COLLABORATIVE STUDY ON THE GENETICS OF ALCOHOLISM (COGA)
DATA FILE

DOCUMENTATION:

The Collaborative Study on the Genetics of Alcoholism (COGA) is a multi-site, multi-disciplinary undertaking with the overall goals of characterizing the familial transmission of alcoholism and related phenotypes and identifying susceptibility genes using genetic linkage. The study is being coordinated by the SUNY Health Science Center at Brooklyn (HSCB) under the leadership of Henri Begleiter. The study was initially funded by the National Institute of Alcohol Abuse and Alcoholism (NIAAA) in 1989.

Interviewing and testing of COGA families is conducted at six university centers: SUNY Healthy Science Center at Brooklyn, University of Connecticut, Indiana University, University of Iowa, University of California in San Diego, and Washington University. All six sites carry out the identical study protocol. These institutions continue to be responsible for maintaining scientific and procedural standards as well as quality control in their areas of expertise including Ascertainment, Psychiatric Assessment, Neurophysiology, Molecular Biology, and Genetic Analyses. The scope and scale of COGA also necessitates a complex data management system. The data management and repository functions are shared based on specialty by Washington University (clinical), Indiana University (pedigree), and SUNY (electrophysiology). Washington University doubles as the overall data management and repository coordinating center. In addition, a cell and DNA repository is managed by Jay Tischfield at Rutgers University.

A description of COGA, instructions for gaining access to clinical data, DNA samples, interview instruments, scoring algorithms, and the distribution agreement, are available on the World Wide Web at http://zork.wustl.edu/niaaa.

This documentation describes data and biomaterials currently available to the scientific community, for WAVE I pedigrees of 105 families described in http://zork.wustl.edu/niaaa.cogatable.html

The specific aims for the Wave I collection were to:

1). Ascertain approximately 100 pedigrees with 3 or more living related individuals affected with alcoholism.

2). Obtain blood samples on affected relatives, unaffected sibs and parents of probands.

3). Design and implement state-of-art standardized and objective procedures for establishing a lifetime diagnosis of alcoholism.

4). Systematically extend pedigrees to relatives affected with alcoholism and related conditions.

5). Maintain longitudinal follow-up with subjects and facilitate ongoing access to sources of information used in making diagnostic estimates, including medical records and family history data systematically collected from relatives.

PEDIGREE ASCERTAINMENT:

Ascertainment in Phase I: Potential probands were identified by random consecutive sampling of inpatient and outpatient alcoholism treatment facilities. Inclusion criteria for probands were a lifetime diagnosis of alcohol dependence by DSM-III-R criteria and definite alcoholism by Feighner criteria. The co-occurrence of these diagnoses is called the "COGA alcohol dependence criteria" and is used throughout the study for ascertainment and analysis. A complete description of ascertainment procedures are available in the Ascertainment Procedures Manual.

Those chosen for participation are divided in to the following groups:

STAGE I Families: Probands and families meeting these criteria were designated STAGE I and all available first-degree relatives were personally interviewed with the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA), Family History Assessment Module and self-administered personality questionnaire.

STAGE II Families: Families with three or more first-degree alcohol dependent relatives meeting COGA criteria were designated STAGE II families. They were invited for more extensive testing, including neurophysiology tests (ERPs and EEGs) and a battery of neuropsychological assessments. Blood was also obtained for biological marker studies and the production of lymphoblastoid cell lines and/or DNA. Pedigrees were extended into second and third degree branches if family history assessment revealed the presence of other affected relatives. A standard STAGE II protocol was administered to all first-degree relatives of affected individuals. Extensions were also made by "leapfrogging". Leapfrogging occurs over a living or dead unaffected member into a branch containing at least two first-degree relatives of the "leapfrogee" who had been implicated as alcohol dependent by family history. Families with evidence of bilinear alcohol dependence were not included in the extended assessment protocol.

CONTROL Families: Community dwelling control families were identified by a variety of sources at different COGA centers including drivers' license record, attendees at medical/dental clinics, advertisements mailed to random university students and random ascertainment of community dwelling families.

