Exhibit 1

UNIVERSITY OF PENNSYLVANIA - SCHOOL OF MEDICINE
Curriculum Vitae

Date: February 2004

Ruben C. Gur, Ph.D.

HOME ADDRESS: 815 Saint George's Rd.
Philadelphia, PA 19119
Voice/Fax (215) 247-2716

OFFICE ADDRESS: Brain Behavior Laboratory
Department of Psychiatry
10 Gates Building
University of Pennsylvania
Philadelphia, PA 19104-4283
gur@bbl.med.upenn.edu
Tel: (215) 615-3604
Fax: (215) 662-7903

SOCIAL SECURITY NUMBER: 373-62-2205

EDUCATION:

1967-70	B.A. Hebrew University of Jerusalem
1970-71	M.A. Michigan State University
1971-73	Ph.D. Michigan State University
1973-74	Postdoctoral Fellowship, Stanford University

POSTGRADUATE TRAINING AND FELLOWSHIP APPOINTMENTS:

Doctorate:

1971	NIMH Summer Traineeship, Psychiatric Clinic Oakland County Juvenile Court Pontiac, Michigan
1971-72	Psychology Internship, Psychiatric Clinic State Prison of Southern Michigan Jackson, Michigan
1972-73	Psychology Internship, Counseling Center Michigan State University East Lansing, Michigan

Postdoctorate:

1973-74	Research Associate, Department of Psychology Stanford University, Stanford, California
1974-76	Postdoctorate supervised clinical experience Department of Psychiatry

MILITARY SERVICE: 1965-67 Israeli Defense Forces

FACULTY APPOINTMENTS:

1974-81	Assistant Professor Department of Psychology University of Pennsylvania
1981-84	Research Associate Professor of Psychology in Neurology and Psychiatry University of Pennsylvania
1984-88	Associate Professor of Psychology in Psychiatry and Neurology University of Pennsylvania
1988-	Professor of Psychology in Psychiatry, Neurology and Radiology University of Pennsylvania

HOSPITAL AND ADMINISTRATIVE APPOINTMENTS:

1974-81	Supervisor, Clinical Training Program Department of Psychology University of Pennsylvania
1982-84	Director of Neuropsychology Department of Neurology The Graduate Hospital
1984-	Director of Neuropsychology and the Brain Behavior Laboratory, Department of Psychiatry Hospital of the University of Pennsylvania

BOARD SPECIALTY CERTIFICATION:

Diplomate in Clinical Neuropsychology
American Board of Professional Psychology

LICENSURE: Licensed Psychologist, Commonwealth of Pennsylvania

AWARDS, HONORS AND MEMBERSHIP IN HONORARY SOCIETIES:

Erickson Award for Scientific Excellence for Writing in Hypnosis
Member, Sigma Xi
Elected to Fellow status, National Academy of Neuropsychologists, 1986
Elected to Fellow status, American Psychological Association, Divisions 6 & 30, 1987
Recipient of 1990 Stephen V. Logan Award, National Alliance for the Mentally Ill (NAMI)
Elected to Fellow status, The American Psychological Society, 1992

MEMBERSHIP IN PROFESSIONAL AND SCIENTIFIC SOCIETIES:

American Psychological Association, FELLOW
American Psychological Society, FELLOW
American College of Neuropsychopharmacology
The John Morgan Society
American Association for the Advancement of Science
International Neuropsychological Society
National Academy of Neuropsychologists
The New York Academy of Science
International Society for Neuroimaging in Psychiatry

EDITORIAL/ADVISORY POSITIONS:

Member, NIH Study Section on Neuroscience, Psychopathology, Addiction and Sleep (2002-2006)
Member, NIH Study Section on Clinical Neuroscience and Biological Psychopathology (1993-1996)
Editorial Board: Schizophrenia Research, Archives of Clinical Neuropsychology (1985-1997), Journal of Mental Imagery (1984-), Brain and Cognition (1989-), Brain and Language (1990-), Neuroimaging and Behavior (1994-). Consultant to Panel on Neurological Aspects of Behavior: Development of a National Research Strategy for NIH (1979).
Advisory Board: The Greenwall Initiative on Imaging and Treating the Human Brain: Ethical and Social Implications. The Center for Bioethics, University of Pennsylvania; Institute for Strategic Threat Analysis and Response (ISTAR), University of Pennsylvania, (2002-)
The Conte Center on the Neurobiology of Suicide, Columbia University (J Mann, MD, PI).
Ad-hoc Reviewer for National Institutes of Health, National Science Foundation, Veterans Administration, Office of Naval Research, Canadian Research Council, The Wellcome Foundation, and professional journals including *Science, Nature, The Lancet; Neuropsychologia; Neuropsychology; Journal of Cerebral Blood Flow and Metabolism; Psychological Review; Journal of Abnormal Psychology; Neuropsychiatry, Neuropsychology and Behavioral Neurology; Archives of General Psychiatry; American Journal of Psychiatry; Neuropsychopharmacology; Biological Psychiatry; NeuroImage; Neurobiology of Aging; Annals of Neurology; Neurology.*
Action Editor: Brain and Cognition (2002-)

ACADEMIC COMMITTEES AT THE UNIVERSITY OF PENNSYLVANIA:

1975-1977 Admissions Committee, Psychology Department

1984-1994 Research Committee, Psychiatry Department
1984-1987 Computer Task Force, Psychiatry Department
1984-1988 Chairman's Council for Planning and Development, Psychiatry Department
1990-1992, 1996-1999 Senate Committee on Academic Freedom and Responsibility
1996-1998 Search Committee for Chair of Radiology
2000- Search Committee for fMRI Physicist
1999- University Scholars Council

MAJOR TEACHING & CLINICAL RESPONSIBILITIES AT THE UNIVERSITY OF PENNSYLVANIA:

A. Teaching:
1. Co-founder and Advisor, Biological Basis of Behavior Undergraduate Major program.
2. Supervisor of postdoctoral Fellows (NIMH Training Grant) and doctoral students
3. Member of dissertation committees.
4. Rounds and teaching conferences for Psychiatry residents and Neuropsychology Fellows
5. Supervisor of undergraduate Honors theses.

B. Clinical:

1. Director of Neuropsychology, Department of Psychiatry, Hospital of the University of Pennsylvania
2. Supervisor of interns and practicum students in neuropsychology.

PRESENTATIONS & LECTURES BY INVITATION: (Outside Philadelphia, Past 5 years)

January 27-29, 1998, "Is there a neuropsychological profile of schizophrenia?" CNS Summit, Palm Spring, California.

February 18, 1998, "Sex differences in brain and behavior: Evidence from structural measures and functional imaging studies with cognitive and emotional." SUNY Health Center at Brooklyn-Graduate School, Brooklyn, NY.

April 6, 1998, "Healthy differences between men and women." Haverford College, Haverford, Pennsylvania

April 16, 1998, "Sex differences in human brain anatomy and physiology: Some implications for psychopathology" University of Michigan, Ann Arbor, Michigan.

December 15, 1998 "Sex differences in healthy aging: Linking anatomy, neurocognition and physiology." American College of Neuropsychopharmacology (ACNP), Las Croabas, Puerto Rico.

May 7-10, 1999, "Methods for studying brain-behavior relations with structural and functional neuroimaging." 10th Nikolas Symposium, Athens, Greece.

May 13, 1999, "Temporolimbic Function in Schizophrenia: An integrative Approach from Behavior to the Molecule." The Society of Biological Psychiatry Annual Meeting - 1999, Washington, D.C.

August 221999, "Gender differences". Europäisches Forum Alpbach: Materie, Geist und Bewusstsein, Alpbach, Austria.

November 8, 1999, "Sex Differences in Learning," Learning & The Brain Conference, Boston, MA.

December 6, 1999, "Studying neurobiology of emotions; glance at sex differences." University of Michigan Mental Health Research Institute, Ann Arbor, MI.

April 4, 2000, "Sex Differences in Brain Structure and Function - What are they and so what?" Washington College, Chestertown, MD.

April 10, 2000, "Emotion and Thinking: Normal Sex Differences and Disturbances in Schizophrenia." Norristown State Hospital, Norristown, PA.

April 27, 2000, "Is Exercising Your Brain Like Exercising Your Muscles? Implications for Education of Similarities and Differences." AERA's Brain & Education Symposium, New Orleans.

May 2, 2000, "Sex Differences in Brain Structure and Function: What could they mean for education?" Archer School for Girls, Los Angeles, CA

November 2, 2000, "Cognitive and emotion processing in schizophrenia: Implication for diagnosis and treatment." Wernersville State Hospital, PA

December 15, 2000. "Combined block design and event-related analysis in functional mri studies of emotion processing." ACNP "Hot topics" session, San Juan, PR

November 18, 2001 "Brain Regulation of Thoughts and Feelings: Cognitive and Emotion Processing Deficits in Schizophrenia." NAMI Montgomery County, Elkins Park, PA.

March 4, 2002. "Imaging Studies of Emotion Processing Examining the Effects of Age, Gender, and Disease." University of Iowa School of Medicine, Research Seminar, Iowa City, IA

March 5, 2002. "Behavioral and Neurobiologic Markers of Brain Dysfunction and Genetic Vulnerability to Schizophrenia." University of Iowa School of Medicine, Grand Rounds. Iowa City, IA

April 14, 2002. "The Neurobiology of Sex Differences in the Symptoms and Course of Schizophrenia" Advocates for the Jewish Mentally Ill, Wynnewood, PA

May 5, 2002. "The Science of Deceit: The Polygraph and its Progeny" Judicial In Service Training, Washington D.C.

May 22, 2002. "Functional and Structural Imaging Studies of Emotion Processing" Massachusetts General Hospital, Grand Rounds, Charlestown, MA

Oct 10, 2002. "What to do About Girls, Boys and Brains: Sex Differences From Phylogeny to Ontogeny" CAIS Commission of Women in Independent School's Annual Conference,

2. Gur RC, Gur RE. Handedness, sex and eyedness as moderating variables in the relation between hypnotic susceptibility and functional brain asymmetry. Journal of Abnormal Psychology, 1974, 83, 635-643.

3. Gur RC. An attention-controlled operant procedure for enhancing hypnotic susceptibility. Journal of Abnormal Psychology, 1974, 83, 635-643.

4. Gur RE, Gur RC, Marshalek B. Classroom seating and functional brain asymmetry. Journal of Educational Psychology, 1975, 67, 151-153.

5. Gur RC, Hilgard ER. Visual imagery and discrimination of differences between altered pictures simultaneously and successively presented. British Journal of Psychology, 1975, 66, 341-345.

6. Gur RE, Gur RC. Defense mechanisms, psychosomatic symptomatology and conjugate lateral eye movements. Journal of Consulting and Clinical Psychology, 1975, 43, 416-420.

7. Gur RE, Gur RC, Harris LJ. Cerebral activation, as measured by subject`s lateral eye movements, is influenced by experimenter location. Neuropsychologia, 1975, 13, 35-44.

8. Gur RC, Sackeim HA, Gur RE. Classroom seating and psychopathology: some initial data. Journal of Abnormal Psychology, 1976, 85, 122-124.

9. Gur RC, Reyher J. The enhancement of creativity via free imagery and hypnosis. American Journal of Clinical Hypnosis, 1976, 85, 237-249.

10. Sackeim HA, Packer IK, Gur RC. Hemisphericity, cognitive set and susceptibility to subliminal perception. Journal of Abnormal Psychology, 1977, 86, 624-630.

11. Gur RE, Gur RC. Sex differences in the relations among handedness, sighting-dominance and eye acuity. Neuropsychologia, 1977, 15, 585-590.

12. Sackeim HA, Gur RC, Saucy MC. Emotions are expressed more intensely on the left side of the face. Science, 1978, 202, 434-436.

13. Gur RC, Sackeim HA. Self-confrontation and psychotherapy. Psychotherapy: Theory, Research and Practice, 1978, 15, 258-265.

14. Sackeim HA, Gur RC. Lateral asymmetry in intensity of emotional expression. Neuropsychologia, 1978, 16, 473-481.

15. Sackeim HA, Gur RC. Self-deception, other-deception, and self-reported psychopathology. Journal of Consulting and Clinical Psychology, 1979, 47, 213-215.

16. Gur RC, Sackeim HA. Self-deception: A concept in search of a phenomenon. Journal of Personality and Social Psychology, 1979, 37, 147-169.

17. Sackeim HA, Nordlie JW, Gur RC. A model of hysterical and hypnotic blindness: cognition, motivation and awareness. Journal of Abnormal Psychology, 1979, 88, 474-489.

18. Gur RC, Reivich M. Cognitive task effects on hemispheric blood flow in humans: evidence for individual differences in hemispheric activation. Brain and Language, 1980, 9, 78-92.

19. Gur RC, Packer IK, Hungerbuhler JP, Reivich M, Obrist WD, Amarnek WS,Sackeim HA. Differences in the distribution of gray and white matter in human cerebral hemispheres. Science, 1980, 207, 1226-1228.

20. Sackeim HA, Greenberg MS, Weiman AL, Gur RC, Hungerbuhler JP, Geschwind N. Hemispheric asymmetry in the expression of positive and negative emotions: Neurological Evidence. Archives of Neurology, 1982, 39, 210-218.

21. Gur RC, Sussman NM, Alavi A, Gur RE, Rosen AD, O`Connor M, Goldberg HI, Greenberg JH, Reivich M. Positron emission tomography in two cases of childhood epileptic encephalopathy (Lennox-Gastaut Syndrome). Neurology, 1982, 32, 1191-1194.

22. Gur RC, Gur RE, Obrist WD, Hungerbuhler JP, Younkin D, Rosen AD, Skolnick BE., Reivich M. Sex and handedness differences in cerebral blood flow during rest and cognitive activity. Science, 1982, 217, 659-661.

23. Sussman NM, Gur RC, Gur RE, O'Connor MJ. Mutism as a consequence of callosotomy. Journal of Neurosurgery, 1983, 59, 514-519.

24. Gur RC, Gur RE, Rosen AD, Warach S, Alavi A, Greenberg J, Reivich M. A cognitive-motor network demonstrated by positron emission tomography. Neuropsychologia, 1983, 21, 601-606.

25. Natale M, Gur RE, Gur RC. Hemispheric asymmetries in processing emotional expressions. Neuropsychologia, 1983, 21, 555-565.

26. Reivich M, Gur RC, Alavi A. Positron emission tomography studies of sensory stimuli, cognitive processes and anxiety. Human Neurobiology, 1983, 2, 25-33.

27. Gur RE, Skolnick BE, Gur RC, Caroff S, Rieger W, Obrist WD, Younkin D, Reivich M. Brain function in psychiatric disorders: I. Regional cerebral blood flow in medicated schizophrenics. Archives of General Psychiatry, 1983, 40, 1250-1254.

28. Gur RE, Gur RC, Sussman NM, O'Connor MJ, Vey MM. Hemispheric control of the writing hand: The effect of callosotomy in a left-hander. Neurology,1984, 34, 904-908.

29. Reivich M, Alavi A, Gur RC. Positron emission tomographic studies of perceptual tasks. Annals of Neurology, 1984, 15, 61-65 (Supplement).

30. Gur RE, Skolnick BE, Gur RC, Caroff S, Rieger W, Obrist WD, Younkin D, Reivich M. Brain function in psychiatric disorders: II. Regional cerebral blood flow in medicated depressives. Archives of General Psychiatry, 1984, 41, 695-699.

31. Youkin D, Hungerbuhler JP, O'Connor M, Goldberg H, Burke A, Kushner M, Hurtig H, Obrist W, Gordon J, Gur RC, Reivich M. Superficial temporal-middle cerebral artery anastomosis: Effects on vascular, neurologic, and neuropsychological functions. Neurology, 1985, 35, 462-469.

32. Gur RE, Gur RC, Skolnick BE, Caroff S, Obrist WD, Resnick S, Reivich M. Brain function in psychiatric disorders: III. Regional cerebral blood flow in unmedicated schizophrenics. Archives of General Psychiatry, 1985, 42, 329-334.

33. Trivedi SS, Gur RC, Gur RE, Skolnick BE, Obrist WD, Reivich M, Herman GT. Imaging regional cerebral blood flow measured by the 133-Xenon technique. rCBF Bulletin, 1986, 9, 175-178.

34. Stern MB, Gur RC, Saykin AJ, Hurtig HI. Dementia of Parkinson's disease and Alzheimer's disease: Is there a difference? Journal of the American Geriatrics Society, 1986, 34, 475-478.

35. Gur RE, Resnick SM, Alavi A, Gur RC, Caroff S, Dann R, Silver F, Saykin AJ, Chawluk JB, Kushner M, Reivich M. Regional brain function in schizophrenia: I. A positron emission tomography study. Archives of General Psychiatry, 1987, 44, 119-125.

36. Gur RE, Resnick SM, Gur RC, Alavi A, Caroff S, Dann R, Silver F, Saykin AJ, Chawluk JB, Kushner M, Reivich M. Regional brain function in schizophrenia: II. Repeated evaluation with positron emission tomography. Archives of General Psychiatry, 1987, 44, 126-129.

37. Trope I, Fishman B, Gur RC, Sussman NM, Gur RE. Contralateral and ipsilateral control of fingers following callosotomy. Neuropsychologia, 1987, 25, 287-291.

38. Gur RC, Gur RE, Obrist WD, Skolnick BE, Reivich M. Age and regional cerebral blood flow at rest and during cognitive activity. Archives of General Psychiatry, 1987, 44, 617-621.

39. Gur RC, Gur RE, Resnick SM, Skolnick BE, Alavi A, Reivich M. The effect of anxiety on cortical cerebral blood flow and metabolism. Journal of Cerebral Blood Flow and Metabolism, 1987, 7, 173-177.

40. Knight H, Millman RP, Gur RC, Saykin AJ, Doherty JU, Pack AI. Clinical significance of sleep apnea in the elderly. American Review of Respiratory Disease, 1987, 136, 845-850.

41. Gur RC, Gur RE, Silver FL, Obrist WD, Skolnick BE, Kushner M, Hurtig HI, Reivich M. Regional cerebral blood flow in stroke: hemispheric effects of cognitive activity. Stroke, 1987, 18, 776-780.

42. Warach S, Gur RC, Gur RE, Skolnick BE, Obrist WD, Reivich M. The reproducibility of the Xe-133 inhalation technique in resting studies: task order and sex related effects in healthy young adults. Journal of Cerebral Blood Flow and Metabolism, 1987, 7, 702-708.

43. Trivedi SS, Gur RC. Computer graphics for neuropsychological data. Proceedings of the

National Computer Graphics Association, 1987, 3, 22-32.

44. Gur RC, Gur RE, Skolnick BE, Resnick SM, Silver FL, Chawluk JB, Muenz L, Obrist WD, Reivich M. Effects of task difficulty on regional cerebral blood flow: relationships with anxiety and performance. Psychophysiology, 1988, 25, 392-399.

45. Schmidt ML, Gur RE, Gur RC, Trojanowski JQ. Intraneuronal and extracellular neurofibrillary tangles exhibit mutually exclusive cytoskeletal antigens. Annals of Neurology, 1988, 23, 184-189.

46. Resnick SM, Gottlieb GL, Gur RE, Gur RC, Forciea MA, Zimmerman RA, Malamut B, Saykin AJ, Reivich M, Alavi A. Identical twins with probable Alzheimer's Disease: behavior, anatomy and physiology. Neuropsychiatry, Neuropsychology and Behavioral Neurology, 1988, 1, 61-72.

47. Gur RC, Trivedi SS, Saykin AJ, Gur RE. "Behavioral imaging" - a procedure for analysis and display of neuropsychological test scores: I. Construction of algorithm and initial clinical evaluation. Neuropsychiatry, Neuropsychology and Behavioral Neurology, 1988, 1, 53-60.

48. Gur RC, Saykin AJ, Blonder LX, Gur RE. "Behavioral imaging": II. Application of the quantitative algorithm to hypothesis testing in a population of hemiparkinsonian patients. Neuropsychiatry, Neuropsychology and Behavioral Neurology, 1988, 1, 87-96.

49. Gottlieb GL, McAllister TW, Gur RC. Depot neuroleptics in the treatment of behavioral disorders in patients with Alzheimer's disease. Journal of the American Geriatric Society, 1988, 36, 619-621.

50. Gottlieb GL, Gur RE, Gur RC. Reliability of psychiatric scales in patients with DAT. American Journal of Psychiatry, 1988, 145, 857-860.

51. Resnick SM, Gur RE, Alavi A, Gur RC, Reivich M. Positron emission tomography and subcortical glucose metabolism in schizophrenia. Psychiatry Research, 1988, 24, 1-11.

52. Trope I, Rozin P, Gur RC. Validation of the lateral limits technique with a callosotomy patient. Neuropsychologia, 1988, 26, 673-684.

53. Blonder LX, Gur RE, Gur RC. The effects of right and left hemiparkinsonism on prosody. Brain and Language, 1989, 36 193-207.

54. Trivedi SS, Gur RC. Topographic mapping of cerebral blood flow and behavior. Computers in Biology and Medicine, 1989, 19, 219-229.

55. Blonder LX, Gur RE, Gur RC, Saykin AJ, Hurtig HI. Neuropsychological functioning in hemiparkinsonism. Brain and Cognition, 1989, 9, 177-190.

56. Saykin AJ, Gur RC, Sussman NM, Gur RE. Memory deficits before and after temporal lobectomy: Effect of laterality and age of onset. Brain and Cognition, 1989, 9, 191-200.

Mood effects on limbic blood flow correlate with emotional self-rating. A PET study with oxygen-15 labeled water. Psychiatry Research, 1995, 61, 265-283.

110. Heimberg C, Gallacher F, Gur RC, Gur RE. Diet and gender moderate clozapine-related weight gain. Human Psychopharmacology, 1995, 10, 367-371.

111. Saykin AJ, Gur RC, Gur RE, Shtasel DL, Flannery KA, Mozley LH, Malamut BL, Watson B, Mozley PD. Normative neuropsycholgical test performance: Effects of age, education, gender and ethnicity. Applied Neuropsychology, 1995, 2, 79-88.

112. Saykin AJ, Stafiniak P, Robinson LJ, Flannery KA, Gur RC, O'Connor MJ, Sperling MR. Language before and after temporal lobectomy: Specificity of acute changes and relation to early risk factors. Epilepsia, 1995, 36, 1071-1077.

