to ingest new datasets for a collaborative project. Additionally, the explicit knowledge of the harmonization decisions allows statistical analysis code to be run on many datasets with little customization. The ability to rapidly inspect and select harmonization rules from prior work promotes consistency through time-savings for subsequent investigators. In other words, it is often easier to use an existing representation of a condition than to create a new one. As a library of harmonization rules grows, the time needed to incorporate new datasets will decrease.

Limitations

A key limitation of ADAPT presently is the inability to express complex mathematical transformations. This is often a core requirement of harmonizing certain datasets, but implementing an integrated mathematical equation platform was beyond the scope of this effort. Currently, complex formulas can be applied prior to harmonization or can be executed using harmonized data, which still simplifies the analysis process. However, a more expressive computation would be an important incremental feature. Although we have made the harmonization activity quite streamlined by reducing code requirements, the process does require manual review of data dictionaries to determine whether variables of interest exist. Automated abstraction of data dictionaries and coding manuals could help make that review more efficient, and ADAPT would allow investigators to express rules of interest. Although we have created

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

thousands of rules for hundreds of clinical concepts, the focus of these rules is still our use case domain of HF. Several of the advantages of ADAPT such as reusability and interoperability would grow rapidly with implementation of common clinical concepts that could then be reused in other fields of study. Ideally, harmonization libraries will be created by subject matter experts in different clinical domains to optimize quality and utility, but it could take some time to achieve widespread use through reusability.

Future directions

The use cases described here harmonized structured data from clinical studies or systematic administrative data. One of the largest barriers to model testing and implementation in clinical care is the difference in quality and veracity between research or administrative data and a primary electronic health record. We believe ADAPT has potential utility in translating certain computational tools developed in research data to active clinical data. We planned for this by establishing the ability to use ADAPT on data stored in the OMOP common data model, which is a widely used clinical data warehouse framework. Although not real-time, OMOP-formatted data nevertheless demonstrates many of the weaknesses of real-world clinical data. We are currently exploring use of ADAPT in real-world data and to determine if and how it could be promoted to a real-time framework.

We believe that a growing user community may ultimately improve standardization, scope, and quality of harmonization libraries. There are no specific constraints on harmonization rules currently, and any user can implement any rule they require. However, it will likely prove easier for researchers to leverage definitions for common features that have already been encoded.

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Since the foundation of those definitions is easy to interrogate, we believe that commonly used concepts will ultimately attain relatively stable definitions based on the viewpoints of the larger community. Widespread reuse of specific rules across many users would effectively result in standardization within harmonization projects, and ultimately a formal strategy for designating standardization could prove useful.

CONCLUSION

Validation, sharing, and scaling applications using data from disparate sources requires confidence in comparability of data between sources. The almost limitless variety of uses for these data means that singular definitions for any one concept are unlikely to meet all users' needs. We have created and demonstrated application of a streamlined, platform-agnostic pipeline that enables rapid, explicit, and transparent definition of harmonization decisions that, using cloud computing platforms, can be scaled at minimum to the order of billions of data points. Using tools like ADAPT, we believe that applying the concepts of Findability, Accessibility, Interoperability, and Reusability to harmonization operations in addition to the data sources will ultimately improve confidence in and performance of algorithms based on aggregated and independent clinical data.

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

ACKNOWLEDGEMENTS

The authors wish to thank Martin O'Connor, Mark Musen, John Graybeal, Mike Bristow, John McMurray, James Lewsey, and JoAnn Lindenfeld for their advice and support.

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

REFERENCES

1. Pan K, Bazzano LA, Betha K, et al. Large-Scale Data Harmonization Across Prospective Studies. *Am J Epidemiol*. 2023;192:2033–2049.

2. Knosp B, Craven CK, Paglione L, Rojevsky S, Schneider J, Wilcox A. Understanding Data Harmonization. In: *Informatics Playbook*. Center for Data 2 Health, 2025.

3. Clinical Research Data Sharing Alliance. Standard for Secondary Analysis of Clinical Study Data. *CRDSA Standard 2001*. 2024.

4. Cheng C, Messerschmidt L, Bravo I, et al. A General Primer for Data Harmonization. *Sci Data*. 2024;11.

5. Almeida JR, Silva LB, Bos I, Visser PJ, Oliveira JL. A methodology for cohort harmonisation in multicentre clinical research. *Inform Med Unlocked*. 2021;27.

6. Stratford JK, Pan HH, Mainor A, et al. Best practices for clinical trials data harmonization and sharing on NHLBI bioData catalyst (BDC) learned from CONNECTS network COVID-19 studies. *J Clin Transl Sci*. 2025;9.

7. Francis GS, Cogswell R, Thenappan T. The heterogeneity of heart failure: Will enhanced phenotyping be necessary for future clinical trial success? *J Am Coll Cardiol*. 2014;64:1775–1776.

8. De Keulenaer GW, Brutsaert DL. The heart failure spectrum: time for a phenotype-oriented approach. *Circulation*. 2009;119:3044–3046.

9. Cahan A, Cimino JJ. Improving precision medicine using individual patient data from trials. *CMAJ*. 2017;189:E204–E207.

10. Wilkinson MD, Dumontier M, Aalbersberg IjJ, et al. Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3.

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

11. National Heart L and BI. Biologic Specimens and Data Repository Information Coordinating Center (BioLINCC) home page. Accessed October 9, 2025. https://biolincc.nhlbi.nih.gov/home/.

12. O'connor MJ, Shankar RD, Parrish DB, Das AK. *Knowledge-Level Querying of Temporal Patterns in Clinical Research Systems Knowledge and database disconnect*. IOS Press, 2007.

13. Musen MA. The Protégé Project. *AI Matters*. 2015;1:4–12.

14. Healthcare Cost and Utilization Project. Statewide Inpatient Databases (SID). 2015.

15. Brämer GR. International statistical classification of diseases and related health problems. Tenth revision. *World Health Stat Q*. 1988;41:32–6.

16. Based on the Recommendation of the Ninth Revision Conference 1975. Manual of the international statistical classification of diseases, injuries, and causes of death. In: World Health Organization, 1977.

17. Wei WQ, Bastarache LA, Carroll RJ, et al. Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. *PLoS One*. 2017;12.

18. Wu P, Gifford A, Meng X, et al. Mapping ICD-10 and ICD-10-CM Codes to phecodes: Workflow development and initial evaluation. *JMIR Med Inform*. 2019;7.

