Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Web Database**

GCSDB: an integrated database system for the Georgia Centenarian Study

Jianliang Dai[2], Adam Davey[1], Ilene C. Siegler[3], Jonathan Arnold[2*] and Leonard W. Poon[2, 4]

[1] Temple University; [2] The University of Georgia; [3] FN Duke University Medical Center; [4]The Georgia Centenarian Study; Jonathan Arnold* - Email: arnold@uga.edu; * Corresponding author

received September 05, 2006;revised October 02,2006; accepted October 02, 2006; published online October 07, 2006

Abstract:

GCSDB is a web-oriented integrated database system for the Georgia Centenarian Study, a phase III, population-based, multidisciplinary study of centenarians. The Study recruited 244 centenarians and near-centenarians (age 98 and older), 80 octogenarians and 400 young controls in Northern Georgia. GCSDB incorporates more than 40 relational tables containing data about the participants including demographics, family longevity, physical health, cognition, neuropsychology, mental health, neuropathology, functional capacity, and genetics. The GCSDB web site includes detailed information about these tables and functions for genetic and other kinds of data analysis. More data and functions will be added as the study progresses. GCSDB provides a resource that could be used to identify what biological, psychological, and social factors as well as their epistatic interactions help these centenarians achieve long life.

Availability: http://qa.genetics.uga.edu (login information can be obtained from authors)

Keywords: centenarian; cognition; database; longevity; mental health; neuropathology; neuropsychology; functional capacity; Single Nucleotide Polymorphism (SNPs)

Background:

Over the past two decades, interest in study of centenarians has increased steadily **[1, 2, 3]**, and several countries have initiated their own centenarian studies, such as the United States **[4, 5, 6]**, Japan **[7]**, Italy **[8]**, Hungary **[9]**, France **[10]**, Sweden **[11]**, Finland, and Denmark. **[12]** The fundamental question is how centenarians live longer and what specific biological, psychological, and sociological factors help centenarians survive to become the oldest of the old. **[1, 2, 13]** Bringing together these data from multiple disciplines in ways that are simple and useful is a common problem that must be solved for research to progress.

Phase 3 of the Georgia Centenarian Study (2001-2007) is a multi-disciplinary program project designed to identify further the roles of biological, psychological, and social factors as well as their epistatic interactions contributing to the successful late-life aging. **[2]** Four specific projects within the study are to examine: 1) genetic structure of the Georgia population with a focus on the oldest-old; 2) relations between dementia and neuropathology among centenarians with additional neuropsychological data collected for persons who give permission for autopsy; 3) determinants of functional capacity from neuropsychological, cognitive, health, and demographic variables; and 4) predictors differentiating centenarians who are independent, healthy, and experience a sense of well-being from those who are dependent, frail, and do not experience a sense of well-being.

A population-based sample of individuals was recruited from 44 counties in Northeast Georgia, consisting of 244 centenarians and near-centenarians (age 98 and older), 80 octogenarians primarily for the cognitive and adaptation studies, and 400 controls in equal numbers from the 2nd to 5th decades of life for the genetic studies. Data were collected using standardized data forms and protocols by interviewers from the Data Acquisition Core, then scanned, checked, corrected, verified, and saved as separate electronic files according to the source of the data using the Teleform software package. **[14]**

For ease of data management and data sharing in a multidisciplinary and multi-institutional setting, a web-oriented, integrated database-GCSDB-was established to serve as a resource for effective collaborative research. This changes the paradigm for most social science studies. In this paper, we report on the status of GCSDB and highlight the database table features and web interface functions that were incorporated or developed for genetic data analysis as well as for other data analyses. The database and its interface are designed to promote cross-project questions and analyses of the data.