113. Turetsky BT, Cowell PE, Gur RC, Grossman RI, Shtasel DL, Gur RE. Frontal and temporal lobe brain volumes in schizophrenia: Relationship to symptomatology and clinical subtype. Archives of General Psychiatry, 1995, 52, 1061-1070.

114. Schneider F, Gur RC, Gur RE, Shtasel DL. Emotional processing in schizophrenia: Neurobehavioral probes in relation to psychopathology. Schizophrenia Research, 1995, 17, 67-75.

115. Schneider F, Gur RE, Alavi A, Seligman MEP, Mozley LH, Smith RJ, Mozley PD, Gur RC. Cerebral blood flow changes in limbic regions induced by unsolvable anagram tasks. American Journal of Psychiatry, 1996, 153, 206-212.

116. Cowell PE, Kostianovsky DJ, Gur RC, Turetsky BI, Gur RE. Sex differences in neuroanatomical and clinical correlations in schizophrenia. American Journal of Psychiatry, 1996, 153, 799-805.

117. Ragland JD, Censits DM, Gur RC, Glahn DC, Gallacher F, Gur RE. Assessing declarative memory in schizophrenia using wisconsin card sorting test stimuli: the paired associate recognition test. Psychiatry Research, 1996, 60, 135-145.

118. Wang GJ, Volkow ND, Fowler JS, Logan J, Gur RC, Netusil N, Hitzemann RJ, Pappas NS. Age associated decrements in dopamine D_2 receptors in thalamus and in temporal insula of human subjects. Life Sciences, 1996, 59, 31-35.

119. Gur RE, Petty RG, Turetsky BI, Gur RC. Schizophrenia throughout life: Sex differences in severity and profile of symptoms. Schizophrenia Research, 1996, 21 1-12.

120. Mozley LH, Gur RC, Gur RE, Mozley PD, Alavi A. The Relationship between verbal memory performance and the cerebral distribution of FDG in patients with schizophrenia. Biological Psychiatry, 1996, 40, 443-451.

121. Kareken DA, Moberg PJ, Gur RC. Proactive inhibition and semantic organization: Relationship to verbal memory in patients with schizophrenia. Journal of the International Neuropsychological Society, 1996, 2, 486-493.

122. Szymanski S, Gur RC, Gallacher F, Mozley LH, Gur RE. Vulnerability to tardive dyskinesia development in schizophrenia : An FDG-PET study of cerebral metabolism. Neuropsychompharmacology, 1996, 15, 567-575.

123. Mozley PD, Kim H-J, Gur RC, Tatsch K, Muenz LR, McElgin WT, Kung M-P, Mu M, Myers AM, Kung HF. [I-123] IPT SPECT imaging of CNS dopamine transporters: Non-linear effects of normal aging on striatal uptake values. Journal of Nucl Medicine, 1996, 37, 1965-1970.

124. Ragland JD, Glahn DC, Gur RC, Censits DM, Smith RJ, Mozley PD, Alavi A, Gur RE. PET regional cerebral blood flow change during working and declarative memory: Relationship with task performance. Neuropsychology, 1997, 11, 222-231.

125. Mozley PD, Gur RC, Swanson CL, Turetsky BI, Gur RE, Nienow T, Alavi A. [Tc-99m] ECD SPECT demonstrates that dopaminergic drugs affect cerebral blood flow in healthy human volunteers. Journal of Nuclear Medicine, 1997, 38, 48.

126. Finkelstein JRJ, Cannon TD, Gur RE, Gur RC, Moberg P. Attentional Dysfunctions in neuroleptic-naive and neuroleptic-withdrawn schizophrenic patients and their siblings. Journal of Abnormal Psychology, 1997, 106, 203-212.

127. Gur RC, Ragland JD, Mozley LH, Mozley PD, Smith R, Alavi A, Bilker W, Gur RE. Lateralized changes in regional cerebral blood flow during performance of verbal and facial recognition tasks: Correlations with performance and"Effort". Brain and Cognition, 1997, 33, 388-414.

128. Mozley PD, Sadek AM, Alavi A, Gur RC, Muenz LR, Bunow BJ, Kim H-J, Stecker M, Jolles P, Newberg A. Effects of aging on the cerebral distribution of [Tc-99m] HMPAO in healthy humans. European Journal of Nuclear Medicine, 1997, 24, 754-761.

129. Censits DM, Ragland JD, Gur RC, Gur RE. Neuropsychological evidence supporting a neurodevelopmental model of schizophrenia: A longitudinal study. Schizophrenia Research, 1997, 24, 289-298.

130. Moberg PJ, Doty RL, Turetsky BI, Arnold SE, Mahr RN, Gur RC, Bilker W, Gur RE Olfactory Identification deficits in schizophrenia correlation with duration of illness. American Journal of Psychiatry, 1997, 154, 1016-1018.

131. Coleman AR, Moberg PJ, Ragland JD, Gur RC. Comparison of the Halstead-Reitan and infrared light beam finger tappers. Assessment, 1997, 4, 277-286.

132. Glahn DC, Gur RC, Ragland JD, Censits DM, Gur RE. Reliability, performance characteristics, construct validity, and an initial clinical application of a visual object learning test (VOLT). Neuropsychology, 1997, 11, 602-612.

133. Gur RC, Ragland JD, Gur RE. Cognitive Changes in Schizophrenia: A Critical Look . International Review of Psychiatry, 1997, 9, 449-457.

134. Schneider F, Grodd W, Weiss U, Klose U, Mayer KR, Nagele T, Gur RC. Functional MRI

reveals left amygdala activation during emotion. Psychiatry Research:Neuroimaging Section, 1997, 76, 75-82.

135. Kohler C, Gur RC, Swanson CL, Petty R, Gur RE. Depression in schizophrenia: I. Association with neuropsychological deficits. Biological Psychiatry, 1998, 43, 165-172.

136. Gur RE, Cowell P, Turetsky BI, Gallacher F, Cannon T, Bilker W, Gur RC. A followup MRI study of schizophrenia: Relationship of neuroanatomic changes with clinical and neurobehavioral measures. Archives of General Psychiatry, 1998, 55, 145-152..

137. Kohler C, Swanson CL, Gur RC, Harper Mozley L, Gur RE. Depression In Schizophrenia: II. MRI and PET Findings. Biological Psychiatry, 1998, 43, 173-180.

138. Ragland JD, Gur RC, Glahn DC, Censits DM, Smith RJ, Lazarev MG, Alavi A, Gur RE. Frontotemporal cerebral blood flow change during executive and declarative memory tasks in schizophrenia: a positron emission tomography study. Neuropsychology, 1998, 12, 399-413.

139. Swanson CL, Gur RC, Bilker W, Petty RG, Gur RE. Premorbid educational attainment in schizophrenia: association with symptoms, functioning, and neurobehavioral measures. Biological Psychiatry, 1998, 44, 739-747.

140. Volkow ND, Gur RC, Wang GJ, Fowler JS, Moberg PJ, Ding YS, Hitzemann R, Smith G, Logan J. Association between decline of brain dopamine activity with age and cognitive and motor impairment in healthy individuals. American Journal of Psychiatry, 1998, 155, 344-349.

141. Erwin RJ, Turetsky B, Moberg P, Gur RC, Gur RE. P50 abnormalities in schizophrenia: relationship to clinical and neuropsychological indices of attention. Schizophrenia Research, 1998, 33, 157-167.

142. Gur RE, Maany V, Mozley D, Swanson C, Bilker W, Gur RC. Subcortical MRI Volumes in Neuroleptic-Naive and Treated Patients With Schizophrenia. American Journal of Psychiatry, 1998, 155, 1711-1717.

143. Moberg PJ, Doty RL, Turetsky BI, Arnold SE, Mahr RN, Gur RC, Bilker W, Gur RE. Deterioration of olfactory identification abilities in patients with schizophrenia. American Journal of Psychiatry, 1998, 155(10), 1463-1464.

144. Coleman AR, Norstrand JA, Moberg PJ, Kohler CG, Gur RC, Gur RE. MMPI-2 characteristics of adults diagnosed with Attention Deficit Disorder. International Journal of Neuroscience, 1998, 96, 161-175.

145. Volkow ND, Wang GJ, Fowler JS, Ding YS, Gur RC, Gatley JS, Logan J, Moberg PJ, Hitzemann, RJ, Smith G, Pappas N. Parallel Loss of Pre and Postsynaptic Dopamine Markers in Normal Aging. Annals of Neurology, 1998, 44, 143-147.

146. Cecil KM, Lenkinski RE, Gur RE, Gur, RC. Proton magnetic resonance spectroscopy in the frontal and temporal lobes of neuroleptic naive patients with schizophrenia.

Neuropsychoparmacology, 1999, 20, 131-140.

147. Ragland JD, Gur RE, Klimas BC, McGrady N, Gur RC. Neuropsychological laterality indices of schizophrenia: Interactions with gender. Schizophrenia Bulletin, 1999, 25, 79-89.

148. Gur RC, Turetsky BI, Matsui M, Yan M, Bilker W, Hughett P, Gur RE. Sex differences in brain gray and white matter in healthy young adults. Journal of Neuroscience, 1999, 19, 4065-4072.

149. Moberg PJ, Agrin, RN, Gur RE, Gur RC, Turetsky BI, Doty RI. Olfactory dysfunction in schizophrenia: A qualitative and quantitative review. Neuropsychopharmacology, 1999, 21, 325-340.

150. Gur RE, Turetsky BI, Bilker WB, Gur RC. Reduced gray matter volume in schizophrenia. Archives of General Psychiatry, 1999, 56, 905-911.

151. Ragland JD, Coleman AR, Gur RC, Glahn DC, Gur RE. Sex differences in behavior relationships between verbal episodic memory and resting regional cerebral blood flow. Neuropsychologia, 2000, 38, 451-461.

152. Glahn DC, Cannon TD, Gur RE, Ragland JD, Gur RC. Working memory constrains abstraction in schizophrenia. Biological Psychiatry, 2000, 47, 34-42.

153. Habel U, Gur RC, Mandal MK, Salloum JB, Gur RE, Schneider F. Emotional processing in schizophrenia across cultures: standardized measure of discrimination and experience. Schizophrenia Research, 2000, 42, 57-66.

154. Mozley PD, Acton PD, Barraclough ED, Plossl K, Gur RC, Alavi A, Mathur A, Saffer J, Kung HF. Effects of age on dopamine transporters in healthy humans. Journal of Nuclear Medicine, 1999, 40, 1812-1817,

155. McBride T, Arnold SE, Gur RC. A comparative volumetric analysis of the prefrontal cortex in human and baboon MRI. Brain, Behavior & Evolution, 1999, 54,159-166.

156. Volkow ND, Logan J, Fowler JS, Wang GJ, Gur RC, Wong C, Felder C, Gatley SJ, Ding YS, Hitzemann R, Papass N. Association between age-related decline in brain dopamine activity and impairment in frontal and cingulate metabolism. American Journal of Psychiatry, 2000, 157, 75-80.

157. Matsui M, Gur RC, Turetsky BI, Yan M, Gur RE. The relation between tendency for psychopathology and reduced frontal brain volume in healthy people. Neuropsychiatry, Neuropsychology, and Behavioral Neurology, 2000, 13, 155-162.

158. Gur RC, Alsop D, Glahn D, Petty R, Swanson CL, Maldjian JA, Turetsky BI, Detre JA, Gee J, Gur RE. An fMRI study of sex differences in regional activation to a verbal and a spatial task. Brain and Language, 2000, 74, 157-170.

159. Kohler CG, Bilker W, Hagendoorn M, Gur RE, Gur RC. Emotion recognition deficit in schizophrenia: Association with symptomatology and cognition. Biological Psychiatry,

2000, 48, 127-136.

160. Ragland JD, Gur RC, Lazarev, MG, Smith RJ, Schroeder L, Raz J, Turetsky, BI, Alavi A, Gur RE. Hemispheric activation of anterior and inferior prefrontal cortex during verbal encoding and recognition. NeuroImage, 2000, 11, 624-633.

161. Gur RE, Cowell PE, Latshaw A, Turetsky BI, Gorssman RI, Arnold SE, Bilker WB, Gur RC. Reduced dorsal and orbital prefrontal gray matter volumes in schizophrenia. Archives of General Psychiatry, 2000, 57, 761-768.

162. Gur RE, Turetsky BI, Cowell PE, Finkelman C, Maany V, Grossman RI, Arnold SE, Bilker WB, Gur RC. Temporolimbic volume reductions in schizophrenia. Archives of General Psychiatry, 2000, 57, 769-775.

163. Kohler CG, Ances BM, Coleman AR, Ragland JD, Lazarev M, Gur RC. Marchiafava-Bignami disease: literature review and case report. Neuropsychiatry, Neuropsychology, & Behavioral Neurology. 2000, 13, 67-76.

164. Matsuzawa J, Matsui M, Konishi T, Noguchi N, Gur RC, Bilker W, Miyawaki T. Age-related volumetric changes of brain gray and white matter in healthy infants and children. Cerebral Cortex. 2001, 11, 335-342.

165. Kurtz, M.M., Moberg, P.J., Harper Mozley, L., Swanson, C.S., Gur, R. C., & Gur, R.E. Effectiveness of an attention and memory training program on neuropsychological deficits in schizophrenia. Neurorehabilitation and Neural Repair, 2001, 15, 23-28.

166. Ragland JD, Gur RC, Raz J, Schroeder L, Smith RJ, Alavi A, Gur RE. Effect of schizophrenia on frontotemporal activity during word encoding and recognition: A PET cerebral blood flow study. American Journal of Psychiatry, 2001, 158, 1114-1125.

167. Kurtz, M.M., Ragland, J.D., Bilker, W.B., Gur, R.C., & Gur R.E. Comparison of two forms of the Continuous Performance Test, with and without working memory demands, in healthy controls and patients with schizophrenia. Schizophrenia Research, 2001, 48, 307-316.

168. Mozley LH, Gur RC, Mozley PD, & Gur RE. Striatal dopamine transporters and cognitive functioning in healthy men and women. American Journal of Psychiatry, 2001, 158,1492-1499

169. Silver, H, Shlomo, N, Turner, T, Gur, RC. Perception of happy and sad facial expressions in chronic schizophrenia: Evidence for two evaluative systems. Schizophrenia Research, 2002, 55, 171-177,

170. Gur RC, Ragland JD,Moberg PJ,Turner TH, Bilker WB, Kohler C, Siegel SJ, Gur RE. Computerized Neurocognitive Scanning: I. Methodology and validation in healthy people. Neuropsychopharmacology, 2001, 25, 766-776.

171. Gur RC, Ragland JD, Moberg PJ, Bilker WB, Kohler C, Siegel SJ, Gur RE. Computerized Neurocognitive Scanning II: The Profile of Schizophrenia. Neuropsychopharmacology,

2001, 25, 777-788,.

172. Hill SK, Ragland JD, Gur RC, Gur RE. Neuropsychological differences among empirically derived clinical subtypes of schizophrenia. Neuropsychology, 2001, 15, 492-501.

173. Moelter ST, Hill SK, Ragland JD, Lunardelli A, Gur, RC, Gur RE, Moberg PJ. Controlled and automatic processing during animal word list generation in schizophrenia. Neuropsychology, 2001, 15, 502-509.

174. Kurtz MM, Moberg PJ, Gur RC, Gur RE. Approaches to cognitive remediation of neuropsychological deficits in schizophrenia: a review and meta-analysis. Neuropsychology Review. 2001, 11,197-210.

175. Gur RC, Gunning-Dixon FM, Turetsky BI, Bilker WB, Gur RE. Brain Region and sex differences in age association with brain volume: A quantitative MRI study of healthy young adults. Am J Geriatric Psychiatry, 2002, 10,72-80.

176. Ragland JD, Turetsky BI, Gur RC, Gunning-Dixon F, Turner T, Schroeder L, Chan R, Gur RE. Working memory for complex figures: an fMRI comparison of letter and fractal n-back tasks. Neuropsychology, 2002, 16, 370-379.

177. Langleben DD, Schroeder L, Maldjian JA, Gur RC, McDonald S, Ragland JD, O'Brien CP, Childress AR. Brain activity during simulated deception: An event-related functional magnetic resonance study. NeuroImage, 2002, 15, 727-732.

178. Gur RC, Sara R, Hagendoorn M, Maron O, Hughett P, Turner T, Bajcsy R, Gur RE. A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. Journal of Neuroscience Methods, 2002, 115, 137-143.

179. Gur RC, Schroeder L, Turner T, McGrath C, Chan RM, Turetsky BI, Alsop D, Maldjian J, Gur RE. Brain activation during facial emotion processing. NeuroImage, 2002, 16, 651-662.

180. Gur RC, Gunning-Dixon F, Bilker WB, Gur RE. Sex Differences in temporo-limbic and frontal brain volumes of healthy adults. Cerebral Cortex, 2002, 12, 998-1003.

181. Gur RE, Gur RC. Gender differences in aging: Cognition, emotions and neuroimaging studies. Dialogues in Clinical Neuroscience. 2002, 4,197-207.

182. Goldstein RZ, Volkow ND, Chang L, Wang GJ, Fowler JS, Depue RA, Gur RC. The orbitofrontal cortex in methamphetamine addiction: Involvement in fear. NeuroReport, 2002, 13, 2253-2257.

183. Hill SK, Ragland JD, Gur RC, Gur RE. Neuropsychological profiles delineate distinct profiles of schizophrenia, an interaction between memory and executive function, and uneven distributions of clinical subtypes. Journal of Clinical and Experimental Neuropsychology. 2002, 24, 765-780.

184. Turetsky BI, Moberg PJ, Mozley LH, Moelter ST, Agrin R, Gur RC, Gur RE. Memory-delineated subtypes of schizophrenia: Relationship to clinical, neuroanatomical and physiological measures. Neuropsychology, 2002, 16, 481-490.

185. Gur RE, McGrath C, Chan RM, Schroeder L, Turner T, Turetsky BI, Kohler C, Alsop D, Maldjian J, Ragland JD, Gur RC. An fMRI study of facial emotion processing in schizophrenia. Am Journal of Psychiatry, 2002, 159, 1992-1999.

186. Rosen AC, Gur RC. Ethical considerations for neuropsychologists as functional magnetic imagers. Brain and Cognition. 2002, 50, 469-481.

187. Gunning-Dixon FM, Gur RC, Perkins AC, Schroeder L, Turner T, Turetsky BI, Chan RM, Loughead JW, Alsop DC, Maldjian J, Gur RE. Age-related differences in brain activation during emotional face processing. Neurobiology of Aging. 2003, 24, 285-95.

188. Kryspin-Exner I, Gur RC, Hoheisel B, Klein M, Six N. Neurobehavioral Probes: Adaptierung und Erweiterung von Verfahren zur computergestützen neuropsychologischen Diagnostik. Verhaltenstherapie und Verhaltensmedizin, 2003, 24, 27-51.

189. Kohler CG, Turner TT, Bilker WB, Brensinger C, Siegel SJ, Kanes, SJ, Gur RE, Gur RC. Facial emotion recognition in schizophrenia: Intensity effects and error pattern. American Journal of Psychiatry, 2003, 160, 1768-1774.

190. Silver H, Feldman P, Bilker WB, Gur RC. Workin memory deficit as a core neuropsychological dysfunction in schizophrenia. American Journal of Psychiatry, 2003, 160, 1809-1816.

191. Indersmitten T, Gur RC. Emotion processing in chimeric faces: hemispheric asymmetries in expression and recognition of emotions. The Journal of Neuroscience, 2003, 23, 3820-3825.

192. Bilker WB, Brensinger C, Kurtz MM, Kohler CK, Gur RC, Siegel SJ, Gur RE. Development of an abbreviated schizophrenia quality of life scale using a new method. Neuropsychopharmacology, 2004, in press.

193. Windischberger C, Barth M, Lamm C, Schroeder L, Bauer H, Gur RC, Moser E. Fuzzy cluster analysis of high-field fMRI data. Artificial Intelligence in Medicine, 2003, in press.

194. Ragland JD, Moelter ST, McGrath C, Hill SK, Gur RE, Bilker WB, Siegel SJ, Gur RC. Levels-of-processing effect on word recognition in schizophrenia. Biological Psychiatry, 2004, in press.

195. Kurtz MM, Moberg PJ, Ragland JD, Gur RC, Gur RE. Symptoms versus neurocognitive test perfomance as predictors of psychosocial status in schizophrenia. Schizophrenia Bulletin, 2004, in press.

196. Gur RE, Kohler CG, Turetsky BI, Siegel SJ, Kanes SJ, Bilker WB, Brennan AR, Gur RC. A sexually dimorphic ration of orbitofrontal to amygdala volume is altered in schizophrenia. Biological Psychiatry, 2004, in press.

197. Keefe RSE, Seidman LJ, Christensen BK, Hamer RM, Sharma T, Sitskoorn MM, Lewine RRJ, Yurgelun-Todd DA, Gur RC, Tohen M, Tollefson GD, Sanger TM, Lieberman JA. Comparative

effect of atypical and conventional antipsychotic drugs on neurocognition in first-episode psychosis: A randomized double-blind trial of olanzapine versus haloperidol. American Journal of Psychiatry, 2004, in press.

198. Ragland JD, Gur RC, Valdez J, Turetsky BI, Elliott M, Kohler CG, Siegel SJ, Kanes SJ, Gur RE. Event-related fMRI of frontotemporal activity during word encoding and recognition in schizophrenia. American Journal of Psychiatry, 2004, in press.

199. Irani F, Dankert M, Brensinger C, Bilker WB, Nair SR, Kohler CG, Kanes SJ, Turetsky BI, Moberg PJ, Ragland JD, Gur RC, Gur RE, Siegel SJ. Patient attitudes towards surgically-implantable, long-term delivery of psychiatric medicine, Neuropsychoharmacology, 2004, in press.

Contributions to peer-reviewed clinical research publications, participation cited but not by authorship:

None.

Research Publications, non-peer reviewed: None.

Abstracts: (Excluding abstracts subsequently published as full-length papers; Past 3 years only)

1. Gur RE, Ragland JD, Turetsky BI, Gur RC. Sex Differences in Temporal-Limbic Function in Schizophrenia: Relation of Memory to Structural and Functional Measures. International Congress of Schizophrenia Research, Baltimore, MD, 1998.

2. Lazarev, Ragland JD, Gur RE, Bilker W, Turetsky BI, Swanson, Gur RC. Correlations Between Frontal and Temporal Lobe Brain and CSF Volume with Resting Glucose Metabolism.