19. Kao D, Jhund P, Hall J. LEveraging Aggregated Data From Open Sources in Acute Heart Failure (LEAD-AHF). *Heart Failure Challenge: Selected Solutions | NHLBI, NIH*. 2020. Accessed August 17, 2025. https://www.nhlbi.nih.gov/grants-and-training/funding-opportunities-and-contacts/NHLBI-heart-failure-data-challenge/selected-solutions.

20. HeartShare. Data Sharing . *The Accelerating Medicines Partnership (AMP) Heart Failure (HF) Program* Accessed November 9, 2025. https://amphf.org/for-scientists/data-sharing/.

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

21. Nan Y, Del Ser J, Walsh S, et al. Data harmonisation for information fusion in digital healthcare: A state-of-the-art systematic review, meta-analysis and future research directions. *Information Fusion*. 2022;82:99–122.

Table 1 – Example code for exact string match harmonization

R

```
harmonized_data <-
unique(
  merge(
   alldata[,
        c("patientid",
         "study",
         "visit_yr",
         "form",
         "field",
         "value,
         "datapoint")],
   criteria_match[,
        c("study",
         "form",
         "field",
         "phenotype",
         "match_string",
         "harmonized_val_numeric",
         "harmonized_val_string",
         "rule_number")],
   by.x=c("study",
          "form",
          "field",
          "value"),
   by.y=c("study",
          "form",
          "field",
          "match_string")]
```

BigQuery

```
create or replace table harmonized_data
as (
  select distinct
    study,
    patientid,
    visit_yr,
    phenotype,
    harmonized_val_numeric
    harmonized_val_string,
    rule_num,
    datapoint,
from adapt.alldata as d
join adapt.criteria_match as c
     on (d.study = c.study,
          d.form = c.form,
        d.field = c.field,
        d.value = c.match_string)
```

MySQL (variant legacy SQL)

```
insert into harmonized_data
select distinct
  d.study,
  d.patientid,
  d.visit_yr,
  c.phenotype,
  c.harmonized_val_numeric
  c.harmonized_val_string,
  c.rule_num,
  d.datapoint,
from criteria_match as c,
join alldata as d
  using (field)
where and c.study = d.study
  and c.form = d.form
  and c.field = d.field
  and c.match_string = d.value
```

Table 2 – Example harmonization criteria table

Study	Form	Field	Concept	Lower bound	Upper bound	Match string	Sex-specific?	Harmonized value	Harmonized string	Rule type	Criteria source
DOSE	a_visitsumm	LL_HMG	Anemia	13			Male	1	No	range	WHO Iron Deficiency Anemia Programme
DOSE	a_visitsumm	LL_HMG	Anemia	12			Female	1	No	range	WHO Iron Deficiency Anemia Programme
DOSE	a_visitsumm	LL_HMG	Anemia		13		Male	2	Yes	range	WHO Iron Deficiency Anemia Programme
DOSE	a_visitsumm	LL_HMG	Anemia		12		Female	2	Yes	range	WHO Iron Deficiency Anemia Programme
ESCAPE	lab	3	Anemia	41			Male	1	No	range	WHO Iron Deficiency Anemia Programme
ESCAPE	lab	3	Anemia	36			Female	1	No	range	WHO Iron Deficiency Anemia Programme
ESCAPE	lab	3	Anemia		41		Male	2	Yes	range	WHO Iron Deficiency Anemia Programme
ESCAPE	lab	3	Anemia		36		Female	2	Yes	range	WHO Iron Deficiency Anemia Programme
SCD-HeFT	baseline_new	gender	Sex			1		1	Male	match	
SCD-HeFT	baseline_new	gender	Sex			2		2	Female	match	
SOLVD	SEF	SEF9	Sex			M		1	Male	match	
SOLVD	SEF	SEF9	Sex			F		2	Female	match	

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Table 3 – Summary of data from BioLINCC HF studies harmonized using ADAPT

Trial	Subjects	Study fields	Harmonized concepts	Time points	Raw data processed	Harmonized observations
ATHENA	360	134	102	7	114,969	69,181
BEST	2708	88	107	10	1,606,613	693,385
CARRESS	188	113	129	10	60,897	59,416
DIG	7788	48	66	4	533,021	379,327
DOSE-AHF	308	110	124	9	88,770	90,963
ESCAPE	433	146	175	13	476,688	203,892
EXACT-HF	253	160	24	4	137,445	83,050
FIGHT	300	171	188	4	220,931	99,188
GUIDE-IT	894	107	182	11	796,552	352,177
HF-ACTION	2130	123	122	8	1,451,701	487,769
INDIE	105	131	144	5	55,935	27,552
IRONOUT	225	116	146	4	94,449	50,703
LIFE	335	107	113	10	313,067	135,275
MOCHA	345	66	112	7	224,271	94,059
NEAT-HFpEF	110	153	171	3	60,677	25,639
RELAX	216	156	185	3	103,487	50,724
ROSE-AHF	360	108	116	5	134,248	93,325
SCD-HEFT	2521	123	96	23	4,822,498	902,582
SOLVD	6797	86	79	18	5,885,741	2,122,563
STICH	2621	143	175	22	2,361,186	834,017
TOPCAT	3445	128	167	16	3,086,502	1,660,995
Total	32442	2517	3252		22,629,648	8,515,782
ARIC	15004	477	299	28	4,424,380	7,149,645
CHS	5888	1841	249	23	30,613,027	5,040,353
FHS	15126	1423	288	36	50,334,253	5,191,332
JHS	3883	281	216	18	6,670,066	1,164,141
MESA	6814	515	207	5	25,416,682	3,293,729
SOLVD Registry	6270	74	104	1	771,044	392,335
Total	52985	4537	524	48	118,229,452	22,231,535
Combined	85427	7519	692		140,859,100	30,747,317

