Environmental influences on Child Health Outcomes (ECHO)

Data Quality Plan

Version 01.10

Date: 11Jan2023

Table of Contents

List of Tables and Figures

1 PURPOSE OF THE DATA QUALITY PLAN

This Data Quality Plan for the Environmental influences on Child Health Outcomes (ECHO) Data Analysis Center (DAC) documents the activities, processes and procedures employed to ensure that data on the ECHO-wide Cohort Data Platform are of high quality. This entails ensuring data are valid (accurately measuring what it is intended to measure), reliable (consistently produce a given result), and complete (non-missing data at item and measure level, timely submission of data, and complete follow-up). Although data quality is a primary aim of the DAC, it is the responsibility of all ECHO components to ensure data are of the highest caliber. The purpose of this plan is to:

- Define acceptable quality standards
- Specify quality control and quality assurance activities
- Define roles and responsibility for quality management
- Specify processes for documentation

Activities in ECHO to enhance the quality of data are described in the document according timing relative to data and specimen collection; those that are undertaken prior to collection, during collection and after collection. For example, development of protocol, forms and training are data assurance activities that are initiated prior to implementation of the protocol. Data systems (Participant Registration, Data Transform, REDCap Central, Bio-Track), review of biospecimens and assay results and monitoring are data quality activities while cohorts collect and share data and specimens, and data queries, administrative reports, implementation fidelity reviews are examples of post-collection data quality activities. We further provide information on measures to enhance the reliability and validity of data analyses such as harmonization, centrally deriving analytical variables, documentation, and training users of the ECHO-wide Analysis Workbench. This document conglomerates the various data quality undertakings for the ECHO-wide Cohort Study but is not intended to provide comprehensive information on each activity; information about where to find more details is provided.

2 CONTACT INFORMATION

Contact the ECHO Help Desk for all questions.

Help Desk: https://echoportal.org/Home/HelpDesk

Email: ECHO-DAC@rti.org

Phone: 877-225-0771

Hours: Mon-Fri, 8 AM – 5 PM ET

(Response within 1 business day)

3 ABREVIATIONS AND ACRONYMS

Table 1: Abbreviations and Definitions

Abbreviation or Acronym	Definition
AP	analysis proposals
CAT	Cohort Advocacy Team (ECHO CC and DAC)
CC	Coordinating Center (ECHO)

Abbreviation or Acronym	Definition
CDD	Custom Data Dictionary
CDM	common data model
CSV	comma separated values
DAC	Data Analysis Center (ECHO)
DCF	data collection form
EAW	ECHO Analysis Workbench
ECHO	Environmental influences on Child Health Outcomes
EWCP	ECHO-wide Cohort Data Collection Protocol
FISMA	Federal Information Security Management Act
GOIT	Goals, Objectives, Indicators, Targets
HHEAR	Human Health Exposure Analysis Resource
ID	participant identification number
IF	Implementation Fidelity subgroup
JHU	Johns Hopkins University, an awardee for the Data Analysis Center
IRB	Institutional Review Board
KB	Knowledge Base
LMS	Learning Management System
MIS	measurement information sheet
MOP	Manual of Operating Procedures
NIH	National Institutes of Health
NIH PO	NIH ECHO Program Office / Program Officer
NIH TB	NIH Toolbox
OIT	ECHO Operation and Implementation Team
PI	principal investigator
PIE	Protocol Implementation & Evaluation Committee
PII	personally identifiable information
PLATIPUS	Platform: Inform and Provide User Search
PRO Core	Person-Reported Outcomes Core (ECHO)
RTI	RTI International, an awardee for the Data Analysis Center
QA	quality assurance
QC	quality control
REDCap	Research Electronic Data Capture
SQL	Structured Query Language
SOP	standard operating procedure

4 DESCRIPTION OF THE PROJECT

A goal of the ECHO Program is to support the ECHO-wide Cohort, composed of multiple, synergistic, longitudinal cohort studies, to investigate the effect of a broad range of early environmental exposures—including physical, chemical, social, behavioral, and biological—on child health and development. In support of that goal, the ECHO DAC – located at Johns Hopkins University (JHU) and RTI International (RTI) – has defined quality objectives in line with the specific aims of the DAC.

The ECHO-wide Cohort Data Collection Protocol (EWCP) will support numerous longitudinal investigations. ECHO cohorts contribute existing and newly collected data elements specified in the EWCP to the ECHO-wide Cohort Data Platform hosted by the DAC. The DAC also maintains the ECHO Analysis Workbench (EAW) for data management and analyses, and creates and places the data in National Institutes of Health (NIH) -sponsored repositories.

A comprehensive quality assurance (QA) and quality control (QC) plan ensures data are of the highest possible quality. For this project and plan, quality assurance activities take place prior to beginning data collection (or data submission for extant data) and focus on preventing data issues, while quality control activities focus on identifying and correcting data issues.

5 QUALITY ASSURANCE

5.1 Protocol, MOPs, DCFs, MIS, SOPs, Resource Library

5.1.1 Protocol

The Protocol Implementation & Evaluation (PIE) Committee supports and monitors successful implementation of the EWCP, develops plans for evaluating the EWCP, oversees processes associated with protocol evaluation, reviews proposals for protocol amendments, maintains records of protocol change history, and disseminates best practice guidance.

PIE's Implementation Fidelity (IF) Sub-Group seeks to monitor and promote successful EWCP data collection and submission activity, to ensure the quality and consistency of EWCP data. The IF subgroup monitors adequacy of sites' training to conduct EWCP measures, documentation of measure administration according to the EWCP, adherence to the EWCP Manual of Operating Procedures (see section 5.1.2 below) and aspects of data quality relevant to implementation fidelity. The IF Sub-Group identifies protocol implementation concerns and communicates to appropriate groups for further investigation and remediation according to defined escalation pathways. See Sections 5.14 and 6.2 below.

The ECHO Coordinating Center (CC) maintains the current and historical versions of the EWCP on the ECHO SharePoint collaboration site.

5.1.2 Manual of Operating Procedures (MOP)

The objective of the EWCP MOP is to facilitate consistency in protocol implementation and data and specimen collection across participants and cohort sites. The use of the MOP increases the likelihood that the program will be scientifically credible. The MOP has a threefold purpose:

1. As a handbook to guide ECHO-wide Cohort study conduct and operations
2. As a reference to follow procedures related to the EWCP
3. As a supplement to the study protocol by detailing:
   a. Regulatory requirements (Chapter 1 and appendices)
   b. Training requirements (Chapter 2 and related guidance)

c. Operational procedures (measurement types, administration, and best practices) (Chapter 3 and related guidelines)
d. Data capture, data management, data quality, and data security procedures (Chapter 4)
e. Biospecimen collection, processing, storage, and shipping procedures (Chapter 5 and appendices)

Study personnel have ready access to the MOP and should be familiar with its contents, as applicable to their designated roles. The EWCP Materials SharePoint page hosts all versions of the MOP and associated protocol materials.

5.1.3 Data Collection Forms (DCFs)

DCFs capture a subset of the essential and recommended data elements using preferred and acceptable measures as listed in the EWCP. DCFs do not include any performance-based assessments or proprietary measures that have copyright restrictions on production of DCFs. Each DCF includes instructions for form-completion that states who should complete the form, the life stages for which the form should be completed, and the participant identification (ID) that should be used (pregnant woman's participant ID or child's participant ID). Each DCF has an accompanying measurement information sheet (MIS) for measure-specific instructions. Only trained, designated study personnel may administer DCFs.

The PIE Committee's DCF sub-group, comprised of DAC, CC, and ECHO Person-Reported Outcomes Core (PRO Core) representation, is responsible for amending and up-versioning DCFs. The DCFs may be amended via the Protocol Amendment Proposal process, to clarify Study Staff Instructions, or if errors are identified during DCF implementation. The sub-group evaluates cohort specific requests related to its impact on previously collected data and harmonization as well as skip logic and the ability to collect robust data, and amends the DCF accordingly.

All Spanish DCF translations undergo independent QC processes by the translation vendor, the CC translations team, and the collaborative cohort review team. The translation vendor verifies that the Spanish DCF is a complete and accurate translation of the original English source and attests to this via Certificate of Translation. The CC reviews translated files for any errors and confirms ECHO version control and readiness for Institutional Review Board (IRB) submission. ECHO cohort reviewers are native Spanish-speakers who may also review translations for accuracy, and can provide ECHO-specific feedback or participant feedback back to the translation vendor and the CC translation team. Once all QC processes are complete, the Spanish DCF(s) are submitted to the IRB for final approval before release to participants.

All DCFs follow versioning procedures to ensure that updated versions of the form can be easily distinguished from prior versions. The numeric version of the DCF and corresponding DCF date are updated, the change is logged in the ECHO-wide Protocol Measures Master List maintained by the DAC and also available on ECHOPortal site. While in some cases the introduction of a new version "retires" the use of a DCF, the original source DCF and all tracked change versions, along with the current version, are maintained on the secure ECHOPortal site.