Methodology:

Database Implementation

GCSDB was developed on a Sun microsystems SunBlade with the Solaris 5.8 operating system; database management was Oracle version 10g. GCSDB was

constructed following the relational database schema from the data collection instruments and booklets. All tables in GCSDB share a common field-the unique participant ID number as the primary key (GCSID) for linking tables generated by multiple projects and/or cores. Web interface to the database is produced from Java Sever Page (JSP) technology and the Struts framework under the SQL query tool. Web pages are served using the Apache-Tomcat web server version 5.0.30. In addition to the relational database, underlying statistical analyses were implemented via Java or FORTRAN.

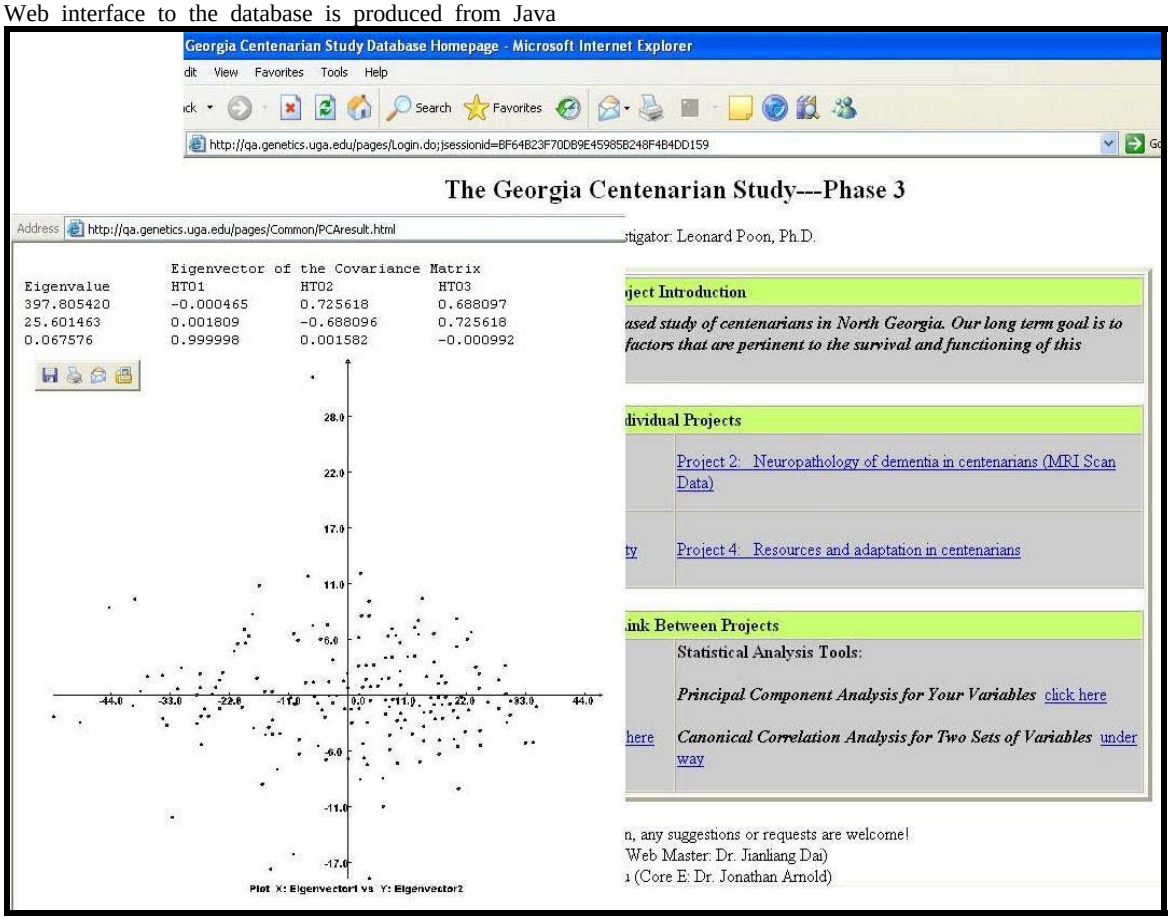

Figure 1: Composite of screen displays demonstrating the homepage of GCSDB. (1) A brief introduction of The Georgia Centenarian Study Project with a link to a detailed introduction. (2) Four links to the web pages of four individual projects (see Figure 2 for Project 3). (3) Methods to cross link between four projects. The lower left-hand corner shows the example result of Principal Component Analysis, PCA

Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Web Database**

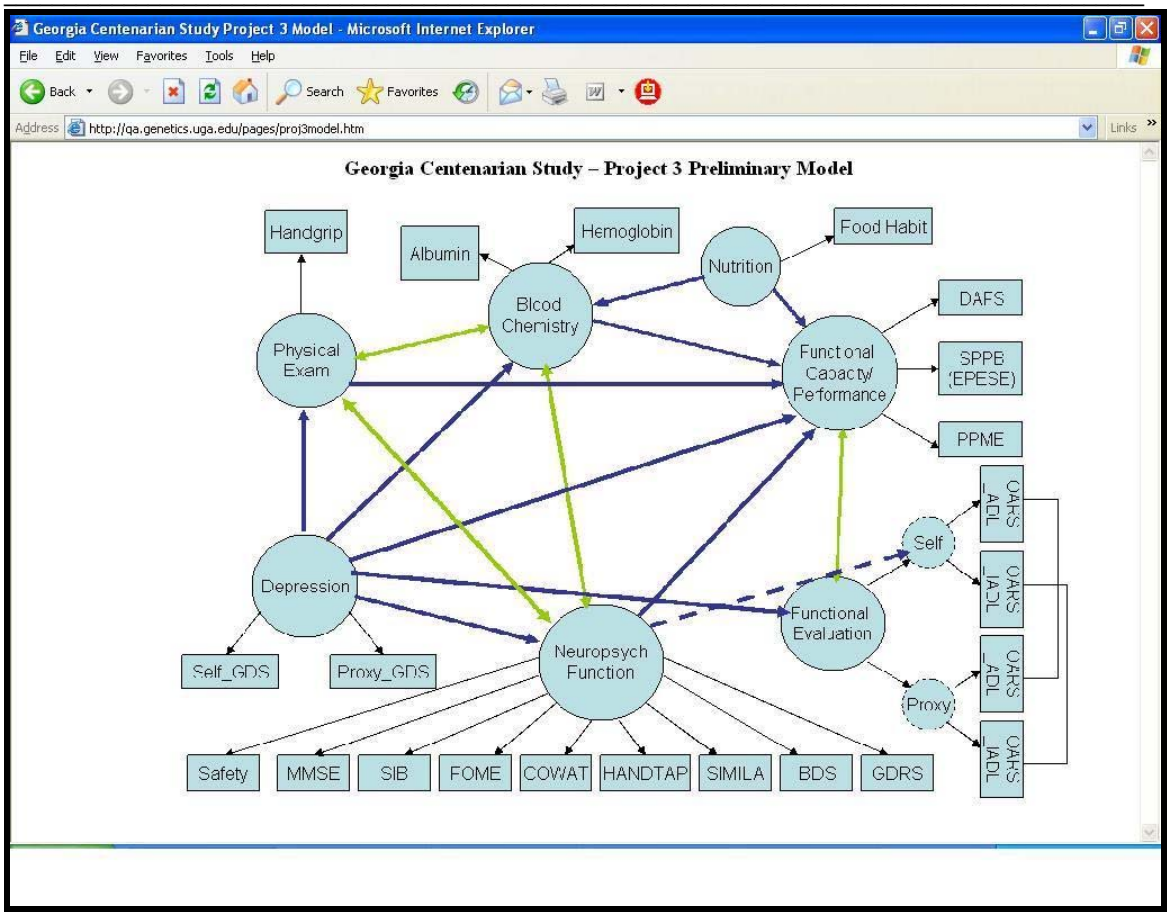

Figure 2: Screen display showing the web page for project 3. The page consists of the proposed project 3 research model which connects the potentially interrelated domain areas together. Clicking the domain area picture in the model brings the user to the web page for all the tables in that area

