Data Users FAQs

What is the Rare Disease Cures Accelerator-Data and Analytics Platform (RDCA-DAP ®) and what does it do?

The Rare Disease Cures Accelerator-Data and Analytics Platform (RDCA-DAP) is an integrated database and analytics framework that is designed to be used in building novel tools to accelerate drug development across rare diseases. It promotes the sharing of existing clinical data and encourages the standardization of the collection of new clinical data. By making such data available in a format suitable for analytics, RDCA-DAP accelerates the understanding of disease progression (including causes for variance in disease progression), clinical outcome measures and biomarkers, and facilitates the development of mathematical models of disease and innovative clinical trial designs. RDCA-DAP is positioned to generate solutions to drug development, which can be made publicly available to researchers in industry, government, regulatory agencies and academia. As such, the utility of the patient-level data will be maximized and used to develop tools that will be accessible to the community to optimize and accelerate drug development across rare diseases.

What is the value of RDCA-DAP?

Integrated, anonymized patient-level data can be used to develop models or other tools that aid in the development of efficient clinical trial protocols. Models can inform the size and length of a trial, its specific arm allocation, the design of comparator groups, frequency of assessments, choice of endpoints and trial population needed to evaluate the efficacy of a drug to modify underlying course of the target indication. Such data can also be used in the development of quantitative tools to support adaptive trial designs, innovative assessments or endpoints and potentially historical controls (in non- platform designs) or shared control patients (in platform designs). Learnings from such data may even be applied across related disease areas and have the potential to significantly speed up patient access to new treatments. Regulators will also use the data and tools to help make data-driven decisions on use of novel trial designs, biomarkers and new drugs.

What data are in RDCA-DAP?

RDCA-DAP contains integrated data from multiple sources (clinical trials, natural history studies, registries, curated data from electronic health records etc.) across multiple rare diseases. Data is contributed from any source (initial sources will include the IAMRARE ® databases from NORD, clinical trial and natural history datasets), and the custodians of these data have agreed to provide them to RDCA-DAP with defined provisions that are addressed in individual Data Use Agreements . Data ownership is maintained by contributor/custodian. As they become available, new datasets will be added to the database. All data is mapped to a common data standard and common terminology prior to integration. Datasets from a single source are not to be identifiable, however a user will be able to identify the data type (clinical trial, registry etc.).

Who has access to the data in RDCA-DAP?
RDCA-DAP platform allows tiered access to qualified researchers to the patient-level data dependent on the specific level of accessibility defined by the individual data custodians (people who share data with the platform). However, all the integrated data is used to advance drug development tools; quantitative tools to understand disease progression, quantify the contribution of biomarkers to optimize clinical trial design, etc. These tools and results from analyses will be made public.

How do I get access to the data in RDCA-DAP?
Qualified researchers may apply for access to data that is permitted for sharing through RDCA-DAP. Researchers will need to provide a reasonable explanation of the intended use of the data and sign the Data Use Agreement for privacy protection, data security, acknowledgement of the source of the data and agree to a preview prior to any scientific publications. All data access requests will go through a review process. After approval and agreement to the DUA, the researcher will be able to access the patient-level data at no cost.

What quality control steps will be in place?
All data contributed to RDCA-DAP is verified by a quality control process to ensure its integrity and validity before being included in the platform. If certain data are deemed invalid, they may be archived rather than being held in active status in the platform. Data is normalized to a common format to allow for aggregation and analysis.

What data sharing standards will be utilized?
Data is mapped to data standards appropriate to the type of data, where possible. Clinical Data Interchange Standards Consortium (CDISC) data standards to support regulatory submissions are utilized where appropriate, and data is tagged with standardized ontologies (e.g., OMIM, Monarch) as appropriate. Some data may remain unstructured but searchable until it can be standardized.

How does the platform account for the likelihood of duplicate records from the same patient in different studies?
This is a likely occurrence in rare disease datasets. RDCA-DAP recognizes the likelihood that individual patients may be represented in multiple datasets that are integrated into the database. In anonymized datasets, if the contributed data have not adopted global identifiers, it is not possible to recognize potential duplicate patients across datasets. However, the potential bias towards an artificially reduced analytics uncertainty can be adequately controlled by standard statistical and mathematical methods. Other quantitative drug development tools built by C-Path using databases with this same issue have accounted for this bias within the development of the models and tools, and this has not affected the utility of the tools developed. Moving forward, we strongly support the use of global identifiers (GUIDS) that can be used in all studies to allow us to link records from individual patients without identifying the person in question.

What kinds of analytics functionalities does the platform have? What is of value across many diverse disease areas? Are these analytics tailored to specific diseases or are they simple canned reports for all disease areas?

Several different tiers of analytics functionalities exist within the platform. At a first level, a user-friendly data interrogator allows researchers to identify, categorize, cross-reference and stratify the variables captured across the database (within diseases and across diseases) and generate summary descriptive statistics for distribution, frequencies, central tendency and dispersion. It also for the development of customized analysis-ready subsets with the variables of interest (within diseases and across diseases). A synthetic data generator is currently being developed to provide the ability to supplement available data for deeper analysis capabilities. At a second level, private research workspaces allow researchers to interface the generated analysis subsets with advanced analytics tools, such as R, Python, and more. This will allow researchers to develop advanced models to quantitatively describe relevant aspects of disease progression within and across diseases, capturing relevant sources of variability. These sophisticated models could potentially be submitted for regulatory review and endorsement as quantitative drug development tools. For such regulatory-endorsed tools, a interface will allow researchers to access dedicated workspace environments for the purpose of running simulations based on those models, intended to optimize clinical trial design.

What are the benefits of the platform analytics for participating groups?

There are many benefits for using the analytics platform – here are a few:

- Access to standardized data in one place;
- An easy-to-use interface to determine if the platform has the data you need and do simple statistical tests;
- Ability to download some data in CDISC format, appropriate for regulatory filing (dependent on the type of data in the database);
- Platform tools to help with simple statistical analysis and data exploration;
- Access to a platform with advanced pharmacometrics tools to develop more complex analyses, where you can add your own software as needed;
- Potential collaboration with the statisticians and pharmacometricians within C-Path's Quantitative Medicine team who have experience with these types of data and solutions for medical product development;
- Potential collaboration with C-Path's regulatory science experts who have experience developing regulatory-grade drug development tools such as outcome measures, biomarkers, natural history and disease progression models.

Who can I contact with questions about the database or suggestions on analytics?

An email can be sent to rdcadap@c-path.org with any questions. Contact information will also be provided on the RDCA-DAP website and within the platform itself.

For more FAQs specific to patients and patient organizations, visit https://c-path.org/programs/rdca-dap/overview/resources/#faqs

FDA Acknowledgment