3. Ragland JD, Lazarev, Gur RC, Klimas B, Gur RE. Word Recency and Recognition Judgements in Patients with Schizophrenia. International Congress of Schizophrenia Research, Baltimore, MD, 1998.

4. Gur RE, Ragland JD, Turetsky BI, Gur RC. Sex Differences in Temporal-Limbic Function in Schizophrenia: Relation of Memory to Structural and Functional Measures. International Congress of Schizophrenia Research, Santa Fe, New Mexico, 1999.

5. Gur RE, Gur RC, Turetsky B, Ragland D, Arnold SE. Temporolimbic Function in Schizophrenia: An Integrative Approach From Behavior to the Molecule. Society of Biological Psychiatry, Washington, D.C., 1999.

Editorials, Reviews, Chapters, including participation in committee reports:

1. Gur RE, Gur RC. Correlates of conjugate lateral eye movement behavior in humans. In S.

Harnad et al (eds.), Lateralization in the nervous system. New York: Academic Press, 1976.

2. Gur RE, Levy J, Gur RC. Clinical studies of brain organization and behavior. In A. Frazer, A. Winokur (eds.), Biological basis of psychiatric disorders. New York: Spectrum, 1977.

3. Sackeim HA, Gur RC. Self-confrontations, self-deception and consciousness. In G.E. Schwartz, D. Shapiro (eds.), Consciousness and self-regulation: advances in research. New York: Plenum Press, 1978.

4. Gur RC. Imagery, absorption and the tendency toward "mind exploration" as correlates of hypnotic susceptibility in males and females. In F.H. Frankel, H.S. Zamansky (eds.), Hypnosis at its bicentennial: selected papers. New York: Plenum, 1978.

5. Gur RC, Gur RE. Handedness and individual differences in hemispheric activation. In J. Herron (ed.), The Neuropsychology of left-handedness. New York: Academic Press, 1979.

6. Levy J, Gur RC. Individual differences in psychoneurologicalorganization. In J. Herron (ed.), The neuropsychology of left-handedness. New York: Academic Press, 1979.

7. Sackeim HA, Gur RC. Facial symmetry, perceiver biases and the communication of emotion. In J.T. Cacioppo, R.E. Petty (eds.), Social Psychophysiology. New York: Guilford, 1983.

8. Gur RC. Measurement and imaging of regional brain function: Implications for neuropsychiatry. In P. Flor-Henry, J. Gruzelier (eds.), Laterality and psychopathology. Amsterdam: Elsevier, 1983.

9. Reivich M, Gur RC. Sensory and psychological effects on local glucose metabolism in humans. In M. Reivich (ed.), Positron emission tomography. New York: Liss, 1985.

10. Gur RC, Gur RE, Sussman NM, Selzer M. Positron emission tomography in epilepsy. In M. Reivich (ed.), Positron emission tomography. New York: Liss, 1985.

11. Gur RC. Measurement of regional brain physiology in humans: early applications in behavioral neurology. In M-M. Mesulam (ed.), Principles of behavioral neurology. Philadelphia: F.A. Davis, 1985.

12. Alavi A, Chawluk JB, Hurtig HI, Dann RW, Saykin AJ, Gur RC, Reivich M. Determination of regional cerebral function and structure in normal aging and dementia with positron emission tomography (PET), magnetic resonance imaging (MRI) and x-ray computed tomography (XCT). In H. Hafner, G. Moschel, N. Sartorius (eds.), Mental health in the elderly. Heidelberg: Springer-Verlag, 1985.

13. Gur RC, Gur RE. Hemispheric specialization and regional cerebral blood flow. In A. Glass (ed.), Individual differences in hemispheric specialization. New York: Plenum Publications, 1987.

14. Gur RC. Imaging the activity of the human brain. National Forum, 1987, 67, 13-16.

15. Gur RC. Integration of multifaceted data on regional brain function: from metabolic images

to behavioral images. In R. Takahashi, P. Flor-Henry, J. Gruzelier, S. Niwa (eds.), Cerebral dynamics, laterality and psychopathology. Amsterdam: Elsevier, 1987.

16. Gur RC. Atlas of brain imaging: The 133-Xenon inhalation technique. Masters in Psychiatry, 1987.

17. Gur RC, Gur RE, Saykin AJ. The neuropsychological study of schizophrenia. In C.A. Tamminga, S.C. Schulz (eds.), Advances in Neuropsychiatry and Psychopharmacology, Vol 1. Schizophrenia Research. New York: Raven Press, 1991, pp.153-162.

18. Gur RC, Gur RE. The use of neuroimaging techniques in brain injury. In J. Dywan, R. Kaplan (eds.), Neuropsychology and the Law. New York:Springer-Verlag, 1991.

19. Gur RC, Gur RE. The impact of neuroimaging on human neuropsychology. In R.G. Lister, H.J. Weingartner (eds.), Perspectives on Cognitive Neuroscience. Oxford University Press, 1991, 23, 417-435.

20. Gur RC, Gur RE. Laterality in schizophrenia: PET studies. In N.D. Volkow (ed), Positron Emission Tomography in Schizophrenia,1991, 3, 47-58.

21. Gur RC, Saykin AJ, Gur RE. Neuropsychological assessment in psychiatric research and practice. In S.B. Guze, J. Helzer (eds), Psychiatry, 1991, 72, 1-16.

22. Gur RC, Gur RE. Neurobehavioral and neuroimaging data in the medical-legal context. In H.V. Hall, R.J. Sbordone (eds.), Disorders of Executive Functions: Civil and Criminal Law Applications, 1993, 4, 107-122.

23. Gur RC. Neuropsychological methods for evaluating regional brain dysfunction. In A. Wilner (ed.), Cerebral Damage Before and After Cardiac Surgery. England: Kluwer, 1993, 8, 101-111.

24. Gur RC, Gur RE, Saykin AJ. Behavioral Imaging: The Neuropsychological Assessment. In K. Maurer (ed.), Imaging of the Brain in Psychiatry and Related Fields. New York: Springer-Verlag, 1993, 351-361.

25. Gur RC, Saykin AJ, Gur RE. Brain function in schizophrenia: application of neurobehavioral studies. In N.C. Andreasen (ed.), Schizophrenia: From Mind to Molecule, 1994, 5, 93-104.

26. Gur RC, Gur RE. Methods for the study of brain-behavior relationships. In A. Frazer, P.B. Molinoff, A. Winokur (eds.), Biological Bases Of Brain Function and Disease. Raven Press. 1994, 15, 261-279.

27. Gur RC, Gur RE. The potential of physiological neuroimaging for the study of schizotypy: experiences from applications to schizophrenia. In A. Raine, T. Lencz, S.A. Mednick (eds.), Schizotypal Personality. Cambridge University Press, 1995, 17, 406-425.

30. Gur RC, Gutierrez JM, Holdnack JA, Mahr RN. Neurobehavioral probes as applied in physiologic neuroimaging studies: Methodologic considerations. In E.D. Bigler (ed.),

Handbook of Human Brain Function: Neuroimaging. New York: Plenum Press, 1996, 10, 199-214.

31. Gur RC, Moberg PJ, Gur RE. Aging and cognitive functioning. In R. Lavizzo-Mourey, M. Forciea (eds.), Geriatric Secrets, 1996, 26, 126-129. 3rd Edition in press.

32. Gur RE, Gur RC. Blood flow and metabolism in schizophrenia. In R. Mathew (ed.), Cerebral Blood Flow in Neuropsychiatric Disorders. Great Neck, NY: PMA Publishing Corp., in press.

33. Gur RE, Gur RC. Schizophrenia: Brain Structure and Function. In H.I. Kaplan, B.J. Sadock (eds.), Comprehensive Textbook of Psychiatry/VII, Philadelphia: Lippincott Williams & Wilkins, 2000.

34. Gur RC, Cowell P, Gur RE. Gender Differences in Neuropsychological Testing. To appear in L.J. Dickstein, B.L. Kennedy (eds.), Gender Differences in the Brain: Linking Biology to Psychiatry. New York:Guilford Publications, Inc.

35. Gur, R.C., Moelter, S.T., Ragland, JD. "Learning and memory in schizophrenia". In, Sharma, T., & Harvey, P. (Eds.), Cognition in Schizophrenia. Oxford University Press: Oxford, G.B., 1999.

36. Gur RC, Gur RE. Neuroimaging applications in the elderly. American Journal of Geriatric Psychiatry, 2002, 10, 5-11.

COMMENTARIES:

1. Gur RC. Measuring hypnotic susceptibility: A guest editorial. American Journal of Clinical Hypnosis, 1979, 21, No. 2 and 3, (October 1978/January 1979). (Two issues devoted to the psychometrics of hypnotizability, edited by RC Gur).

2. Gur RE, Gur RC. A note on Levick and Voneida: Eye movements in schizophrenics vs. normal subjects. Archives of General Psychiatry, 1979, 36, 493-494.

3. Sackeim HA, Gur RC. Asymmetry in facial expression. Science, 1980, 209, 834-836.

4. Sackeim HS, Gur RC. Voice recognition and the ontological status of self-deception. Journal of Personality and Social Psychology, 1985, 48, 1365-1368.

5. Gur RE, Skolnick BE, Gur RC. Gruzelier's "reconsideration" considered. (Response to J. Gruzelier's reconsideration of Gur et al [1983] conclusions). Archives of General Psychiatry, 1985, 42, 633.

6. Saykin AJ, Gur RC. A review of Neuropsychological Assessment of Neuropsychiatric Disorders. In I. Grant, K.M. Adams (eds.), New York, Oxford University Press, 1986.

7. Gur RC. Do we have a research method for studying self reports? A review of A. Giorgi (ed.), Phenomenology and psychological research. Contemporary Psychology: A Journal of Reviews, 1987, 32, 547-548.

8. Gur RC, Gur RE. Hypofrontality in schizophrenia: RIP. The Lancet, 1995, 345, 1383-1384.

9. Lenkinsky RE & Gur RC. MRS: A novel tool for studying brain function. Contemporary Psychology: A Journal of Reviews, 1997, 42, 351-352.

10. Moberg, P.J., Doty, R.L., Turetsky, B.I., Arnold, S.E., Mahr, R.N., Gur, R.C., Bilker, W., & Gur, R.E. . Olfactory identification abilities deteriorate in patients with schizophrenia, even for those with relatively recent onset [letter; reply]. American Journal of Psychiatry, 1998, 155, 1463-1464.

11. Gur RC &.McBride, T. Toward a unitary description of neuropsychological functions: a review of Handbook of Clinical and Experimental Neuropsychology Edited by Gianfranco Denes and Luigi Pizzamiglio. Contemporary Psychology, 2000, 45, 682-683.

PATENT:

Gur RC, Gur RE, Trivedi SS. "Behavioral Imaging: Topographic Display of Neuropsychological Data, U.S. Patent No. 4862359.

INVESTIGATOR ON GRANTS (Active Support):

1. NIMH RO1 MH60722 – The Neurobiology of Affective Dysfunction in Schizophrenia
   Co-PI (with Raquel E. Gur, MD, PhD)
   Project Period: 12/1/00-11/30/05; Annual Costs $252,097

2. NINDS RO1 NS39135 Quantitative MRI and 1H-MRS in Traumatic Brain Injury
   Robert Grossman, MD, PI; Investigator
   Project Period: 4/01/00-03/31/05; Annual Direct Costs: $427,468;

3. NIA RO1 AG17524 STN Stimulation: DA Transporters & Outcomes In Parkinson's
   P. David Mozley, MD, PI; Investigator
   Project Period: 2/15/00-01/31/05; Annual Direct Costs: $240,246

4. NIH RO1 DC04278 Olfaction in Epilepsy.
   Richard Doty, PhD, PI; Investigator
   Project Period: 1/1/00-12/31/04; Annual Direct Costs: $216,253

5. NIH RO1 MH42191 – A Neurobehavioral Family Study of Schizophrenia,
   Co-PI (with Raquel E. Gur, MD, PhD)
   Project Period: 3/1/2001-2/28/2006; Annual Direct Costs: $416,030

6. NIH P50 MH64045 Center: The Neurobiology of Stimulus Encoding in Schizophrenia
Racquel E. Gur, MD PhD; PI; Investigator
Project Period: 08/01/01-06/30/06; Annual Direct Costs: $1,426,622

	Total 20%	
Sub-project: Core A	$190,233	5%
Sub-project: Project II	$194,777	15%

7. T32 MH19112-10 Schizophrenia: A Neuropsychiatric Perspective
$250,000/yr.

Exhibit 2

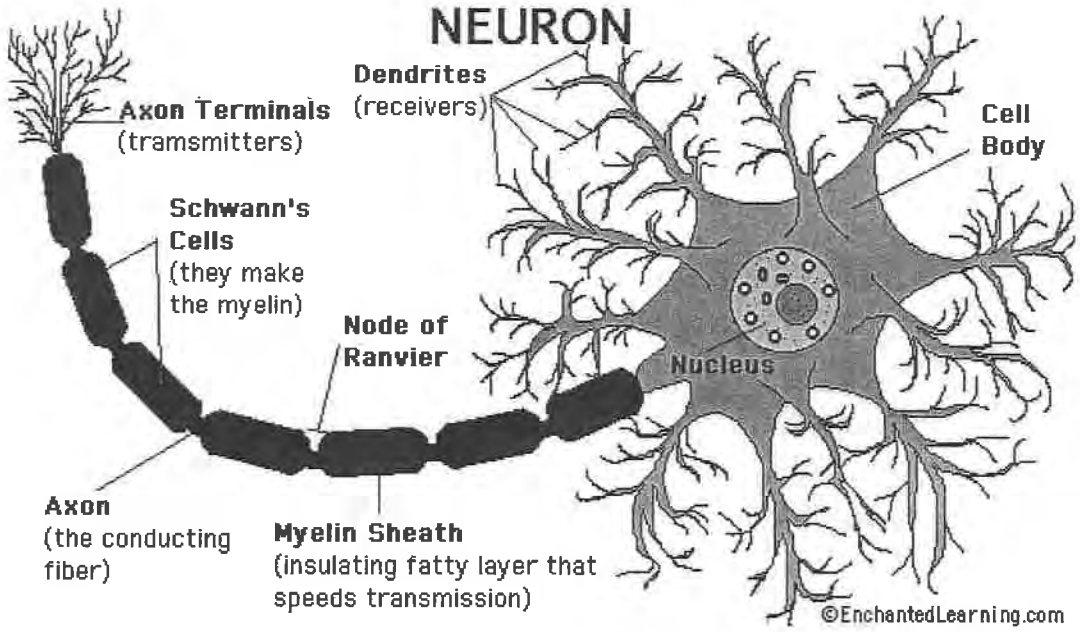


EXHIBIT
DR. GUR
2
3/19/04 FK

FILED
NOV 23 2004
Clerk of the Courts
Rec'd By

Exhibit 3

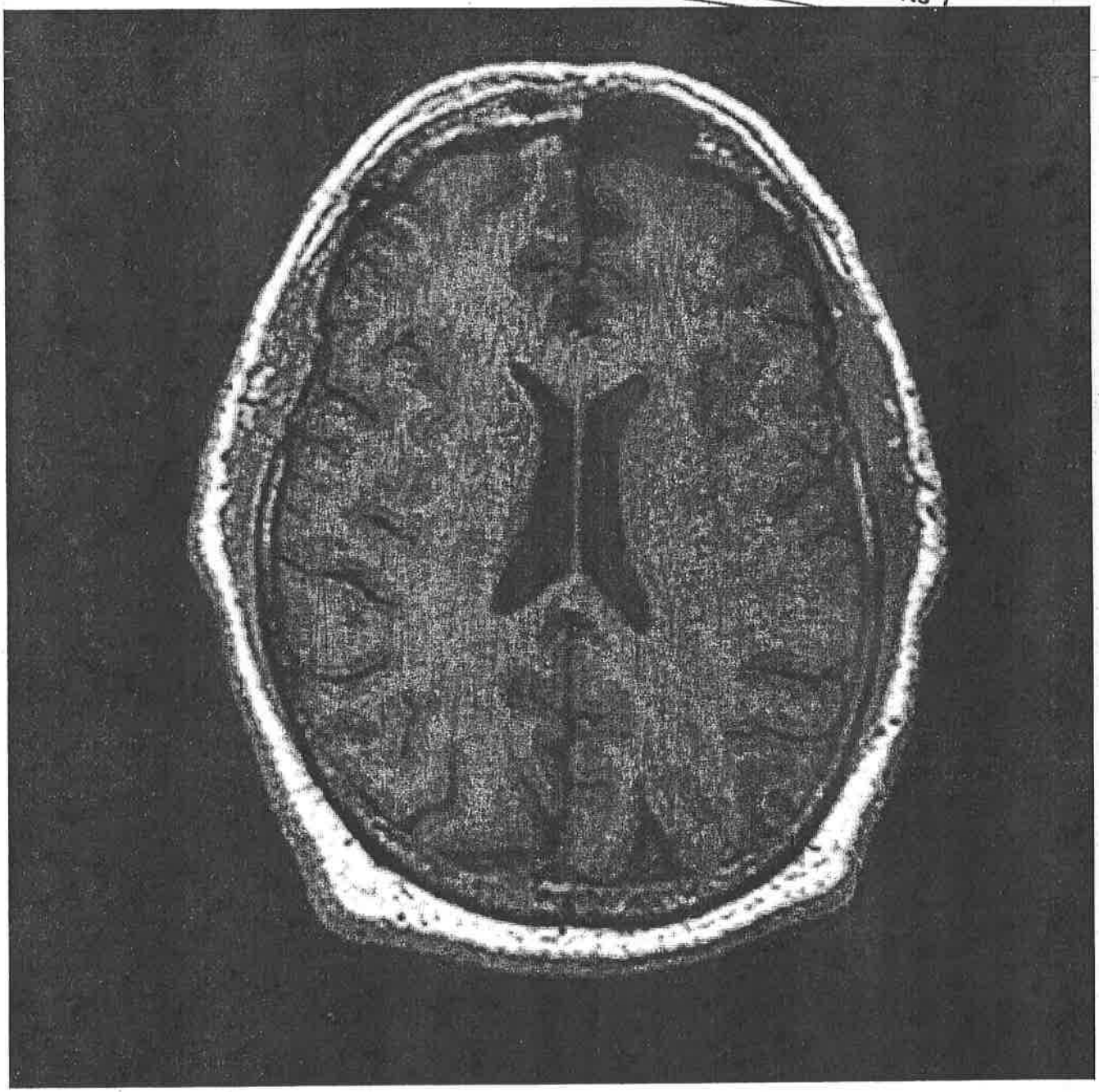
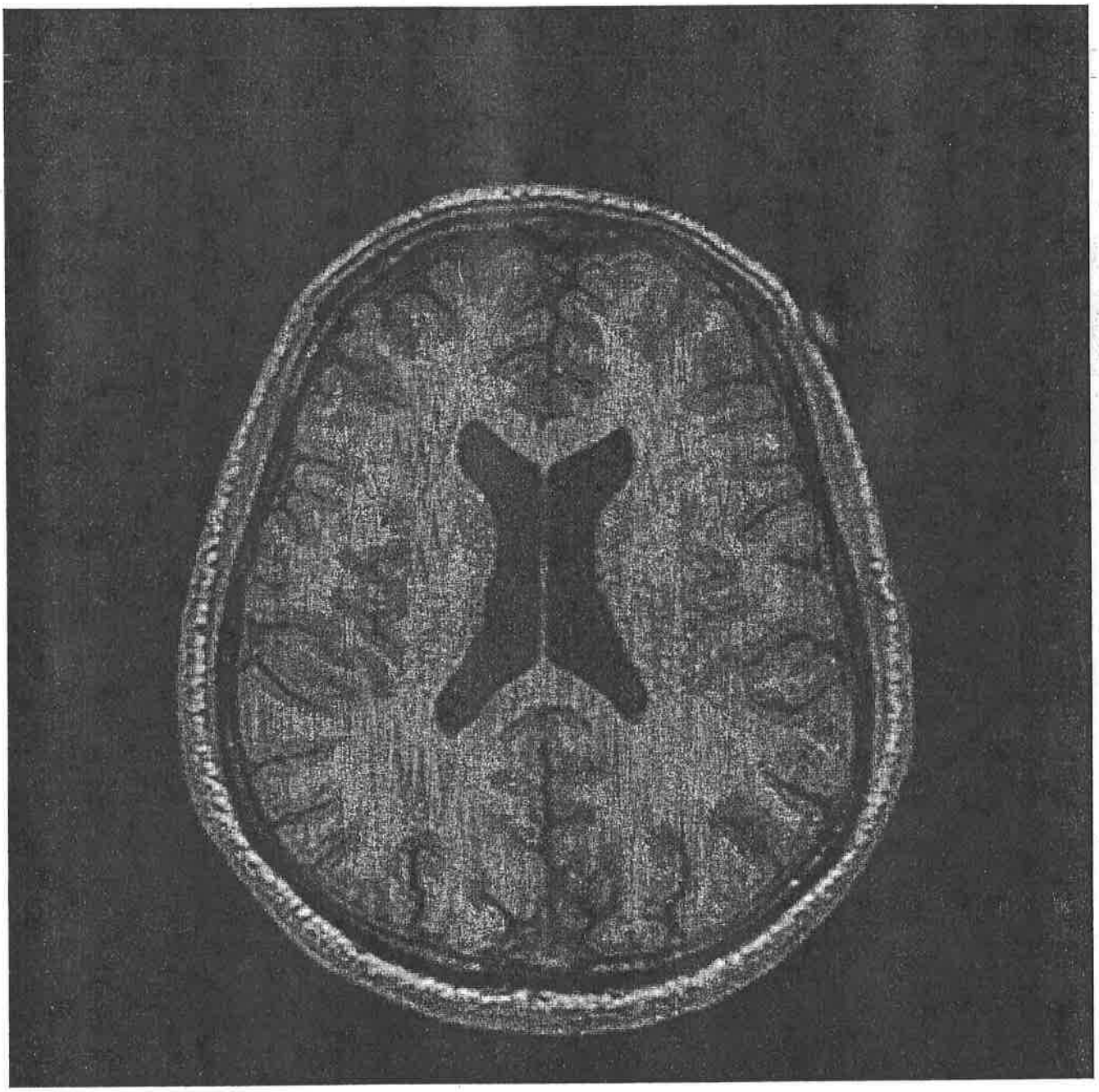
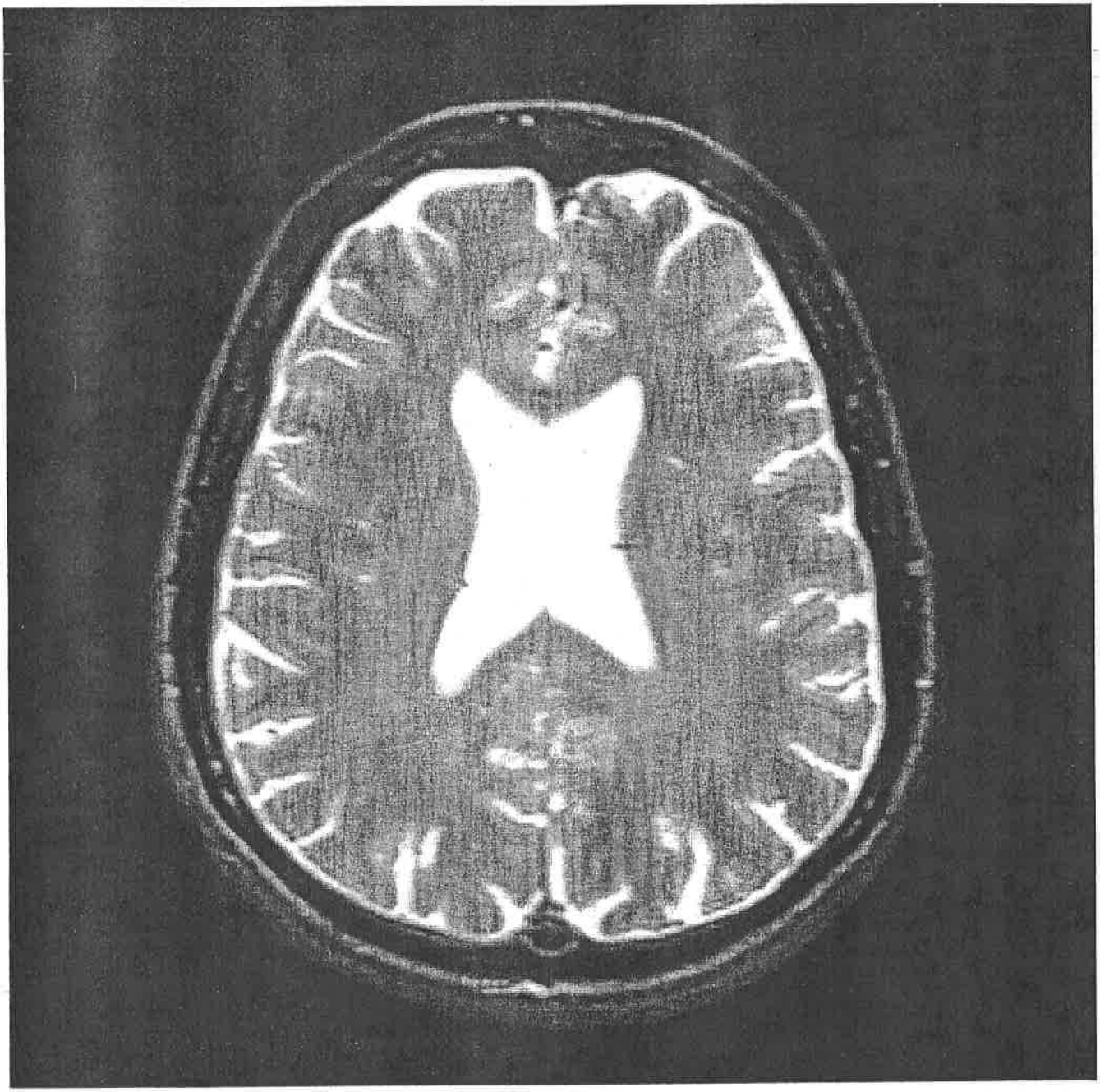