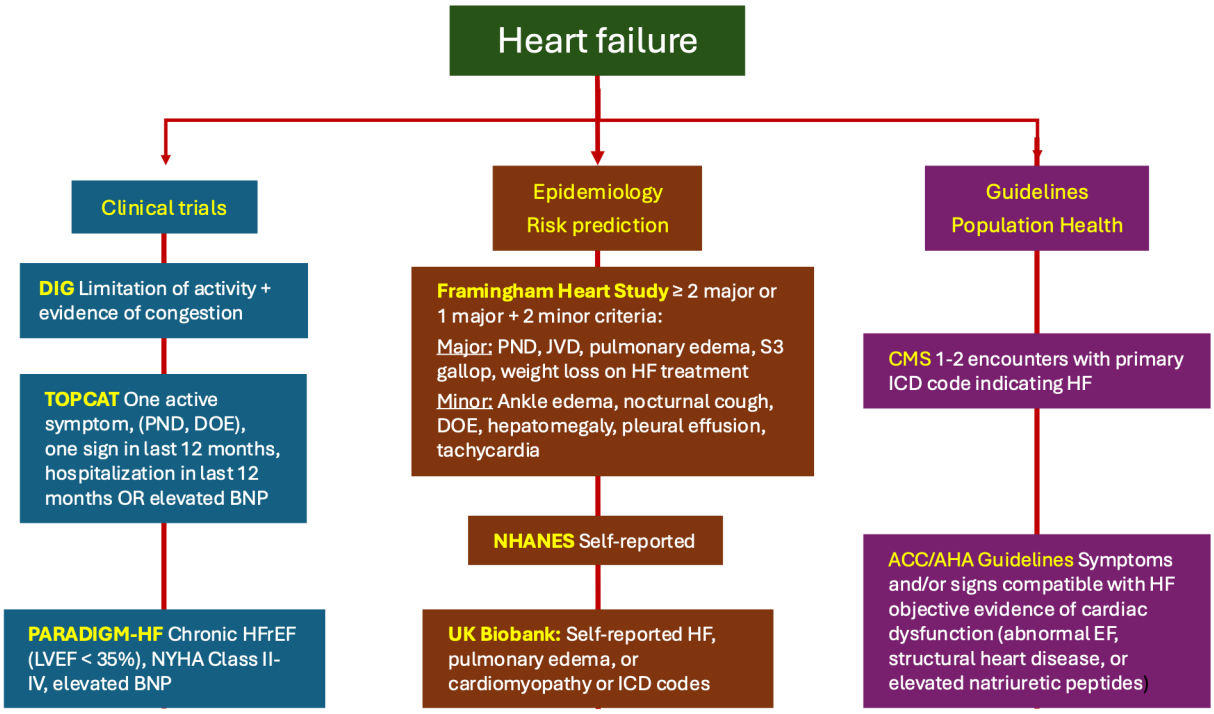
medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Table 4 – Administrative hospital discharge data

	Core	Diagnosis	Procedure	Phecodes
Rules	15,641	7942	954	84532
Raw data	8,742,661,928	2,033,064,349	442,330,213	2,033,064,349
Harmonized data	7,887,443,398	2,033,064,349	954	1,192,264,729
Time elapsed, sec	106	17	6	25

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 1 – Divergent, context diagnostic criteria for heart failure



medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 2 – Conceptual ADAPT harmonization framework