Database Content

GCSDB contains more than 40 relational tables for 4 interrelated projects. Number of columns in these tables ranges from 23 to 519. Column attributes are scale (double), ordinal (integer), or nominal (String). Each data entry in every table corresponds to one study participant (centenarian, octogenarian, or young control). Depending upon a participant's level of involvement in the project, these tables include the participants' information in the following tests or domain areas:

(1) Demographics and Family Longevity. Data include age, date of birth, gender, racial and ethnic backgrounds, handedness, years of education, county and place of residence, and occupational history. Family longevity includes ages and causes of death of mother, father, grandparents, siblings, number of siblings, lifetime residential locations, and military services.

(2) Physical Health. The protocol is composed of 3 domains: (a) Physical Examination and Health History, including vital signs, presence/absence of certain diseases, anthropometrics, neurological/musculoskeletal measures and current medications; (b) Blood Chemistry Profile, including glucose, BUN, creatinine, albumin, complete blood count, , C-reactive protein, hemoglobin A1c, ferritin, thyroxine, thyroid stimulating hormone, vitamin B12, and vitamin D; and (c) Physical Function Assessment includes measures of bed mobility, bed to chair transfer skills, standing balance, walking, step-up, and chair standing abilities;

(3) Cognition, Neuropsychology, Mental Health, and Neuropathology. Tests and assessments include: the Mini-Mental State Examination (MMSE) **[15]**, Global Deterioration Scale **[16]**, and Severe Impairment Battery **[17]**, Fuld Object Memory Evaluation (FOME) **[18]**, Wechsler Adult Intelligence Scale-III, Similarities sub-test, Letter Number Sequencing sub-test, and Matrix Reasoning

subtest **[19]**, Behavioral Dyscontrol Scale (DBS) **[20]**, ILS Health & Safety Scale **[21]**, COWAT **[22]**, Clock Drawing Test **[23, 24]**, Geriatric Depression Scale **[25]**, CERAD battery **[26]**, and Clinical Dementia Rating Scale. **[27]** Magnetic Resonance Imaging data of the voluntarily donated brain tissue were also available for some centenarians who passed-away before the completion of the project.

(4) Functional Capacity and Independence. The basic functional capacity was assessed using both a self-report as well as performance-based measures of basic/physical and instrumental activities of daily living (BADL and IADL). Physical functional capacity was measured with NIA Short Performance Battery (SPPB) **[28]** and Physical Performance and Mobility Examination (PPME). **[29]** The performance-based measures include selected subtests of the Direct Assessment of Functional Status-*Revised* (DAFS-*R*). **[30]** The measures and Performance Rating Scale was taken from OARS. **[31]** In addition, there are other tables summarizing personality, life events, social support, and economic resources.

(5) Genetics. The Single Nucleotide Polymorphism (SNPs) table includes 15 SNPs from three candidate longevity genes, i.e., *APOE* **[32, 33]**, *HRAS1* **[34]**, and *LASS1/LAG1* **[35, 36]** for all of the participants. The SNPs are located in exon or promoter regions in these genes.