5.1.4 Measurement Information Sheets (MISs)

MISs are available for each essential (preferred and acceptable) and recommended measure for the EWCP. MISs serve as guidance documents to facilitate implementation and consistency in data collection across cohorts.

MISs include the following information related to measurement administration:

- Life stages of administration
- Target participant (i.e., the target participant may be different from the participant completing the measure; e.g., biological mother, primary caregiver if not the biological mother)
- Person completing the measure (i.e., "reported by," e.g., biological mother, biological father, other caregiver)
- Mode of administration (e.g., questionnaire, self-administered or interviewer-administered, performance-based, or observation)
- Available translations
- Protocol measurement type (essential—preferred or acceptable, and recommended—preferred)
- Brief explanation of the measure
- Equipment and supplies required to administer the measure
- Estimated time to complete the measure
- Number of items (if applicable)
- Procedures and instructions
- Reference documents
- Links to the Data Collection Forms (DCFs)

The CC maintains all MISs on the EWCP Materials SharePoint page.

5.1.5 Standard operating procedures (SOPs)

SOPs are developed and maintained to document routine operations associated with analysis, data management, systems, and software development. For example, the DAC maintains SOPs for

- ECHO-DAC communications
- ECHOPortal Systems Security
- ECHO-wide Cohort DCF guidelines
- ECHO-wide Cohort Data Platform Geospatial Metadata Standards
- Research Electronic Data Capture (REDCap) Best Practice Documentation Standards
- Guidelines for Data Harmonization
- Data Submission (including extant genetic and epigenetic data, bioassay results, etc.)
- DAC responsibilities on writing teams.

SOPs, technical overviews, and systems manuals are made available via the Resource Library on ECHOPortal, a web-based gateway for ECHO data systems and resources.

5.1.6 Resource Library

The Resource Library, found under the Administrative panel on ECHOPortal, is a central location for viewing and downloading REDCap and printable data collection forms; specimen tracking and information forms and their instructions; training webinars, technical overviews, and manuals related to DAC systems; data dictionaries, and "Ask the DAC" session recordings and materials. Additional systems information and tutorials are located in the Knowledge Base (KB).

5.2 Change Management

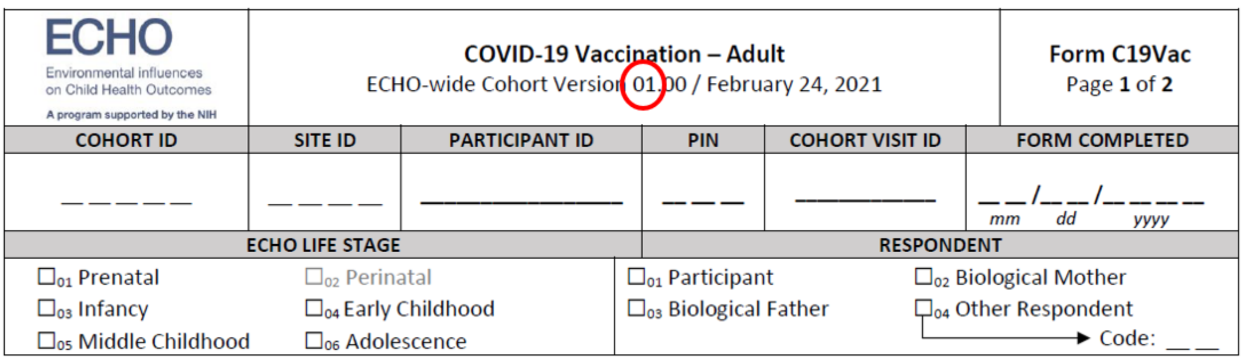
5.2.1 Protocol Change Processes

The Protocol Implementation and Evaluation (PIE) Committee facilitates protocol amendments using a multi-step process. First, cohorts propose protocol changes through an established communications pathway. Next, the PIE Committee reviews proposed changes and solicits evaluation of the proposed changes from ECHO stakeholders (e.g. ECHO Working Groups, Committees, Program Office). The PIE Committee presents agreed upon proposed changes to the ECHO Steering Committee for their review and approval. Finally, the PIE Committee incorporates approved changes into the protocol. The CC oversees regulatory submissions of the revised protocol. Protocol amendments are communicated using the Change Memo process, see section 5.2.3.

5.2.2 Form Versioning

DCFs are version controlled with version number and date indicated at the top of the form. The first final version a new form is assigned Version 01.00.

Major revisions to a form will increment the version number by 1 and reset numbers to right of decimal (e.g., a major revision to a form with Version 01.23 will become a final Version 02.00). Each major form version exists in different database tables on the ECHO-wide Cohort Data Platform. **Figure 1** displays an example.


ECHO
Environmental influences on Child Health Outcomes
A program supported by the NIH

COVID-19 Vaccination – Adult
ECHO-wide Cohort Version 01.00 / February 24, 2021

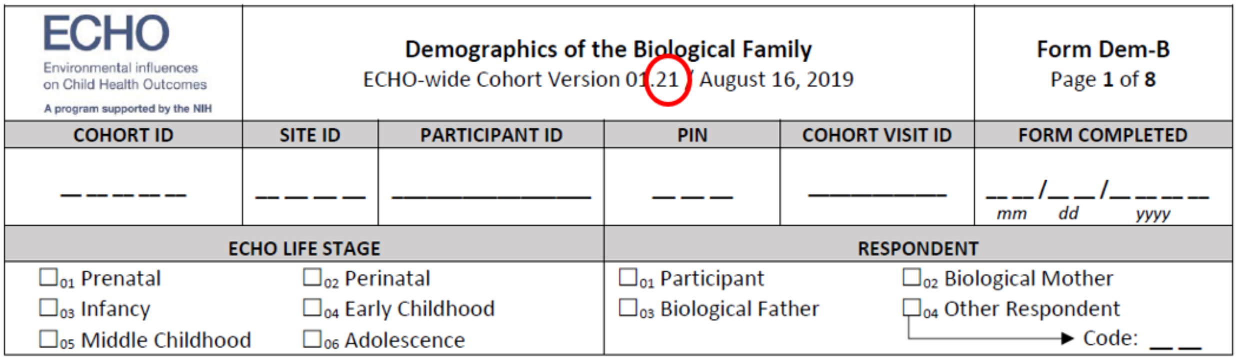
Form C19Vac
Page 1 of 2

COHORT ID	SITE ID	PARTICIPANT ID	PIN	COHORT VISIT ID	FORM COMPLETED
__ __ __ __ __	__ __ __ __	______________	__ __ __	__________	__ __/__ __/__ __ __ __ mm dd yyyy

ECHO LIFE STAGE		RESPONDENT	
☐01 Prenatal	☐02 Perinatal	☐01 Participant	☐02 Biological Mother
☐03 Infancy	☐04 Early Childhood	☐03 Biological Father	☐04 Other Respondent → Code: __ __
☐05 Middle Childhood	☐06 Adolescence		

Figure 1: Major Version Example

Minor revisions to a form will increment the version number by 0.01 (e.g., a minor revision to a form with Version 01.21 will become a final Version 01.22). Data collected under forms with minor version changes are stored in the same database table on the ECHO-wide Cohort Data Platform. **Figure 2** displays an example.



ECHO Environmental influences on Child Health Outcomes A program supported by the NIH	Demographics of the Biological Family ECHO-wide Cohort Version 01.21 / August 16, 2019				Form Dem-B Page **1** of **8**
COHORT ID	**SITE ID**	**PARTICIPANT ID**	**PIN**	**COHORT VISIT ID**	**FORM COMPLETED**
_ _ _ _ _ _	_ _ _ _ _	____________	_ _ _	__________	_ _/_ _/_ _ _ _ mm dd yyyy

ECHO LIFE STAGE		**RESPONDENT**	
☐01 Prenatal	☐02 Perinatal	☐01 Participant	☐02 Biological Mother
☐03 Infancy	☐04 Early Childhood	☐03 Biological Father	☐04 Other Respondent
☐05 Middle Childhood	☐06 Adolescence		Code: _ _

Figure 2: Minor Version Example

All revisions to a form also have a corresponding date of release, indicated in the header. It should be noted that REDCap electronic versions of forms always match the DCF version number and this data is hard coded into the form to ensure that records collected using REDCap can always be attributed to a specific DCF version.

The versioning process described above was put into effect in 2021 with the release of protocol version 2.1. Prior to this approach, new versions of DCFs were assigned a version number based on the current active protocol. This introduced complication and thus the new approach was adopted. All changes to forms are documented through posting the prior version with "tracked changes" in the Resource Library under Printable Data Collection Forms.