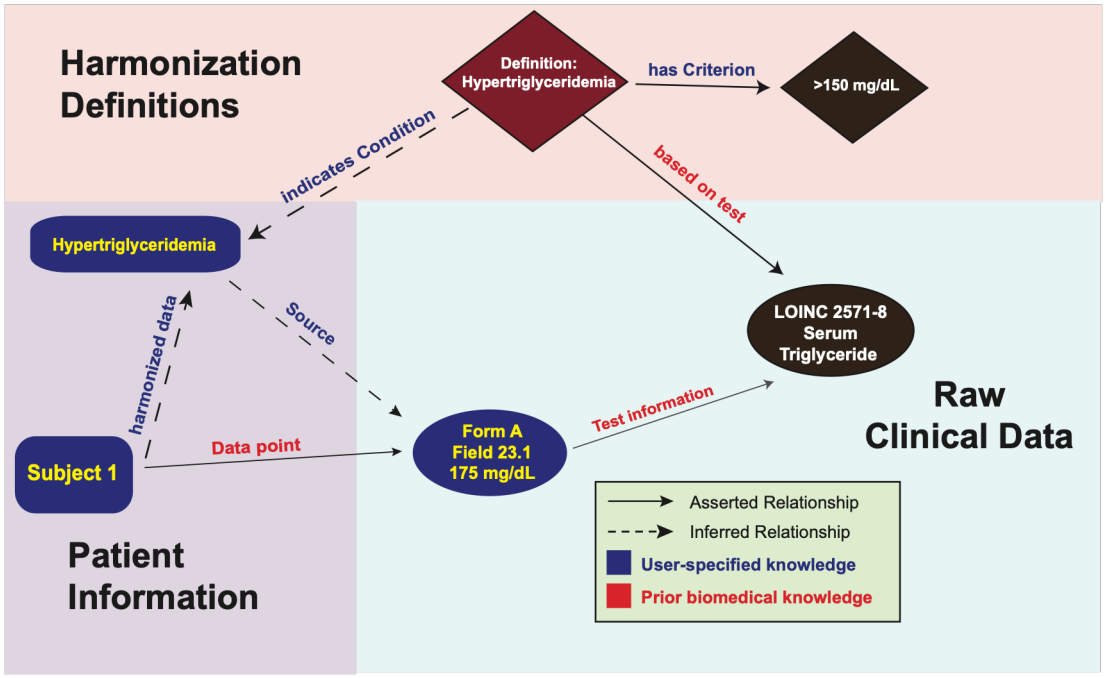


Figure 3 ADAPT harmonization pipeline components

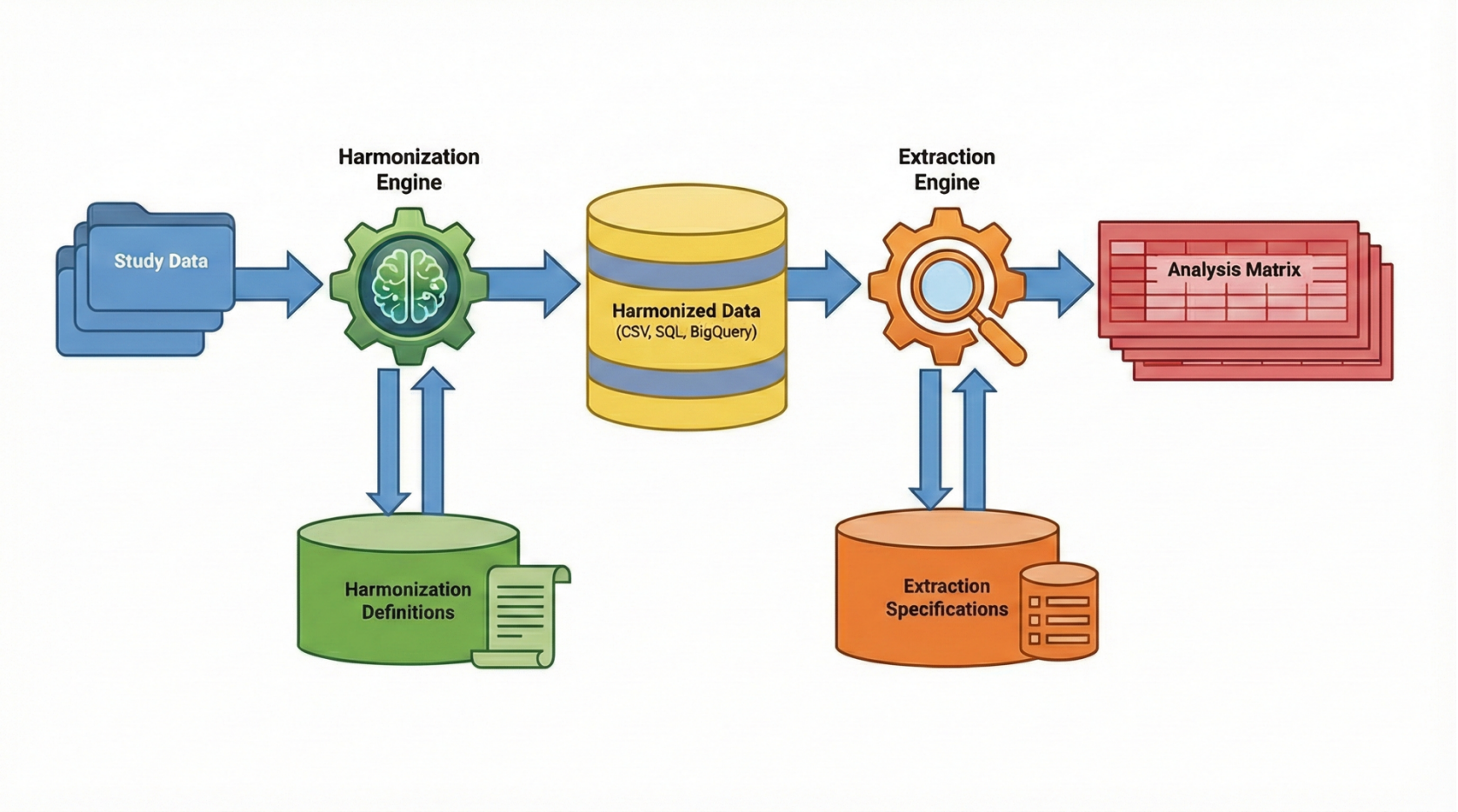


Figure 4: Application of ADAPT harmonization criteria

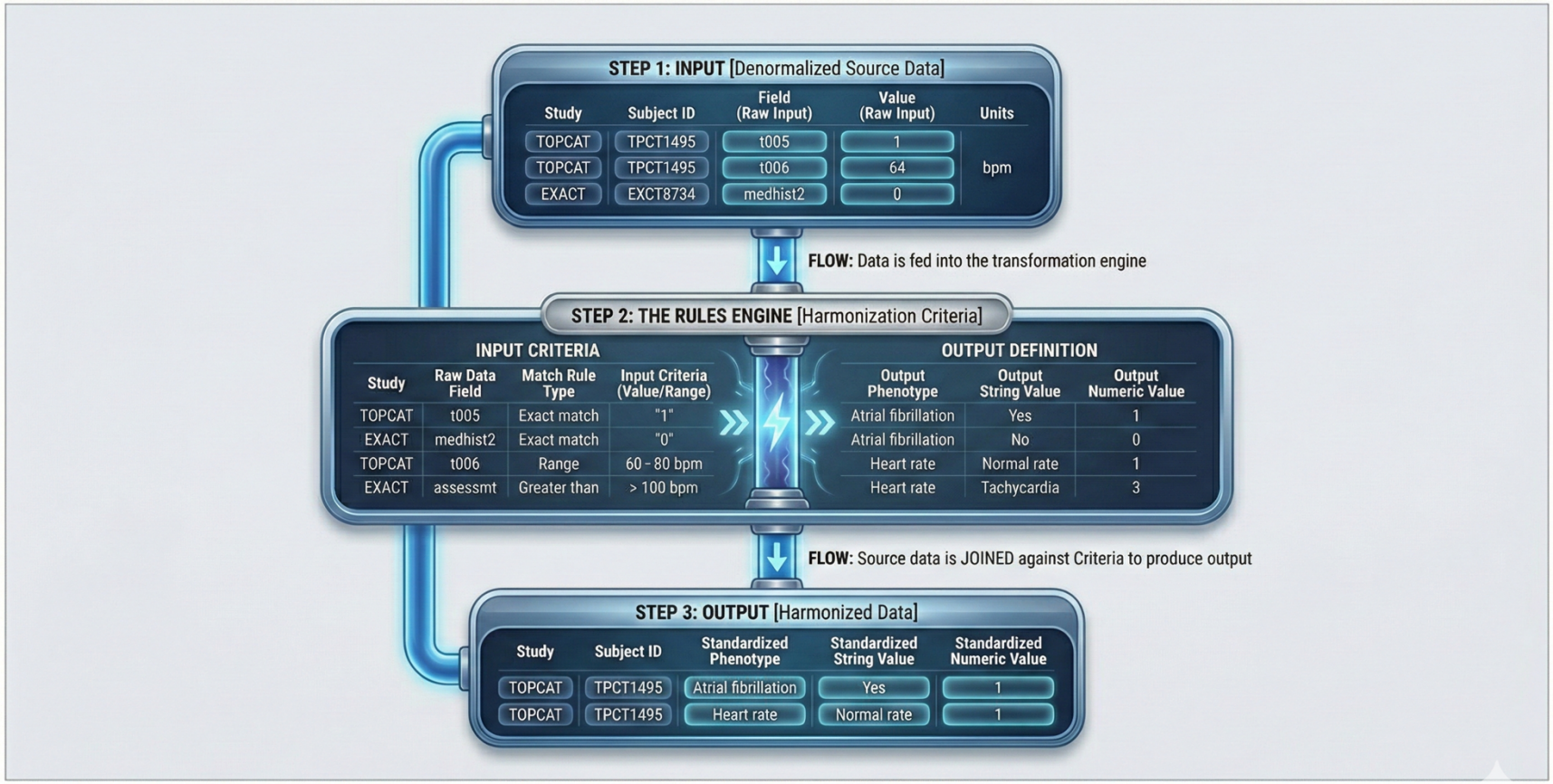


Table S1: Baseline characteristics of BioLINCC heart failure trials using harmonized dataset