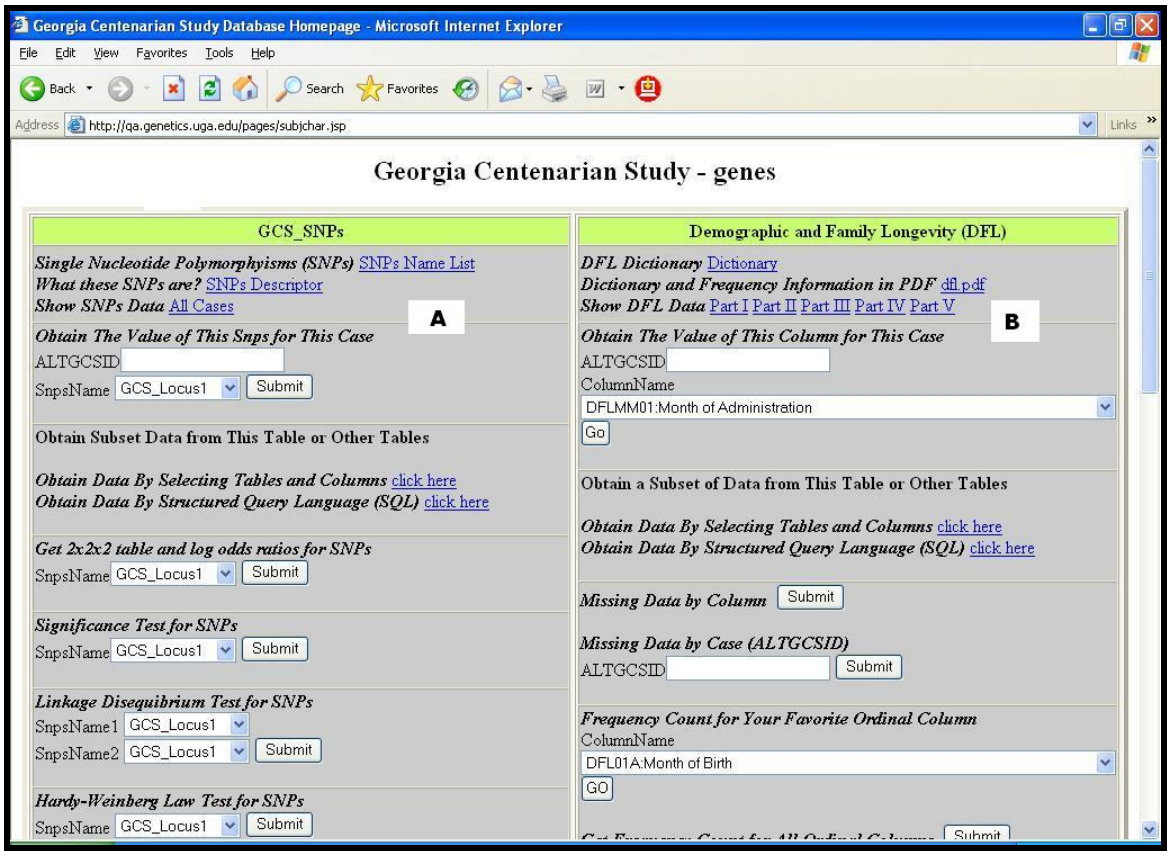

Figure 3: A portion of the web page for the SNPs and Demographics and Family Longevity data. (A) Information and functions provided for SNPs data analyses. (B) Information and functions provided for preliminary analyses of Demographics and Family Longevity data as well as other data

Database Interface

The GCSDB homepage contains three sections (Figure 1): an introduction to the Georgia Centenarian Study; links to web pages for the four individual projects (Figure 2); ways to cross link among the tables (users may retrieve data either by selecting table names and column names or directly by SQL; a plot function is available to obtain a preliminary plot for any two column variables of interest from any table in the study to help investigators choose variables for further statistical analysis; Regression, Principal Component Analysis, and

canonical correlation analysis tools are available to identify relationships among set of variables of interest.