5.2.3 Change Memos

Change Memos track and communicate new and revised documents in a uniform way, notify the ECHO Program of changes to documents, and provide a consistent method for the ECHO Community to identify new and revised protocol-related documents needing regulatory oversight, training, and implementation.

Change Memos are distributed to the ECHO Community via email and associated supporting documents are available on the ECHO SharePoint collaboration site and ECHOPortal. Change Memos may include protocol amendments, new DCFs/DCF updates, MOP revisions, new MISs or MIS updates, other participant-facing materials, communications to participants, new or updated study staff training, and any other miscellaneous ECHO items being added or updated.

Change Memos are quality checked and reviewed by multiple stakeholders prior to release, including: the CC Learning Management System (LMS) team, CC IRB team, Document Owners, CC Communications team, Hub reviewers, and Change Memo team. Cohorts are asked to maintain their own document history according to established cohort procedures and maintain all Change Memos in their site regulatory file for reference. Site staff training must be completed prior to performing any associated study-related activity. The CC recommends that sites complete training on the LMS within two weeks of a Change Memo.

5.3 Systems

5.3.1 ECHOPortal

ECHOPortal is a secure web portal that provides an entry point where the ECHO community (investigators and staff) can access tools for data capture and data management, and an administrative space for sharing of study materials and reports.

The ECHOPortal environment has substantial security and compliance features. Most tools available via ECHOPortal were developed to comply with the Federal Information Security Management Act (FISMA) low-level information security requirements for federal information security systems, but the virtual-computing data enclave (that houses all ECHO Cohort data) is certified and accredited to implement FISMA moderate security controls. In the FISMA low environment, it is sufficient to have cohort site staff assigned individual user IDs and passwords (1-factor authentication) to gain access. Additional security procedures and controls are in place for the FISMA moderate computational environment that houses the virtual-computing data enclave where software and tools are provided to enable multiple users to analyze ECHO-wide cohort data without moving those data outside the confines of ECHOPortal. To access the virtual enclave, a user must have had human subjects protection training, and login using 2-factor authentication (i.e., identity must be proven with a username/password, plus a one-time use numerical password that is dynamically generated by a security application on a mobile device before a user can gain access). All user actions while working within the environment are tracked.

Below we provide a brief overview of the features and the tasks you can perform within ECHOPortal. ECHOPortal functions allow users to:

- Register participants;
- Capture new data using the centralized REDCap system;
- Batch submit new and existing data to the DAC;
- Specify how data captured using a cohort-specific data capture system can be mapped to the ECHO-wide cohort common data model (CDM);
- Download data captured using the centralized REDCap system;
- Track biospecimens and manage shipments to the biorepository;
- View reports and documents; and
- Correct data.

The DAC Systems webinar trainings shown in **Figure 3** below, and available under the Resource Library, provide a general overview of systems for participant registration, data collection and data transfer, and biospecimen tracking for ECHO Cohorts.

DAC Systems Support Materials

Show 10 entries Search:

Select All	Module	File Name	Version	Version Date
☐	Webinar	ECHOPortal	v1.0	03/21/2019
☐	Webinar	Participant Registration	v1.0	03/21/2019
☐	Webinar	REDCap Central	v1.0	03/21/2019
☐	Webinar	BioTrack	v1.0	03/21/2019
☐	Webinar	DataTransform	v1.0	03/21/2019

Figure 3: DAC Systems Webinar Trainings Available in ECHOPortal Resource Library

5.3.2 Participant Registration

Cohorts must register a participant in ECHOPortal before entering or submitting data for that participant into the centralized data base. There are different levels of registration, but only after consenting participants to the EWCP, may a cohort collect new data. Cohorts provide dates of consent in Participant Registration. For participants who have not yet been consented, cohorts may submit either de-identified, limited, (dates only, but no geographic information), or identifiable data based on the IRB review of cohort-specific consents. This level of identifiability is specified on Data Use Agreements signed by cohort local institutions and the DAC institutions.

5.3.3 REDCap Central

REDCap Central is the centralized data collection system provided by the DAC to support cohort and site data collection for the EWCP. REDCap Central is hosted on ECHOPortal and is interconnected with other ECHOPortal systems to streamline well-structured, qualitative data capture and aggregation. For example, cohort projects in REDCap Central are created from the data collection schedule and measures each cohort has specified in the Cohort Visits and Cohort Forms utilities in Data Transform.

The development of electronic forms to capture EWCP data follows best practices to ensure a high degree of fidelity with source DCFs. Once a DCF is finalized and approved for data collection, a REDCap developer is assigned the task of creating the form, based on the DCF. The DAC uses Jira project tracking software (Atlassian) to document the entire process, assign to a developer, and track special instructions. Once complete, an independent reviewer is assigned to test the new electronic data capture form, and ensure at a minimum that the form matches the paper-based DCF, variable names and attributes are accurate, skip logic is performed correctly, all header and footer variables are included properly, and range checks or other required validations are accurate or reasonable. All development is done on a development server, tested on a staging server, and only released to the production server when all requirements are met.

REDCap Central data is exported weekly to the EWCP data platform in the expected CDM format, therefore, no additional mapping is necessary. Warnings and errors resulting from the automated processing, are reviewed in the same manner as data uploaded by cohorts.

REDCap forms developed for use in REDCap Central are made immediately available to cohorts for adoption into local systems. Most cohorts utilize these pre-programmed forms to lower the burden of replicating in their own systems. By adopting these forms, most of the constraints placed on data (valid ranges, data types, validation checks) will also be inherited by local systems. Additionally, all of the quality measures in place in REDCap Central forms are aligned with data upload pipeline processing, which means the same constraints for receipt of uploaded data are in place as in REDCap Central, when receiving data collected in local systems.

5.3.4 Data Transform

Data Transform is the application tool developed by the DAC, to specify how to transform cohort-specific data to the ECHO-wide Cohort CDM. It has reporting features to help with pipeline processing errors and QA/QC post-processing data queries. The same system is used to transform both new and existing data. Each mapping protocol, extant or new, is defined with the data collection visits, the planned forms/measures to be collected at each of those visits, data dictionaries with validations specified, cohort to CDM data mapping, and numerous reporting tools. Data Transform supports the metadata necessary for both cohort uploads and REDCap Central data, to be processed through the data pipeline.

5.3.5 Bio-Track

The DAC has developed a Biospecimen Tracking System (Bio-Track) to maintain information about research specimens collected for the ECHO-wide Cohort Data Collection Protocol. Bio-Track performs several functions to log and track specimens and related information from collection, local processing, and storage through shipment to the specimen repository at Fisher BioServices. Fisher BioServices uses customized Core LIMS to manage and track all ECHO specimens both shipped by and received by the repository and is integrated with Bio-Track to communicate information on shipments, allow linkage to the inventory, and specimen reconciliation. Specimens are visually inspected and barcodes are scanned at Fisher and compared against the following specimen information in the Core LIMS database provided via Bio-Track API:

- Container barcode and the Specimen ID,
- Specimen attributes and characteristics,
- Physical vial type, sample coloration, and sample volume.

Empty and or low volume samples are confirmed and verified by a second individual and reported to the ECHO CC via the ECHO Discrepancy Log. If a shipment is delayed in transit (> 24 hours), an e-mail notification from Fisher is submitted to the ECHO CC. The DAC performs periodic reviews of specimen shipment status in Bio-Track against the data stored in Core LIMS. If a discrepancy is detected, the DAC notifies Fisher BioServices and the CC facilitates meetings as needed to resolve the discrepancy.

5.3.6 ECHO Learning Management System (LMS)

Cohorts use the ECHO LMS to train users in specific EWCP skills and proficiencies, and to track training compliance. The CC maintains all global and cohort-specific EWCP course content, (excluding DAC systems training) in the LMS. The ECHO CC manages the ECHO LMS content including courses, curricula, and user accounts and privileges. The CC uses established guidelines, procedures, and tools for managing new user access and reviewing and maintaining site content. CC also provides phone and email support for account creation and end-user assistance. Instructions and an LMS User Training Guide are located in the LMS section of the Training Corner on SharePoint.

5.3.7 ECHO SharePoint Collaboration site

The ECHO CC maintains the ECHO Program SharePoint site, a controlled access site with over 1,500 active ECHO Program staff user accounts. All ECHO Working Group and committee meeting calendars, agendas, minutes, and presentations are available on this site. The site also includes collaboration spaces for ECHO groups, the ECHO Program Manual, and all EWCP materials not housed within the ECHOPortal Resource Library. The CC uses established guidelines, procedures, and tools for managing new user access and reviewing and maintaining site content. CC also provides phone and email support for account creation and end-user assistance.