STAGE IV Genotyped Sample: Stage II families were reviewed to select those that were suitable for genetic linkage and candidate gene analyses. After the families were pruned to eliminate uninformative individuals and branches, their DNA was sent to COGA labs for genotyping. These families have been designated STAGE IV and two separate samples were selected for genetic studies. The first sample of 105 multigenerational families (992 genotyped individuals) was used to conduct genome wide genetic linkage

and candidate gene studies of alcohol dependence and related phenotypes (Reich et al,., 1998).

STRUCTURED CLINICAL INTERVIEW - SSAGA I:

The Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA), expressly developed for COGA, is a polydiagnostic psychiatric interview that covers the major psychiatric disorders in DSM-III-R and provides complete diagnosis in DSM-III-R and ICD-10 as well as diagnoses for Substance Dependence in Feighner and DSM-IV.

Companion instruments for use with children, ages 7-12 (C-SSAGA-C), adolescents, ages 13-17 (C-SSAGA-A) and for interviewing parents about their children (C-SSAGA-P) are also available.

The reliability of the SSAGA has been assessed in relation to both rater test-retest (Bucholz et al., 1994) as well as a comparison of raters across COGA Centers (Bucholz, Hesselbrock, Shayka, et al., 1995). Test-retest reliabilities for lifetime DSM-III-R alcohol and other drug dependencies as well as major depressive disorder and the antisocial personality disorder were high, with agreement ranging from Kappa .70-.90. The cross-center agreement was also acceptable for alcohol and other drug dependencies, with Kappa ranging from .57 to 1.00, except for stimulant dependence (K=.44).

A second study (Hesselbrock, M. et al., 1999) examined the concurrent diagnostic validity of the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) across alcohol and drug dependencies, major depression, anxiety disorders, and ASPD in relation to The Schedule for Clinical Assessment in Neuropsychiatry (SCAN; WHO). Concordance between the two instruments for the diagnoses surveyed was good. Kappas for substance dependence range from K=.63 for alcohol dependence to K=.85 for cocaine and stimulant dependence. Kappa for major depression was .71 for ASPD K=.70, and Kappa =.62 for panic disorder. These data, combined with results from two previous studies which examined reliability, indicate that the SSAGA is a highly reliable and valid instrument for use in studies of a variety of psychiatric disorders, including alcohol and drug dependence.

Unique features of the SSAGA compared to other structured research interview include attention to making diagnoses according to several criteria systems, the addition of nondiagnostic items for phenotyping of alcoholism, and attention to comorbidity of alcohol/substance diagnoses in relation to other non-substance abuse disorders.

The SSAGA interview, Specifications, and a copy of the SSAGA with variable names are available at the website http://zork.wustl.edu/niaaa.

DIAGNOSTIC CATEGORIES:

Diagnostic categories used to characterize the sample were:

1). COGA criteria = positive for both Feighner and DSM-III-R
2). DSM-IV
3). ICD-10

THE GENETIC SAMPLE:

On two occasions, in 1996 and in 1997, all Phase I STAGE IV sample of families were reviewed by the COGA Ascertainment Committee to select families and branches of families that were suitable for the genetic linkage and candidate gene studies. Pedigrees were pruned by eliminating uninformative (for linkage studies) individuals and branches of these pedigrees. Average pedigree size was reduced approximately 25% by this process (from 12.43 to 9.0 individuals). Phenotypes used to select these samples included Alcohol Dependence by COGA and ICD-10 criteria, being "unaffected" (defined as drinking alcohol but having no symptoms of any form of alcohol abuse or dependence), and having had a EEG and ERP neurophysiological assessment. Individuals were also retained in the sample to maximize the use of identity by descent (IBD) statistics in affected sibling and other relative pairs. DNA from the selected individuals in the genetic study was then sent to COGA genetic laboratories at Washington and Indiana Universities. The Wave 1 (1996) sample comprises a total of 105 pedigrees including 176 nuclear families and 979 genotyped individuals. They have all been genotyped across the entire genome.