Characteristic	BEST (N=2708)	DIG (N=7788)	DOSE (N=308)	ESCAPE (N=433)	EXACT (N=253)	GUIDE-IT (N=894)	HF-ACTION (N=2130)	NEAT-HFpEF (N=110)	RELAX (N=216)	ROSE (N=360)	SCD-HeFT (N=2521)	SOLVD (N=2568)	STICH (N=2621)	TOPCAT (N=1767)
HF category														
HFpEF	1 (0)	614 (8)	71 (24)	0 (0)	0 (0)	0 (0)	3 (0)	106 (96)	207 (96)	94 (26)	0 (0)	0 (0)	30 (1)	1364 (77)
HFmrEF	0 (0)	1027 (13)	30 (10)	3 (1)	0 (0)	0 (0)	57 (3)	3 (3)	9 (4)	42 (12)	2 (0)	0 (0)	152 (7)	402 (23)
HFrEF	2707 (100)	6147 (79)	199 (66)	419 (99)	253 (100)	847 (100)	2066 (97)	1 (1)	0 (0)	222 (62)	2518 (100)	2567 (100)	1954 (91)	0 (0)
EF, %	23.0±7.3	32.0±12.5	34.6±17.5	19.3±6.6	24.5±8.2	24.3±8.3	25.2±7.5	60.5±6.5	61.3±6.8	36.5±16.9	23.8±6.9	24.9±6.7	27.9±8.9	58.2±7.8
Age, yrs	60.3±12.4	63.910.9	66.0±13.6	56.1±13.9	61.9±13.6	61.5±13.9	58.6±12.7	69.3±9.2	68.5±10.3	69.5±12.2	59.0±11.9	60.4± 9.9	61.6± 9.8	71.5± 9.7
Female	593 (22)	1926 (25)	82 (27)	112 (26)	45 (18)	286 (32)	599 (28)	63 (57)	104 (48)	96 (27)	588 (23)	504 (20)	381 (15)	882 (50)
Race														
White	1896 (70)	6660 (86)	222 (72)	258 (60)	169 (67)	459 (52)	1322 (62)	98 (89)	197 (91)	272 (76)	1932 (77)	2060 (80)	2019 (77)	1384 (79)
Black	627 (23)	1128 (14)	76 (25)	119 (27)	74 (29)	318 (36)	669 (31)	9 (8)	8 (4)	74 (21)	425 (17)	396 (15)	75 (3)	302 (17)
Hispanic	143 (5)	0 (0)	2 (1)	46 (11)	0 (0)	58 (7)	0 (0)	3 (3)	0 (0)	3 (1)	111 (4)	0 (0)	114 (4)	50 (3)
Asian	21 (1)	0 (0)	0 (0)	0 (0)	0 (0)	27 (3)	107 (5)	0 (0)	0 (0)	0 (0)	30 (1)	0 (0)	244 (9)	19 (1)
Other	21 (1)	0 (0)	8 (3)	10 (2)	10 (4)	26 (3)	32 (2)	0 (0)	11 (5)	9 (3)	23 (1)	112 (4)	165 (6)	4 (0)
DM	964 (36)	2218 (28)	158 (51)	140 (32)	138 (55)	410 (46)	678 (32)	43 (39)	93 (43)	200 (56)	767 (30)	663 (26)	996 (38)	788 (45)
HTN	1596 (59)	3674 (47)	246 (80)	203 (47)	198 (78)	706 (79)	1260 (59)	99 (90)	183 (85)	298 (83)	1400 (56)	1085 (42)	1606 (61)	1588 (90)
Medications														
β-blocker	1354 (50)		256 (83)	269 (62)	241 (95)	845 (95)	2012 (94)	77 (70)	164 (76)	267 (74)	1738 (69)	197 (8)	1826 (70)	378 (21)
ACE-I/ARB	2603 (96)	7274 (93)	197 (64)	391 (90)	214 (85)	708 (79)	2011 (94)	70 (64)	152 (70)	207 (57)	2432 (96)	1285 (50)	1891 (72)	1395 (79)
MRA	91 (3)		86 (28)	186 (43)	133 (53)	444 (50)	962 (45)	0 (0)	23 (11)	80 (22)	484 (19)			886 (50)
Measurements														
HR, bpm	82±13	78±13	76±13	82±16	73±12	77±15	71±11	71±12	70±12	75±14	75±14	80±13	74±14	69±11
SBP, mm Hg	118±19	127±20	116±20	106±16	110±16	116±20	114±18	130±16	128±17	111±19	120±19	125±18	121±18	128±16
est. GFR	65±23	59±22	54±23	57±25	56±22	57±24	68±24	59±20	60±22	43±15	88±39	63±18	72±23	64±22

Table S2: Number of hospital records harmonized using ADAPT.

State	Patients	Core features	Diagnoses	Procedures	External injury	Phecodes
AR	2006783	84824145	18696012	2871403	192426	15644266
AZ	12417678	516405842	98355450	17954835	2220299	59522124
CA	47442141	927070285	304470812	75515713	12287987	124448181
CO	5925870	172530761	55489321	9330887	518013	34749739
DE	185774	3113218	1256172	399863		679336
FL	49615882	1990739061	368463894	65467944	6205070	230879283
MD	2416671	97722600	32545439	4230288		29072073
NC	5584549	233300482	69878275	9199740	411734	57202065
NH	2485046	47429817	17310824	4134940		10904724
NJ	20860903	700420692	106318507	33127965	2075904	48177828
NV	3599047	32123985	36858299	5536887		26498627
NY	48998544	2003280807	305582190	102948036	5838256	183976066
OK	6370567	127103812	61038193	9949185	1364611	44452251