For each table, the web page provides the following information and functions:

1. a dictionary listing the data-type, data length, column name, full name, and each category value for ordinal columns;
2. a PDF file showing frequencies and descriptive statistics for most columns;
3. a raw data table and function giving values for the user selected column for a given participant;
4. missing value counts for every column and a function shows what columns are missing for a given participant;
5. frequency counts for a user selected ordinal column and frequency counts for ordinal columns;
6. maximum, minimum, and mean values and number of participants with values less or larger than mean values for all scale columns;
7. number of participants with values larger or less than a given cut-off value for the user selected scale column;

For the SNPs table (Figure 3), the following important information and genetic analyses are available;

1. SNPs descriptor includes the name, identity, cytogenetic map, chromosomal and physical location for every SNP;
2. the 2 x 2 x 2 contingency table (age, race, SNP allele) and log odds ratios for each SNP **[37]**;
3. exact tests of association for each SNP **[37]**;
4. exact test of Hardy-Weinberg Law for each SNP;
5. Linkage Disequibrium Test for pair-wise SNPs;
6. haplotype analysis results for SNPs from APOE, HRAS1 and LASS1, respectively and for all the SNPs. **[38]**

Data dictionaries, forms, code books, questionnaires, and other relevant metadata are on the Web site or are being added as requested. The database is passworded to protect the anonymity of centenarians, and centenarian names have been replaced with a GCSID.

Future Development:

Data acquisition for functional capacity, and adaptation and resource projects has been completed; genetic data analysis is still under way from the collected blood samples. Neuropathology collection is expected to continue until 2007. GCSDB is updated as data become available. Functions and tools are added as data reduction and analysis begin. The database/Web interface constitutes a virtual resource for interested researchers. At present, GCSDB is only available to our project personnel, but will be released to the research community one year following completion of our project according to the data sharing plan on the Web site. It is our obligation and desire to share data, cell lines, and brain tissues with qualified researchers at that time. This data archive is part of making results available online. We expect that these rare resources can be used to test new hypotheses to gain new knowledge on contributors for the extreme longevity of these centenarians and that our system can serve as a model to other cross-disciplinary projects.

Acknowledgement:

The Georgia Centenarian Study is funded by 1P01-AG17553 from the National Institute on Aging. We also thank the UGA College of Agricultural and Environmental Sciences for their support. [4]The Georgia Centenarian Study (Leonard W. Poon, PI) is funded by 1P01-AG17553 from the National Institute on Aging, a collaboration among The University of Georgia, Louisiana State University Health Sciences Center in New Orleans, Boston University, University of Kentucky, Emory University, Duke University, Rosalind Franklin University of Medicine and Science, Iowa State University, and University of Michigan. The authors acknowledge the contributions of the Study's project and core leaders to this paper: L.W. Poon, S. M. Jazwinski, R. C. Green, M. Gearing, W. R. Markesbery, J. L. Woodard, M. A. Johnson, J. S. Tenover, I. C. Siegler, P. Martin, M. MacDonald, C. Rott, W. L. Rodgers, D. Hausman, J. Arnold, and A. Davey. We also acknowledge M. A. Batzer, E. Cress, and L. S. Miller for their contributions. Authors acknowledge the valuable recruitment and data acquisition effort from M. Burgess, K. Grier, E. Jackson, E. McCarthy, K. Shaw, L. Strong, and S. Reynolds, data acquisition team manager; S. Anderson, E. Cassidy, M. Janke, and T. Savla, data management; M. Durden for project fiscal management.

References:

[01] U. Lehr, *Zeitschrift Fur Gerontologie*, 24:227 (1991) [PMID: 1957533]

[02] http://www.geron.uga.edu/research/centenarianstudy.php

[03] J. W. Vaupel, *et al.*, *Science*, 280:855 (1998) [PMID: 9599158]

[04] T. T Perls, *Medical Hypothesis*, 49:405 (1997) [PMID: 9421805]

[05] L. W. Poon, *et al.*, *International Journal of Aging & Human Development*, 34:1 (1992) [PMID: 1737657]

[06] L. W. Poon, *et al.*, *International Journal of Aging and Human Development*, 34:31 (1992) [PMID: 1737659]

[07] Y. C. Chan, *et al.*, *Journal of Nutritional Science and Vitaminology*, 43:73 (1997) [PMID: 9151242]