Exhibit 4



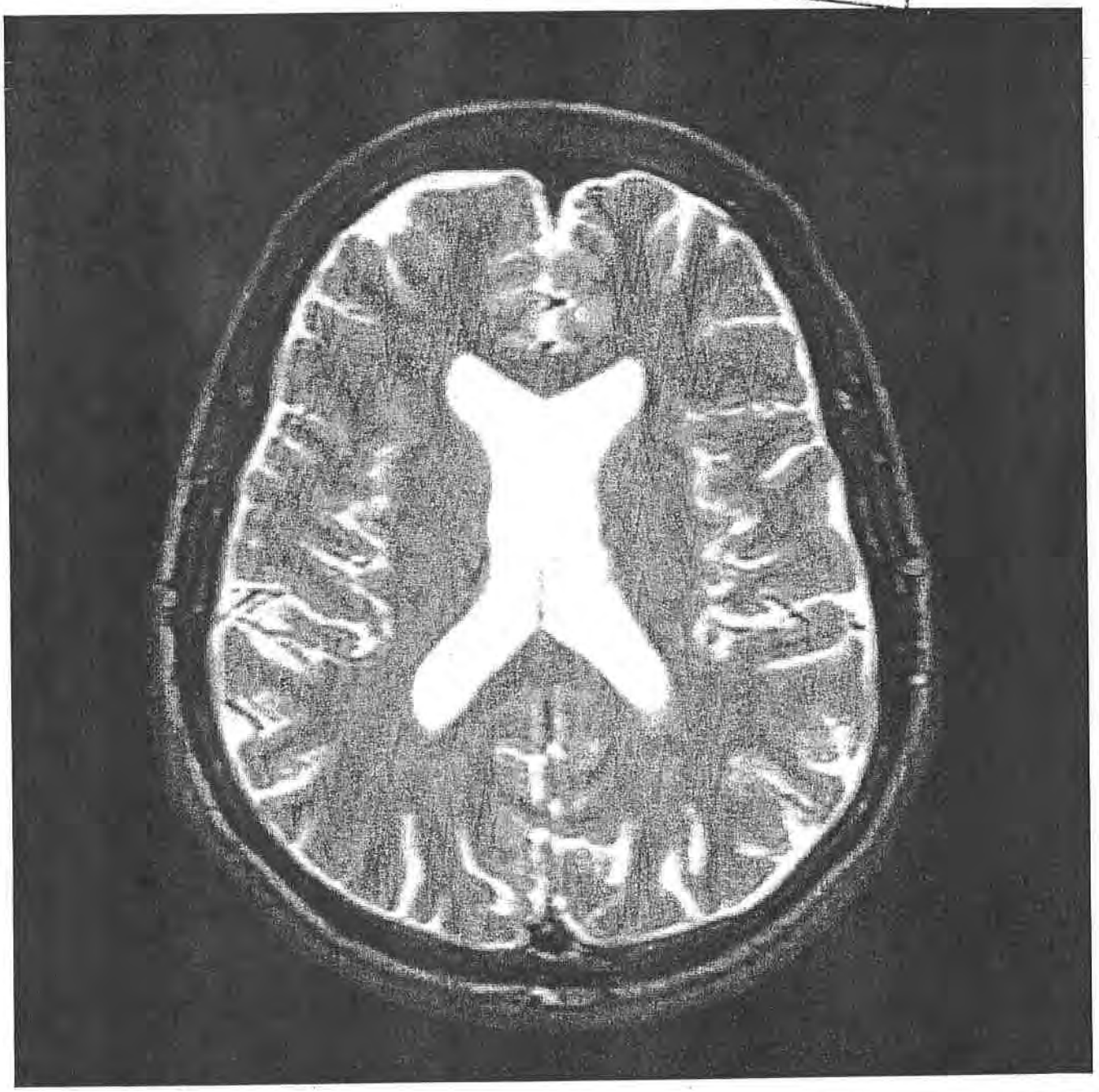
ML014

Exhibit 5



MLOY

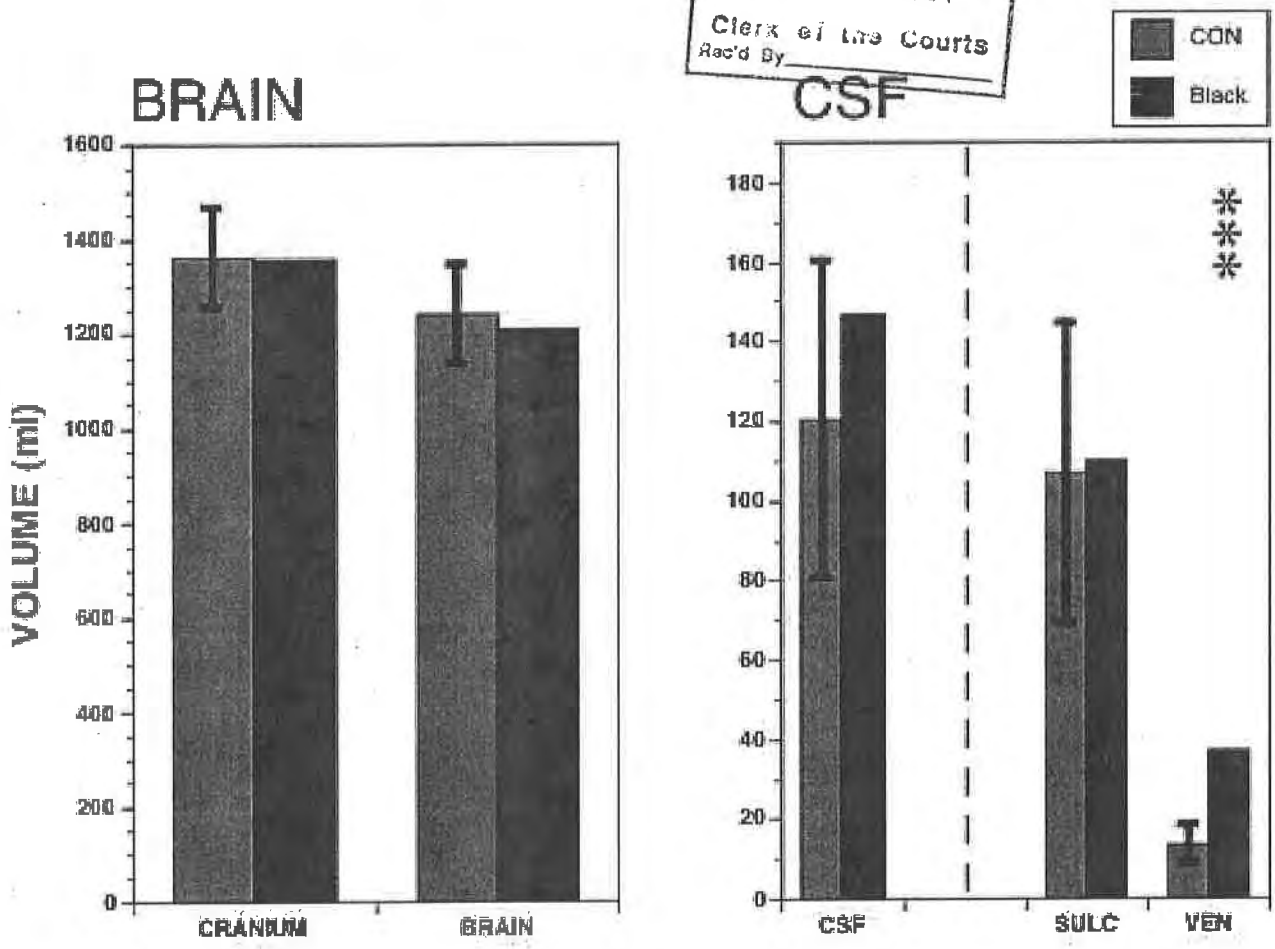
Exhibit 6



ML 014

Exhibit 7

Figure qMRI-1. Volume in milliliters (ml=cm^3) for a sample of healthy control men (CON_M, N=79, in gray bars ±SD) and Mr. Byron Black (black bars)



*** Significantly abnormal**

Exhibit 8

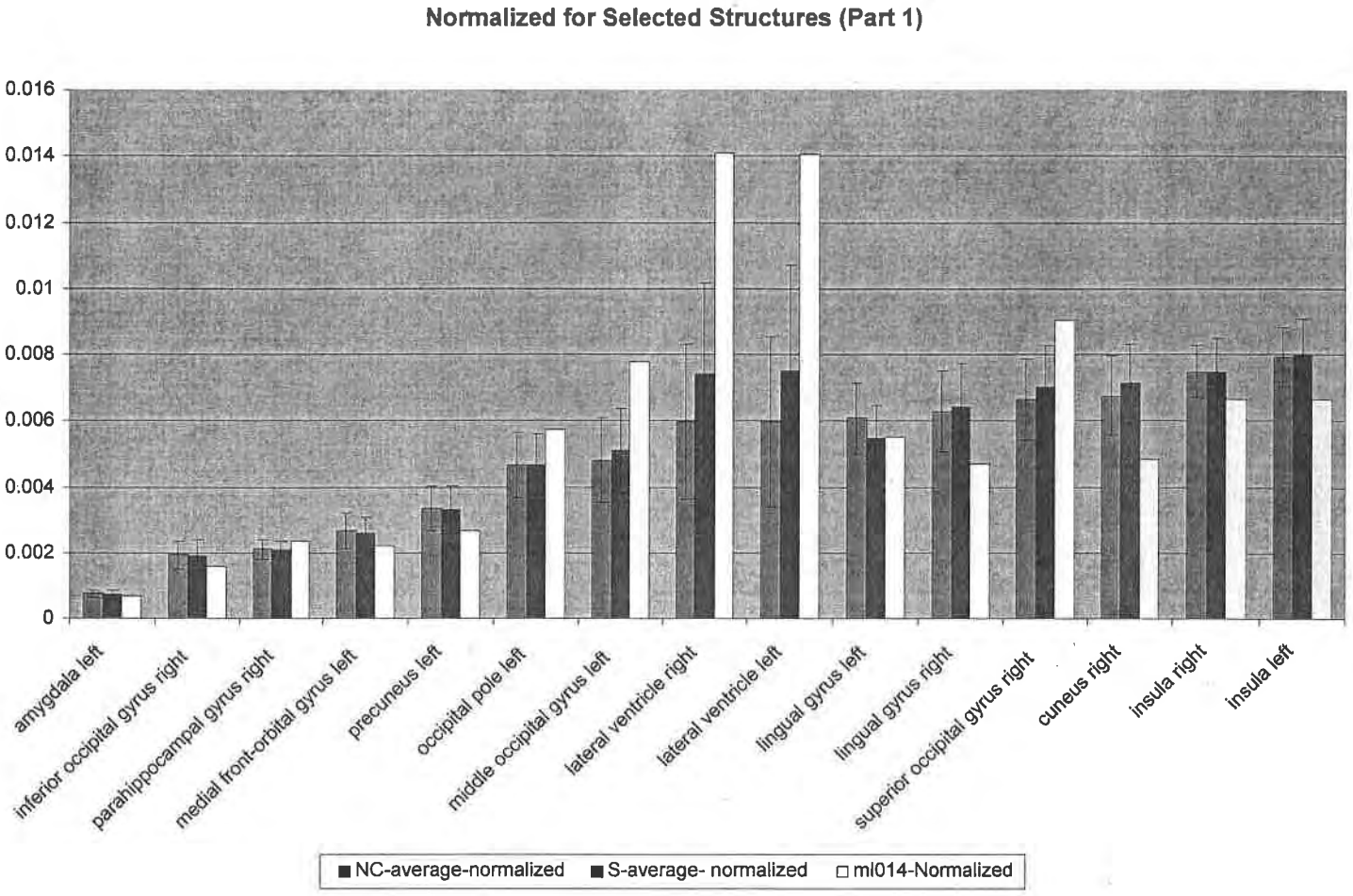
Report on GUR_ML014, by Christos Davatzikos, Ph.D

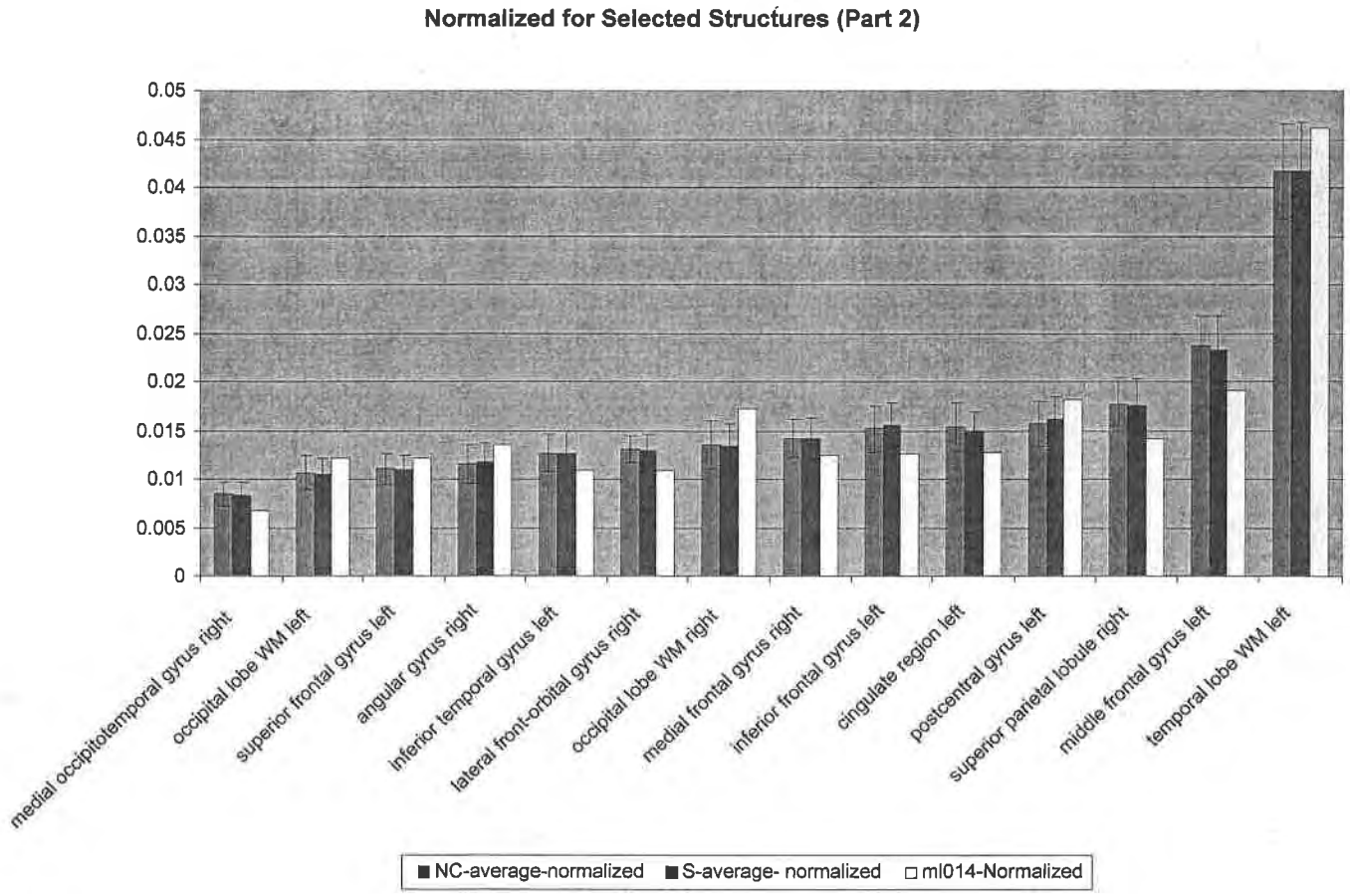
The images of GUR_ML014 were examined via delineation of 92 regions of interest (ROI), which was assisted by a semi-automated template warping algorithm.

The overall evaluation resulted in the conclusion that GUR_ML014 has a highly unusual brain structure, when compared to 79 healthy men. Moreover, GUR_ML014 was compared to a group of 69 patients with schizophrenia and was found to be completely different from them as well. Therefore, GUR_ML014's abnormality must be related to some condition other than schizophrenia. In order to evaluate "how normal" GUR_ML014 is, we constructed a principal component statistical model from our 79 healthy men. Using cross-validation (a sample-reuse technique for robust parameter estimation evaluating each individual on statistical models that use other individuals), we found that 100% of normal controls were more similar to the average normal brain than GUR_ML014 was. Thus, if we were to randomly select an individual from the general normal population, we would have 0% chance of finding an individual that has brain structure as unusual as that of GUR_ML014. The same was true for schizophrenia patients, i.e. GUR_ML014 would appear to be 100% different from randomly selected schizophrenia patients.