5.4 Formal testing of systems

Software development teams utilize best practices when modifying the ECHO data systems. These techniques include unit testing, code reviews, user interface testing, and system regression testing against verified test data. In addition, multiple environments are set up for initial development, quality assurance testing, staging for production, and production. These environments allow all aspects of software development and infrastructure software to be evaluated, prior to releasing deployments to the production environment. A final change control step serves as an extra quality assurance step in the process.

5.5 Training and certification

During study planning, the ECHO Operation and Implementation Team (OIT), comprised of all ECHO Components, made key decisions on how role-based training would take place across the ECHO Cohorts, to support protocol implementation. The team developed global and cohort-specific training courses on each protocol measures, excluding alternative measures, for the ECHO Training Course Library (for internal CC use only). The ECHO Global Training Matrix and the Cohort Forms by Life Stage reports specify role-based training requirements for site activation and implementation. ECHO utilizes several types of training, including, web-based, document review, video presentations, in-person, and virtual.

The ECHO DAC is responsible for training activities and documentation related to transferring extant data and data collection systems and the PRO Core is responsible for training activities and documentation related to NIH Toolbox (NIH TB). CC houses TB related training materials and other resources in the ECHO Training Corner. The ECHO CC is responsible for all other training activities and documentation, including Anthropometric and Physical Measures.

5.5.1 ECHO SharePoint Training Corner

The ECHO Training Corner is a resource page within the ECHO SharePoint Collaboration site that houses all training-related materials for Cohort Trainers and other study personnel to support training compliance, data quality, and protocol integrity.

5.5.2 Cohort Trainer Training

Each cohort PI designates one or more staff member to serve in the Cohort Trainer role for the cohort site(s). The CC and PRO Core Training Teams conducted in-person training sessions for available designated Cohort Trainers, using a “Train-the-Trainer” approach, from February through May 2019. The goals of in-person training were to:

- Provide information and tools to successfully train site staff on the EWCP
- Deliver hands-on training for anthropometric, physical, and NIH TB assessments to standardize data collection across ECHO sites, participants, and life stages
- Connect cohort trainers with Cohort Advocacy Teams (CATs) and each other to share experiences and best practices for data collection, visit choreography, and more

After attending the in-person Cohort Trainer training session (including anthropometric and physical measures competency assessments) and receiving NIH TB certification from the PRO Core, the Cohort Trainer is considered trained and certified.

Ongoing Cohort Trainer training is conducted by previously trained and certified Cohort Trainers or the CC/PRO Core when indicated. If a Cohort Trainer leaves a cohort, the cohort PI or designee must follow processes outline in MOP Chapters 1 and 2. Replacement Cohort Trainers receive in-person or virtual training by another trained and certified Cohort Trainer, following the same procedures in which they were trained. Specifically, the trained and certified existing Cohort Trainer will:

- Train the replacement Cohort Trainer on anthropometric and physical measures and complete competency checklists,
- Train the replacement Cohort Trainer on NIH TB, and

- Direct the replacement Cohort Trainer to PRO Core for NIH TB certification.

The CC and PRO Core coordinate training when an existing trained and certified Cohort Trainer is not available. These trainings combined with practice on site allow newly trained staff to demonstrate competency in anthropometric and physical measures and apply for NIH TB certification using the existing processes.

5.5.3 Site Training Responsibilities

Chapter 2 of the EWCP MOP outlines all required training for site staff and covers:

- Training requirements for Cohort Trainers, Research Administrators (data collectors), and all other delegated study personnel
- The ECHO LMS
- The ECHO Global Training Matrix
- Data Transform Cohort Forms by Life Stage Report
- Cohort Trainer responsibilities
- Maintaining training compliance

The ECHO Global Training Matrix and current Data Transform "Cohort Forms by Life Stage" Report are used to identify training courses that should be completed based on delegated roles and tasks performed and enrolling life stage. Global training courses are those that are required by all study personnel. Role-based training requirements are aligned to the roles and tasks performed, as delegated by the cohort site PI. Cohort site-specific data measures and biospecimen training courses must be completed by delegated study personnel prior to performing the activity.

ECHO utilizes the Train-the-Trainer approach. Therefore, the ECHO Cohort Trainer has many responsibilities such as coordinating new, revised, and ongoing training for site staff, conducting quality assurance activities, and maintaining documentation of site training to ensure the cohort site staff are adequately trained. The ECHO MOP, Chapter 2 outlines these responsibilities.

5.5.4 Data System Training Webinars

The DAC invited Cohort Data Managers to two-day training sessions to learn and practice using the DAC-developed systems, including how to map cohort-specific data to the Common Data Model, transfer data, and use REDCap Central for data collection. In addition, the DAC created a training module for cohorts.

Cohort Trainers must complete DAC ECHO-wide Cohort Data Systems and Processing training and certifications. The DAC recommends that at least one data collector, biospecimen collector, data entry staff member, and data manager pass certification requirements for use of the various DAC systems. However, this requirement will be at the discretion of the Cohort Trainer who is certified in all DAC systems needed by the cohort. Certification involves receipt of a DAC certification packet with instructions for completing the applicable exercises. The person being certified will complete the following tasks as applicable for their role on the study:

- Registering mock participants
- Mapping mock data
- Uploading mock data
- Entering data in REDCap Central (if applicable)
- Registering biospecimens

- Creating a shipping manifest for Fisher BioServices

The DAC orchestrates data systems training and certification. Delegated study personnel document training course completion locally and with the DAC.

Subsequent to the initial training, the DAC hosts weekly and now bi-weekly Ask the DAC interactive sessions to relay information about system updates and clarify any system issues posed by Cohorts.

5.5.5 Annual Training and Proficiency Testing

Cohort Trainers are required to complete annual Anthropometric and Physical Measures training by completing a “read and review” LMS course and achieving 100% on a post-course knowledge assessment. Additionally, Cohort Trainers complete an NIH TB Refresher Course annually, beginning in the calendar year following their certification. These are tracked by the CC to ensure compliance.

5.5.6 Training meetings

The ECHO Component Training Team, comprised of CC, DAC, and PRO Core representatives, meets monthly and the CC Training Team meets weekly, to discuss training-related processes, issues, and resolutions.

During and after site activation, the Component Training Team held bi-monthly Cohort Trainer Webex Sessions to encourage Cohort Trainer collaboration. These meetings were retired and replaced with bi-monthly “Ask Abby” sessions with Abigail Sivan, PhD, focused on Cohort Trainer questions for administering the NIH TB.

Additionally, the CC Training Team attends calls with individual cohorts and their CC CATs to answer training-related questions when necessary.

5.6 Data Pipeline Processing

The Data Pipeline involves a multi-step process to help ensure the quality of incoming data. These steps include: file uploads verified against a list of allowable files (the contract), file ingestion to extract zip file contents, variable and data transformation per data mapping, special transformation functions (if needed), data validations against the data dictionary, and generation of an emailed pipeline processing report to indicate summary information on successes, warnings, and failures of submitted data files. Additional messages related to pipeline processing with associated personally identifiable information (PII), are reviewed in the online Reports section of Data Transform.

Some of the specific quality checks on incoming data include:

- Validate the participant against the participant registration;
- Ensure that the participation level allows data to be submitted for the participant;
- Validate data ranges, dates, and discrete value lists; and
- Ensure that duplicate data rows are rejected.

5.7 Bioassay Pipeline Processing

The Bioassay Pipeline processes laboratory specimen data provided by cohorts for incorporation on the ECHO-wide Cohort Data Platform. It applies to specimen information for samples to be sent or already sent to the Human Health Exposure Analysis Resource (HHEAR) or other laboratories for new assay results on Type A and B samples as part of approved Analysis Proposals (AP). The Bioassay Pipeline is being further developed to process extant non-omics bioassay data. To ensure quality of incoming data, file uploads are checked for proper naming conventions and content, data is validated against the data dictionary, and a processing report is emailed to indicate summary information on successes, warnings, and failures of submitted data.

Some of the specific quality checks on incoming data include:

- Validate the participant against the participant registration;
- Ensure that the participation level allows data to be submitted for the participant;
- Check specimen types and sample counts to ensure specimens sent to the lab adhere to the goals of the approved AP;
- Validate data ranges, dates, and discrete value lists;
- Ensure that duplicate data rows are rejected;
- For Type A and B specimens, Specimen Plans must be approved before HHEAR-ECHO Specimen IDs are assigned and specimens are shipped to the HHEAR laboratory; and
- Multi-variate checks have been implemented to validate conditional requirements (e.g., required data elements based on specimen type).

5.8 Cohort Advocacy Teams (CATs)

Each Cohort Awardee has designated Cohort Advocacy Teams (CAT) as main points of contact. The ECHO-wide Cohort Data Collection Protocol Awardee, Cohort, and Site List documents all ECHO Cohorts with assigned CC and DAC CATs and their respective responsibilities.