SD	532351	20606827	5599299	795664	35166	4752713
TX	60914776	1354721286	454771131	91085140	12001421	263410345
VT	1125189	18954228	10148233	3790222	440912	6537528
WA	10730635	468636862	78840071	15485316	1325273	49222827
WI	2978837	110523368	34520925	4334120	241392	29272565
WV	1321429	17172237	8188358	1569238		3898781
Total	285512672	8926680315	2068331405	457727386	45158464	1223301322

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Table S3a: Peer-reviewed publications produced using ADAPT-generated datasets:

1. Kao DP, Wagner BD, Robertson AD, Bristow MR, Lowes BD. A personalized BEST: Characterization of latent clinical classes of nonischemic heart failure that predict outcomes and response to bucindolol. PLOS One. 2012:7(11):e48184. PMC3492337. link
2. Kao DP, Hsich E, Lindenfeld J. Characteristics, adverse events, and racial differences among delivering mothers with peripartum cardiomyopathy. JACC:Heart Failure 2013; 1(5):409-416. PMC3806506. link
3. Katz DF, Turakhia MP, Sauer WH, Tzou WS, Heath RR, Zipse MM, Aleong RG, Varosy PD, Kao DP. Safety of ventricular tachycardia ablation in clinical practice: findings from 9,699 hospital discharge records. Circ Arrhythm Electrophysiol 2015;8(2):362-70. PMC9893573 link
4. Kao DP, Lewsey JD, Anand IS, Massie BM, Zile M, Carson PE, McKelvie R, Komajda M, McMurray JJV, Lindenfeld J. Characterization of subgroups of heart failure patients with preserved ejection fraction with possible implications for prognosis and treatment response. Eur J Heart Fail. 2015; 17(9):925-35 PMC4654630. link
5. McIlvennan CK, Lindenfeld J, Kao DP. Sex differences and in-hospital outcomes in patients undergoing mechanical circulatory support implantation. J Heart Lung Transplant. 2017;36(1):82-90. PMC7061938. link
6. Kao DP, Stevens LM, Hinterberg MA, Görg C. Phenotype-specific association of single nucleotide polymorphisms with heart failure and preserved ejection fraction: a genome-wide association analysis of the Cardiovascular Health Study. J Cardiovasc and Transl Res. 2017;10(3):285-294. PMC5894815. link
7. Johnson-Sasso CP, Tompkins, Kao DP, Walker LA. Marijuana use and short-term outcomes in patients hospitalized for acute myocardial infarction. PLoS One 2018;13(7):e0199705. PMC6040751. link
8. Jarvie JL, Metz TD, Davis MB, Ehrig JC, Kao DP. Short-term risk of cardiovascular readmission following a hypertensive disorder of pregnancy. Heart. 2018;104(14);1187-94. PMC7061962. link
9. Merrill M, Sweitzer NK, Lindenfeld J, Kao DP. Sex differences in outcomes and responses to spironolactone in heart failure with preserved ejection fraction – a secondary analysis of the TOPCAT trial. JACC: Heart Failure. 2019;7(3) 228-38. PMC6817333. link
10. Aleong RG, Tompkins C, Zipse MM, Aftab M, Varosy P, Sauer W, Kao DP. Analysis of outcomes in 8304 patients undergoing lead extraction for infection. J Am Heart Assoc. 2020;9(7):e011473. PMC7428595. link
11. Flint KM, Shah SJ, Lewis EF, Kao DP. Variation in clinical and patient-reported outcomes among complex HFpEF phenotypes. ESC Heart Fail. 2020;7(3):811-824. PMC7261552. link
12. Davis MB, Jarvie J, Gambahaya E, Lindenfeld J, Kao D. Risk prediction for peripartum cardiomyopathy in delivering mothers: a validated risk model. J Card Fail. 2020;S1071-9164(20)31574-3. link
13. Stevens LM, Linstead E, Hall JL, Kao DP. The association between coffee intake and incident heart failure risk – a machine learning analysis of the Framingham Heart,

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Atherosclerosis Risk in Communities, and Cardiovascular Health Studies. Circ Heart Fail. 2021:14(2):e006799. PMC8081115 link

14. Singh I, Duca L, Kao D, Chatfield K, Khanna A. Outcomes in hospitalizations of women with Turner Syndrome compared to women without Turner Syndrome. Cardiol Young. 2021;31(10):1667-74. link
15. Gupta P, Brinza E, Khazanie P, Peterson PN, Ho PM, Kao DP. Forecasting heart failure: seasonal alignment of heart failure outcomes in New York. ESC Heart Fail. 2024 [in press].

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Table S3b: Peer-reviewed abstracts made using ADAPT-generated datasets (n=49):

1. Kao DP, Kreso E, Fonarow G, Krantz M. Anemia, Kidney Disease, and Inpatient Outcomes for Heart Failure in California 2000-2006. 13th Annual Scientific Meeting of the Heart Failure Society of America, Sept 13-16, 2009; Boston, MA. J Card Fail 2009:15(6S):S102-3. link
2. Kao DP, Raff H, Peterson P. Gender Differences in Use of Implantable Cardiac Defibrillators and Resynchronization Therapy for Heart Failure in California, 2000-2006. 13th Annual Scientific Meeting of the Heart Failure Society of America, Sept 13-16, 2009; Boston, MA. J Card Fail 2009:15(6S):S79-80. link
3. Heath R, Kao DP, Belardi D, and Varosy PD. Mortality and Complication Rates of Catheter-based Ablation Procedures Are Lower at High Volume Centers. Heart Rhythm 2010;7(5, Supplement 1):PO1-56. link
4. Dattilo PB, Kao DP, Messenger JC, Casserly IP. Packed Red Blood Cell Transfusion is Associated with Adverse Outcomes in Patients Undergoing Endovascular Therapy of Peripheral Arterial Disease. Annual Scientific Session & Expo of Am Coll Cardiol, April 2-5, 2011, New Orleans, LA. J Am Coll Cardiol 2011;57(14S):E1493. link
5. Kao DP, Kreso E. Gender and Racial Differences in Demographics and Outcomes in 800 Inpatient Admissions for Takotsubo Cardiomyopathy. Annual Scientific Session & Expo of Am Coll Cardiol, April 2-5, 2011, New Orleans, LA. J Am Coll Cardiol 2011;57(14S):E264. link
6. Kao DP, Wagner BD, Roberston AD, Kittelson JM, Bristow MR, Lowes BD. High-dimensional clinical phenotype analysis predicts mortality and response to beta-blocker therapy in nonischemic heart failure. Annual Scientific Session & Expo of Am Coll Cardiol, April 2-5, 2011, New Orleans, LA. J Am Coll Cardiol. 2011;57(14S):E1266. link
7. Simmons RP and Kao DP. Characteristics and short-term outcomes of necrotising fasciitis. Infectious Disease Society of America Annual Meeting Oct 20-23, 2011, Boston, MA.
8. Kao DP, Hsich E, Lindenfeld J. Characteristics of 2067 patients hospitalized with peripartum cardiomyopathy. Annual Scientific Session & Expo of Am Coll Cardiol. J Am Coll Cardiol. 2012:59(13S):E1572 link
9. Kao DP, Lewsey JK, Massie BM, McMurray J, Carson PE, Anand IS, Lindenfeld J. Characterization of heart failure patients with preserved ejection fraction in the I-PRESERVE trial who have improved outcomes with irbesartan therapy. European Society of Cardiology Heart Failure Congress. Eur J Heart Fail. 2012;11(Suppl 1):61115. link
10. Kao DP, Varosy P, Aleong R. Impact of payor status on presentation, therapies, and outcomes in 31,614 patients hospitalized with Wolff-Parkinson-White Syndrome. American Heart Association Annual Scientific Sessions. Circulation. 2012:126:A14800. link
11. Kao DP, Varosy P, Nguyen DT, Tzou W, Katz D, Schuller J, Sung R, Steckman D, Sauer WH, Aleong R. Temporal trends in mortality and adverse events in 24,890 patients undergoing lead extraction over a decade. American Heart Association Annual Scientific Sessions.Circulation 2013;128(Suppl 22):A16352. link

