ASH Research Collaborative Data Hub Protocol
A Multi-Center Data Hub of Individuals Living with Hematologic Disease

Version: 2.4
Date: July 16, 2024
Funding: The American Society of Hematology, the U.S. Department of Health & Human Services, and the Centers for Disease Control and Prevention

A. DATA HUB OVERVIEW

Background and Significance
Benign and malignant hematologic diseases are relatively rare conditions within the spectrum of medical practice in any one site of care. Nonetheless, recent research in hematologic conditions from basic, translational, clinical and population perspectives offer the possibility of improving the way that these diseases are treated, and the outcomes experienced by patients. A repository that aggregates and validates this data across institutions and other practice settings is needed in order to identify variation in care, new findings, and further research.

Purpose of the Data Hub
The ASH Registry, doing business as the ASH Research Collaborative (ASH RC), is a non-profit organization established by the American Society of Hematology (ASH) in 2018 that aims to improve the lives of those affected by blood diseases by fostering collaborative partnerships to accelerate progress in hematology. The foundation of the ASH RC is its Data Hub, a technology platform that facilitates the exchange of information by aggregating in one place, and making available for inquiry, research-grade data on hematologic diseases.

The primary goal of the Data Hub is to further the scientific knowledge base for the diagnosis, understanding, and management of benign and malignant hematologic conditions by assembling data collected in routine clinical care and closed clinical trials. Secondary goals are to characterize and study practice patterns for benign and malignant hematologic conditions in clinical practice, and to aggregate patient-reported data to further understand and improve the patient experience. These objectives will be fulfilled by amassing data from patients' electronic medical records and other data sources within institutions and networks to support prospective data collection efforts, such as those that include patient reported outcomes (PROs).

Current data collection efforts are limited to specific hematologic diseases and are usually narrow in scope. Efforts that are larger in scope usually do not include patient-level data with the detail that is required to accelerate basic, translational, and clinical science research in the way that is envisioned by the Data Hub. Furthermore, centers and networks that have been queried have indicated willingness to contribute detailed, patient-level hematologic data to a large and centralized data collection effort. These standard of care observations underlie the rationale for the development of the Data Hub.

The purpose of this umbrella Data Hub Protocol is to permit the creation of a patient-level data repository obtained from individuals, or the records of individuals, with benign or malignant hematologic diseases. Specific practices related to research studies that intend to use Data Hub data, including study procedures, informed consent documents, and other patient-facing materials, will be submitted for IRB approval in separate amendments to this Data Hub Protocol or under separate protocols. The first two disease areas of focus include sickle cell disease and multiple myeloma.

B. METHODS
This is a multi-center, retrospective, long-term data collection of patients with benign or malignant hematologic diseases. Information will be collected on patient demographics, disease characteristics, genomic and molecular data, laboratory data, pathology, radiographic reports, clinical status, quality of life, medications, and dosing information.

Data are collected for all patients who meet the inclusion criteria as outlined in this protocol. Additional interventions or testing outside of the normal course of treatment requires that the patient or his/her/their legal guardian sign a separate informed consent document.

The following sections describe the data that will be collected, aggregated, stored, and managed by Data Hub staff and IQVIA, the ASH RC's technology vendor.

Patient Selection and Inclusion Criteria
To be included in the Data Hub, patients must have documented benign or malignant hematologic disease. There will be no discrimination or bias with respect to inclusion on the basis of sex, race, or religion.

Patient Recruitment
Patients who are receiving or seeking medical care for specified hematologic diseases at participating institutions may be invited to participate in the Data Hub. Participation means that the patient (or the patient's parent/legal guardian) agrees to have his/her/their fully identifiable information submitted to the Data Hub, to be contacted for future research, to permit his/her/their data to be linked to external sources, and to have access to a web-based patient portal for longitudinal engagement.