ECHO CC CATs are responsible for providing and coordinating operations, procedures, and activities related to site management. Cohort and site staff are encouraged to contact the Lead Clinical Research Associate for questions related to the protocol or general program operations and to contact the Clinical Coordinator for questions regarding regulatory documents. In addition to the general CC email (echocc@duke.edu) and regulatory document specific email (ECHORegDocs@duke.edu), sites are encouraged to contact their designated CC CAT representatives directly or via the ECHO DAC Help Desk.

ECHO DAC CATs assist the cohorts and sites with technical questions about DAC-provided systems, including: Participant Registration; REDCap Central; Data Upload; Data Transform, and Bio-Track. They assist with mapping new and existing cohort data to the ECHO common data model using Data Transform and assist with problem-solving for data that cannot be mapped. They monitor and evaluate data quality. RTI-DAC data managers are responsible for day-to-day operations; JHU-DAC analysts support these activities, particularly with respect to data mapping and report generation. Cohorts and sites are encouraged to contact their designated DAC CAT representatives directly or submit questions via the ECHO Help Desk Web Form.

5.9 ECHO Help Desk

The ECHO Help Desk was implemented by the DAC in March 2019 and now serves as a centralized help desk with members from the CC, DAC, and PRO Core working within the application to address tickets that are related to processes and systems that their component manages or owns.

Use of the help desk application has allowed for streamlining of many CC-, DAC-, and PRO Core-managed requests and processes, which has reduced response times. Members from the

CC and DAC who oversee help desk operations monitor response times and notify responsible CC and DAC members of any outstanding tickets (i.e., tickets open for more than two weeks).

Enhancements to the help desk have been made over time, and will continue, to ensure the application meets the evolving needs of the project and its stakeholders, such as the ECHO cohorts.

5.10 Knowledge Base (KB)

The ECHO KB, a dynamic, searchable application of answers to frequently asked questions that have come through the help desk and on calls with the cohorts and other members of the ECHO community, such as the weekly “Ask the DAC” sessions, was launched in February 2020.

New content continues to be reviewed on a periodic basis by CC, DAC, and PRO Core subject matter experts and is then made available to the ECHO cohorts and others via ECHOPortal.

CC and DAC members who manage review and release of new content continue to identify innovative ways to share information, including brief systems demonstrations, such as Data Transform QA/QC queries video tutorials.

5.11 Ask the DAC

ECHO “Ask the DAC” sessions began in March 2019, typically last 90 minutes, and occur semi-monthly. Sessions are attended by approximately 75 to 100 staff from the CC, DAC, ECHO Cohorts, and NIH Program Office (PO). Members from the DAC often include leadership, data systems developers, CATs, data analysts, and DAC investigators. The meetings are a forum for providing important DAC data systems updates and demonstrations, which are announced in advance of each session, and offer an opportunity for cohorts to raise questions and concerns. Ask the DAC meeting materials are archived within the Resource Library.

5.12 Coordinator calls

ECHO Study Coordinator Collaboration Calls are monthly meeting attended by approximately 150 to 200 staff from the CC, DAC, ECHO Cohorts, and NIH PO. The meetings allow an open forum to discuss recent program updates and for cohorts to ask questions.

The CC organizes, hosts, and facilitates the meetings, develops and distributes agendas, identifies presenters, and compiles and distributes post-meting materials. The DAC provides brief presentations of DAC-related topics and answer DAC-related questions.

Meetings occur monthly and typically last 60 minutes. Study Coordinator Collaboration Call meeting materials are archived on the Study Coordinator Corner SharePoint page.

5.13 Tracking of DAC System changes

The DAC at RTI maintains an internal ECHO Quality Management Plan that defines quality standards for key informatics and infrastructure tasks and the processes and procedures, activities, responsible person and frequency for ensuring those standards are met. In particular, the Quality Management Plan specifies that system design documentation and the JIRA management system are used to track development of new system features, requested changes, decisions, approvals (as needed), and status of completion. The DAC System lead oversees change management activities associated with each new release of system.

5.14 Compliance Checklists

The CC and the IF subgroup of the PIE Committee ask each EWCP site to complete Compliance Checklists by responding to semiannual surveys that confirm practices and procedures implemented at each institution (ECHO Compliance Checklist – Fall 2021, ECHO Compliance Checklist – Spring 2022). The purpose of the checklist is to ensure protocol implementation quality, reproducibility, and reliability, by confirming appropriate processes implemented at sites participating in the EWCP. Specifically, the checklist enables the IF subgroup and the CC to report on:

- Adequacy of training at each site to conduct EWCP measures
- Documentation of measure administration according to the EWCP
- Adherence to EWCP MOP

One staff member from each site completes the checklist in communication with the site PI. CC CATs monitor completion of the Compliance Checklist. The CC and the IF subgroup review results and take actions, as defined in the below escalation pathway (**Figure 4**).

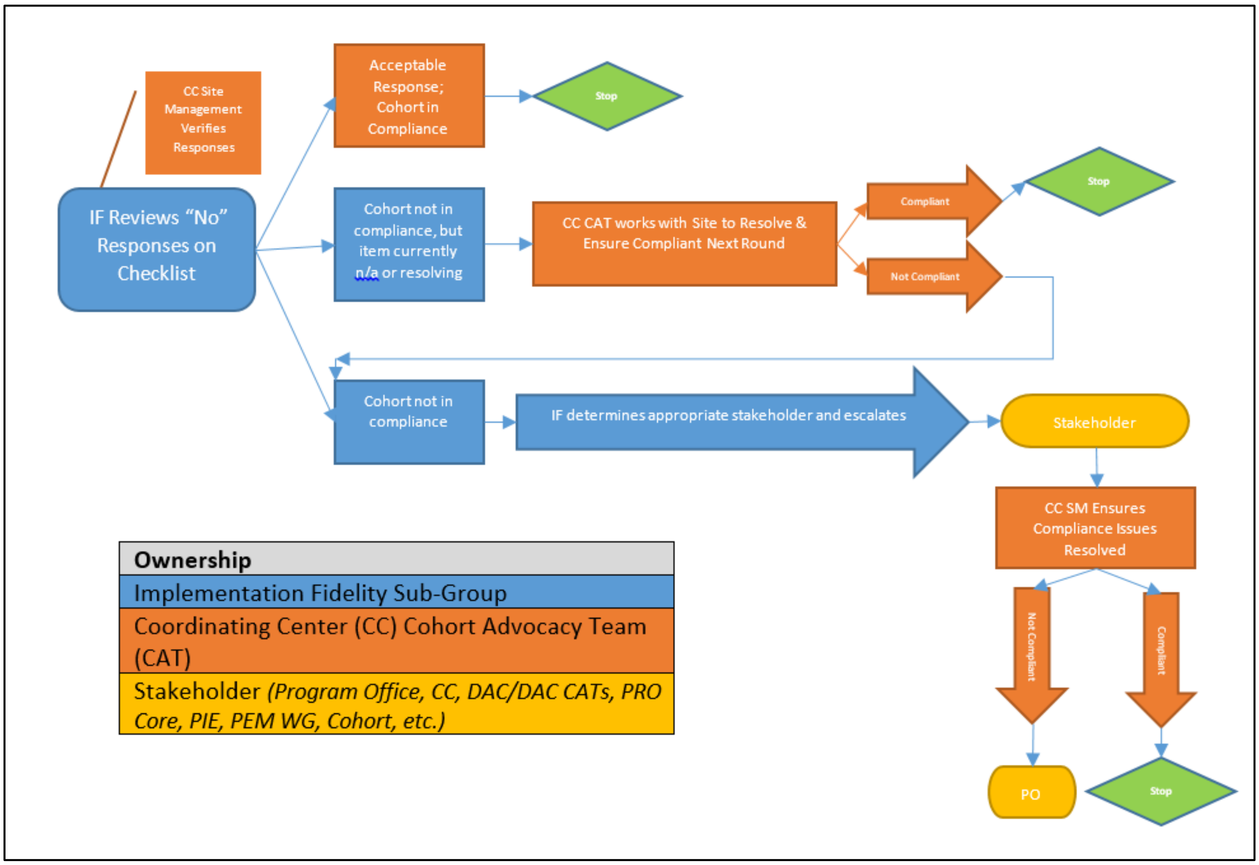


Figure 4: Implementation Fidelity Sub-Group Escalation Pathway for Cohort Compliance Checklists

6 QUALITY CONTROL

There are multiple mechanisms for assuring the data loaded to the ECHO-wide Cohort data platform are valid, reliable, and complete. Pipelines perform validation checks at the time of data submission to ensure that data adhere to the rules/requirements established for the project. When a validation error occurs in the data pipeline, processing continues but the erroneous file/record/variable is rejected, and the error is documented in a report that cohorts can review. This process protects the downstream database from storing data with errors, while ensuring error-free data keep flowing. Not all data received at the DAC flow through data pipelines, but rather undergo manual review with feedback provided to submitter through email or data queries. Once the data exist in the database, queries are the primary means of identifying potential errors or inconsistencies that might be difficult check for in a data pipeline (e.g., some complex queries that look for outliers within longitudinal data, or inconsistencies on data submitted on different forms or data sources). Finally, reports that summarize data at the awardee, cohort, cohort site, and ECHO-wide levels are helpful in identifying random errors and logical inconsistencies as well as systematic variation in the data. Each of these 4 mechanisms for monitoring the quality of data being submitted to the DAC and existing on the ECHO-wide Cohort data platform at described below. We also describe quality control procedures for derived variable development (through harmonization of CMD and CDD variables) and data analysis.

