

RESEARCHERS DATA DICTIONARY

# Genetic Data (RDD-Gen)

Copyright© 2015 University of Washington

Created and published by the National Alzheimer's Coordinating Center (Walter A. Kukull, PhD, Director). All rights reserved. This publication was funded by the National Institutes of Health through the NIH/ National Institute on Aging (Cooperative Agreement U01 AG016976).

## Introduction

The *Researcher's Data Dictionary-Genetic Data* (RDD-Gen) is intended to be the primary resource for researchers interested in identifying UDS and/or MDS subjects for whom genetic data are available. The RDD-Gen describes variables that contain either genetic data (APOE genotype) or information about the availability of genetic data that can be obtained, by request, from the Alzheimer's Disease Genetics Consortium (ADGC), the National Institute of Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS), or the Database of Genotypes and Phenotypes (dbGaP).

Note that updates to these data and inclusion of genetic data for additional UDS subjects are obtained per the discretion of ADGC and the individual Alzheimer's Disease Centers (ADCs).

### Definitions

- Variables with source type **ADGC** are coded exactly as they are provided by ADGC.
- **Derived variables** are developed by NACC. These variables provide information that is collected indirectly from data in the Uniform Data Set (UDS), Neuropathology (NP) Data Set, the individual ADCs, the National Cell Repository for Alzheimer's Disease (NCRAD) and ADGC — for example, **NACCAPOE** provides the APOE genotype for each subject, when available. This information is obtained from ADGC, the individual ADCs, NCRAD, and the NACC Neuropathology Data Set. This derived variable combines the four sources of information so that one variable captures all available data.

### Requesting genotype and sequencing data available at ADGC, NIAGADS, or dbGaP

Genotype data may be obtained from ADGC or NIAGADS; however, exome genotype data may be obtained from ADGC only. Exome sequencing data may be obtained from dbGaP. Please contact these groups directly for access to the genotype and sequencing data.

Alzheimer's Disease Genetics Consortium . . . . . <http://www.adgenetics.org/>  
 NIA Genetics of Alzheimer's Disease Data Storage Site . . . . . <https://www.niagads.org>  
 Database of Genotypes and Phenotypes . . . . . <https://www.ncbi.nlm.nih.gov/gap>

### ABBREVIATIONS

<b>NACC</b>	<b>National Alzheimer's Coordinating Center</b>
<b>ADGC</b>	<b>Alzheimer's Disease Genetics Consortium</b>
<b>NCRAD</b>	<b>National Cell Repository for Alzheimer's Disease</b>
<b>NIAGADS</b>	<b>National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site</b>
<b>dbGaP</b>	<b>Database of Genotypes and Phenotypes</b>
<b>ADC</b>	<b>NIA/NIH Alzheimer's Disease Center</b>

## Table of Variables

	Variable name	Short descriptor	Data type	Data source
1	<b>ADGCGWAS</b>	Genotype data are available from ADGC (y/n)	Numeric cross-sectional	ADGC
2	<b>ADGCRND</b>	ADGC data-selection round	Character cross-sectional	ADGC
3	<b>ADSPSEQ</b>	ADSP sequencing completed (y/n)	Numeric cross-sectional	ADGC
4	<b>ADSPWGS</b>	ADSP whole genome sequencing completed (y/n)	Numeric cross-sectional	ADGC
5	<b>ADSPWES</b>	ADSP whole exome sequencing completed (y/n)	Numeric cross-sectional	ADGC
6	<b>NACCGDS</b>	Genotype data are available from NIAGADS (y/n)	Numeric cross-sectional	ADGC
7	<b>NACCDBGP</b>	Exome sequencing data are available from dbGaP (y/n)	Numeric cross-sectional	ADGC
8	<b>NACCNCRD</b>	Samples are available from NCRAD (y/n)	Numeric cross-sectional	NCRAD
9	<b>NACCAPOE</b>	APOE genotype	Numeric cross-sectional	NACC derived
10	<b>NACCNE4S</b>	Number of APOE e4 alleles	Numeric cross-sectional	NACC derived