Informed Consent
Patients who meet the inclusion criteria may be approached by a member of the study team in-person, by telephone, or electronically. Patients may consent in-person or via an electronic consent portal in which they watch a series of educational videos and take a short quiz to ensure comprehension. **Medical record data for patients who have not yet consented will be submitted to the Data Hub in the form of a limited data set as defined by HIPAA.**

The site is responsible for verifying that if collecting consent, it will be done so from adult patients, or the parent/legal guardian of children, according to IRB requirements and applicable local regulations. Verbal and/or written assent will be obtained from children based on local IRB requirements. Children who reach the age of 18 years, or the legal age of majority according to local regulations, during their participation in the Data Hub, will be asked to provide informed consent to continue participation in the Data Hub.
Consenting to participate in the Data Hub includes the following:

- that a patient's physician may submit his/her/their fully identifiable data to the Data Hub,
- that a patient's limited data may be used for clinical trials or medical research studies,
- that a patient may be contacted for future research,
- that a patient's Data Hub data may be linked to external sources (i.e., insurance payer databases for reimbursement purposes), and
- that participation is voluntary, and he/she/they can withdraw at any point in time.

Electronic Informed Consent
The ASH RC created a streamlined, single-platform, centralized Participant Portal to provide Data Hub sites with capabilities to obtain informed consent electronically (eConsent). The Portal includes animated videos to help patients better understand the Data Hub program and what it means to participate. Patients will have the ability to consent at their leisure and download, or print, a copy of the consent form. In the future, the eConsent Participant Portal may also include educational materials, patient reported outcome surveys, and a data dashboard.

No countersignature (i.e., the principal investigator) will be required when a patient provides his/her/their consent using the eConsent Participant Portal. By supplying this tool, participating Data Hub sites will have the ability to track consent and properly perform participant matching to support future ASH RC-related activities.

Withdrawal of Consent
Patients may decide to withdraw their data from the Data Hub at any time. Data shared for research prior to a

patient's decision to withdraw from the Data Hub cannot be destroyed or be removed; however, data that has not yet left the Data Hub or has not yet been used for research will not be used in any future studies following a patient's decision to withdraw consent for participation. Additionally, if requested; a patient's data housed in the Data Hub may be deleted. Patient withdrawal must be documented in his/her/their electronic medical record and the site investigator is responsible for making a similar notation in their study files or electronic medical record. Patients will also have the ability to withdraw from the Data Hub via the eConsent Portal.

Once a patient withdraws consent, transmission of his/her/their PHI from the electronic health record to the Data Hub will immediately cease.

Patient Costs and Compensation
There is no cost to the patients for participating in the Data Hub and similarly, no compensation will be offered to patients or families for their participation in the Data Hub.

Patient Risk/Benefit Assessment
There is no added procedural risk to patients through involvement in the Data Hub. No testing, time, risk, or procedures beyond those required for routine care will be imposed. The primary risk associated with this project is the potential for a breach of patient confidentiality. IQVIA (see: Section C), the Data Hub's technology partner and technical architect of the Data Hub, has established a robust plan for ensuring appropriate and commercially reasonable physical, technical, and administrative safeguards are in place to prevent, preempt, and mitigate any potential risks the Data Hub and its participants may face.

There is no direct benefit to patients for participating in the Data Hub. Indirect benefits, obtained through the contribution of data, may include the development of new drug therapies, clinical practice guidelines, and an overall better understanding of hematological diseases.

Oversight
Management of the Data Hub will be conducted in accordance with applicable FDA Regulations and guidelines, as well as all other applicable national and local laws and regulations.