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

12. Kao DP, Varosy P, Nguyen DT, Tzou W, Katz D, Schuller J, Sung R, Steckman D, Sauer WH, Aleong R. Predictors of adverse events in 24,890 patients undergoing lead extraction. American Heart Association Annual Scientific Sessions, Nov. Circulation 2013;128(Suppl 22):A16362. link
13. McIlvennan CK, Lindenfeld J, Kao DP. Sex differences in characteristics and in-hospital outcomes in patients undergoing mechanical circulatory support implantation, 1995-2011. J Am Coll Cardiol 2014;63(12 Suppl):A814. link
14. Kao DP, Lindenfeld J. A risk score for predicting mortality in 10,582 patients hospitalized with takotsubo cardiomyopathy. American Heart Association Scientific Sessions. Circulation 2014;130(Suppl 2):A20617. link
15. Kao DP, Lindenfeld J. Gender and race differences in patient characteristics and outcomes in 10,582 hospitalizations for takotsubo cardiomyopathy. American Heart Association Scientific Sessions. Circulation 2014;130(Suppl 2):A20533. link
16. Tompkins CT, Varosy P, Tzou W, Aleong RG, Nguyen D, Katz DF, Schuller JL, Sung RK, Sauer WH, Kao DP. Gender differences in procedure-related complications following ablation of ventricular tachycardia. Heart Rhythm Society Scientific Sessions. Heart Rhythm 2015; 12(5 Suppl):PO05-143. link
17. Aleong RG, Tompkins C, Varosy PD, Katz DF, Schuller JL, Nguyen D, Tzou WS, Sung RK, Sauer WH, Kao DP. Risk score to predict adverse events in 57,220 patients undergoing lead extraction. Heart Rhythm Society Scientific Sessions. Heart Rhythm 2015;12(5 Suppl):AB08-02. link
18. Aleong RG, Tompkins C, Varosy PD, Tzou WS, Katz DF, Schuller JL, Sung RK, Nguyen D, Sauer WH, Kao DP. Analysis of outcomes in 8304 patients undergoing lead extraction for infection. Heart Rhythm Society Scientific Sessions. Heart Rhythm 2015;12(5 Suppl):AB08-04. link
19. Tompkins CT, Varosy P, Aleong RG, Sauer WH, Katz DF, Nguyen D, Schuller JL, Sung RK, Tzou W, Kao DP. Gender differences in procedure-related complications following lead extraction. Heart Rhythm Society Scientific Sessions, Boston, MA Heart Rhythm 2015:12(5 Suppl):PO06-33. link
20. Davis MB, Gambahaya E, Lindenfeld J, Kao DP. Prediction of peripartum cardiomyopathy at time of childbirth: a validation of associated characteristics among 3,474,622 delivering mothers. European Society of Cardiology Heart Failure Congress. Eur J Heart Fail. 2015:17(Suppl 1):P299. link
21. Kollengode MS, Kao DP, Khanna AD. Characteristics and in-hospital outcomes of adults undergoing Fontan operation: Insights from administrative data. 25th International Symposium on Adult Congenital Heart Disease. June 3-6, 2015. link
22. Kao DP, Katz DF, Tzou W. Trends, Predictors, and Regional Variation in Use of Intracardiac Ablation Among 184,443 Hospitalizations for Ventricular Tachycardia. American Heart Association Scientific Sessions. Circulation. 2015; 132(Suppl 3):A18718. link
23. McIlvennan CK, Kao DP. Characteristics and In-Hospital Outcomes of Adults Undergoing Extracorporeal Membrane Oxygenation. J Heart Lung Transp. 2016 Apr;35(4, suppl):S152–3. link
24. Johnson-Sasso CP, Kao D, Walker LA. Marijuana use and short-term outcomes in patients hospitalized for acute myocardial infarction. J Am Coll Card. 2016;67(13):569. link