6.1 Pipelines

6.1.1 Data Pipeline for Common Data Model or Custom Data Dictionaries

The Data Pipeline Processing Report is generated by the Data Pipeline when it processes data that has been mapped to the Common Data Model (CDM) or Custom Data Dictionaries (CDDs). It includes statistics for the entire batch of uploaded file(s), giving header information about the upload itself, followed by summary level details for each file processed. Another section provides summary statistics on the CDM and CDD forms to which the data is mapped. This report gives success, warning, and error codes and messages to help the cohort with initial resolution of any data issues. This information includes file level, row level, and mapping feedback. It also includes some special information about variables included in the uploaded file, but not mapped to the EWCP data platform.

When row level messages occur, the report lists the first ten rows in the file containing the information being flagged. Since the point of this report is to alert the cohort about the completion of their file processing by the pipeline, it is emailed to them at the end of processing. Only summary level statistics and information are provided, without any participant data. In cases not involving file level or row level errors, the specific participant data issues can be reviewed in the Data Pipeline Messages Report found in Data Transform. The Technical Overview - Data Pipeline Processing and Message Summary Reports describes the Processing Report Codes, Data Pipeline Message Summary Codes, and Codes Requiring DAC Support.

The Data Pipeline Messages is a special module in Data Transform, under the Reports section. This module allows a cohort to find the errors and warnings needing to be addressed, per the results of pipeline processing. Different error codes and other filters, allow the cohort to select a set of detail messages to be reviewed. In addition to some standard information for each pipeline message detail row including the error/warning and the Participant ID, the row can be expanded by clicking on a control to the left side of the row. Once opened, this provides additional information which includes the specific variable and its value, followed by the transformed variable and its value. This information enables the cohort to view the exact issue being flagged by the pipeline process. Once the data has been corrected by the cohort and reprocessed successfully, the pipeline messages for that data issue will no longer display.

6.1.2 Bioassay Pipeline

The Bioassay Pipeline Processing Report is generated by the Bioassay Pipeline when it completes processing a single CSV file, or a zip file containing CSV files. The report includes statistics for the entire batch of uploaded file(s), giving header information about the upload itself, followed by summary level details for each CSV file processed for adherence to the SOPs and validation checks.

This report gives success, warning, and error codes and messages to help the cohort with initial resolution of any data issues. This information includes file level, row level, and multi-variate checks.

When row level messages occur, the report provides the number of occurrences and lists the first 10 row numbers in the CSV file containing the information being flagged. Since the point of this report is to alert the cohort about the completion of their file processing by the pipeline, it is emailed to them at the end of processing. Only summary level statistics and information are provided, without any participant data.

Processed Specimen Plans are made available to authorized ECHOPortal users via Data Transform's Reports menu (under the Assays header). Downloaded files will contain approval codes and dates to indicate the status of each specimen included in the file.

6.2 Manual QC

6.2.1 Bioassay data

Manual QC of bioassay data is occasionally used for some files that need to be re-submitted under older SOP versions. For example, a cohort may be asked to make a correction and resubmit their data, but the structure of the files no longer matches the current pipeline specifications. The DAC may agree to manually process the uploaded data file using the Manual QC process rather than burden the cohort with a larger set of edits to upgrade their files to a newer SOP version.

Manual QC is performed to check that data adheres to the SOPs and validation check requirements. In this case, the term “Manual QC” refers to usage of Structured Query Language (SQL) queries or SAS merges to review the datasets. This manual process has been helpful in refining the validation requirements because visual inspection of the data can inform decision making. The results of the manual QC also served as confirmation of the automated pipeline validation checks through comparison of the results.

6.2.2 ECHO Eating Habits Food Frequency Questionnaire Data Quality Checks

6.2.2.1 Site/Cohort QC

The overall roles and responsibilities of site and cohort staff when quality checking the ECHO Eating Habits Food Frequency Questionnaires are detailed in the MISs related to these proprietary measures (Eating Habits Questionnaire - Ages 8-17 and Eating Habits Questionnaire - Ages 2-7).

Sites and cohorts are broadly responsible for entering booklet tracking data in the Food Frequency Tracking System on ECHOPortal, performing a thorough review of the participant completion/quality of the data entry on the booklets before shipping booklets to the CC, and responding to any queries from the CC or DAC.

6.2.2.2 CC QC

The CC responsibilities when quality checking the ECHO Eating Habits Food Frequency Questionnaires are detailed in the MISs related to these proprietary measures (Eating Habits Questionnaire - Ages 8-17 and Eating Habits Questionnaire - Ages 2-7).

The CC is broadly responsible for tracking and collating received shipments from the sites and cohorts, performing a thorough review of the participant completion/quality of the data entry on the booklets before shipping booklets to the publisher, contacting cohorts and sites with any queries, and updating the Food Frequency Tracking System to reflect receipt date and shipped date to the publisher for analysis. The CC follows the ECHO CC Management of Eating Habits Questionnaires Guide to complete the QC processes.

6.2.2.3 DAC QC

The DAC validates incoming data uploaded in batches from the publisher after analysis has been completed. The checks ensure participant IDs exist in Participant Registration at level of participation that indicates data can be shared, key fields are non-missing, no duplicates exist, and the data conform to data dictionary specifications from the publisher regarding variable types and valid ranges (if applicable).

6.2.3 NIH Toolbox (TB) submissions Data Quality Checks

Cohort data managers must follow post-processing instructions (available in the Resource Library on ECHOPortal) to submit data exported from iPads to the DAC. During this step, cohorts are expected to remove exported records that are not intended for ECHO. The DAC validates incoming files if they contain records not found in the Participant Registration system or contain records for participants with a participation level of –1. The DAC also confirms that the “name” column in the exported Registration file does not contain PII, and generates a query when key fields are missing, duplicates exist, and the data do not conform to data dictionary specifications. The PRO Core analyzes the data to identify issues such as systematic variability over time, and by cohort and mode of administration.

6.3 Data Query System

The QA/QC module in the Reports section of Data Transform, includes Query Definitions and Query Details. When an analyst within the DAC or PRO Core evaluates data for a single form, or compares data across multiple forms, the resulting set of detailed information can be formed into a Query. For example, an analyst may create a Query to check for missing data that is essential to an analysis, or to examine inconsistencies between related data elements. The Query is loaded into the system to share with cohort staff, and to track their responses to the Query Detail.

Utilizing this concept, an analyst can develop a “definition” using a standardized Excel template, of the information they intend to make available in support of the specific data rows being addressed within the Query. The DAC analyst processed that information using SAS or R, which the resulting Query Definition and Query Details into Data Transform for the cohort.

The cohort staff may then download their information for the Query and apply standardized response codes (see valid response codes below) and free-form text, indicating what actions they can or cannot take to resolve the data in question. The updated spreadsheet is then loaded back into Data Transform by the cohort, with their responses. The QA/QC dashboard can be used by the cohorts and by the CC or DAC staff, to monitor progress towards resolving the data rows initially addressed by the query.

Table 2: Data Query Response Codes, Descriptions, and Statuses

Query Response Code	Description	InProgress[a]	OnHold[b]	Disregarded[c]	Resolved[d]
100	Correction – pending upload	✔			
200	No action possible at present		✔		
300	Corrected	✔			
310	DAC follow-up required	✔			
400	Verified – no action needed			✔	
500	No action possible ever / ignore			✔	
700	Resolved				✔
800	Exclude				

a. InProgress: Latest response code indicates no correction occurring
b. OnHold: Latest response code indicates no corrective action can be taken at this time
c. Disregarded: Data verified and remains unchanged, or no further corrective action is possible
d. Resolved: Data corrected and validated by the system through a reprocessed query

As Queries are reprocessed, some weekly and some on a longer schedule, the system automatically resolves data that is no longer showing up in the Query, e.g., such as when the cohort submits new corrected data. Likewise, any new data that meets the criteria being addressed by the Query, will be automatically added to the system for cohort review and response. Not all edit queries may be resolved, which is also reflected in the response by the cohort. The QA/QC Query Metrics dashboard, as shown in **Figure 5**, reflects these updates to the system on a regular basis.