The ASH RC has a Board of Directors to oversee the operations of the Data Hub. This expert panel directs the efforts of ASH RC staff dedicated to the handling of the data in a manner that preserves patient trust and complies with all international, federal, state, and local guidelines. The Board is responsible for:

- Developing strategic frameworks for evaluating requests to use Data Hub data for research proposals, and prioritizing proposals based on ASH's overall scientific agenda.
- Ensuring the interests of benign and malignant hematological diseases are adequately addressed in existing and future efforts.
- Assuring requests for data from the Data Hub include documentation of IRB approval (when required).
- Managing policies related to removal of direct identifiers from information contained within the Data Hub for research.
- Documenting investigator access to the Data Hub.
- Ensuring the security of the database linking the Data Hub linkage codes with participant identifiers and the documentation of investigator access to this database.
- Reporting any conditions that may negatively impact the confidentiality of information contained within the Data Hub.
- Granting approval for investigators to access de-identified Data Hub data for research studies.
- Disease specifications will be developed on a project-by-project basis as disease modules are sequentially onboarded within the Data Hub. Separate amendments to this protocol or additional protocols will be developed to detail data collection mechanisms for specified diseases, including study procedures and all patient-facing materials.

Data Hub Design and Procedures
Data Hub staff will be responsible for managing the operational aspects of the Data Hub. Data included in the Data Hub will be used for research about hematologic diseases and quality improvement purposes. Separate data collection protocols, informed consent documents, and any other documents as appropriate must be approved by the IRB as amendments or new protocols.

Site Training
All participating sites will be trained on the Data Hub protocol, logistics, and data collection methods. Retraining will be conducted every three (3) years or more frequently, if needed. Additionally, updated diagnostic criteria and other Data Hub-specific information will be communicated to the sites.

Site Monitoring
Data Hub personnel or designees will perform onsite monitoring/remote monitoring contacts (e.g., via phone follow-up) consistent with the Guidance for Industry Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment (GPPPA). In the event that site monitoring identifies noncompliance issues, additional monitoring and site education will be performed.

Participating Site Statement of Agreement
All responsible parties provide a written statement agreeing to the content of the proposal and the confidential nature of the documentation made as part of this Data Hub and acknowledging that the ASH RC has the right to discontinue this Data Hub at any time and/or amend this Data Hub Protocol as appropriate.

Ethical Considerations and Privacy of Personal Data
Patient information collected in the Data Hub will comply with the standards for protection of privacy of individually identifiable health information according to local privacy laws. Informed consent and assent forms, as applicable, will be prepared according to the institutional requirements for informed consent and the applicable regulations.

C. DATA COLLECTION AND MANAGEMENT

IQVIA
IQVIA serves as the ASH RC's technology vendor. IQVIA has extensive experience managing and integrating complex data sets for biomedical and health research and has designed the ASH RC Data Hub to facilitate common legacy data migration use cases as well as prospective data collection.

Data Collection
Data collection will be done from patients' existing medical records and/or manual data entry into an electronic data capture form. Participating sites will submit HIPAA-compliant retrospective limited datasets of a patient's electronic medical record on an ongoing basis for all patients who meet the inclusion criteria prior to the patient's provision of consent. Until a patient consents, data submitted to the Data Hub is in the form of a limited data set.

The frequency in which the site submits data will vary; however, it is required that data be submitted, at a minimum, on a quarterly basis. As the data models are finalized for the Data Hub programs, the data collection forms will be available on the ASH RC's website.

The ASH RC will ensure that a Data Hub Site Agreement (DHSA) and DUA are in place with all Data Hub participating institutions prior to commencing transmission of limited datasets and patient consent for fully identifiable datasets.

Data Model and Transmissions
The Data Hub will define a canonical data structure or set of structures, according to the requirements of each stored data set. The Data Hub will establish a primary data acquisition application programming interface (API) that securely identifies the requested data structures for capture. This approach allows the Data Hub to implement appropriate

constraints for each of the project's data structures in one location, which facilitates the application of appropriate data quality and completeness checks at the moment incoming data reaches the Data Hub system, using the project's own canonical definitions of each relevant data structure. Where appropriate, data structures may be based on a combination of Fast Healthcare Interoperability Resources (FHIR), Consolidated-Clinical Data Architecture (C-CDA), IQVIA's open-source Research Instrument Open Standard (RIOS) for form-based data, and/or client-specific structure definitions.

