

Dementia DataHub

Annual One-Year Research Pilot Awards

Response to Applicant Questions – Year 2 Cycle

Posted August 4, 2026

Compiled from applicant inquiries received following the RFP release. Responses below are intended as a single group reply to all inquirers. Individual responses to project-specific questions are being handled separately.

Eligibility

Are there citizenship, residency, or visa requirements?

There are no citizenship or visa requirements for applicants to the DDH pilot awards. The award is made to your home institution, and your institution is responsible for ensuring that your visa allows you to be productive on the project for the full award period. We encourage you to apply.

Can an assistant professor past the 10-year Early Stage Investigator window apply?

Our preference is for applicants who meet the NIH definition of an Early Stage Investigator. That said, the ESI window is not a hard eligibility bar. If you are new to dementia research or to Medicare claims research and believe you are at a comparable career stage in substance if not in years, you may apply and should briefly explain your circumstances in your cover page. The review committee will consider it.

Can a collaborator at another institution participate?

Yes. The award itself goes to a single institution, and we are not able to split it. However, there is nothing preventing a co-investigator at another institution from participating in the work, joining calls with the paired DDH research scientist, or taking part in the analytic discussions. Name your collaborator in the application and describe their role.

How the awardee institution chooses to use its funds is up to them. These are modest awards, and we would rather see two engaged people working on a good question than turn away a productive collaboration on administrative grounds.

Budget and Indirect Costs

Is there a cap on indirect costs?

There is no separate indirect cost cap because there is no separate indirect cost category. The award amount is a fixed total that must cover all costs to your institution, including any institutional overhead. You propose a total amount, we negotiate it, and payments are then made against milestones on a reimbursable basis.

We recognize that institutions with high overhead rates will have less purchasing power at a given award amount. We do not have a mechanism to supplement awards to offset this, and we are not able to require or request that institutions waive overhead. Applicants should propose a

total amount that lets them accomplish the work they are proposing at their institution. If that number is higher than what the mechanism can support for the scope described, the project may not be feasible at this scale, and we would encourage you to consider a more tightly scoped question.

Public Use Files: Access and Timing

How do I request the public use files?

Use the [DDH data request form](#). No data use agreement is required. The public use files are public, and there are no application materials, eligibility requirements, or approvals involved. We typically respond within a few business days.

We strongly encourage requesting and reviewing the files before you write your proposal.

Which files are available now, and when will the rest be released?

The 2020, 2021, and 2022 annual files are available now, along with their codebooks.

The pooled 2020 to 2022 summary file is expected to be approved for release before the second week of August 2026. This file aggregates three years of data to reduce cell suppression and supports analysis at finer geographic levels, including hospital referral region and hospital service area.

The pooled comorbidity file is not yet available and is expected by September at the latest.

Both pooled files will be available to awardees during the project period. If your proposal depends on one of them, say so in your approach section and describe how you would proceed if the release timing shifted.

Restricted Medicare Data: What Is Available

Several applicants have asked what is actually in the restricted files. Here is what we can tell you.

Years available. For purposes of pilot projects, Medicare claims are available from 2018 forward. This is a wider window than the public use files, which currently cover 2020 through 2022. If your question benefits from a pre-pandemic baseline, that is available to you.

File types. The restricted data include 100 percent Medicare fee-for-service claims (Parts A, B, and D), Medicare Advantage encounter data, and the Master Beneficiary Summary File.

Assessment files. Several applicants asked about CMS assessment data. Confirmed accessible files and year ranges are:

Assessment file	Setting	Years accessible
MDS 3.0	Nursing facility	2019 to 2025
OASIS	Home health	2020 to 2023
IRF-PAI	Inpatient rehabilitation	2020 to 2023

Medicaid claims data are available from 2018 through 2023.

The Hospice Item Set is not available. Hospice utilization can still be studied through the hospice claims in the fee-for-service files, so what you lose is the assessment items rather than the ability to examine hospice use.

Assessment files are large and linking them to claims adds meaningful cohort construction time. If your design depends on assessment data, scope accordingly and say so in your approach section.

Residential geography. The beneficiary file includes state and county FIPS at monthly resolution, and a single ZIP code drawn from the mailing address. Because these are annual files, a beneficiary who moved across years will show different values across years, so you do have a residential history rather than a single fixed baseline. County is the finer-grained option temporally. ZIP is available but at annual resolution only. If your project involves assigning geographic exposures, design around county unless you specifically need ZIP-level precision, and be aware that mailing address is not always residence.

Variable lists. We are not able to distribute a NORC-specific data dictionary, but the underlying files are standard CMS research files and ResDAC publishes full documentation for all of them. Start here:

- [Find the CMS data file you need](#) for an overview of each file and what it contains
- [Master Beneficiary Summary File Base A/B/C/D](#) for enrollment, demographic, and geographic variables
- [Inpatient](#), [Carrier](#), and [MedPAR](#) for the claims files most pilot projects use

You can also search individual variables through the Search Data Variables tool on the ResDAC site. Reviewing this documentation before you write your approach section will help you scope realistically.

What VRDC Programming Support Can and Cannot Do

This is the question we have received most often, and it is the one most likely to determine whether a proposal is fundable. Please read this section before finalizing your approach.

What is supported

Programmers can construct beneficiary-level analytic cohorts, produce aggregated tabulations, and fit standard regression models. That includes linear and logistic regression, Poisson and negative binomial models, and survival models such as Cox proportional hazards. Interaction terms and stratified models are fine. These are ordinary procedures and they run quickly.