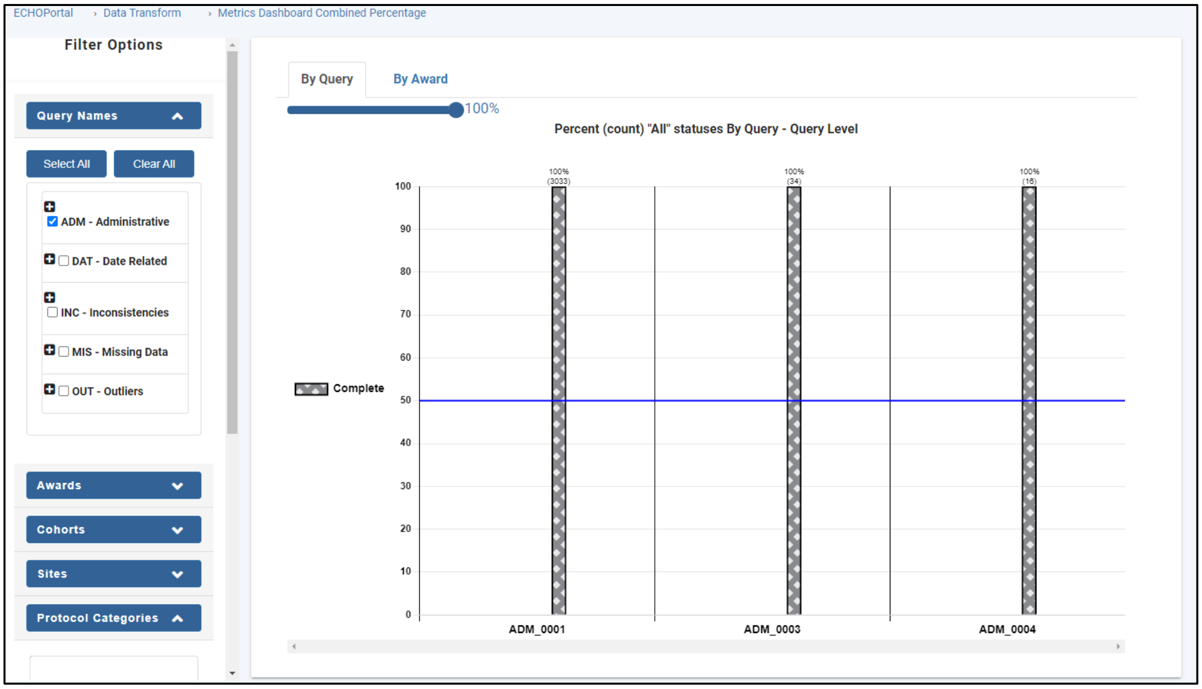


Figure 5: QA/QC Query Metrics Dashboard Example

6.4 Reports

The DAC creates and updates Administrative Reports related to cohort participation levels, registrations, REDCap Central data, Bio-Track status and Data Transfer disposition weekly and posts on ECHOPortal. Cohorts use these reports frequently as they interact with the DAC and troubleshoot areas of concern.

Goals, Objectives, Indicators, Targets (GOIT) were established in Year 5 of the ECHO program to set criteria for success, and monitor extent of achievement. The Year 6 GOIT includes four overall goals, each having three or more objectives. Each objective has one or more indicator and target. Y6 GOIT include a total of 19 targets. Many of these targets are monitored through administrative reports produced by the DAC. The CC uses several of the administrative reports to maintain GOIT Dashboards.

The Implementation Fidelity Sub-Group of PIE identified ten EWCP measures representative of potential data quality issues (e.g., long questionnaires, skip logics, etc.) and reviews reports produced by the DAC related to missing data; accuracy, precision, reliability of physical measures; and data usability.

Table 3: Key Data Quality Issue Reports

Key Data Quality Issues Identified in Reports	Details
Missing data	Aspects of missing data: • Entirely missing forms or measures, overall and by individual cohorts • Completeness within forms/measures (e.g., individual items on a questionnaire, physical measures, etc.) • Patterns of missing data Reasons for variability in missing data may include:

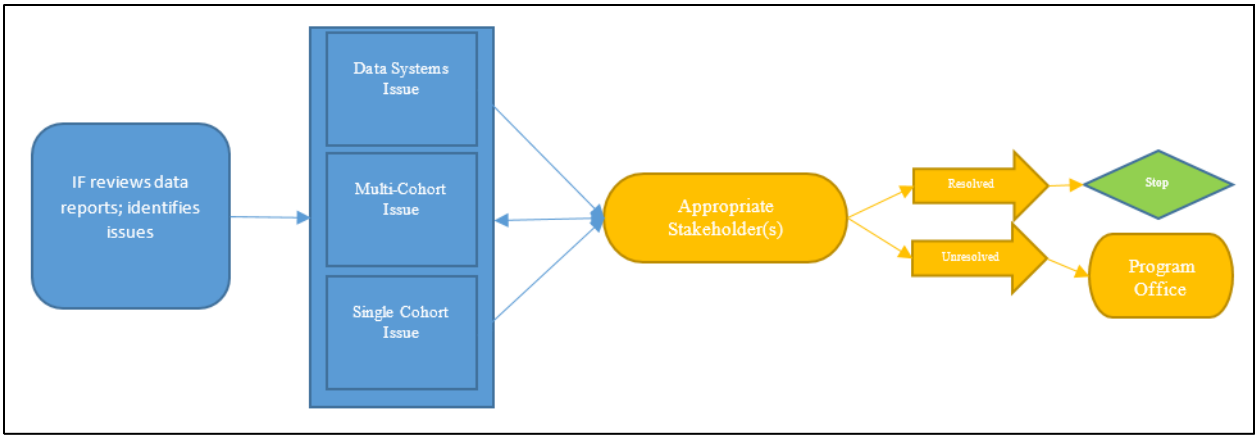
Key Data Quality Issues Identified in Reports	Details
	• Impact of data systems: REDCap vs. local systems • Mode of Administration (e.g., self vs. interviewer-administered measures) • Remote vs. in-person measures • Parent vs. caregiver/child responses.
Concerns for Physical Measures	Accuracy, prevision, reliability of physical measures • Digit preference, rounding, incorrect units • Protocol for repeated measures not followed • Large differences in repeated measures • Implausible change when longitudinal data are available
Data Usability	"Don't know" and "Prefer not to answer" responses

As outlined in **Table 4**, the PIE IF Sub-Group established metrics for various data quality issues presented in the IF reports and established appropriate thresholds in conjunction with the PIE Committee to trigger further exploration and escalation. **Figure 6** displays the escalation pathways. The IF reports are available on ECHOPortal with other Administrative Reports.

Table 4: Thresholds for Key Data Quality Issues Identified in Reports

Quality Issue	Metric	Threshold
Entirely missing forms or measures	For Essential elements, any entirely missing forms should be justified	None
Completeness: Missing data within forms/measures (e.g. items on a form or questionnaire, physical measures)	% complete (no missing values), overall and by cohort	10%
Patterns of missing data	% missing for each item on a form, overall and by cohort. Review patterns: meaningful declines with larger number of items, specific questions more likely to be missed (e.g., sensitive questions? Skip pattern not understood?)	Will depend on the problem
Data collection: REDCap vs local systems	Completeness by data entry system, overall and item-specific, Difference in rate of missing data between the systems	> 5 percentage point difference
Self vs interviewer administered measures	Completeness by mode of administration (when available), overall and item-specific AND by cohort Difference in rate of missing data by mode of administration	> 5 percentage point difference
Remote vs. in-person measurement	Completeness by remote vs in-person administration (when available), overall and item-specific AND by cohort Difference in rate of missing data by mode of administration	> 5 percentage point difference

Quality Issue	Metric	Threshold
Differences in repeated measures	Correlation of repeated measures, mean difference between measures, overall AND by cohort	Correlation>0.9
Usability of data	% of Don't Know or prefer not to answer within and between questionnaires, overall and by cohort	10%



Ownership
Implementation Fidelity Sub-Group
Coordinating Center (CC) Cohort Advocacy Team (CAT)
Stakeholder *(Program Office, CC, DAC/DAC CATs, PRO Core, PIE, PEM WG, Cohort, etc.)*

Figure 6: Implementation Fidelity Sub-Group Escalation Pathway for Key Data Quality Issues

6.5 Quality Control of Data Harmonization

The ECHO Data Harmonization Working Group Process on the Data Harmonization SharePoint page describes the formal review process that variables undergo prior to being posted to PLATIPUS. There are two phases to the process for reviewing Data Harmonization Reports that document the work to harmonize elements and constructs for analysis – a Technical Review and a Conceptual Review. There are additional sources of quality control in the data harmonization process prior to the submission of Data Harmonization Reports. The DAC harmonizes data to derive new analytical variables using a team approach; DAC Biostatisticians and investigators, together with writing teams collaborate on establishing relevant keywords to use to search the database, review of identified information, and creation of new variables using established literature and principles of inclusivity. In addition, new variables are presented at internal meetings for review prior to submitting Data Harmonization reports to the Working Group. Frequencies of the new variables overall and by cohort are provided in the reports as well as the DCF and cohort-custom tables that were used to capture the data, thus allowing cohorts to review and assess whether their data were identified and used, and correspond to their knowledge about their cohort's data.

