TriNetX Explore real world, real-time global data

Best Practices in Study Design and Publishing using TriNetX

What research questions is TriNetX suitable for?
How can you best use the Query Builder to create research cohorts?
How can you best use the Advanced Analytics tools for your research?

July 2024

CONTENTS

INTRODUCTION

The TriNetX platform offers a unique and convenient way to access and use Real-World Data, with over 120 healthcare organizations (HCOs) contributing data to at least one of our research networks. This guide will cover the most common use cases for TriNetX data to prepare you to publish your results with confidence.

TRINETX USE CASES

There are several use cases for the TriNetX data and platform including natural history of diseases, safety and efficacy studies, treatment utilization patterns, and more. The most important questions to ask yourself when determining if TriNetX data are a good fit for your research question are:

1. Can I answer my question using only structured data elements found in the EMR?

TriNetX EMR data are supplemented in several ways (natural language processing of provider notes from a subset of HCOs, links to tumor registries from a subset of HCOs, links to closed claims data on the specialized Linked network, etc.), however, the core of the TriNetX Data Model is structured EMR data. There are certain data types that are not available within the TriNetX Data Model at this time:

- Patient Reported Outcomes (PROs) – While some common PROs are included in the TriNetX data model, these are not always recorded as standard of care and not all HCOs record these in a structured field of the EMR.
- Imaging – TriNetX does not ingest imaging data including the images themselves or imaging reports. It is possible to Query for the procedural code(s) associated with an imaging procedure, and inferences can be made based on the diagnoses in the patients' charts following an imaging procedure, but specifics like the size of a lesion, the exact nature of a fracture, or the percent occlusion of an artery which may be contained in an imaging report cannot be Queried.
- Pathology – TriNetX collects histology and staging data for oncology patients if the data is contained in a structured EMR field or if the HCO allows linking of patients to tumor registries. However, more in-depth details often found in pathology reports such as overall dimensions of a tumor after resection, tumor margins, or other histological details are not currently supported.

It's always important to consider the opportunities and challenges of any data source and real-world EMR data is no exception. Here are a couple of things to keep in mind when using EMR real-world data:

- Missing data and selection bias examples:
    - Patients may see certain healthcare providers that are not in the TriNetX network even if they are also seen by one or more providers in the network. While many clinical facts should still be captured in the EMR of the TriNetX HCO (medications the patient reports, any diagnoses from outside providers that the patient reports, etc.), others might not be (labs may not be input in a structured field if they are faxed from an outside source, codes for procedures done elsewhere will not appear in the EMR, etc.). Even facts that are added to structured EMR fields, like medications, may have a different start date than the actual start date, because the first instance of the term visible by TriNetX will be the first date that the patient reports this medication during an encounter with the TriNetX HCO.
    - The absence of a code in the EMR does not always equate to the absence of a condition or symptom. Unlike in prospective clinical trials where patients are actively asked about what has happened since their last visit to record adverse events, only what the patient reports and what the physician records in a structured field is captured in our structured EMR data. Clinical facts such as occasional over the counter medications, mild illnesses that did not prompt a healthcare visit, and specific symptoms may be underreported.
- Additionally, there are a few publications on this topic linked below that may be useful in referencing as you consider TriNetX as your data source:
    - Opportunities and challenges in using real-world data for health care - PMC (nih.gov)
    - Real-World Evidence | FDA
    - Real-World Evidence: Considerations Regarding Non-Interventional Studies for Drug and Biological Products Guidance for Industry (fda.gov)

2. Are all the variables for my study able to be used in Advanced Analytics or a downloadable dataset?

TriNetX has a duty to protect patient privacy and anonymity. Because certain events can make a patient more easily re-identifiable, we impose restrictions on the use of certain codes in the Advanced Analytics tools and downloadable datasets.

Sensitive terms include diagnosis codes related to specific birth outcomes, body mass index (BMI) over 50, and events that may have been publicly reported like assaults, vehicle accidents, or military operations. All these terms have the potential to raise the likelihood of re-identification to an unacceptable level, and therefore any Query that uses one of

these terms to define a cohort cannot be used in any Advanced Analytics tool or for a downloaded dataset. Additionally, sensitive terms cannot be used as an outcome in any Advanced Analytics tools.

You can find a current list of all Sensitive terms in this article in our FAQ.

CHOOSING A NETWORK

TriNetX offers several networks optimized for use in RWE studies, each suited for specific scenarios:

- **Dataworks-USA and Dataworks-Global** – These networks are available for Industry users. These networks offer unrounded patient counts, but do not offer the ability to see which HCOs patients are coming from. These networks are well suited both for the TriNetX Advanced Analytics tools and for downloadable datasets.
- **Research Network** – This network is available for HCO users and is comparable to the Dataworks networks.
- **Linked** – This network links EMR data to closed claims data. This contains data from US HCOs only.
- **HCO self-networks** – These networks are available to any HCO user to Query the data from their own institution.
- **HCO collaborative networks** – Some HCOs have data-sharing agreements with other HCOs on the TriNetX network and have created networks that are a combination of their self-networks for collaboration.
- **Analytics regional networks** – These networks are available for Industry users. They offer un-rounded patient counts for use in our Advanced Analytics suites, but they do not allow dataset downloads. The most used Analytics network is the Analytics EMEA network; we do not offer downloadable dataset including any of our EMEA HCOs, but this network gives the opportunity to perform RWE studies in the TriNetX platform using our EMEA data.