Data Hub participants will submit patient EHR data in the agreed upon format using the secure data submission portal developed by IQVIA. IQVIA will provide each Data Hub participant with credentials and access to a secure data holding zone, with initial data transfers facilitated and supported by IQVIA.

System Integration Solution
The preferred method for sending electronic medical record data to the Data Hub is through data exports produced according to a site's preferred format and established data ingestion channel. Currently, the Data Hub has ingestion channels for OMOP- and FHIR-structured data. For Data Hub participants without OMOP or FHIR data export capabilities, a centralized data entry mechanism will be provided.

In order for the Data Hub to provide quality improvement services, such as outcomes reports, Data Hub sites are asked to submit limited data sets as defined by the Health Insurance Portability and Accountability Act ("HIPAA") (i.e. no direct identifiers are submitted to the Data Hub for those patients) on all patients previously seen and currently seen at their institution that meet the inclusion criteria outlined in this protocol on an ongoing basis. In some instances, sites may choose to only submit limited data sets on patients who meet the inclusion criteria under the Data Use Agreement (DUA) in place between ASH RC and the participating sites.

The Data Hub accepts data conforming to HL7 FHIR R4. FHIR data is securely accessed by the Data Hub via facility approved application programming interfaces (APIs). These APIs are often modeled to comply with US Federal regulations on health data interoperability (see 45 CFR 170, 171), which are intended to enable direct access to electronic health information by patients and healthcare providers. As such, the APIs typically require the exchange of PHI, as defined by HIPAA, during the data request and/or data receipt phases. In these cases, IQVIA will receive and process PHI solely for the purposes of (1) facilitating data request and receipt, (2) mastering received data to ensure uniqueness and correct linkages, and (3) generating Limited and/or De-Identified Data Sets for submission to the Data Hub, as defined by HIPAA.

Before any received data is made available in the ASH RC Data Hub, IQVIA removes all relevant PHI from patient records, as needed, in accordance with the prevailing research protocol and patient consent . IQVIA generates limited data sets for patients without an applicable consent on file and generating identifiable data sets for patients with an applicable consent on file to be used for Data Hub as described in the Purposes of the Data Hub section within section A of this protocol.

Electronic Data Capture
For sites participating in the Multiple Myeloma Data Hub Program, an eCRF-based data ingestion pathway will be utilized. The eCRF allows sites to enter data specifically tailored to the harmonized data model. The tool is delivered on the same platform used to manage the System Integration process and requires the installation of a small hosting utility on the servers or desktops of the site. This utility then affords users with the secure, encrypted connectivity necessary to populate Data Hub information on an electronic analog of the paper-based data collection form. Data is transmitted to the data mart where it is aggregated with other data streams in preparation for final upload to the Data Hub's data warehouse.

If it is determined that additional data is needed outside of what is available by way of a patient's medical record, but also consistent with standard care, an electronic data entry mechanism will be created for supplemental data collection.

Other Data Sources
In addition to the aforementioned data sources, the Data Hub will seek to collect data from single center or multi-center closed, de-identified clinical trial data sets stored inside or outside of the U.S. The Data Hub will ingest the entirety of these clinical trials data sets within specified malignant or benign hematologic diseases, so as long as these data sets contain appropriate data elements and meet data quality standards as specified by the Data Hub.

Data Verification
The ASH RC understands the need to ensure the data submitted to the Data Hub is complete and accurate. Direct electronic health record extraction of data elements for the calculation of clinical performance and measures adherence is novel, and thus, quality assurance standards are not yet well established. Upon receipt of the Data Hub data submission, IQVIA reviews the data electronically and reports back any coding errors. Such electronic checks verify consistency and completeness and allows Data Hub participants to have ample opportunity to submit data until it satisfies the schema criteria. In addition to coding errors, IQVIA compares the data submissions to a series of validation rules that assess format, completeness, and range consistency on key variables. Participants will be contacted if any discrepancies are identified.