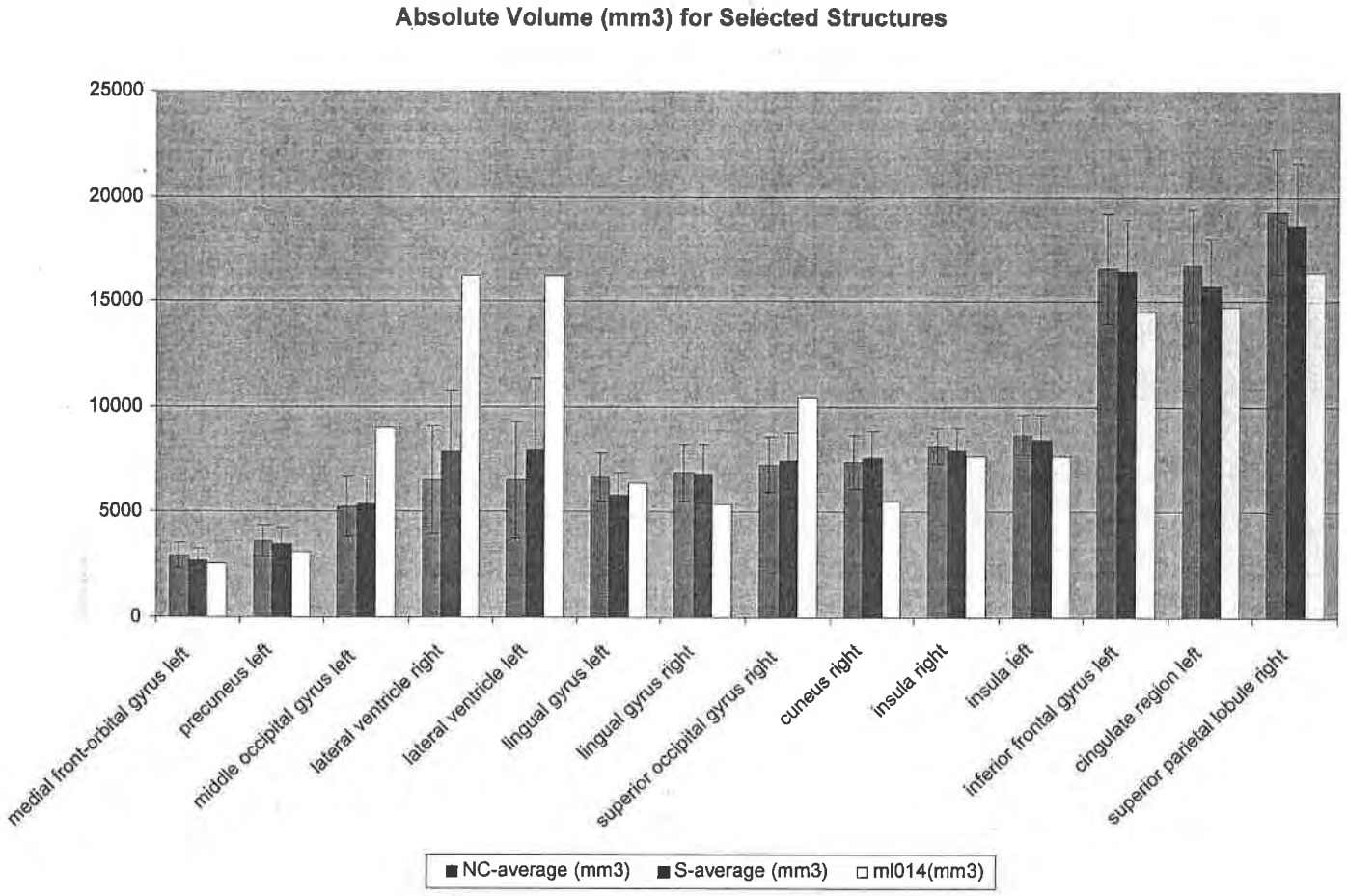
In order to localize GUR_ML014's structural abnormalities, we examined the 3 principal components - after excluding the first two principal components that accounted for overall size variability, which seemed to be most informative of GUR_ML014's brain abnormality. These components showed a striking ventricular enlargement. Since ventricular expansion typically occurs when brain tissue is lost to a variety of reasons, or when brain tissue fails to form during development, we can conclude that GUR_ML014 had significantly lower amount of brain tissue, relative to normal individuals. However, this was the case only in certain brain regions, and it was not an effect present throughout the entire brain. In particular, the aforementioned principal components revealed that the combination of ventricular enlargement, along with atrophy or malformation that was particularly pronounced in the right medial orbitofrontal, middle frontal, and inferior frontal gyri, the right cingulated and precuneus, the left inferior frontal, middle frontal and cingulate gyri were the most characteristic structural abnormalities in GUR_ML014. The relative large volume of primary visual and sensory-motor areas further supports the conclusion that underlying structural abnormalities in regions mediating executive functions and emotional modulation were the result of a targeted brain damage or malformation. This most important principal component was approximately 2.4 standard deviations away from the average normal profile. The subsequent two principal components were 1.4 and 2.6 standard deviations away from the average normal profile, respectively, further increasing the overall level of GUR_ML014's abnormal brain structure, and revealing additional abnormalities in the entorhinal and perirhinal areas.

Examination of volumes of individual structures may help appreciate the magnitude of these abnormalities. We present them below, stressing the caveat that even if an individual has measurements that fall within normal range, this individual might be highly abnormal if a certain combination of these measurements, which is analogous to the aforementioned principal components, is highly abnormal. For example, one can look at an eye of relatively small size and blue color, and conclude that this eye is normal. Then, the observer can look at another eye of larger size and green color, which might also appear to be normal. However an observer presented with these two eyes belonging to the same person will conclude that the combination of these two eyes is a highly unusual pattern for a normal individual.


Normalized for Selected Structures (Part 1)
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
amygdala left
inferior occipital gyrus right
parahippocampal gyrus right
medial front-orbital gyrus left
precuneus left
occipital pole left
middle occipital gyrus left
lateral ventricle right
lateral ventricle left
lingual gyrus left
lingual gyrus right
superior occipital gyrus right
cuneus right
insula right
insula left
NC-average-normalized
S-average- normalized
ml014-Normalized

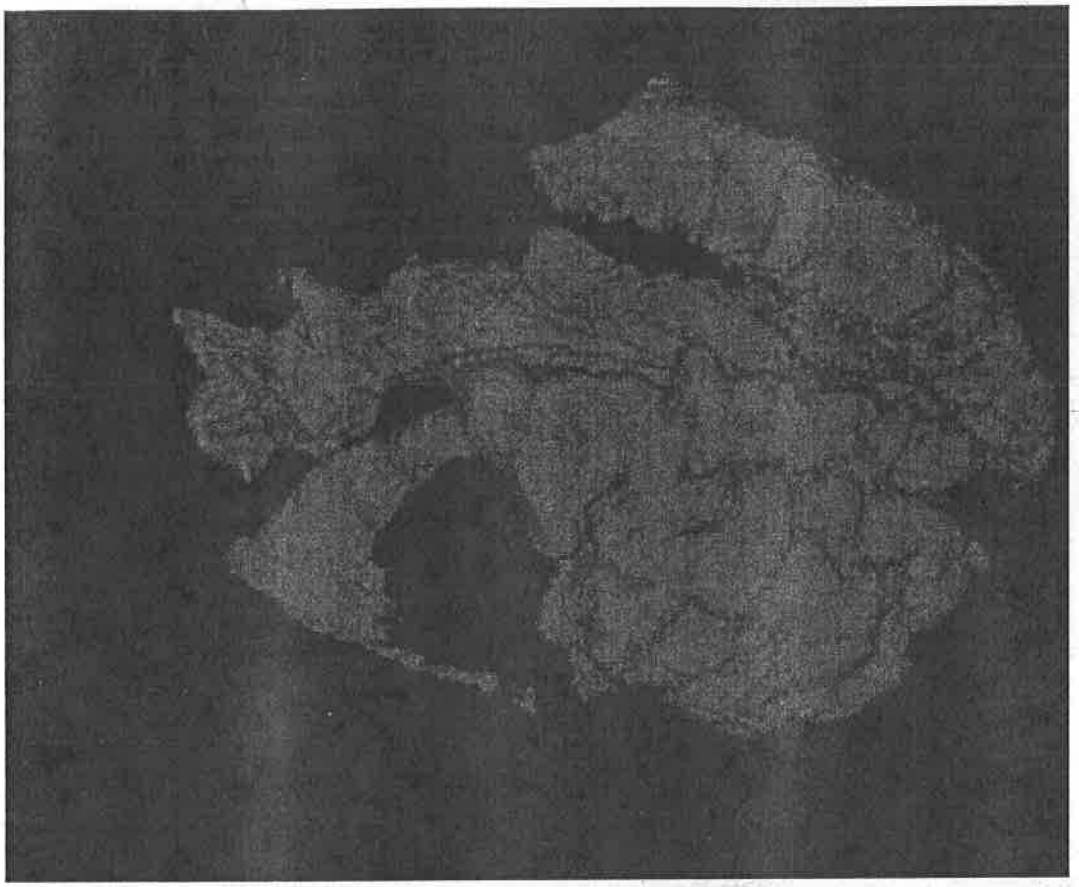


Above are volumetric measurements of ROIs, after they were normalized to total brain+ventricular volume. Therefore, a large number reflects a relatively preserved structure, and a small number reflects a relatively damaged, or potentially underdeveloped structure. The bars correspond to 1 standard deviation from the average value of the respective group. Blue=79 normal controls, purple=69 schizophrenia patients, yellow=GUR_ML014.



Absolute volumetric measurements, i.e. exact volumetric measurements without any normalization for global brain volume.

Exhibit 9



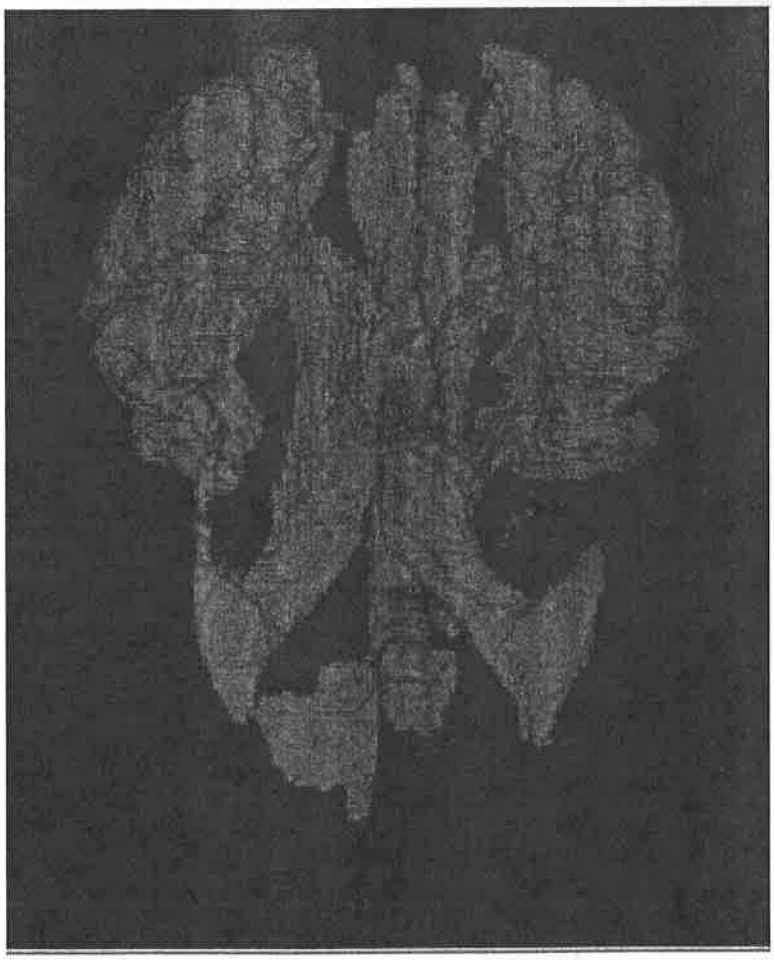


Figure qMRI-4. A 3-dimensional overlay rendering of regions showing volume loss in Mr. Byron Black relative to healthy brains. Top and bottom views are provided.

Exhibit 10

Ruben C. Gur PhD ABPP/CN
Neuropsychology

Voice/Fax: (215)247-2716 email: gur@bbl.med.upenn.edu

815 Saint George's Rd.
Philadelphia, PA 19119

QUANTITATIVE STRUCTURAL BRAIN IMAGING CONSULTATION
DRAFT

Patient's name: **Byron Black**
Date of birth: March 23, 1956
Date of report: March 17, 2004
Referring source: Don Dawson, Esq.

Summary of qMRI quantitative analysis for Mr. Byron Black

MRI scans were acquired on a GE Signa 1.5 Tesla scanner (Milwaukee, Wisconsin). T1-weighted images were obtained with a spoiled gradient-recalled pulse sequence (SPGR) using flip angle 35o; TR 35ms, TE 6ms, field of view 24cm, NEX 1, 1mm slice thickness and no interslice gaps. Transaxial images were acquired in planes parallel to the orbitomeatal line with in-plane resolution of 0.9375x0.9375mm. No parenchymal lesions or skull abnormalities were evident neuroradiologically.

The brain volumes were extracted by automated procedures described earlier[1]. Prefrontal subdivisions were derived with neuroradiologic and neuroanatomic input, using topographical triangulation and tissue segmentation techniques to maximize the precision and reliability of region delineation. Subfrontal regions were parcellated manually by an expert (Bruce I. Turetsky, MD) who had achieved a reliability > .90 calculated by the intraclass correlation (ICC) method for whole brain and frontal and temporal volume quantification (as described in Cowell et al[2] and Gur et al[3]). The orbital prefrontal region includes the rectal, medial orbital and suborbital gyri, the ventral portion

---

[1] Kohn MI, Tanna NK, Herman GT, Resnick SM, Mozley PD, Gur RE, Alavi A, Zimmerman RA, Gur RC. Analysis of brain and CSF volumes from magnetic resonance imaging: methodology, reliability and validation. Radiology, 1991, 178, 115-122; Gur RE, Turetsky BI, Bilker WB, Gur RC. Reduced gray matter volume in schizophrenia. Arch Gen Psychiatry, 1999, 56, 905-911.

[2] Cowell PE, Turetsky BT, Gur RC, Grossman RI, Shtasel DL, Gur RE. Sex differences in aging of the human frontal and temporal lobe. The Journal of Neuroscience, 1994, 14, 4748-4755.

[3] Gur RE, Cowell PE, Latshaw A, Turetsky BI, Gorssman RI, Arnold SE, Bilker WB, Gur RC. Reduced dorsal and orbital prefrontal gray matter volumes in schizophrenia. Archives of General Psychiatry, 2000, 57, 761-768.

of the mesial superior gyrus, and the anterior, posterior and lateral orbital gyri. The lateral portion of the orbital region includes Brodmann area (BA) 47, lateral portions of BA11, and inferolateral portions of BA10. The medial portion of the orbital region corresponds to BA12, BA25, medial BA11, inferomedial BA10, and ventral BA32 and BA24. The MRI data were then submitted to semi-automated analysis using semi-automated template warping algorithms developed by Christos Davatzikos, PhD[4]. Data are available for comparison on a large normative sample of healthy people, and Mr. Black's data were compared to those of 79 healthy young men at his age range.

The results of the quantitative analysis indicated a cranial volume (1356.2ml) in the normal range, as well as normal volume of supertentorial whole-brain tissue (1209.4ml). However, as can be seen in **Figure qMRI-1**, there is an abnormally high volume of ventricular cerebrospinal fluid (CSF) for the whole brain. The automated parcellation procedure (see enclosed report by Dr. Davatzikos) confirmed the finding of enlarged ventricles and noted pervasive abnormalities consistent with atrophy or malformation, which was particularly pronounced in the right medial frontal, orbitofrontal, inferior frontal, middle frontal and cingulate gyri, the right precuneus, and the left inferior frontal, middle frontal and cingulate gyri (**Figures qMRI 2-4**).

Interpretation

The volumetric analysis of the MRIs revealed pervasive abnormalities in the whole-brain CSF volume, which in ventricles was almost twice than that of healthy men at his age range. Regional analysis indicated abnormally low volumes in the right medial frontal and orbitofrontal regions and bilaterally in the inferior and middle frontal gyri and in the cingulate gyrus. The nature and extent of abnormality indicate very early trauma and in all likelihood Mr. Black was born with most of these abnormalities.

The etiology of these abnormalities is uncertain, but most consistent with parenchymal loss caused by poor intrauterine conditions (e.g., illness of the mother during pregnancy, a diabetic mother, or a mother who is abusing alcohol or some other toxins). Indeed, the pattern of large ventricles associated with callosal anomalies has been quite strongly linked to the fetal alcohol syndrome. The other possibility is that the parenchymal loss reflects head injury, most probably in early childhood. While head injury alone could not explain all the anomalies, head injuries could have exacerbated them by inducing further tissue loss in frontal and ventricular regions that are most susceptible to the effects of impact and shearing.

Abnormalities of this extent and location will have major behavioral consequences. Tissue loss in the inferior, mesial frontal and orbitofrontal system would severely impair Mr. Black's ability to control emotion and aggressive impulses and to

---

[4] e.g., Shen D, Davatzikos C. Very high-resolution morphometry using mass-preserving deformations and HAMMER elastic registration. Neuroimage, 2003, 18, 28-41.

consider the context for his actions. Loss in these regions would also be associated with deficits in declarative memory and reality orientation. Tissue loss in the cingulate gyri impairs his ability to resolve conflict and integrate his actions with situational demands.

These abnormalities in volume would explain the patient's poor executive behavior, his loss of impulse control, and the neurocognitive deficits documented in neuropsychological testing. The MRI findings also help interpret the PET results, suggesting structural underpinnings to regions of reduced metabolic activity. A worrisome finding is that some regions showing volume reduction also show abnormally elevated levels of cerebral glucose utilization. The increased metabolic activity in the cingulate gyri seems to have induced increased glycolysis in temporal regions susceptible to the kindling of seizure foci. Whether this grim prognosis is warranted depends on the quality of Mr. Black's cerebral circulation, which has not been assessed (see PET report).

Prepared by

Ruben C. Gur Ph.D., ABPP
Professor
Director of Neuropsychology and the Brain Behavior Laboratory

Exhibit 11

DECLARATION OF RUBEN C. GUR, PH.D.

I, Ruben C. Gur, Ph.D., state the following:

1. I am a neuropsychologist, with a special focus on imaging applications to the diagnosis and study of people with severe behavioral disturbances associated with brain dysfunction. Counsel for Byron Black requested that I evaluate records and results of preliminary psychological and neuropsychological testing, perform my own evaluation, and render an opinion regarding his diagnosis and competency and recommend additional steps needed to diagnose his condition.

2. Having done so, it is my expert opinion that there is sufficient evidence in the available data to indicate brain dysfunction. This declaration details my background, identifies the bases for my opinion, and presents my expert opinion.

Synopsis of Curriculum Vitae

3. My Curriculum Vitae is attached to this declaration. My qualifications for the opinions I state in this declaration include the following:

a. I have been licensed as a psychologist in Pennsylvania since 1976. I received a B.A. from the Hebrew University of Jerusalem and an M.A. and a Ph.D. in Clinical Psychology from Michigan State University. I completed Postdoctoral Fellowships at Stanford University and at the University of Pennsylvania.

b. I am a Diplomate on the American Board of Professional Psychology, with Specialty in Clinical Neuropsychology (ABPP/CN).

c. I am, or have been, a member of the American Psychological Association, Division of Physiological and Comparative Psychology (Fellow), Division of Neuropsychology (Fellow), the American Psychological Society (Fellow), the American College of Neuropsychopharmacology (Fellow), the American Association for the Advancement of Science, the International Neuropsychological Society, the National Academy of Neuropsychologists, the New York Academy of Science, and the John Morgan Society.

d. Among other honors, I have received the Erikson Award for Scientific Excellence and the 1990 Stephen V. Logan Award from the National Alliance for the Mentally Ill. I have authored or co-authored refereed publications in peer-reviewed journals, made numerous national and international presentations in the field of imaging and brain dysfunction, have served and am serving on Editorial Boards of professional journals, have served on Search Committees for journal Editorship, and have reviewed manuscripts for leading journals in the areas of imaging, brain and behavior, and schizophrenia, have served on Advisory Panels and Study Sections of the National Institutes of Health and currently serve on the NIH Review Group on "Clinical Neuroscience and Biological Psychopathology." I have contributed chapters to textbooks and other scholarly volumes on the topic of brain imaging, neuropsychology and schizophrenia.

e. I have the academic rank of Professor (with Tenure) on the Standing Faculty of the University of Pennsylvania, with a primary appointment in Psychiatry and secondary appointments in Neurology and in Radiology. I am currently the Director of Neuropsychology, Department of Psychiatry at the Hospital of the University of Pennsylvania. I am Principal Investigator of the Neuropsychology Core and the Functional Imaging Core of the Federally funded Schizophrenia Center, Co-PI of the functional MRI project of the Conte Center for Neurosciences, and Co-PI and investigator on several individual NIH grants (RO1s) on brain imaging and psychopathology. I also supervise interns and practicum students in neuropsychology. I am the Co-founder and advisor of the Biological Basis of Behavior Undergraduate Major Program at the University of Pennsylvania. Additionally, I am a supervisor of postdoctoral Fellows and doctoral students in Psychology and Neuroscience and Co-PI of a Federally funded Training Program in the behavioral neurosciences.

f. I have participated in the diagnosis of hundreds of individuals where issues similar to this case were raised requiring neuropsychological testing and neuroimaging.

g. I have been recognized as an expert and allowed to testify with respect to my expert opinions in the specialty of Neuroimaging and Neuropsychology in state and Federal

courts.

Bases of Opinion--Background and Clinical History

4. I have reviewed the following documents:

Birth Certificate of Byron Black
Hospital Birth Records of Byron Black
Educational Records of Byron Black
Medical Records of Byron Black
- Baptist Hospital
- Meharry Hospital (General Hospital formerly)
- Metro Health Records
- Riverbend Maximum Security Prison Health Records
- Vanderbilt Clinic & Hospital Records

Incarceration Records of Byron Black
Psychological Records and Transcript of Testimony
- Kenneth Anchor, Ph.D. ABPP Licensed/Board Certified and Clinical Psychologist
- Pamela Auble, Ph.D. Clinical Neuropsychologist
- William Bernet, M.D. Psychiatrist
- Gillian Blair, Ph.D. Licensed Psychologist
- DeDe Wallace Center Competency Records
  - Calvilyn Y. Allmon, M.S.S.W.
  - Bradley Diner, M.D.
  - Leonard Morgan, Jr., Ph.D. Clinical Psychologist
- Pat Jaros, M.A. Licensed Psychological Examiner
- William Kenner, M.D. Psychiatrist
- Patti van Eys, Ph.D. Licensed Clinical Psychologist

Transcript of Competency Hearing Byron Black
Mackey v. State 537 S.W.2nd 704 (TN 1975)
First Degree Murder Statute
Mental Retardation Statute 39-13-203 pages 46-47
Mitigation Statute 39-13-204 page 25
Interview by Libby Moore April 23, 1997 of Julia Mai Black, Finis Black, Dan Black and Alberta Black Crawford.
Declaration of Connie Westfall
Interview of Lynette Childs Black 04/26/97 by Connie Westfall
Declaration of Gaye Nease
Interview of Jackie M. Thomas 09/26/01 by Gaye Nease
Interview of Alberta Black Crawford 03/19/01 by Gaye Nease
Interviews of Lynette Childs Black 03/24/01 & 11/10/01 by Gaye Nease
Interview of Johnny Moore 08/15/01 by Gaye Nease
Interview of Mary Frances Coplan 11/05/01 by Gaye Nease
Interview of Finis Black 03/23/01 by Gaye Nease
Interview of Mary C. Harrison 03/15/01 by Gaye Nease
Interview of Arleta Black Swanson and Karen Black Greer 10/18/01 by Gaye Nease
Interview of Richard Corley 10/11/01 by Gaye Nease

Interviews of Melba Black Corley 03/22/01 & 10/10/01 by Gaye Nease
Interview of Freda Black Whitney 03/17/01 by Gaye Nease
Miranda Warning information
Transcript of Evidence State of Tennessee v. Walter R. Kendricks, Case # 92-C-1496 pgs 73-152
Medical and Death Information on Julia Mai Black
Declaration of Ross Alderman

5. The documents I have reviewed provide a moderately high index of suspicion that Mr. Black suffers from a brain disorder. The main factors in support of this possibility are: I. Reports of head injuries; II. Reports of exposure to neurotoxins; III. Reports of physicians and family members; IV. Performance on the psychological and neuropsychological tests; V. Behavior during trial and appeals; VI. Behavior in prison; VII. Behavior during interview

6. The behavioral effects of brain injury are assessed with neuropsychological testing. These procedures provide measures of performance on major behavioral domains that can be linked to brain systems. There are several standardized "Neuropsychological Batteries" and several such batteries have been administered to Mr. Black. They reveal significant deficits indicative of brain dysfunction. The areas of deficits, combined with my own testing and information from the records I have reviewed, indicate damage in frontal and temporal lobe functions, particularly those related to the limbic system. Against a background of low intellectual abilities, deficits are particularly pronounced in executive functions, memory and emotion processing. This conclusion is buttressed by the "behavioral imaging" algorithm[1]

---

[1] Gur RC, Trivedi SS, Saykin AJ, Gur RE. "Behavioral imaging" - a procedure for analysis and display of neuropsychological test scores: I. Construction of algorithm and initial clinical evaluation. Neuropsychiatry, Neuropsychology and Behavioral Neurology, 1988, 1, 53-60.