For more detailed information about regional availability of certain data types and how you can best query patients in each region, check out the Best Practices for Querying Global Data Strategy Guide.

If you are interested in using any of these networks for your research but do not currently have access, please reach out to your TriNetX Account Manager (for Industry users) or Healthcare Partnerships Manager (for HCO users) to learn more.

ADVANCED ANALYTICS OVERVIEW

TriNetX currently offers 7 Advanced Analytics tools on the platform:

- **Analyze Outcomes** – This tool can be used on a single Cohort from a Query and allows the user to select an Index Event. The user can explore patient characteristics in a customizable time window prior to the Index Event and outcomes in a customizable time window following the Index Event. Statistical measures available for outcomes are overall percentage of patients with the outcome, Kaplan-Meier survival analysis, number of instances of the outcome, and lab results.
- **Compare Outcomes** – This tool is very similar to Analyze Outcomes, however, it allows for two Cohorts to be compared. This tool includes the option for 1-to-1 Propensity Score Matching prior to analyzing outcomes.
- **Compare Cohorts** – This tool can be used to compare clinical facts from two Cohorts, cross-sectionally without defining an Index Event.
- **Advanced Explore Cohort** – This tool can be used on a single Cohort and allows the user to select an Index Event and a customizable time window for analysis with reference to the Index Event. There are two main differences between this tool and Analyze Outcomes: First, all patients in the cohort are considered when calculating percentages of patients with a particular code no matter what the time window is, whereas characteristics prior to the Index Event may be calculated from a subset of patients in large Cohorts in Analyze Outcomes. Second, the time window can straddle the Index Event, rather than strictly separating pre-existing characteristics from outcomes.
- **Competing Risks** – This tool can be used on a single Cohort and allows the user to select an Index Event and several outcomes that may be mutually exclusive. The tool then produces an Aalen-Johansen cumulative incidences plot of the different outcomes.
- **Incidence and Prevalence** – This tool can be used on a single Cohort and customizable date ranges. The user can choose one or more outcomes of interest and calculate the incidence and prevalence of the outcomes during the specified date ranges.
- **Treatment Pathways** – This tool can be used on a single Cohort and allows the user to see how patients move through different treatment regimens over time.

You can find recorded webinar videos along with our in-depth Advanced Analytics strategy guide here in our Training Center.

DOWNLODABLE DATASETS OVERVIEW

Downloadable datasets are available and offer patient-level de-identified data. These are a great option if you need additional statistical analysis beyond what is offered in our Advanced Analytics tools or if additional specification of your cohort(s) is needed beyond the logic that is possible in the Query Builder.

If you are interested in learning more about downloadable datasets, please reach out to your TriNetX Account Manager (for Industry users) or Healthcare Partnerships Manager (for HCO users).

Data Dictionaries for TriNetX datasets can be downloaded from the Help & Training Center here.

ETHICAL REVIEW

The data available on the TriNetX platform is de-identified and stripped of data elements that are considered PHI (i.e., age information for patients over 90) or that could otherwise be considered too identifiable and put patient anonymity at risk (i.e., the sensitive terms discussed above). This means that analysis on the TriNetX platform or offline analysis of downloaded datasets is typically not considered "human subjects research" and would typically be exempt from IRB review. However, it is always best practice to submit an exemption request to your local IRB for an official determination. Additionally, this official exemption may be required by journals as part of a manuscript submission.

For HCO users, it may be possible to use TriNetX to identify a patient population and have an honest broker at your institution re-identify the patients in your query for a chart review study. Because this involves the identification of patients, this is considered "human subjects research" and will always require IRB approval prior to the re-identification of patients. For more information on how to work with an honest broker to re-identify patients from your query, please reach out to your Healthcare Partnerships Manager.

WHAT TO INCLUDE IN YOUR PUBLICATION

Be sure to cite TriNetX as the source of the data in your methods section. The following details should be included in the methods section or a supplemental methods section of your publication to maximize reproducibility of your research:

- Description of TriNetX. TriNetX is a Global federated real-world data and analytics platform for research. This publication can be included in the Materials and Methods section to help describe TriNetX.

- The name of the network on which queries were run, along with information like how many HCOs were part of the network at the time of your analysis, as this can change over time.
- The codes used to create your query and at least a basic description of the logic applied to those codes.
- Whether a downloaded dataset or the Advanced Analytics tools were used, and which Advanced Analytics tools were used.
- **Note:** It is recommended that once you have your queries and analyses all set up, you rerun everything on the same date. This may include all Base Analytics (Explore Cohort, Analyze Criteria, etc.) as well as all parts of your analysis. TriNetX networks are frequently updated with new data and there is currently no way to freeze cohorts so that they don't change over time, so running everything at once ensures you have all necessary information, which is helpful when writing your manuscript or responding to reviewer comments. It is also important to note how many patients and HCOs were online at the time the analysis was run.