PLATIPUS – The ECHO Cohort Platform: Inform and Provide User Search – is a searchable web-based program on ECHOPortal that provides a gateway for metadata about the ECHO-wide Cohort Study based on the data in the database. In addition to showing the numbers of individuals and cohorts contained in tables based on the common data model, it also uses the derived data resulting from the DAC's harmonization efforts and contains all the Data Harmonization Reports which were reviewed by the Data Harmonization Working Group. Cohorts can also use PLATIPUS to review their content, ensuring full inclusion of their Custom Data Dictionaries in the harmonization process.

6.6 Quality Control of Data Analyses

All individual-level data analyses are conducted on the EAW, the FISMA moderate cloud-based platform, developed and maintained by the RTI DAC. Quality control of data analysis involves the application of methods and processes that enhance the reproducibility and transparency of results, match research questions and available data to appropriate statistical methods and reduce opportunity for error. Quality control of data analysis includes documentation and principles of reproducibility, exploration of data before and after data manipulation, understanding missing data, implementation of proper analytical methods, and evaluation and interpretation of cohort-specific results.

6.6.1 Documentation and Reproducibility

Reproducibility describes the ability to utilize data and software code to recreate the analytic findings of an ECHO AP. It is a hallmark of rigorous science, and a necessary feature of all ECHO projects. A folder is automatically created for each AP, containing nested folders to organize code, output, and other project materials. Individuals should use the preferred EAW directory structure to enhance the reproducibility and sharing of analyses. Each analyst should maintain a README document, stored within the AP folder, to describe the project goals, program names and purposes, and key decisions. Annotation of analytic code is complementary to a README file.

6.6.2 Analyses

Publication Committee review of analysis proposals, DAC workshops, and ECHO investigator expertise should be utilized to identify and properly apply the appropriate statistical methods. DAC-created statistical programming code, located in the Shared Resources directory on the EAW, can also be used to reduce potential coding errors. Peer-review of data analyses is encouraged. The DAC team presents DAC-conducted analyses for peer-review by DAC methodologists at various stages – development of the analytical proposal and during the analysis phase at research in-progress meetings.

6.6.3 Data manipulation

Most analyses will use harmonized versions of key variables available in the EAW. If the AP requires additional data manipulation and variable creation, it is important to describe the data before and after any data manipulation and to view a few randomly selected observations before and after the data manipulation to determine if the manipulation achieved the expected and desired result.

6.6.4 Cohort adjustment

The ECHO Pediatric Cohorts reflect different underlying populations, recruited with different sampling criteria, and used varying study designs and methods to collect data prior to the initiation of ECHO. This level of heterogeneity may lead to biased results if the data are naively pooled together. It is critical to evaluate and understand potential cohort effects. There are numerous approaches to understand and account for cohort differences including meta-analysis, mixed and fixed effects models, and leave-one out analyses. Analysts email ECHO Help Desk for guidance.

6.6.5 Missing data

The ECHO-cohort data platform has missing data. There is both general missing data and unbalanced missing data. Consideration of missing data and its implications is necessary. The application of multiple imputation for missing data should be evaluated for its applicability to a given research question and the interpretation of results.

7 QUALITY ROLES AND RESPONSIBILITIES

The overall roles and responsibilities for overseeing EWCP data quality activities, processes, and procedures are summarized below.

The **ECHO DAC** is responsible for developing and maintaining the environment to host, analyze, and control access to ECHO-wide Cohort data. The DAC is also responsible for tracking biospecimens and their associated metadata in the ECHOPortal Biospecimen Tracking system (Bio-Track), developing, validating and applying QA/QC methodologies, and providing statistical analysis support for primary and secondary ECHO analyses.

The **ECHO CC** is responsible for maintaining the current version of the ECHO-wide Cohort Data Collection Protocol, for developing and maintaining the ECHO MOP, collaborating with DAC and other to provide protocol training, and for collaborating with ECHO DAC and others regarding data quality assurance by ensuring administration of outcome tools and acquisition of biospecimens is appropriate and consistent across all research sites.

The **ECHO PRO Core** is responsible for efforts to enable, curate, and, where needed, enhance existing or develop new information about exposures, covariates, and outcomes obtained through self- or proxy-report, observation, performance-based testing, or personal sensor devices. The ECHO PRO Core also designs and validates instruments, conducts quality checks on PRO data, works with cohort sites to resolve issues of missing data, etc., and conducts analysis of datasets for evaluation of core element PRO data.

ECHO Cohorts are responsible for ensuring staff are fully informed about the conduct of the processes within the protocol, collecting and sharing protocol data to the ECHO-wide platform and biospecimens to the ECHO Biorepository, resolving data and biospecimen queries, and retaining source documents and other relevant study files.

The **ECHO PIE Committee** supports and monitors successful implementation of the EWCP, develops and oversees plans and processes for evaluating the EWCP, develops and implements plans for protocol amendments, and maintains records of protocol change history.

- The **PIE IF Subgroup** monitors and promotes successful administration of the EWCP and helps to assure consistency of high quality EWCP data. The IF Subgroup recommends specific tasks, timelines, and mode of reporting to the ECHO components, that ensure implementation fidelity. Topics may include adequacy of site training to conduct EWCP measures, documentation of measure administration according to the EWCP, and other aspects of data quality relevant to implementation fidelity

The **NIH ECHO PO** sets scientific and collaborative vision, ensures that components adhere to NIH policies, provides funding, and facilitates collaborative research. Through the ECHO CC and DAC, the NIH ECHO PO monitors accurate protocol implementation and internal quality assurance across the EWCP.

8 REVIEW AND REVISION

The Data Quality Plan is a dynamic document that the DAC will update throughout the conduct of the study to reflect protocol amendments as well as refinement of study procedures, best practices, and guidance. The ECHO CC, PRO Core, and NIH ECHO PO support its development and maintenance.

All ECHO Program Manual documents are to be reviewed at least every two years; however, more frequent reviews may be warranted as processes change or major edits are identified.

Table 5: Summary of Changes

Version	Date	Summary of Changes
01.10	11Jan2023	All hyperlinks removed for version to be made available on ECHO public website
01.00	02Aug2022	Original document

Approval Page

Name / Title	Approval Date	Signature (signed electronically)
Lisa Jacobson, ScD, ScM ECHO DAC PI Johns Hopkins University	Jan 25, 2023	Lisa Jacobson Electronically signed by: Lisa Jacobson Reason: Approved Date: Jan 25, 2023 09:14 EST
Diane Catellier, DrPH ECHO DAC PI RTI International	Jan 24, 2023	Diane Catellier Electronically signed by: Diane Catellier Reason: Approved Date: Jan 24, 2023 13:20 EST

Appendix A
Referenced Documents

ECHO-wide Cohort Data Collection Protocol (EWCP)

EWCP Manual of Operating Procedures (MOP)

Data Collection Forms

ECHO-wide Protocol Measures Master List

Measurement Information Sheets (MISs)

Change Memos

ECHO Data Use Agreements

LMS User Training Guide

ECHO Global Training Matrix

Cohort Forms by Life Stage reports

In-person Cohort Trainer training materials

NIH Toolbox training materials and other resources

ECHO-wide Cohort Data Collection Protocol Awardee, Cohort, and Site List

ECHO Help Desk Web Form

Ask the DAC meeting materials

Study Coordinator Collaboration Call meeting materials

Compliance Checklists

- ECHO Compliance Checklist – Fall 2021
- ECHO Compliance Checklist – Spring 2022

Administrative Reports, including Implementation Fidelity Sub-Group Reports

Technical Overview - Data Pipeline Processing and Message Summary Reports

Eating Habits Questionnaire - Ages 8-17 MIS

Eating Habits Questionnaire - Ages 2-7 MIS

ECHO CC Management of Eating Habits Questionnaires Guide

ECHO Data Harmonization Working Group Process

Data Harmonization Reports

Appendix B
Referenced Systems and Sites

ECHO Help Desk / https://echoportal.org/Home/HelpDesk

ECHO SharePoint collaboration site

EWCP Materials SharePoint page

ECHOPortal

Resource Library

Knowledge Base (KB)

ECHO Learning Management System (LMS)

REDCap Central system, available on ECHOPortal

Participant Registration system, available on ECHOPortal

Data Transform system, available on ECHOPortal

ECHO SharePoint Training Corner

GOIT Dashboards

PLATIPUS