Robert Wood Johnson Foundation

AcademyHealth

TransUnion Data Description

This is supporting information for the 2020 Health Data for Action (HD4A) Call for Proposals. To apply, or for more information about the funding opportunity, please visit **www.rwjf.org/cfp/hd4a3**.

TransUnion is a global information and insights company that makes trust possible in the modern economy. We do this by providing a comprehensive picture of each person so they can be reliably and safely represented in the marketplace. As a result, businesses and consumers can transact with confidence and achieve great things. We call this Information for Good®.

A leading presence in more than 30 countries across five continents, TransUnion provides solutions that help create economic opportunity, great experiences and personal empowerment for hundreds of millions of people.

TransUnion's Social Determinants of Health (SDOH) research datasets are comprised of depersonalized, event-driven individual-level data and aggregated, neighborhood-level data; providing researchers the opportunity to study the impact of changes in socioeconomic and financial circumstances on health outcomes over time.

Listed below are two (2) individual-level data sets and one (1) aggregated data set that have potential applications for healthcare research. For HD4A research proposals, TransUnion requires researchers to select either individual-level data or neighborhood-level data sets depending on what best suits the research topic. A data dictionary, which contains a list of variables for each category, is available for review after entering into a confidentiality and non-disclosure agreement with TransUnion.

Individual Level Data

Depersonalized SDOH Risk Attributes. TransUnion's depersonalized, non-credit based SDOH risk attributes are derived from over 10,000 data sources, from which data is linked to an identity using our proprietary linking logic. The database from which these attributes are derived is event-driven (not self-reported) and built on a foundation of credit header demographic data and then augmented with proprietary data sources, such as asset ownership information, and public record information, such as information about interactions with the judicial system. In addition, TransUnion has created custom healthcare attributes, which provide insight into housing instability, social isolation, transportation risk and adverse life events.

Depersonalized Archived Credit Histories. TransUnion's depersonalized, archived credit history database contains historical credit information on over 200 million credit active US adults, providing insight into consumers' changing credit utilization and financial situation over time. The credit database comprises of over 180 attributes and proprietary TransUnion scores, with archived historical data available from 2010 onwards and updated quarterly.

Researchers should limit their research proposals to 500,000 consumers, with no more than twelve (12) historical files over a maximum ten (10) year period, for example, eight (8) annual files and two bi-annual files from 2010 – 2020 or twelve (12) quarterly files from 2015 – 2018.

For file delivery, typically, researchers will provide a demographic input file of the study population, and TransUnion will return de-identified data appended via a secure file transfer process.

With TransUnion approval, in some situations, researchers may be permitted to merge external healthcare data files with TransUnion data. In situations where the data may be merged, TransUnion requires researchers to consult with TransUnion about the merging process to ensure adequate safeguards are in place to protect consumer privacy and to prevent reverse engineering of the data to identify any individual consumer. TransUnion can suggest potential third party intermediaries to assist, if required, but there may be an additional cost associated.

Neighborhood Level Data

Aggregated Credit Data (ACD) is a national database containing the median credit information of all credit-active consumers living in a geographic area including ZIP code, ZIP+2 or ZIP+4. The file contains aggregated, modeled credit attributes and proprietary TransUnion risk scores designed to provide insight into neighborhood level credit utilization and financial risk.

Historical aggregated credit data files are also available to researchers upon request, which allows for research into changes in credit utilization and financial distress at the neighborhood level over time. For the HD4A research project, TransUnion proposes to make available two (2) national ACD files, one from Q4 2019 and another from Q4 2020, as the availability of this data may allow researchers to investigate the impact of COVID on neighborhood credit utilization and financial distress. Researchers can request ACD information by providing a list of states, counties or ZIP codes, which are then delivered via secure file transfer.

General information:

For HD4A proposals, TransUnion will license data directly to academic institutions or research-focused organizations for non-commercial use only. Successful applicants are required to complete a Research Data Services Agreement granting access to the TransUnion data for twelve (12) months following file delivery. At the end of 12 months, successful applicants will have the option to renew their data license at current rates or, alternatively, if researchers choose not to renew the license then all data files provided will need to be destroyed. TransUnion requires the right to review and approve all research for accuracy and bias prior to publication.

Research Topics and Proposals:

TransUnion data can support a broad range of topics, including, but not limited to, the following:

- Socioeconomic characteristics and disparities in health
- Socioeconomic risk factors impacting mental health, suicide attempts or opioid addiction
- Analysis of the socioeconomic impact of an SDOH related intervention, e.g. housing or job assistance intervention
- Socioeconomic factors influencing medication adherence
- Geographic analysis of COVID spread and neighborhood credit utilization
- Geographic analysis of whether negative health outcomes correlate to neighborhood economic decline.

Possible research questions might include:

- Investigating whether COVID-19 spread is more prevalent in low-income neighborhoods
- Investigating whether more financially distressed neighborhoods are disproportionally impacted by certain health conditions
- What are the individual and group socioeconomic risk factors that impact healthcare utilization and at what level (e.g. access care, avoid care, seek ED, do not seek PCP, medication adherence, etc.)?
- What chronic diseases have the greatest impact on income and wellbeing?
- Which high cost drugs or treatment paths (i.e. cancer, trauma, etc.) are most correlated with socioeconomic risk?

The above examples are purely for illustrative purposes. We encourage other topics and research questions that are not listed. However, only proposals that adhere to TransUnion's mission of using 'Information for Good' will be considered.

Please contact Jake Carl, Market Leader, Emerging Technologies (Jake.Carl@transunion.com) with any additional questions about the data delivery process or to request a confidentiality and non-disclosure agreement to access the data dictionary.