What is not supported

The environment is not suited to methods that require large numbers of model refits. Bootstrapped confidence intervals, causal mediation analysis with bootstrap-based indirect effects, simulation-based estimation, and machine learning approaches involving cross-validation or tuning are all beyond what the available programming support can absorb. If your analytic plan depends on one of these, the project is unlikely to be feasible at this scale.

Where the budget actually goes

Applicants often assume the model is the expensive part. It is not. Constructing the cohort is the expensive part. Identifying an eligible population, applying enrollment and washout criteria, following beneficiaries across years, and building analytic variables from 100 percent Medicare claims will consume most of the available programming allocation before a single model runs. Once the cohort exists, tabulation and estimation are inexpensive.

The practical implication is that you should propose one well-specified analytic file supporting one well-specified question, rather than a family of related analyses.

Design toward coefficients, not tables

Everything extracted from the VRDC goes through CMS disclosure review. Cells containing fewer than 11 beneficiaries cannot be released, and neither can cells from which a count below 11 could be recovered by combining other released cells. This second requirement is the one that surprises people. A large stratified table can fail review even when most of its cells are large.

Regression coefficients, standard errors, and model fit statistics are not counts and generally release cleanly. A detailed cross-tabulation stratified by state and demographic group often does not.

If your question can be answered either by tabulating variation or by modeling it, model it. You will get a more publishable result and a much easier disclosure path. We can usually release a national-level descriptive table alongside the model, which is what journals want anyway.

Plan for a slow clock

Every extraction goes through disclosure review, and the calendar cost of a round trip is larger than the dollar cost. Plan on two or three extraction cycles across the project year. Do not build a proposal that assumes you will iterate six times.

Consider whether you need the restricted data at all

Some of the strongest questions we have seen can be answered using the public use files alone, or the public files linked to other county-level public sources. If your analysis works at the county level, you may not need the VRDC. That path is faster, cheaper, and entirely under your own control. It is also fully competitive for this award. We are not scoring proposals on whether they use restricted data.

Linking DDH or Medicare Data to Other Sources

Already approved for linkage. NORC's data use agreement covers a set of external files that can be linked to the restricted Medicare data without additional approvals. If your question can be answered using one of these, that is by far the fastest path and we encourage you to design toward it. The approved files and their linkage keys are listed below.

Other publicly available data. Other public data can generally be brought into the VRDC, but each new source must be added to our DUA first. This takes time, and the timeline is not fully within our control. If your project depends on a public source not already on the list, tell us in

your letter of intent or in your application so we can start that process early. Do not assume it can be added mid-project.

Non-public data. If you want to bring in data that is not publicly available, you would need to obtain approval under your own data use agreement with the data holder and provide us documentation of your permission to bring it into the VRDC. This is possible but adds meaningful time and risk to a one-year project.

Linking at the county level using public DDH files. If your analysis works at the county level, you may not need the VRDC at all. The DDH public use files can be linked to ACS, County Health Rankings, or any other county-level public source on your own machine, with no DUA involvement. Several published studies have taken this approach.

Approved files for linkage

Approved file for linkage	Linkage key
American Community Survey (ACS)	County
American Hospital Association Annual Survey	Hospital name or Medicare Provider Number
AHRQ Compendium of U.S. Health Systems	Hospital name or Medicare Provider Number
Provider, Enrollment, Chain and Ownership System (PECOS)	PECOS ID, Medicare Provider Number, NPI
Medicare Inpatient Impact File	Medicare Provider Number
Revalidation Clinic Group Practice Reassignment Data	NPI
Medicare Fee-For-Service Public Provider Enrollment	NPI
HRSA Office of Pharmacy Affairs (OPAIS) 340B Program data	Medicare Provider Number and/or provider name
NIC Map Vision Seniors Housing Locations	ZIP code
AHRQ Social Determinants of Health data	ZIP code
Medicare Provider of Services Public Use File	Medicare Provider Number
Medicare Hospital Cost Reports	Medicare Provider Number
Medicare Physician and Other Supplier Public Use File	Medicare Provider Number
NPPES	Medicare Provider Number / NPI
PitchBook	(Derived) Medicare Provider Number
Other public Medicare fee schedule rulemaking and impact files	Medicare Provider Number
Medicare Advantage rate book, plan benefit package, and other public files	Medicare Part C Plan ID
Publicly available MCO information	Medicaid MCO Plan ID and plan name
UW Population Health Institute County Health Rankings and Roadmaps	County FIPS
FCC Fixed Broadband Deployment Dataset	ZIP code

Race and Ethnicity Detail

Can specific Asian subgroups be identified?

No. Medicare race and ethnicity data use the RTI race code categories, which include a combined Asian and Pacific Islander category. Asian Indian is not a separate category and cannot be identified at the individual level in these data. The same limitation applies to other Asian subgroups.

There is no supplemental CMS file or linked source that resolves this. Individual-level Asian subgroup identification is not available.

What you can do instead?

An ecological approach is feasible and can be scientifically defensible if framed carefully. Because American Community Survey data are already approved for linkage at the county level, you can attach county-level ACS estimates of the share of residents reporting a specific ancestry and use that share as a covariate or stratifier in an analysis of Asian and Pacific Islander beneficiaries.

This estimates something different from what individual-level identification would estimate, and the difference matters. You would be describing how outcomes among Asian and Pacific Islander beneficiaries vary with the ancestry composition of the county where they live. You would not be describing outcomes among individuals of a particular ancestry. A proposal taking this approach should say so plainly and should not describe its findings in individual terms.