Gur RC, Saykin AJ, Blonder LX, Gur RE. "Behavioral imaging": II. Application of the quantitative algorithm to hypothesis testing in a population of hemiparkinsonian patients. Neuropsychiatry, Neuropsychology and Behavioral Neurology, 1988, 1, 87-96.

Gur RC, Saykin AJ, Benton A, Kaplan E, Levin H, Kester DB, Gur RE. "Behavioral imaging": III. Inter-rater agreement and reliability of weightings. Neuropsychiatry, Neuropsychology, and Behavioral Neurology, 1990, 3, 113-124

applied to the available neuropsychological data (see Figure enclosed). Such deficits indicate cerebral dysfunction even when abnormalities are undetected by MRI or PET. They are seen in people with head injuries and in schizophrenia.

7. These deficits result in difficulty focusing and sustaining attention. Byron Black is unable to persevere with a task for a sustained period of time. He is easily distracted by irrelevant stimuli. His deficits have a significant deleterious effect in many aspects of real world functioning such as participating meaningfully in a courtroom setting. A person with his deficits is likely to misinterpret or miss altogether significant portions of courtroom proceedings.

8. Mr. Black's brain is damaged in those areas responsible for impulse control and inhibition. A person with his deficits is likely to jump to conclusions, misinterpret emotional expressions of others be unusually emotional himself. He is prone to act impulsively and to have significant difficulty controlling his behavior.

9. Mr. Black will have unusual difficulty learning to associate causes and effect and will have a significantly diminished ability to recognize or avoid undesirable consequences.

10. Mr. Black is likely to misinterpret the statements and actions of others and to act impulsively on that misinterpretation.

11. His brain impairments are clearly revealed in tests of memory functioning both verbal and non verbal. It affects his short term recall, as well as long term. He has significant deficits in non verbal memory and his performance declines markedly as the complexity of the task increases. His overall performance on the Halstead-Reitan is in the moderately severe impairment range, which is significant.

12. These impairments effect all aspects of his ability to problem solve and process information. His impairment would seriously interfere with his ability to keep pace with courtroom proceedings.

13. Byron Black was exposed to neurotoxins *in utero* and as a small child. It is

documented that Mr. Black's mother drank throughout pregnancy. Additional exposures include the fact that Mr. Black was at high risk for lead poisoning and likely exposed to lead. Some risk factors for lead poisoning which are relevant to Mr. Black include that he grew up in a house that was built before 1950 and in zip code where more than 27% of the housing was built before 1950 (lead paint and lead water pipes), he ate and chewed on non-food items such as paint (as a baby Byron chewed on an old, wooden, shellacked crib) or dirt (yard area was dirt, Byron liked to play in the dirt), had a family member that worked in house construction or repair, (as a young child, Byron followed his grandfather to work doing house repairs in the neighborhood) and belonged to a high risk group, e.g. poverty level. In addition, Byron had iron deficiency anemia as an infant. We know that iron deficiency can increase gastrointestinal absorption of lead. Finally, Mr. Black has been an avid football player at varsity level and has suffered several head injuries, some requiring stitches. While there has not been a formal diagnosis of concussion, such head injuries, individually and cumulatively, are likely contributors to some of the symptoms of brain damage displayed in his behavior and testing.

14. Each of these exposures can contribute to impairment of frontal lobe functioning, including poor impulse control and emotional disinhibition. Exposure to these toxins causes structural damage to the brain, including orbital frontal and temporal lobes that contribute to attentional disorder and motor impairment.

15. Byron also demonstrates a symptom complex associated with serious psychiatric disorders. The symptoms include: paranoid and delusional beliefs, as well as negative symptoms of schizophrenia. These symptoms produce attentional problems as well as misinterpretations of environmental stimuli, such as courtroom proceedings. These psychiatric symptoms coupled with frontal, temporal, and limbic system impairment compound his inability to understand and appreciate reality. He is unable to distinguish between reality and his delusions and is unaware that he suffers from psychiatric illness.

16. His neurologically based impulse control deficits, his inability to control his behavior in order to avoid undesirable consequences, his unusually high distractability and his

impaired and paranoid ideation, greatly compromise his capacity for careful thought and weighing of consideration for and against a proposed course of action. His perception of self interest was severely distorted by his underlying psychiatric disorder and damaged brain. Byron lacks the abilities to make decisions based on their long term consequences.

17. The brain is a complex organ, as could be expected from the complexity of human behavior, and lacking tools for studying the living brain has made scientific progress slow and laborious. However, methods developed in the '70s and implemented in the '80s have yielded powerful tools for obtaining reliable measures of brain structure, function and behavior. These methods have become standard in the assessment of brain disorders.

18. Magnetic Resonance imaging (MRI) has become the major method for assessing the structural integrity of the brain, namely brain anatomy. Many brain disorders, formerly requiring expert and sometimes ingenious clinical procedures for diagnosis, can now be diagnosed by visual inspection of the MRI scans. MRI also permits highlighting of specific features of the brain by controlling scan parameters and using contrast agents. Clinical reading of the printed images is sometimes, unfortunately, insufficient to detect effects of some disorders, particularly those associated with diffuse or subtle loss of tissue. Such disorders require the use of reliable methods for soft tissue segmentation and volumetric analysis, which permit accurate quantitation of global and regional gray matter, white matter, and cerebrospinal fluid compartments.

19. Brain anatomy can be intact yet the brain is still dysfunctional because of aberrant activity. This is seen in many brain disorders, including cases of epilepsy, Parkinson's disease, and early dementia. Brain activity is associated with physiologic change, and positron emission tomography (PET) provides the most accurate quantitative measures of several parameters important for assessing brain physiology. Although PET is a versatile method enabling the measurement of parameters related to both energy metabolism and neurotransmitter function, most relevant for assessing brain dysfunction is the ability to measure local glucose metabolism using ^{18}F-fluoro-d-2-deoxyglucose (FDG). Normative data are available to detect and document

abnormalities. Ideally, the procedure should include quantitative measurement using arterial or arterialized blood samples. However, the data can be useful even if the study is done in a facility that has not been certified for arterial modeling studies. It is also helpful, although not always essential, to perform measurement of cerebral blood flow (CBF) using an appropriate ligand (e.g., $^{15}OH_2$). Ideally, CBF should be obtained both at rest and during activation with neurobehavioral probes. However, even resting baseline values will help establish areas of uncoupling that could be both diagnostic and prognostic.

20. Integration of the clinical, neuroanatomic, neurophysiologic and neuropsychological data is required to determine competency, diagnosis, and the extent to which brain impairment may have caused Mr. Black's behavior. I have experience in such integration, have done it in other capital cases (including death penalty appeals). I have been consulting with Dr. Robert Kessler, a neuroradiologist at Vanderbilt University, in an effort to coordinate a comprehensive neuroimaging study on Mr. Black. Unfortunately, time constraints have prevented us from being able to conduct the testing prior to today's filing deadline. After reviewing the neuroimaging results, I will be able to render a final opinion.

DATE: 11/15/2001

Dr. Ruben Gur, Ph.D.

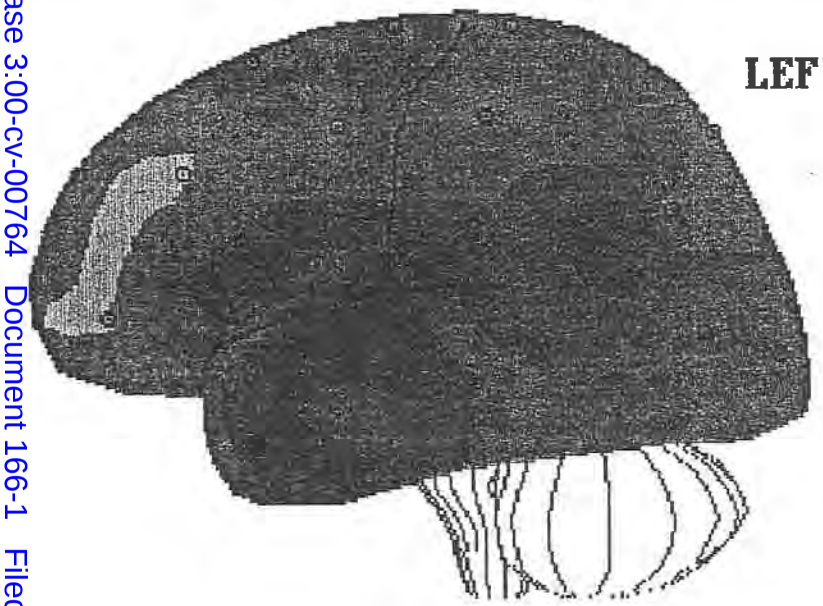
MAIN File Edit Select Calculate Display

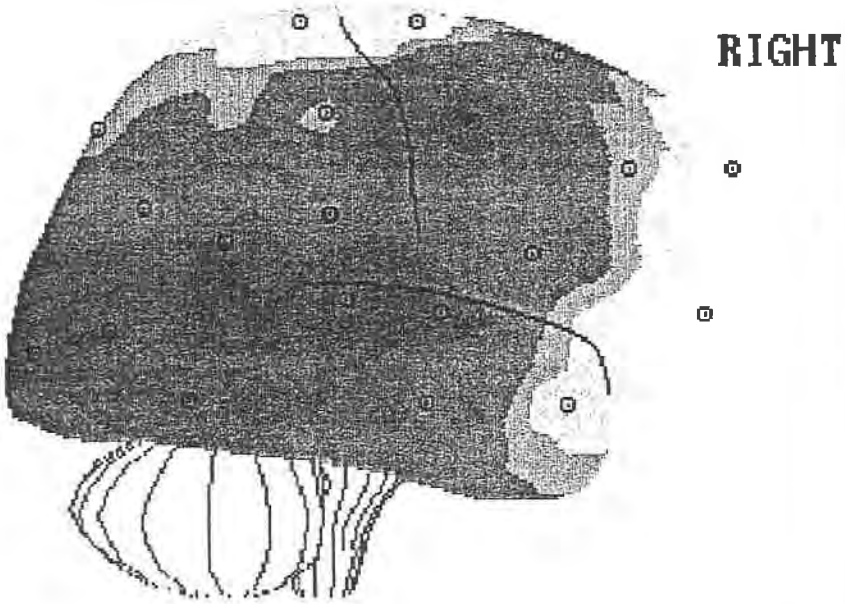
Bio-logic®

BI Version P2.35

I.D. bb001
Weight Table: AVERAGE

Note:





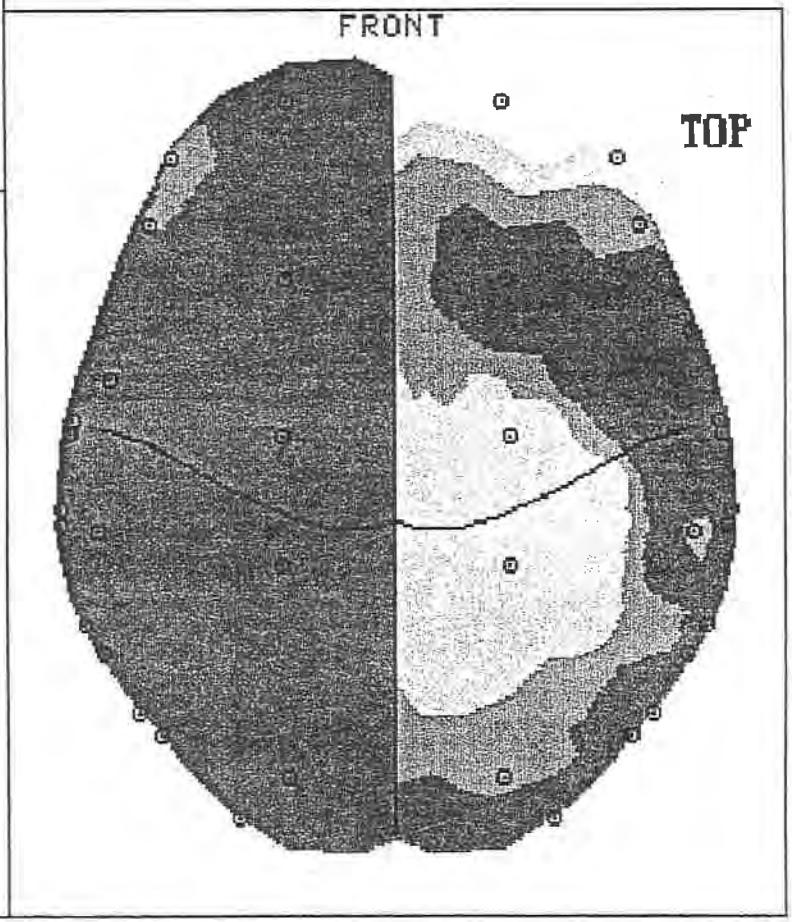


Exhibit 12

Ruben C. Gur PhD ABPP/CN
Neuropsychology
Voice/Fax: (215)247-2716 email: gur@bbl.med.upenn.edu

815 Saint George's Rd.
Philadelphia, PA 19119

QUANTITATIVE FUNCTIONAL BRAIN IMAGING CONSULTATION
DRAFT

Patient's name: **Byron Black**
Date of birth: March 23, 1956
Date of report: February 29, 2004
Referring source: Don Dawson, Esq.

Summary of PET quantitative analysis for Mr. Byron Black

The PET ^{18}F-FDG study of cerebral metabolic rates for glucose (CMRgl) was deemed technically sound. A quantitative analysis of the PET data was requested to evaluate whether regional activity values deviate statistically from normal.

To perform the quantitative analysis, count data were downloaded electronically to a SUN Microsystems workstation and a standard set of regions of interest (ROIs) was applied using validated procedures[1]. Dr. J. Daniel Ragland performed placement of ROIs and the region-to-whole-brain (R/WB) ratio values were transferred electronically to a Macintosh computer for graphical display using Deltagraph® (v4.5 for Mac). Two graphs were generated, one to evaluate regional hypo- or hyper-metabolism and the other to focus on any asymmetries.

Examination of the R/WB regional values (**Figure PET-1**) showed abnormal relative metabolism in a significant number (15 out of 35, $p < 0.001$) regions. Notably,

---

[1] Gur RE, Resnick SM, Alavi A, Gur RC, Caroff S, Dann R, Silver F, Saykin AJ, Chawluk JB, Kushner M, Reivich M. Regional brain function in schizophrenia: I. A positron emission tomography study. Archives of General Psychiatry, 1987, 44, 119-125.
Gur RC, Mozley LH, Mozley PD, Resnick SM, Karp JS, Alavi A, Arnold SE, Gur RE. Sex differences in regional cerebral glucose metabolism during a resting state. Science, 1995, 267, 528-531
Ragland JD, Glahn DC, Gur RC, Censits DM, Smith RJ, Mozley PD, Alavi A, Gur RE. PET regional cerebral blood flow change during working and declarative memory: Relationship with task performance. Neuropsychology, 1997, 11, 222-231.

focal decrease in metabolism was documented in Insula (IN), Orbital Frontal (OF), Rectal Gyrus (RG), Corpus Callosum – Anterior (C1), Corpus Callosum – Posterior (C2), Lenticular – Medial [Globus Pallidus] (LM), while increases were notable in Dorsal Prefrontal – Medial (DM), Occipital cortex, Medial (OM), Occipital cortex, Lateral (OL), Occipital Temporal (OT), Mid–Temporal (MT), Parahippocampal Gyrus (PH), Cingulate Gyrus – genu (CG), Cingulate Gyrus – Posterior (CP), and Pons (PO). Reduced insular and orbitofrontal, as well as rectal gyrus activity could be a result of head injuries. However, reduced callosal activity is consistent with neurodevelopmental anomalies. Pending comparison with structural MRI, it could also reflect partial volume effects of enlarged ventricles. The abnormally increased activity in some temporal, occipital and frontal regions could reflect epileptiform or psychotic process. This and the finding of reduced white matter callosal activity could also relate to shearing associated with closed head injury. The findings of cortical hyper-metabolism are worrisome because they may lead to further tissue deterioration, unless there is also increased cerebral blood flow, which we were not able to determine.

Hemispheric asymmetry measures (**Figure PET-2**) indicate that the increased activity in the Occipital Temporal and the reduced activity in the rectal gyrus are lateralized to the left. The increased activity in the left cerebellum would be normal if Mr. Black is left-handed. This can be verified against neuropsychological test data.

Interpretation

Quantitative examination of the PET data indicates an abnormal shift in regional activity, with hypo-metabolism in regions associated with arousal and emotional control, and hyper-metabolism of glucose in several cortical regions. Abnormal asymmetries in these regions were also documented. While the abnormalities in regional CMRgl are not themselves diagnostic in that they do not converge to support a specific etiology, they do establish with statistical certainty that Mr. Black has abnormal brain function. The extent to which the increased glucose metabolism in cortex, particularly temporal areas, is cause for concern depends of the adequacy of cerebral perfusion. This can be established with a PET study using a tracer such as $^{15}OH_2$ for measuring regional cerebral blood flow (rCBF). The interpretation of all these PET abnormalities also depends strongly on the anatomic analysis (pending).

Prepared by

Ruben C. Gur Ph.D., ABPP
Professor
Director of Neuropsychology and the Brain Behavior Laboratory

Abbreviations in Figures:

SF = Superior Frontal; DL = Dorsal Prefrontal – Lateral; DM = Dorsal Prefrontal – Medial; MF = Mid–Frontal; IF = Inferior Frontal; SM = Sensorimotor; SP = Superior Parietal; SG = Supramarginal Gyrus; OL = Occipital cortex, Lateral ; OM = Occipital cortex, Medial; LI = Lingual Gyrus; FG = Fusiform Gyrus; OT = Occipital Temporal; ST = Superior Temporal; MT = Mid–Temporal; IT = Inferior Temporal; TP = Temporal Pole; PH = Parahippocampal Gyrus; HI = Hippocampus; AM = Amygdala; IN = Insula; OF = Orbital Frontal; RG = Rectal Gyrus; CA = Cingulate Gyrus – Anterior; CG = Cingulate Gyrus - genu; CP = Cingulate Gyrus – Posterior; C1 = Corpus Callosum – Anterior; C2 = Corpus Callosum – Posterior; CN = Caudate Nucleus; LM = Lenticular – Medial [Globus Pallidus]; LL = Lenticular – Lateral [Putamen]; TH = Thalamus; MI = Midbrain; PO = Pons; CE = Cerebellum.