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

25. Kao DP. Association of anemia and transfusion with adverse outcomes in 22,291 patients hospitalized with takotsubo cardiomyopathy. European Society of Cardiology Heart Failure Congress. Eur J Heart Fail 2016;18(Suppl 1):P927. link
26. Kao DP, Gambahaya ET. Race differences in timing of presentation in 1137 peripartum cardiomyopathy patients compared with 2.1 million delivering mothers without peripartum cardiomyopathy. European Society of Cardiology Heart Failure Congress. Eur J Heart Fail 2016;18(Suppl 1):P294. link
27. Stevens L, Hinterberg MA, Görg C, Kao DP. Shared single nucleotide polymorphism may explain the relationship between clusters of comorbid risk factors and incident heart failure in the Framingham Heart Study. American Heart Association Scientific Sessions. Circulation. 2016:134(Suppl 1):A18089. link
28. Wang ET, Kao DP. Hospitalized heart failure patients of Native American, Pacific Islander, Asian, Hispanic, and African American background have significantly different baseline characteristics and outcomes compared with white patients. American Heart Association Scientific Session. Circulation. 2016;134(Suppl 1):A18108. link
29. Stevens L, Hinterberg MA, Görg C, Kao DP. Characterization of molecular profiles in complex phenotypes associated with heart failure with preserved ejection fraction. Cardiovascular Genome-Phenome Study Awardee Evaluation, AHA Precision Medicine Institute. American Heart Association Scientific Sessions Nov. 2016.
30. Kao DP, Flint K, Merrill M. Latent class echocardiographic phenotypes predict adverse outcomes in heart failure with preserved ejection fraction – a secondary analysis of TOPCAT. European Society of Heart Failure Congress. Eur J Heart Fail. 2017:19(Suppl 1):503. link
31. Flint K, Merrill M, Kao DP. Sex differences in Long-Term Health Status in Heart Failure with Preserved Ejection Fraction. European Society of Heart Failure Congress. Eur J Heart Fail. 2017:19(Suppl 1):327. link
32. Kao DP, Flint K, Merrill M. Spironolactone reduces all-cause mortality in women but not men with heart failure with preserved ejection fraction enrolled in TOPCAT from the Americas. European Society of Heart Failure Congress, Paris, FR, May 2017. Eur J Heart Fail. 2017;19(Suppl 1):530. link
33. Stevens LM, Görg C, Kao DP. Coffee intake affects heart failure and stroke survival and is significant in predicting heart failure and stroke risk. American Heart Association Scientific Sessions. Circulation 2017;136(Suppl 1):A21081. link
34. Doran B, Roeder C, Merrill M, Jhund P, Stevens L, McMurray J, Görg C, Kao DP. Latent cluster analysis using clinical and functional characteristics identifies novel phenotypes of patients with reduced ejection fraction enrolled in the HF-ACTION trial with differential association with mortality. American Heart Association Scientific Sessions. Circulation 2018;138(Suppl 1):A12890. link
35. Merrill M, Roeder C, Butler M, Doran B, Stevens L, Görg C, Kao DP. Complex heart failure phenotypes differ in response to medical therapy and exercise training. American Heart Association Scientific Sessions. Circulation 2018;138(Suppl 1):A16910. link
36. Butler M, Roeder C, Doran B, Merrill M, Jhund P, Kao DP, Görg C, Stevens L. Complex heart failure phenotypes may occur across the spectrum of LVEF with

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

important implications for etiology and clinical outcomes. American Heart Association Scientific Sessions. Circulation 2018;138(Suppl 1):A13171. link

37. Purohit SN, Merrill M, Kao DP. Echocardiographic latent class analysis phenotypes predict adverse outcomes and treatment response in heart failure with preserved ejection fraction. American Heart Association Scientific Sessions. Circulation. 2018; 138(Suppl 1):A11813. link
38. Agrawal Vineet, Kao DP. Pulmonary hypertension and right ventricular dysfunction identify subgroups with HFpEF that benefit from aldosterone antagonist therapy. American Heart Association Scientific Sessions. Circulation. 2018;138(Suppl 1):A15911. link
39. Riordan M, Kao DP. Clinical phenotype differences in outcomes and responses in spironolactone in heart failure with preserved ejection fraction. Circulation. 2020;142:15229. link
40. Purohit SS, Jhund PS, Lindenfeld DP. Latent class analysis of echocardiographic phenotypes predicts adverse outcomes and treatment response in heart failure with Doran B, Leyba K, Kao DP. Marriage and the strong of heart: An investigation of the relationship between marital status and new-onset heart failure in the Multi-Ethnic Study of Atherosclerosis. European Society of Cardiology Heart Failure Congress. May 20-24, 2022. Eur J Heart Fail. 2022;24(S2):67. link
41. Leyba K, Doran B, Kao DP. Widowed status at heart failure onset is associated with mortality in elderly women but not men – a secondary analysis of the Cardiovascular Health Study. European Society of Cardiology Heart Failure Congress. May 20-24, 2022. Eur J Heart Fail. 2022;24(S2):67. link
42. Leyba K, Kao DP. Lifelong bachelor status is associated with increased mortality in men with heart failure - a secondary analysis of the Multi-Ethnic Study of Atherosclerosis. Journal of the American College of Cardiology. 2023;81(8, suppl A):307. link
43. Riordan M, Kao DP. Clinical and functional differences in heart failure phenotypes across the ejection fraction spectrum provides prognostic information. Journal of the American College of Cardiology. 2023;81(8, suppl A):589. link
44. Leyva K, Kao DP. Lifelong bachelor status is associated with increased mortality in heart failure: a secondary analysis of the Atherosclerosis Risk In Communities Study. European Society of Cardiology Heart Failure Association Congress. May 19-23, 2023, Eur J Heart Fail. 2023;25(Suppl. S2):244. link
45. Brinza EK, Kao DP. Comparative seasonal variation in outcomes among patients hospitalized for heart failure in cold- and warm-climate states. American Heart Association Scientific Sessions, Nov 11-13, 2023. Circulation. 2023;148(suppl_1):18447. link
46. Shalowitz E, Kao DP. Elevation of patients hospitalized with heart failure: the sky's the limit? European Society of Cardiology Heart Failure Association Congress. May 11-14, 2024. Eur J Heart Fail. 2024;26(Suppl S2):617. link
47. Leyba K, Roeder CR, Görg C, Jhund P, Kao DP. Latent class analysis reveals a novel phenotype of heart failure that spans the left ventricular ejection fraction. European Society of Cardiology Congress 2024 August 30, 2024. London, United Kingdom. link
48. Leyba K, Roeder CR, Görg C, Jhund P, Kao DP. Latent class analysis reveals a novel phenotype of heart failure that spans the left ventricular ejection fraction.

medRxiv preprint doi: https://doi.org/10.64898/2026.07.13.26357501; this version posted July 16, 2026. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

European Society of Cardiology Congress 2024 August 30, 2024. London, United Kingdom. link

49. Leyba K, Kao D. Relationship status is associated with sex differences in social and health-seeking behaviors among heart failure patients. American Heart Association Scientific Sessions. November 16-18, 2024 [Accepted for presentation]