For studies that are exempt from IRB review, a suggested adequate description of why the study is exempt is:

> "This retrospective study is exempt from informed consent. The data reviewed is a secondary analysis of existing data, does not involve intervention or interaction with human subjects, and is de-identified per the de-identification standard defined in Section §164.514(a) of the HIPAA Privacy Rule. The process by which the data is de-identified is attested to through a formal determination by a qualified expert as defined in Section §164.514(b)(1) of the HIPAA Privacy Rule. This formal determination by a qualified expert refreshed on December 2020."

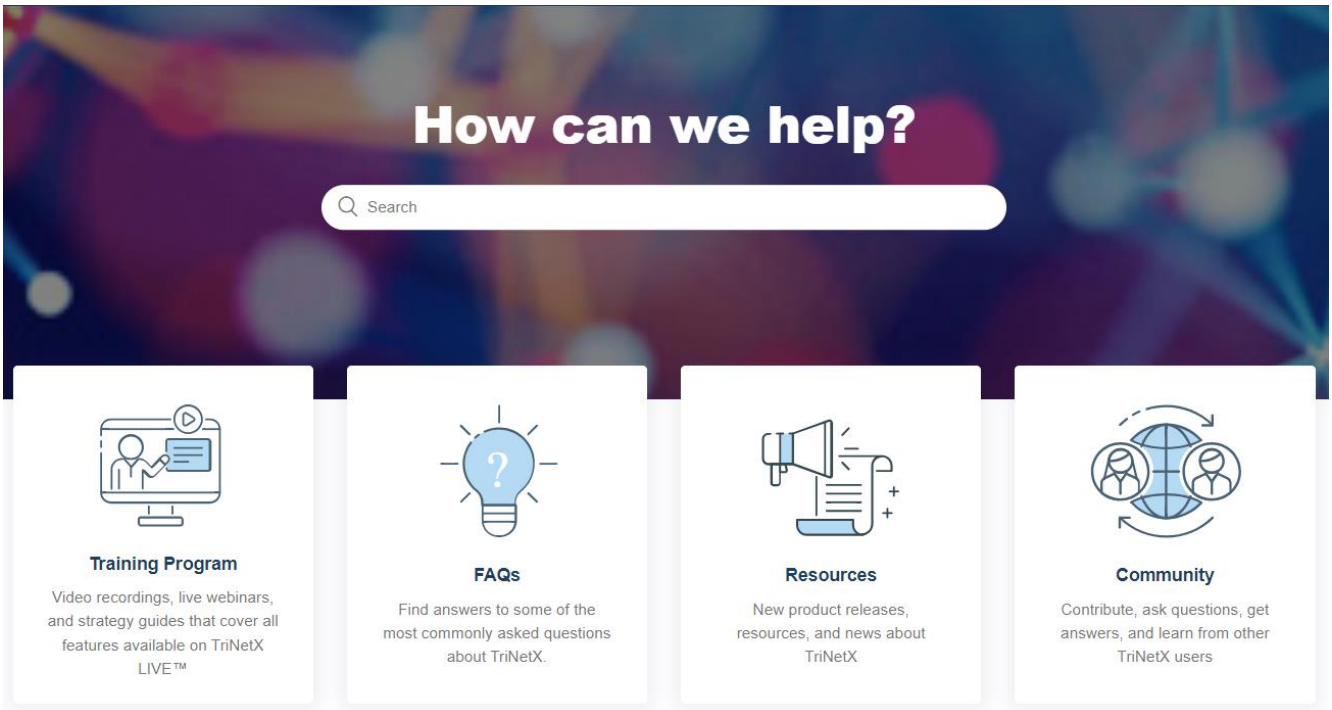
In addition to what should be included, it is important to note what information is considered proprietary and should **not** be included in any published materials:

- HCO names – Unless the publication is performed by researchers at an HCO and uses only data from their own institution, HCO names should not be disclosed.
- Platform screenshots – Screenshots of any part of the TriNetX platform may not be used in published materials without the express written permission of TriNetX. In the case that permission is given, TriNetX will provide appropriate language to be used relating to the screenshot(s).

For additional information on these topics, please refer to our Publication Guidelines webpage.

OTHER RESOURCES

As always, there are additional resources if you require more assistance. The best place to start is the Help Center, where you can find training videos, FAQs, new product releases, and more.



The webinar titled TNX 500 Best Practices in Study Design and Publishing in the **Training Center** goes over the strategies featured in this guide.

Design Assistance—located within the platform—is a great resource to receive assistance specifically relating to your queries. Simply type your question into the design assistance field and your question and study will be sent to an analyst on the TriNetX Clinical Sciences team. You will receive a response within 1 business day.