PAR = Parietal
OCC = Occipital
CC = Corpus callosum
BG = Basal ganglia
BS = Brainstem

Exhibit 13

MAIN File Edit Select Calculate Display

Bio-logic®

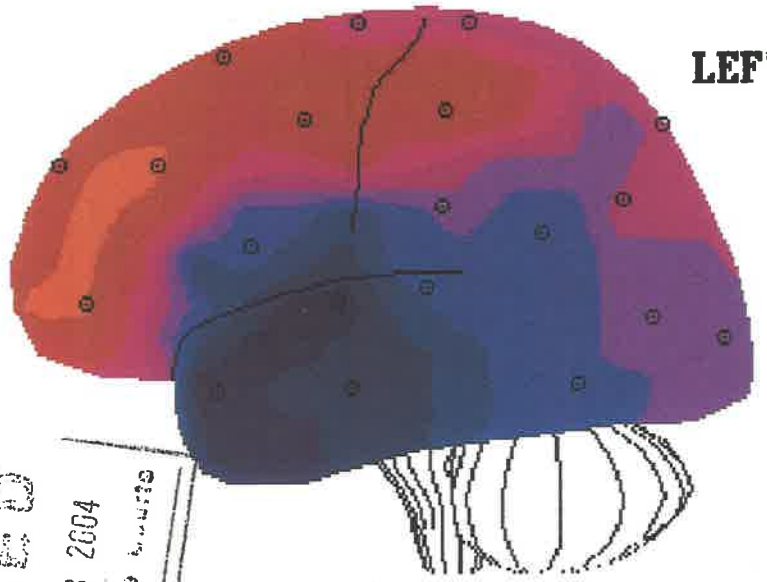
BI Version P2.35

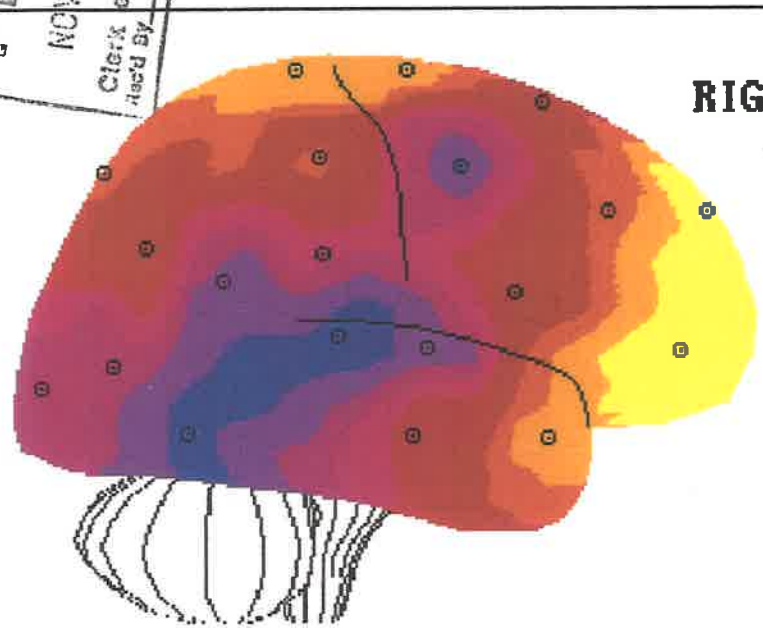
I.D. bb001

Weight Table: AVERAGE

EXHIBIT
Dr. Gur
#13
3/19/04 PK

Note:





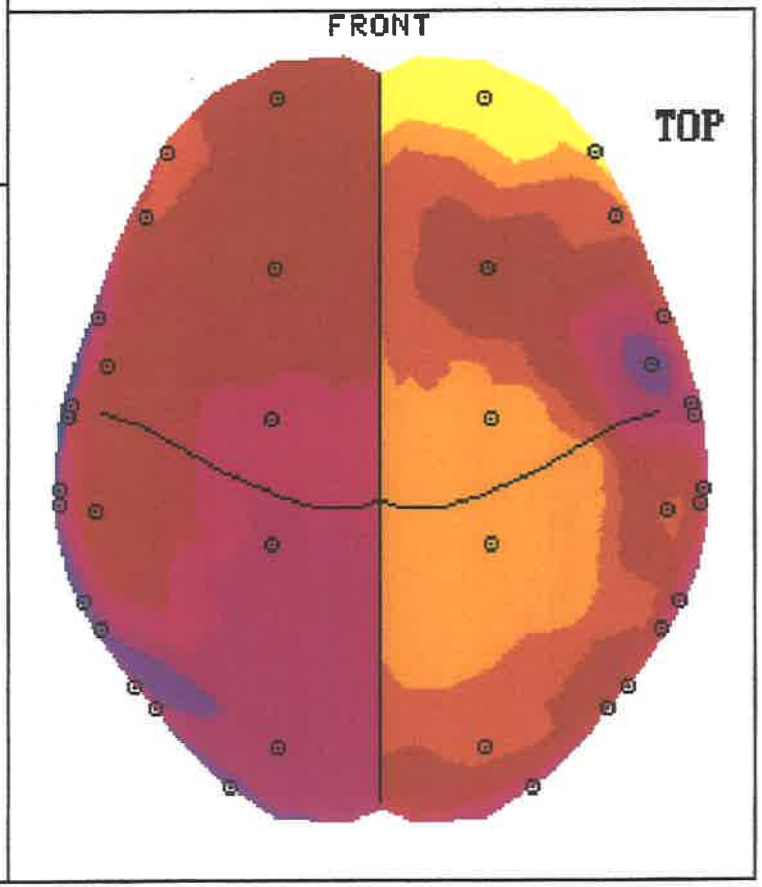


Figure 1 PET regional ration (R/WB) results for MEAN ±SD of cerebral metabolic rates for glucose (CMRgl) in healthy males and Mr. Byron Black (BB).

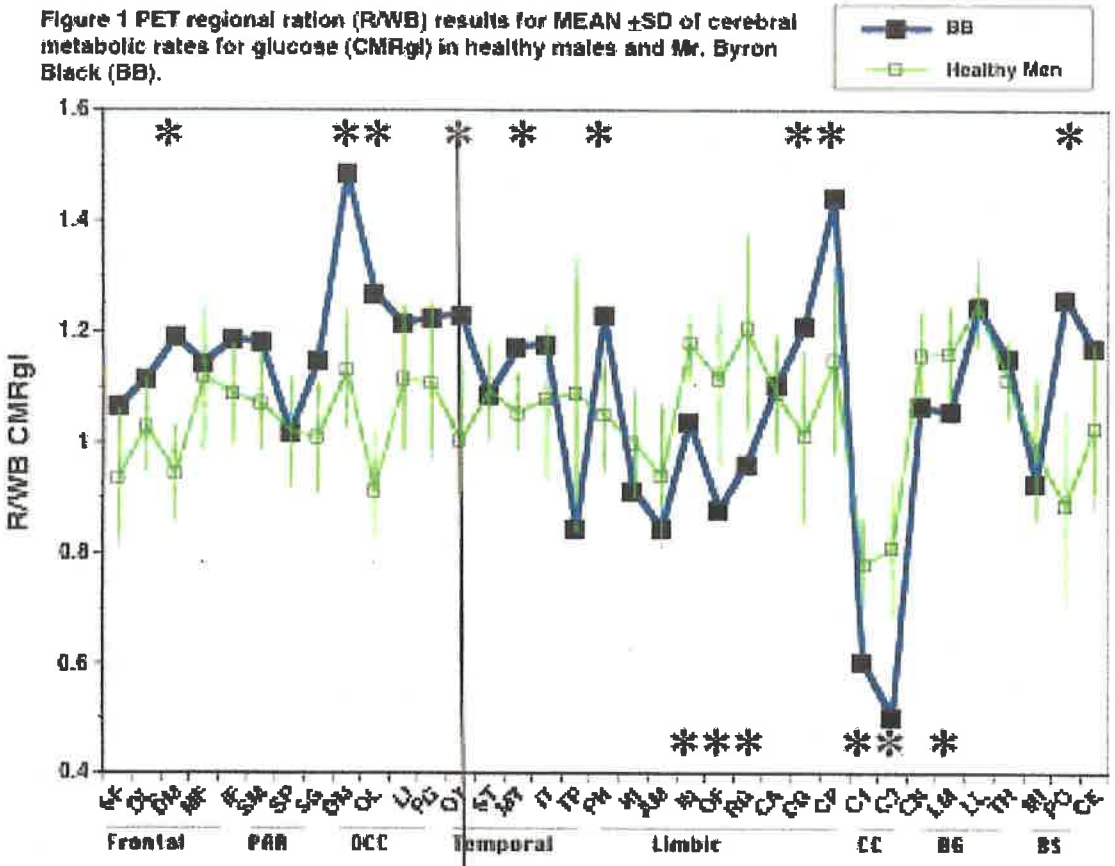


Figure 2 PET regional ration (R/WB) results for MEAN ±SD LATERALITY of cerebral metabolic rates for glucose (CMRgl) in healthy males and Mr. Byron Black (BB).

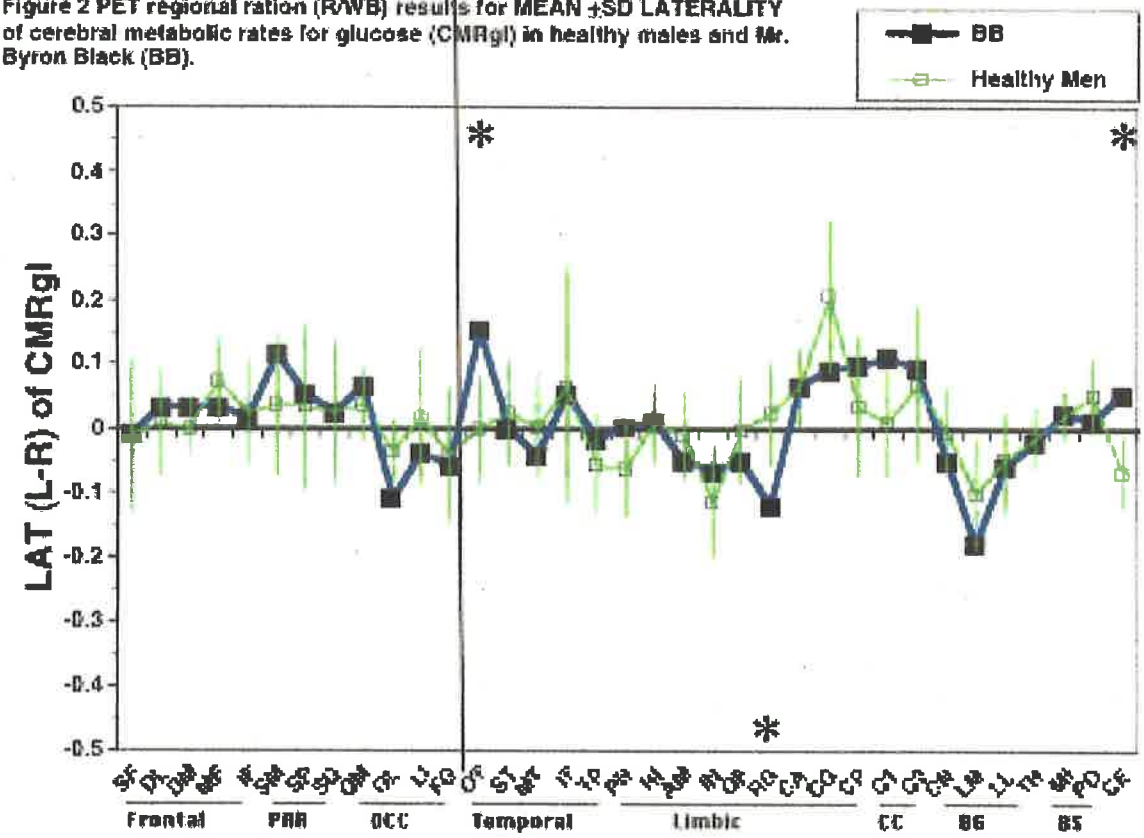


Exhibit 14

NeuroImage 15, 233–251 (2002)
doi:10.1006/nimg.2001.0977, available online at http://www.idealibrary.com on IDEAL®

Corpus Callosum Shape and Neuropsychological Deficits in Adult Males with Heavy Fetal Alcohol Exposure

Fred L. Bookstein,*,[1] Ann P. Streissguth,† Paul D. Sampson,‡ Paul D. Connor,† and Helen M. Barr†

*Institute of Gerontology, University of Michigan, Ann Arbor, Michigan 48109; and †Department of Psychiatry and ‡Department of Statistics, University of Washington, Seattle, Washington 98105

Received April 8, 2001

Persons with brain damage consequent to prenatal alcohol exposure have typically been diagnosed with either fetal alcohol syndrome (FAS) or fetal alcohol effects (FAE), depending on facial features. There is great variability of behavioral deficits within these groups. We sought to combine neuroanatomical measures with neurocognitive and neuromotor measures in criteria of greater sensitivity over the variety of consequences of alcohol exposure. To this end, midline curves of the corpus callosum were carefully digitized in three dimensions from T1-weighted MR scans of 15 adult males diagnosed with FAS, 15 with FAE, and 15 who were unexposed and clinically normal. From 5 h of neuropsychological testing we extracted 260 scores and ratings pertaining to attention, memory, executive function, fine and gross motor performance, and intelligence. Callosal midline shape was analyzed by new morphometric methods, and the relation of shape to behavior by partial least squares. The FAS and FAE subgroups have strikingly more variability of callosal shape than our normal subjects. With the excess shape variation are associated two different profiles of behavioral deficit unrelated to full-scale IQ or to the FAS/FAE distinction within the exposed subgroup. A relatively thick callosum is associated with a pattern of deficit in executive function; one that is relatively thin, with a deficit in motor function. The two combine in a very promising bipolar discrimination of the exposed from the unexposed in this sample. Thus there is considerable information in callosal form for prognosis of neuropsychological deficits in this frequently encountered birth defect. © 2002 Elsevier Science

Key Words: **corpus callosum; fetal alcohol syndrome; fetal alcohol effects; partial least squares; procrustes shape coordinates; singular warps.**

INTRODUCTION

Alcohol is a teratogen over the course of prenatal human development, and the brain is the fetal organ most sensitive to the damage it induces. Yet contemporary protocols for diagnosis make no reference to brain images, prenatal or postnatal, in classifying persons as damaged or undamaged by the prenatal exposure. This paper demonstrates how an unusual style of morphometric analysis of the adult brain image, in combination with conventional observations of cognitive and motor performance, might lead to a considerable increase in the sensitivity with which alcohol effects can be diagnosed and to a potentially important discrimination among two new subtypes of that diagnosis.

A diagnosis of fetal alcohol syndrome (FAS) entails evidence of a compromised central nervous system (CNS), growth deficiency, and uniquely characteristic facial stigmata subsequent to heavy fetal alcohol exposure (Stratton *et al.*, 1996). The facial features include short palpebral fissures, flat philtrum, thin upper vermilion, and flat midface. Adults with this diagnosis typically have at least a moderate intellectual deficit and a highly variable assortment of deeper specific deficits in attention, memory, motor function, and judgment. The related diagnosis of "fetal alcohol effects" (FAE) has been applied to patients with CNS compromise and a history of exposure without the full set of physical findings. Evidence is accumulating that the sociobehavioral prognosis for patients diagnosed with FAE is as severe as that for FAS. In one large database (Streissguth *et al.*, 1996), the prevalence of "secondary disabilities" (like dropping out of school or going to jail) in patients diagnosed FAE (mean IQ, 90) is no less severe than in those diagnosed FAS (mean IQ, 79). When psychometric scores in the same study were adjusted for IQ, profiles of deficit in the two diagnostic groups were identical. Similar findings have been reported by others (Mattson *et al.*, 1997, 1998). Yet though the clinical behavioral sequelae of the two diagnoses appear indistinguishable, the FAS diagnosis

[1] To whom correspondence should be addressed at Institute of Gerontology, University of Michigan, 300 North Ingalls Building, Ann Arbor, MI 48109-2007. Fax: (734) 936-2116. E-mail: fred@brainmap.med.umich.edu.

1053-8119/02 $35.00

is more distinctive and less controversial than the diagnosis of FAE (Stratton *et al.,* 1996). We recently estimated population prevalence of FAE or alcohol-related neurodevelopmental disorder in Seattle for the cohort born in 1975 (Sampson *et al.,* 1997a) as twice that of FAS (about 6 per thousand, versus 3). It follows that the facial diagnosis of FAS, even though it has 100% specificity, has unacceptable sensitivity for the task of detecting the full range of alcohol teratogenesis.

As alcohol can affect brain development throughout gestation, whereas the period of facial vulnerability is relatively short (Sulik *et al.,* 1981), one might reasonably suspect that behavioral abnormalities would be related to structural brain characteristics more strongly than to facial anomalies. In that case, studies of brains might generate a prognostically useful differential diagnosis not afforded by facial measurements however conscientious. This paper attempts to lay the groundwork for a quantitative approach of far greater sensitivity, without loss of specificity, by combining laboratory tests of neuropsychological function with direct quantitative measurements of three-dimensional MR brain images. For this first analysis we assessed only one structure: the outline of the corpus callosum in its midline. Our interest arises from a consistent finding in the literature associating callosal dysplasia with alcohol exposure. In an early study of this kind (Mattson *et al.,* 1994), of four children diagnosed FAS/"PAE" (our FAE) three showed dysgenesis or agenesis of the corpus callosum. There are now at least 11 reported cases of callosal agenesis in FAS/FAE as viewed in MRI (Mattson and Riley, 1997a; Swayze *et al.,* 1997), and disproportionate reductions are seen in the mean callosa of exposed cases even when grossly normal (Riley *et al.,* 1995; Mattson and Riley, 1997b).

In a thoughtful recent discussion of the joint role of neuroanatomy (possibly quantitative) and neuropsychological testing in psychopathology, James Harris writes:

> The adult neuropsychological database was largely established through the evaluation of adults with known brain damage (strokes, seizures, tumors, penetrating wounds, head trauma). . . . [But lesions of this type] occur far less commonly in children where congenital malformations related to pre- and postnatal insults . . . are more common. A majority of the neuropsychological dysfunctions of early life are not then associated with known brain insults, nor are they associated with lesions demonstrable on routine neuroimaging studies. . . . Neuropsychological testing, [therefore, should] integrate psychiatric and psychological information on behavior and the mind with neurological information on the brain and nervous system. (Harris, 1998, vol. ii, p. 23, p. 20)

The damage caused by prenatal alcohol is best described as diffuse damage that has been hitherto associated with "lesions" in only a small fraction of cases. In this sense, although alcohol is known to be the teratogen in prenatal alcohol exposure, we do not know where the "insult" is. Except for the literature's hint about the corpus callosum, we would know neither where to look in the brain for signs of specific damage nor which behaviors to examine that would assess the function of those specifically damaged systems. In other words, neither approaches from neuroanatomy nor approaches from developmental neuropsychology, separately, promise to be particularly helpful at this time. Progress is likelier by the integration of the two information streams, the behavioral and the neurological.

The present study is intended as a prototype for such investigations, combining a neuroanatomical channel (the corpus callosum) and a neuropsychological channel (a wide variety of neuropsychological laboratory tasks known to be sensitive to prenatal alcohol exposure). Each set of measurements formally serves both to assess group differences in its own domain and to help focus our attention on salient aspects from the other domain. Specifically, we measured callosal outline and neuropsychological function in 45 adult males, 30 of whom were diagnosed as affected by prenatal alcohol exposure. Our hypothesis is that callosal shape will be differently distributed between unexposed and exposed in ways that correlate with clinically meaningful differences in behavior. In fact, the combined data set supports a remarkably specific association of callosal shape variation with a bipolar profile of behavioral anomaly for which the FAS–FAE distinction supplies no information.

The callosal outline data here were previously exploited by Bookstein *et al.* (2001) in combination with data from 33 landmark points in a sample including adult females as well as these males. The present paper extracts much more information from the same outline resource by combining it with neuropsychological measurements. In particular, the major finding of this paper, as conveyed in the final figures, involves the correlation of structural and behavioral information; it would not be accessible from neuroanatomical data alone.

METHODS

Sample

We studied 45 Seattle-area males aged 18 years and over: 15 unexposed normals and 30 with alcohol teratogenesis, collectively referred to here as "the exposed." Prior to enrollment in the Seattle FAS Followup Study database (Streissguth *et al.,* 1996, 1997), these men had been diagnosed as either FAS (N = 15) or FAE (N = 15). Diagnosis was by D. W. Smith or one of his fellows (usually S. K. Clarren) between 1972 and 1995. A diagnosis of FAS combines findings of maternal alcohol abuse, CNS manifestations, growth deficiency, and specific facial anomalies. Patients with evidence of the first two of these criteria but not all of the physical

findings fall into the FAE group. FAS and FAE diagnoses were similarly ascertained, and each required a maternal alcohol exposure history in accordance with the Institute of Medicine (IoM) description: "a pattern of excessive intake characterized by substantial regular intake, or heavy episodic drinking" (Stratton *et al.*, 1996, p. 6). Diagnostically, all would fall within the IoM's categories of FAS, partial FAS, or alcohol-related neurodevelopmental disorder (ARND). For the actual diagnostic criteria used, see Streissguth *et al.* (1991). Normals were recruited from employees and their children at local health care facilities and educational institutions. The subjects were approximately group-matched for age and ethnicity. All three groups averaged 23 years of age (range, 18–36). Subjects were white except for 4 blacks (2, 2, 0 by diagnostic group) and 13 Native Americans (3, 4, 6). Clinically, all MR images were grossly normal except that our neuroradiologist noted two occurrences of callosal hypoplasia (one an FAS patient, the other an FAE). Subjects were *not* group-matched for IQ, of course, as IQ deficits are entailed in the diagnosis of the alcohol damage itself. Mean IQ scores (normals, FAE, FAS) were 113, 87, and 84. Three patients (2 FAS, 1 FAE) had IQ scores under 70; none were lower than 65. A total of 9 subjects (1 normal, 3 FAE, 5 FAS) were left-handed (measured as in Reitan, 1974). It is not feasible to match for socioeconomic status (SES) in studies of adult patients with fetal alcohol damage, as one characteristic of these patients is the great number of living situations in which they reside over the course of development, likely spanning a wide range of SES levels (Streissguth *et al.*, 1996). We did, however, control the "E" of SES to some extent, by excluding anyone who had completed a college education from the pool of potential unexposed subjects.

MR Images and Derived Data

T1-weighted sagittal SPGR images were acquired over 12 min in a GE 1.5-T Signa scanner at the University of Washington. TE was 8 ms, TR 29 ms, flip angle 45°. The resulting $256^2 \times 124$ arrays of $0.86^2 \times 1.50$ mm^3 voxels were processed by Edgewarp 3D software (Bookstein and Green, 1998; Bookstein *et al.*, 2001). From the MR images were located four landmark points (anterior and posterior commissure, tip of fourth ventricle, and rostrum) along with 39-point semilandmark tracings (see below) of the callosal midline. Morphology was digitized (by one of the authors, F.L.B.) blind to diagnostic group and behavior. Starting at rostrum, callosa were traced as 40-point polygons in space. (That is, digitizing took place in 3D, not in any single plane, with all three coordinates varying together.) A typical set of digitized locations, projected onto a convenient parasagittal image plane, is on the left in Fig. 1. They are spaced roughly inverse to curvature on a reference form (the first one digitized). On the right is a typical digitizing scene: a section perpendicular to the actual callosal midline through a candidate point along the lower border of isthmus. Every point digitized lay precisely on the "vertical" (axis of symmetry) of an image like this at the apparent boundary between callosum and CSF. If the tissue adjacent to callosum was not CSF, perhaps fornix or cavum pellucidum, the point digitized was taken as the (visual) extrapolation of the callosum–CSF boundary on this same axis.