The ASH RC is committed to the development of innovative and reliable data quality assurance and validation enhancements. The Data Hub team is currently exploring future methodologies in order to ensure completeness and validity. Until those methodologies are established, however, the Data Hub will rely on individuals at the participating sites to verify that the data being submitted is complete and accurate.

Physical Security and Access Control
The Data Hub provides end-to-end protection for all data, both in transit and at rest. Data Hub application environments use GPG with AES 256 encryption; all end-user connections to Data Hub portals are encrypted using SSL; file transfers are conducted over SFTP or HTTPS; and all received FHIR messages and data files are individually encrypted with GPG before archiving in an AES 256 encrypted environment. Data is also archived to PR's secure cloud backups service. Everything contained on the Data Hub server is encrypted (e.g., GPG, AES256), and all data sent to and from the server is encrypted. All backups are encrypted with GPG before transmission. IQVIA will be responsible for access control oversight and security of the data stored within the Data Hub. Additional information on data protection and security is available upon request.

Equipment
By default, systems run on highly reliable hardware. In the unlikely event of hardware failure, the system can be completely rebuilt and restored from backups via automation. By default, there is a minimum of 98% uptime.

Some other data elements to be acquired by the Data Hub, particularly in the setting of automated data collection protocols, are predefined as versioned HL7 FHIR Resources, which are referenced by versioned Data Hub- and/or disease area-specific FHIR API Profiles, templates for structured data file submissions, and templates for spreadsheet import and electronic data capture interfaces; this approach enables IQVIA to apply consistent data validation and quality assurance logic across disparate data sources and acquisition modalities, and to confidently integrate and repurpose acquired data (as appropriate). Acquired data is indexed and tagged for repurposing in a clinical data warehouse (CDW) using an industry- standard PostgreSQL database backend.

The Data Hub automatically "unpacks" relevant component data from acquired sources into specialized data marts, on demand, to support specific data visualization, reporting, and analytic functions; these data marts are built using auditable configuration logic and persist as PostgreSQL databases; all analytic results are stored in the CDW.

Transportation and Transmission of PHI
If migrating data from research instruments, the Data Hub can automatically create Comma Separated Values (CSV) templates and apply instrument-specific data validation rules upon import (based entirely on the instrument's configuration within the system).

Quality Assurance
Upon receipt of the Data Hub data submission, IQVIA reviews the data electronically and reports back any coding errors. Such electronic checks verify consistency and completeness and allow sites to have ample opportunity to submit data until it satisfies the Data Hub's criteria. IQVIA assists participants with this process by providing the following:

- Program-level and technical documentation describing data submission requirements. The documentation will include extensive FAQs and an onboarding video.
- Data quality reports documenting feedback on their data submission.
- Ability to submit written questions to the integration team.
- Direct one-on-one assistance from integration team, as needed.
- Ability to attend weekly webinar "office hours" hosted by the integration team where they will have the opportunity to ask further questions and receive targeted assistance.

External Access to Data Hub Data
All data shared for research purposes will either be de-identified or will constitute a limited dataset. Prior to providing any data, however, the ASH RC follows a formal review and approval process in which all requests for data access must be submitted through the ASH RC Data Hub Research and Analytics Request Form.

ASH RC contracts with data analytic centers to conduct such analyses. These analytic centers are required to sign a contractual instrument that holds them to the same standard regarding protection of protected health information as are outlined in the DHSA/DUA between ASH RC and the participating site with a separate Data Use Agreement (included with this protocol).

Data Storage and Confidentiality
Original, immutable source data (e.g., FHIR message, structured file, tabular file) is archived for the maximum time allowed by law or ASH policy. All data validation logic is defined in an auditable Data Validation Utility; validation results are stored in the CDW.