For the public release, 14 subjects refused to share DNA with other than COGA investigators, and these individuals are excluded from the distribution file.

A table describing the current database is available on the web site (http://zork.wustl.edu/niaaa/)

AVAILABLE DATA SETS:

Distribution File:

AVAILABLE DATA SETS

Distribution file (cogaped)
Contains basic demographic and family structure information, alcohol diagnoses, and summary scores for personality assessments. A complete description of the contents is available at the link "Contents of Distribution File".
For the alcohol diagnoses, 0 represents missing information. All other variables are blank when missing.
This file is available both in ASCII format and as a SAS transport file.

Other data sets (all available as SAS transport files)

SSAGA-I (ssaga):

Contains the complete SSAGA-I database. A copy of the instrument, with variable names specified, is available at the link ASSAGA-I with Variable Names". (This document is organized by section.)

For missing single digit variables the value 9 is used. Missing multiple digit variables are represented by the code -9.

Diagnosis file (dx_ssaga):

This is a database of all diagnoses derived from the SSAGA-I..

A description of the contents is available at the link "Contents of Diagnosis File".

In general, the coding of diagnoses is as follows:

1=not affected

3=affected, but dx might be due to some other disorder or condition

5=affected

9=uncertain (crucial information was missing)

. (missing): The relevant interview questions were not asked

For the Feighner alcohol diagnosis, 9 = 'Probable'.

Neuropsychological file (neuropsy):

Contains summary scores derived from the neuropsychological battery.

Missing values (all numeric) are represented by the SAS . convention.

A description of the contents can be found at the link AContents of Neuropsychiatric File@.

Visual ERP file (Vp3):

Contains amplitude and latencies at 19 leads for target and non-target peak p3 and N1. A description of the contents can be found at the link "Contents of ERP (Visual) File". Data are available on 598 individuals.

Auditory ERP file (aod):

Contains amplitudes and latencies at 19 leads for target and non-target peak P3, N1 and P2. A description of the contents can be found at the link "Contents of ERP (Auditory) File". Data are available on 564 individuals.

LIST OF VARIABLES IN THE DISTRIBUTION FILE

****

Column	Variable	Coding/Description
1-8	Unique subject ID	Subject ID number
10-14	Unique family ID	
16-23	Unique father ID	Father ID number (= 0 for founders , marry-ins)
25-32	Unique mother ID	Mother ID number (= 0 founders, marry-ins)
34	Sex	M = male, F = female
36-43	Cell Repository ID	Blank if no DNA available
45	Vital Status	1 = dead, blank = alive
47-48	Age in years	This is the age at SSAGA administration (if given)
50-53	Year of birth	
55	Proband status	1 = proband, otherwise 0
57	Ethnicity (self-reported)	0 = no info 1 = American Indian 2 = Asian 3 = Pacific Islander 4 = Black, non-Hispanic 5 = Black, Hispanic 6 = White, non-Hispanic 7 = White, Hispanic 8 = other
59	Twin Status	M = Monozygotic D = Dizygotic Blank if singleton

61	ALDX1	Alcohol Dependence (DSM-III-R plus Feighner) 0 = no info 1 = never drank 2 = pure unaffected 3 = unaffected with some symptoms 4 = affected
63-64	ALDX1AO	Age of onset for DSM-III-R
66	ALDX2	Alcohol dependence - DSM-IV Coded same as ALDX1
68	ALDX2AO	Age of onset of ALDX2
71	ALDX3	Alcohol dependence - ICD-10 Coded same as ALDX1
73	ALDX3AO	Age of onset of ALDX3
75-76	TPQ_HA	TPQ Harm avoidance subscale
78-80	TPQ_NS	TPQ Novelty seeking subscale
82-83	TPQ_RD	TPQ Reward dependence sub-scale
85-86	SSV	Zuckerman sensation seeking scale score
88-92	MAO Activity	