[08] A. Capurso, *et al.*, *Archives of Gerontology and Geriatrics*, 25:149 (1997)

[09] O. Regius, *et al.*, *Zeitschrift für Gerontologie*, 27: 456 (1994) [PMID: 7871878]

[10] M. Allard, Les 120 ans de Jeanne Calment. Doyenne de l'humanite. Paris: Le Cherche Midi Editeur (1994)

[11] S. M. Samuelsson, *et al.*, *International Journal of Aging & Human Development*, 45:223 (1997) [PMID: 9438877]

[12] B. Jeune, *Population studies of aging Number 15*. Odense, Denmark, Odense University (1994)

[13] G. E. Vaillant & K. Mukamal, *American Journal of Psychiatry*, 158:839 (2001)

[14] Cardiff Software, Inc. Cardiff Teleform [Computer Software]. Vista, CA (2001)

[15] M. F. Folstein, *et al.*, *Journal of Psychiatric Research*, 12:189 (1975) [PMID: 1202204]

[16] B. Reisberg, *et al.*, *American Journal of Psychiatry*, 139:1136 (1982) [PMID: 7114305]

[17] J. Saxton, *et al.*, *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 2:298 (1990)

[18] P. A. Fuld, *The Fuld Object-Memory Evaluation.* Chicago: Stoelting Instrument Company (1981)

[19] D. Wechsler, *Wechsler Adult Intelligence Scale (WAIS-III) (Third ed.).* San Antonio, TX: The Psychological Corporation (1997)

[20] J. Grigsby, *et al.*, *Perceptual and Motor Skills*, 74: 883 (1992) [PMID: 1608726]

[21] P. A. Loeb, *Independent Living Scales manual*. San Antonio, TX: The Psychological Corporation (1992)

[22] A. Benton & K. Hamsher, Multilingual Aphasia Examination. Iowa City: University of Iowa (1997)

[23] M. Freedman, *et al.*, Clock drawing: A neuropsychological analysis. New York: Oxford (1994)

[24] A. Rouleau, *et al.*, *Brain and Cognition*, 18:70 (1992)

[25] J. A. Yesavage, *Journal of Psychiatric Research*, 17:37 (1983)

[26] J. C. Morris, *et al.*, *Neurology*, 39:1159 (1989) [PMID: 2771064]

[27] J. C. Morris, *Neurology*, 43:2412 (1993) [PMID: 8232972]

[28] J. M. Guralnik, *Journal of Gerontology: Series A*, 49:85 (1994) [PMID: 8126356]

[29] C. H. Winograd, *et al.*, *Journal of the American Geriatric Society*, 42:743 (1994) [PMID: 8014350]

[30] D. A. Loewenstein, *Journal of Gerontology*, 4:114 (1989) [PMID: 2738312]

[31] G. Fillenbaum, Multidimensional functional assessment of older adults. Hillsdale, New Jersey: Erbaum (1988)

[32] K. Kervinen, *et al.*, *Atherosclerosis*, 105:89 (1994) [PMID: 8155090]

[33] F. Schachter, *et al.*, *Nature Genetics*, 6:29 (1994) [PMID: 8136829]

[34] M. Bonafè, *et al.*, *Gene*, 286:121 (2002) [PMID: 11943467]

[35] N. P. D'mello, *et al.*, *Journal of Biological Chemistry*, 269:15451 (1994) [PMID: 8195187]

[36] N. K. Egilmez, *Journal of Biological Chemistry*, 264:14312 (1989) [PMID: 2668285]

[37] J. Dai, *et al.*, Exact sample sizes needed to detect dependence in 2 x 2 x 2 tables. *Biometrics*, (submitted)

[38] D. Fallin, & N. J. Schork, *Am J Hum Genet.*, 67:947 (2000)

Edited by M. K. Sakharkar

Citation: Dai *et al.*, Bioinformation 1(6): 214-219 (2006)