

ARIC / ARIC-NCS Data Sharing

The ARIC Study (including ARIC-NCS) has a history of being very collaborative. Data sharing is with individuals, data repositories, and research consortia. For the latter, in some instances it is just summary statistics that are shared while for others it is individual-level data.

1. Data sharing with individual investigators

Any scientist who is willing to partner with an ARIC investigator may submit a manuscript proposal or an ancillary study proposal. Once the Publications Committee or Ancillary Studies Committee, respectively, approves the proposal, the investigator and the ARIC Coordinating Center enter into a Data and Materials Distribution Agreement (DMDA). (For an ancillary study requiring use of biospecimens the DMDA also needs to be signed by NHLBI.) After the signing of the DMDA, the Coordinating Center distributes the requested data to the investigator. A nominal fee is charged for the cost of preparing the datasets. The ARIC Publications Committee typically reviews about 200 new manuscript proposals each year. Publications Committee review of manuscript proposals (and completed manuscripts before journal submission), includes those for ARIC-NCS and other ancillary studies.

2. Data sharing with data repositories

A requirement of the ARIC contract is that data be provided to the NHLBI's data repository, BioLINCC. This is mandated not just for data collected under the contract but also for data from ancillary studies, including ARIC-NCS. Visit, annual/semi-annual follow-up call, and surveillance data are submitted to BioLINCC 3 years after the closure of the relevant datasets, so as to give ARIC investigators protected time to analyze the data and publish results. Genotype and relevant phenotype data are also required to be submitted to dbGaP.

The other avenues of data sharing do not impose a similar time constraint before data may be shared. Further, once data are submitted to BioLINCC or dbGaP, those repositories make decisions about data sharing without consulting ARIC investigators. Thus the Coordinating Center takes additional steps to de-identify data sent to these repositories and applies the most stringent versions of participants' informed consent choices, such as removing participants who specify that their data may be used only for CVD research. ARIC is happy to join other repositories including ones specific to aging and dementia.

3. Summary data sharing with consortia

For some consortia, typically those involved in genomic research, individual studies in the consortia undertake their own analyses (using a specified analysis plan). Summary statistics from these analyses are then shared with a group that meta-analyses the results from the various studies. The NHLBI-funded Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium is a notable example that has used this approach very successfully. CHARGE has a large number of relatively autonomous working groups, many with ARIC participation, and has been extremely productive in terms of scientific publications. ARIC has also been involved in a number of manuscripts published by the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) consortium and the Genetic Investigation of ANthropometric

Traits (GIANT) consortium. ARIC is joining new consortia continually with promising new ones including MARKVCID and GAAIN.

4. Individual-level data sharing with consortia

Some consortia request participant-level data from participating studies. Such consortia often undertake pooled analyses rather than meta-analyzing results from the individual studies. ARIC has contributed data and ARIC investigators have been on manuscripts published by the Emerging Risk Factors Collaboration (ERFC), a consortium of more than 130 studies from more than 30 countries, coordinated by the University of Cambridge in Britain. A more recent collaboration of this type is the NHLBI Pooled Cohorts Study, coordinated by Columbia University in New York. This group recently published its first manuscript, <https://www.ncbi.nlm.nih.gov/pubmed/29982273>, with co-authors including ARIC PIs Drs. Folsom and Couper.

5. Hybrid models

The Chronic Kidney Disease Prognosis Consortium, led by ARIC PI Dr. Coresh at Johns Hopkins, is a mix of types 3 and 4 above. Some participating studies provide individual data. Those studies unable to send the individual participant level data (de-identified) are sent a standard computer program that is designed to automatically save all output needed for meta-analyses. Through Dr. Coresh this consortium has access to all the individual-level ARIC data.

Another type of hybrid is the Global Alzheimer's Association Interactive Network (GAAIN) coordinated by the University of Southern California (USC). In this network a program running on a computer at USC has links to data that are housed locally at each study's site. A GAAIN investigator can query the consortium database to obtain information and summary statistics on data at each "data partner". In order to do any analyses the investigator then contacts each of the data partners that has relevant data, with a request for the data (after obtaining approval from the data partner's Publications Committee, where relevant). ARIC has signed a Memorandum of Understanding with GAAIN as a data partner and the next step is to prepare a dataset that can be accessed by the USC computer.

Somewhat similar to GAAIN is the European Union's "Joint Programme - Neurodegenerative Disease Research (JPND)". The JPND hosts a web site with a "Global Cohort Portal" – a searchable catalogue of cohort studies, with information about the studies but no data. Interested investigators then contact each relevant study to request data (after obtaining approval from the study's Publications Committee, where relevant). ARIC has submitted information about the cohort to JPND.