# Variable Definitions

1	Variable name	<b>ADGCGWAS</b>
	Short descriptor	Genotype data are available from ADGC (y/n)
	Data type	Numeric cross-sectional
	Data source	ADGC
	Allowable codes	0 = No 1 = Yes
	Description / derivation	Indicator of whether genotype data are available from ADGC.
2	Variable name	<b>ADGCRND</b>
	Short descriptor	ADGC data-selection round
	Data type	Character cross-sectional
	Data source	ADGC
	Allowable codes	ADC 1 = Round 1 ADC 2 = Round 2 ADC 3 = Round 3 ADC 4 = Round 4 ADC 5 = Round 5 ADC 6 = Round 6 ADC 7 = Round 7 ADC 8 = Round 8 ADC 9 = Round 9 AA = African American round 88 = Not applicable/ no genotype data available 99 = Missing/ could not be determined
	Description / derivation	This variable indicates the GWAS round in which this subject's sample was analyzed.
3	Variable name	<b>ADSPSEQ</b>
	Short descriptor	ADSP sequencing completed (y/n)
	Data type	Numeric cross-sectional
	Data source	ADGC
	Allowable codes	0 = No 1 = Yes
Description / derivation	Indicator of whether ADSP (either whole genome or exome) sequencing completed.	
4	Variable name	<b>ADSPWGS</b>
	Short descriptor	ADSP whole genome sequencing completed (y/n)
	Data type	Numeric cross-sectional
	Data source	ADGC
	Allowable codes	0 = No 1 = Yes
Description / derivation	Indicator of whether ADSP whole genome sequencing completed.	

5	Variable name	<b>ADSPWES</b>
	Short descriptor	ADSP whole exome sequencing completed (y/n)
	Data type	Numeric cross-sectional
	Data source	ADGC
	Allowable codes	0 = No 1 = Yes
	Description / derivation	Indicator of whether ADSP whole exome sequencing completed.
6	Variable name	<b>NACCNGDS</b>
	Short descriptor	Genotype data are available from NIAGADS (y/n)
	Data type	Numeric cross-sectional
	Data source	ADGC
	Allowable codes	0 = No 1 = Yes
	Description / derivation	Indicator of whether genetic data are available from NIAGADS.
7	Variable name	<b>NACCNCRD</b>
	Short descriptor	Samples are available from NCRAD (y/n)
	Data type	Numeric cross-sectional
	Data source	NCRAD
	Allowable codes	0 = No 1 = Yes
	Description / derivation	Indicator of whether samples are available from NCRAD.
8	Variable name	<b>NACCCDBGP</b>
	Short descriptor	Exome sequencing data are available from dbGaP (y/n)
	Data type	Numeric cross-sectional
	Data source	ADGC
	Allowable codes	0 = No 1 = Yes
	Description / derivation	Indicator of whether genetic data are available from dbGaP.
9	Variable name	<b>NACCAPOE</b>
	Short descriptor	APOE genotype
	Data type	Numeric cross-sectional
	Data source	NACC derived
	Allowable codes	1 = e3,e3 2 = e3,e4 3 = e3,e2 4 = e4,e4 5 = e4,e2 6 = e2,e2 9 = Missing/ unknown/ not assessed
	Description / derivation	APOE genotype is run independently by the ADC and reported to NACC on the NACC Neuropathology Form. APOE genotype is also reported by ADGC and NCRAD. In the rare case that the ADC-reported genotype and the genotype reported by ADGC are not the same, the genotype is set to <b>9 = Missing</b> for that subject.

10	Variable name	<b>NACCNE4S</b>
	Short descriptor	Number of APOE e4 alleles
	Data type	Numeric cross-sectional
	Data source	NACC derived
	Allowable codes	0 = No e4 allele 1 = 1 copy of e4 allele 2 = 2 copies of e4 allele 9 = Missing/ unknown/ not assessed
	Description / derivation	APOE genotype is run independently by the ADC and reported to NACC on the NACC Neuropathology Form. APOE genotype is also reported by ADGC and NCRAD. In the rare case that the ADC-reported genotype and the genotype reported by ADGC are not the same, the genotype is set to <b>9 = Missing</b> for that subject. We used the code for APOE genotype to create a new variable indicating the number of e4 alleles.