Morphometric Tools

Multivariate analysis of these data was by the newly standardized biometric approach to sets of labeled points in two or three dimensions (Dryden and Mardia, 1998; Bookstein, 1996a,b, 1997a, 1998, 1999). *Shape* is the information left in such a figure after we ignore location, orientation, and scale. For multivariate analysis of shape and its correlates, one uses *Procrustes analysis,* which can be thought of in the following four steps. First, for any single specimen's digitized points (x_i, y_i), $i = 1, \ldots, k$, center them by moving the average location or centroid $(\sum_{i=1}^{k} x_i, \sum_{i=1}^{k} y_i)/k$ to (0, 0), and then, after the specimen has been centered on (0, 0) in this way, divide out the scale factor called *Centroid Size,* square root of the sum of squared distances of the points from (0, 0). After this scaling, the sum of those squared distances is precisely 1.0. Second, for any two sets of points with the same labels that have been centered and scaled in this way, define the *Procrustes distance* between them as the square root of the sum of squared Euclidean distances between corresponding points when one form is rotated with respect to the other just so as to minimize this sum of squared distances. Third, from two or more of these sets of centered and scaled points, compute the *Procrustes average shape* of the sample: the set of points that has the least summed squared Procrustes distance to all the forms of the original sample. The Procrustes average shape is taken as having been centered and scaled already, and by convention it is graphed with its longest axis (larger principal moment of inertia) horizontal. Finally, for each original figure in the data set, superimpose it over the form of the sample average in exactly the position that minimized the sum of squared distances between corresponding points, the sum of squares we already used in the course of determining that this particular average shape minimized their total. The positions at which the original landmark locations arrive after this superposition are called the *Procrustes shape coordinates* of the original figures. For the theorems justifying all of these arbitrary-seeming manipulations—for instance, to understand why there is not a separate

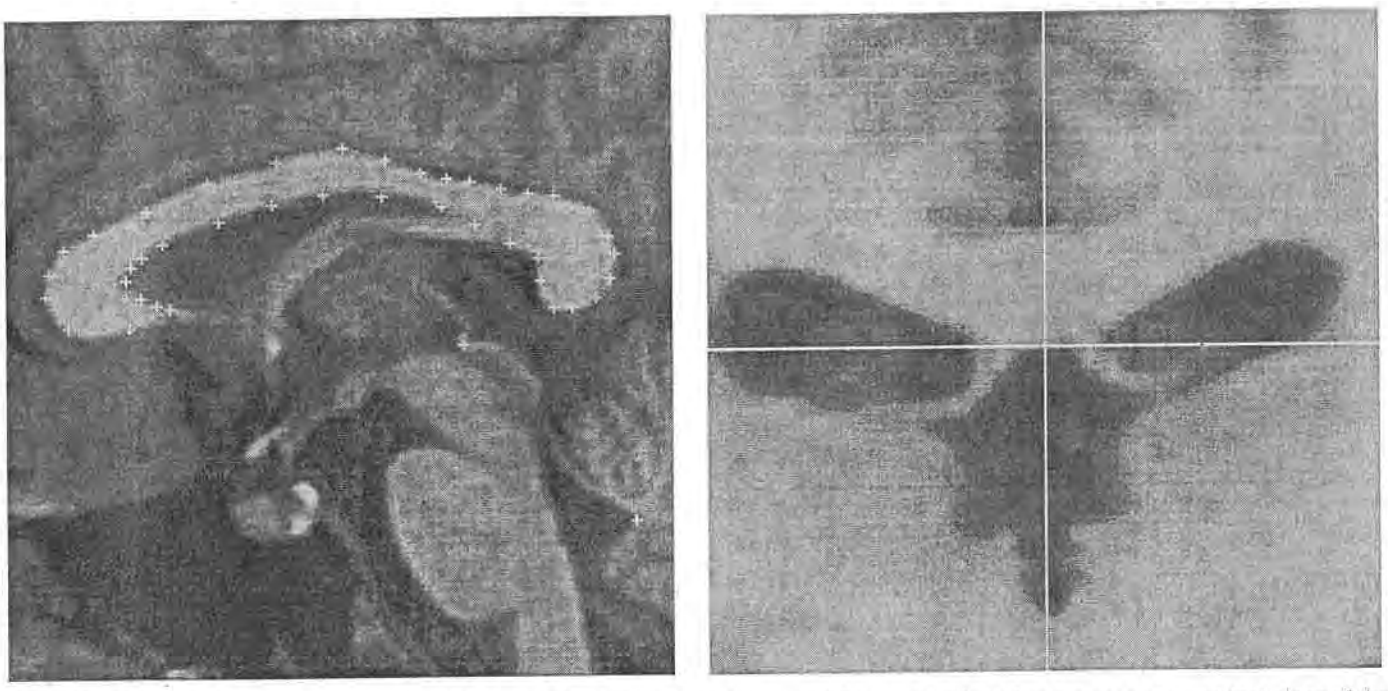


FIG. 1. Aspects of digitizing the corpus callosum in 3D. (Left) Full outline, one subject, as projected onto a near-parasagittal plane of the image. Except for rostrum, all of the points on the outline have been allowed to slip to minimize bending energy with respect to a template form. Other landmarks: anterior commissure, posterior commissure, tip of fourth ventricle, and left and right brain boundaries (not shown) at posterior commissure. (Right) A typical point of the outline (semilandmark 28, the one used in Fig. 4) for one subject, showing how approximate symmetry is used to determine the point digitized. The section here is perpendicular to the estimated tangent line of the outline in its vicinity, and the point digitized is slipped along that tangent line until it lies on the "midline" located here. Digitizing does not take place in any single plane, but in three dimensions, after a variety of sections such as this one verify that the candidate point is an appropriate one. From Bookstein *et al.* (2001).

centered-and-scaled Procrustes average form for each subgroup—see, for instance, Kent and Mardia (2001).

When empirical curves, such as the callosal outlines here, are digitized as discrete point series, individual points are not claimed to be homologous (to correspond) from subject to subject, and consequently, variability along the tangent direction is not informative. There are various versions of the Procrustes method that slide these points along the tangent direction (line or plane) so as to remove this variation for the purposes of averaging shapes and representing their variation and covariation. For instance, the tangential variation can be removed by minimizing bending energy with respect to an average (Bookstein, 1997b, 1998, 1999) or perhaps by perpendicular projection onto a template (Andresen *et al.*, 2000). Points arising from curves or surfaces by sliding this way are called *semilandmarks*, and analysis of semilandmark configurations is in terms of shape coordinates after the Procrustes superimposition over their average, as before. (The actual algorithm we used alternates the Procrustes averaging with an incremental sliding along the discretized curve.) As the present application deals with an almost planar curve, we projected these points onto a single synthetic midplane prior to further analysis. There resulted a total of 80 coordinates, two (one x-coordinate and one y-coordinate) for each of the 40 points of the outlines, as in Fig. 2a. We will refer to these two-dimensional forms as *midline callosal outlines*.

It is often useful to compute *principal components* of shape when it has been represented by Procrustes shape coordinates in this way. These components should not be thought of as factors somehow causally responsible for the observed variation, but instead as specifying the low-dimensional summaries of the data that, taken together, best predict the individual values of all of the shape coordinates simultaneously. Two other interpretations may be helpful: these principal components are the dimensions that best reproduce the observed (Procrustes) distances between all the forms of the data set using linear combinations of the original coordinates (the interpretation as *principal coordinates*, see Reyment and Jöreskog, 1993); also, they are the linear combinations of the shape coordinates that have the greatest variance in the data compared to the variance they would have if the data arose from a model of pure gaussian digitizing noise of the same small variance at every landmark in every direction (Bookstein, 1997a). The components called for by any

quate reliability for the distinctions we are pursuing, which involve large-scale features of this shape that shift by substantial extents in the Procrustes shape space, as will be shown in the grid images that follow. One of the reasons for this perhaps surprisingly high reliability is that along many arcs of this midline, the intersection of the symmetry line with the callosal surface typically lies near the bottom of a fairly sharp groove that twists out-of-plane in an unpredictable fashion as it winds around the callosum. In these regions, the reliability of the semilandmark representation taken in 3D is likely to be better than that of curves traced in any single plane section of the image, however carefully that section is chosen.

Average Size and Shape

In the Procrustes approach, one analyzes size and shape separately. "Size" is measured as the square root of the sum of squared distances of all the labeled points from their center of gravity case by case, the centroid size introduced above (Bookstein, 1991; Dryden and Mardia, 1998). Callosal outlines of the FAE and FAS groups average 92 and 94% of normal mean size, respectively, when size is quantified in this way prior to any Procrustes analysis. These are not different from each other but, as a pool, are significantly different from normals ($P \sim 0.002$ by t test). For arch-shaped forms like these, the formula for centroid size reduces more or less to the long diameter of the callosal form, the distance from genu to splenium. Midline callosal areas (area of these midline callosal outlines), likewise computed before any scaling, average 91 and 95% of the normal mean for FAE and FAS, respectively, but none of these differences are significant, owing to the fact that callosal area has a larger coefficient of variation than the long diameter of its arch in this sample. For additional discussion of these and other size measures for this data set, see Bookstein *et al.* (2001).

As Figure 2a shows, there is considerable variation in the locations of the semilandmarks within the full sample, but no suggestive or statistically significant differences in the average shapes of the callosum among the three diagnostic groups (Fig. 2b), as computed by any of four separate permutation tests of Procrustes distance: normal versus FAE, normal versus FAS, FAE versus FAS, or normal versus exposed.

Shape Variation

Despite the absence of average shape differences, there is a remarkable difference between exposed and unexposed in *variability* of shape, which can be examined either globally or locally. Globally, we use the version of principal components analysis that applies to labeled point data, as reviewed under Methods. Figure 3 shows the first three principal components for the full 45-subject data set of callosal shapes (eigenvalues 0.031, 0.029, and 0.022 out of a total Procrustes variance of 0.177; note the near circularity of the first two). As deformations, these principal components are large-scale aspects of shape: height–width ratio, splenium height–genu height ratio, and general thinning/thickening of the arc. From the corresponding scatters of scores (Fig. 3, bottom row), there is substantially more variance on principal component 3 for the exposed than for the unexposed ($F_{29,14} = 3.45$, $P \approx 0.01$), but there is no mean difference in scores for any of these principal components among the groups, nor any differential clustering between the FAS and FAE subsets. The fourth component (eigenvalue 0.018) shows neither mean differences nor variance differences, and so we stop this report at the third.

Likewise, local variability of shape is higher in the exposed than in the unexposed. Over all 40 semilandmarks, the greatest ratio of variances between groups in the direction perpendicular to the average curve is 6.73 at semilandmark 28, near the spring of fornix on the isthmus (Fig. 4). Geometrically, the pattern in the figure suggests a considerably more regulated alignment of splenium with isthmus in the unexposed. A discrimination of exposed from unexposed by this single coordinate detects all but 4 of the 30 exposed cases with only 2 false positives. This is a *quadratic* discriminator in that it attends to unsigned (bidirectional) distance from the average. Notice that its distribution in Fig. 4 does not distinguish FAS from FAE. It is this combination of excess shape variation in the exposed both globally and locally that constitutes the structural aspect of our findings.

Recall the first part of this study's hypothesis: that callosal shape is differently distributed between exposed and unexposed subsamples. If the finding concerns a ratio of variances, then we should be testing that ratio by a permutation of group label over the set of 45 outline shapes. The best of these ratios for a single point, 6.73, is nearly double that for the best principal component (the ratio of 3.45 for PC3, Fig. 3, as already noted), and so it is that maximum we test. Although the finding deals only with semilandmark 28, the appropriate distribution is of the maximum of this ratio over all the semilandmarks, not just the 28th. Owing to the obvious presence of a few outlying points in Fig. 2, we work with ranks of the signed deviations (recall that transformation to ranks is a common maneuver in the world of permutation testing generally). In 2000 random permutations of exposure status over the callosal outlines, a ratio greater than that in Fig. 4 is found at any semilandmark only seven times. As the standard error of that count of 7 is almost exactly its square root, and $7 + \sqrt{7} < 10$, we can reasonably take the significance level of this variance ratio finding as $P < (10/2000) = 0.005$. (This simple application of permutation methods to testing differences in variances is appropriate whenever the subsample means

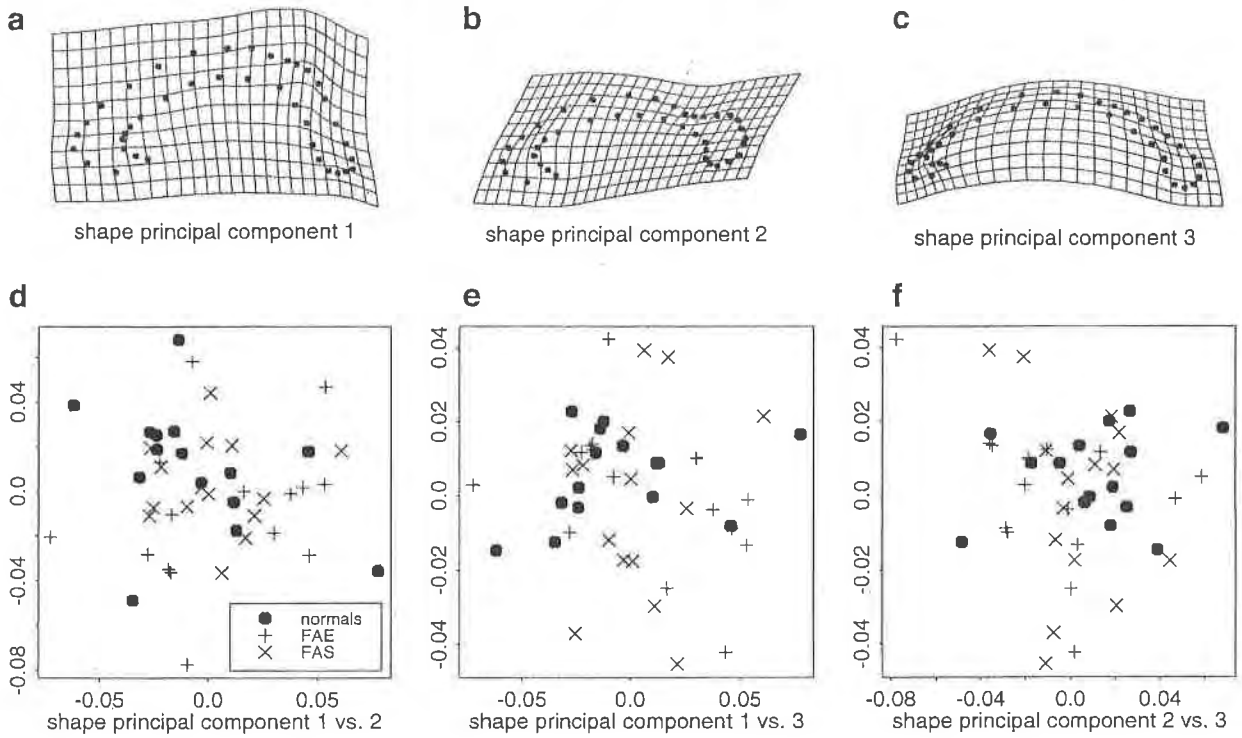


FIG. 3. Shape principal components and scores. (Top row) The first three principal components of the callosal outline shapes in Fig. 2, diagrammed as deformations of the average outline by arbitrary multiples. Eigenvalues: 0.031, 0.029, and 0.022. (Bottom row) Pairwise scatters of the first three principal component scores, showing excess variability of the third in the exposed subgroups. Group is coded by plotting symbol: •, unexposed; +, FAE; ×, FAS. Scatter of the unexposed subgroup is vertically constrained in (e) and (f).

are "known to be equal" (Good, 2000, p. 40), an assumption that seems justifiable in this case. One reviewer kindly pointed out that this test presumes also that the distributions differ only by a scale factor. By a theorem of Romano, 1990, in any event the resampling distribution of the variance ratio corresponds to a blended distribution. The assumption of proportionality is not obviously violated in the first few principal components of the joint distribution, Fig. 3.)

Brain–Behavior Analysis

The covariance matrix of the ranks of 260 neurocognitive and neuromotor indicators against the shape coordinates of Fig. 2 (a matrix 260 × 80) was subjected to PLS analysis. Singular values were 4.47, 3.03, 2.13, ... Figure 5 shows the first two callosal shape scores and neuropsychological latent variable scores that result, and Fig. 6 shows the first two singular warps (latent variables of shape interpreted as splined deformation grids). Singular warp 1 represents general vertical expansion/compression mildly intensified in the isthmus. Singular warp 2 entails relative thinning/thickening of the callosum almost uniformly along its arc. Note also that the thin arc appears as if it might penetrate further "into the frontal lobe" than the thick. Latent variable pairs beyond the second show no useful patterns by group, and so are neither figured nor scattered here.

Figure 5a shows that the unexposed and the exposed subsamples have the same range of scores on singular warp 1 (vertical expansion/contraction) but differ considerably in average values of the corresponding neuropsychological latent variable, NLV1. This is the pair of normalized dimensions (one of shape and one of neuropsychological performance) that have the highest pooled sample covariance (namely, 4.47), that is, the strongest predictive relationship from form to function. The corresponding correlation is 0.43 (0.30 in the unexposed subsample alone, but 0.53 in the exposed; the covariance is thus not particularly dependent on the mean shift in NLV1 between groups). NLV1 is highly IQ-loaded. Of the 10 items having the highest saliences (coefficients B_1 of the previous equations), 4 are IQ subscales or subsummaries (Picture Assembly, Block Design, the Performance IQ subscale, and Picture Completion), another is Word Attack, 2 are APT success scores (selective and alternating attention, numbers correct), and the rest are Wisconsin Card Sort error scores. The correlation of this summary neuropsychological LV1 with full-scale IQ is −0.743 within the unexposed subsample and, owing to systematic deficits of IQ in the fetal alcohol disorders, a full

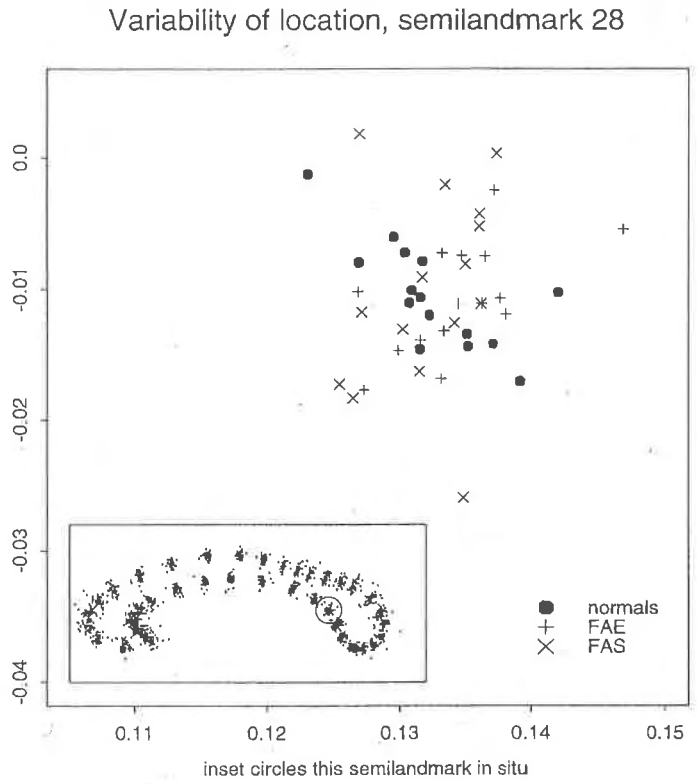


FIG. 4. The exposed subgroups show much greater variability of the location of semilandmark 28, shown here in context between isthmus and splenium (inset) and in magnification. The uniform component of variation from the average callosal shape has been removed from the Procrustes coordinates pictured on the left in Fig. 2. The coordinate perpendicular to the average curve (northeast–southwest direction) is a very effective quadratic discriminator of exposure.

−0.885 in the full sample of 45 men; the 2 exposed subjects who fall clearly among the unexposed (+ and × at lower left in the figure) are the two of highest IQ. The corresponding dimension of shape, which itself correlates −0.292 with IQ, is as shown in Fig. 6c: a lower, flatter arch goes with high IQ. There is no mean difference between FAS and FAE on NLV1 in this sample.

The second latent variable pair, Fig. 5b, displays a huge correlation, more than 0.8, between an aspect of callosal shape (nearly uniform thinning/thickening) and the corresponding neuropsychological latent variable, NLV2. The exposed subgroup (FAS and FAE thoroughly intermingled) is scattered well past the unexposed *on both ends of the distribution.* In the scatter of NLV1 versus NLV2, Fig. 5c, the unexposed are in a fairly tight cluster at left center; the exposed, regardless of diagnosis, are distributed with 8.79 times as much variance, mostly in the vertical band at right. Note that the top 12 points on NLV2 were all exposed (plotted as + or ×) and likewise the bottom 10. The resulting pair of apparent clusters is separated by 2.57 standard deviations of the pooled NLV2 score. This bipolarity in the scatter for the exposed strongly suggests two additional clinical entities that will be characterized separately in the course of the discussion.

The second part of our hypothesis, that callosal shape differences correlate with differences in behavior, can be tested by applying permutation procedures to this PLS analysis. When the matching of images to test scores was randomized 500 times, no correlation among the first three pairs of latent variable scores was ever found larger than 0.45, and an F ratio of greater than 8.79 for the ratio of subgroup variances, exposed over unexposed, was encountered only once. Taken together, the two tests show the PLS analysis of callosal shape against neuropsychological outcome to be highly statistically significant as a whole. In order to estimate sensitivity and specificity of the paired extreme clusters (see Discussion), we need the standard error of the cluster separation emerging in Fig. 5c. For this purpose, we fixed the membership of these clusters (the cases with the top 12 NLV2 scores and the cases with the lowest 10 scores) and then recomputed the formula for NLV2 from first principles for the 45 different data sets of 44 subjects (leaving out each subject in turn). This is called a *jackknifing* of the interesting statistic, the difference between them. In the resulting recomputations, the lower cluster of 10 averaged −1.25 when NLV2 is z-scored, versus −1.38 in the complete analysis, and the upper cluster of 12 averaged 1.26, versus 1.19 in the complete analysis. Their separation thus averages 2.51 units of NLV2, with a jackknifed standard error of 0.44. We conclude that the two poles of NLV2 are stable; they will be characterized in connection with Table 2 in the Discussion.

At this point, then, both clauses of our hypothesis, the difference of distributions of callosal shape between exposed and unexposed and the correlation between callosal shape and neuropsychological performance, have been shown to be highly statistically significant. None of the main multivariate findings can be plausibly attributed to chance, nor can the sorting of most of the exposed cases into the extremes of NLV2 as shown in Fig. 5.

DISCUSSION

Midline Curves as Data

The findings here in respect of size have been anticipated in earlier studies by other methods. Our finding of 7% diminution in midline callosal area among the fetal-alcohol-exposed is in the direction expected from earlier reports in humans (Mattson *et al.,* 1994; Riley *et al.,* 1995; Swayze *et al.,* 1997, Sowell *et al.,* 2001a) and is in keeping with volumetric studies of fetal alcohol effects upon other regions of the brain (Archibald *et al.,* 2001; Sowell *et al.,* 2001b). Other methods of studying the information content of solid brain images, such as

tell to what extent this divergence of findings owes to differences in image-analytic or biostatistical methodology or neuropsychological breadth and to what extent it expresses true differences between their sample and ours. (The Sowell *et al.* sample averages 13 years of age and is almost equally divided between males and females, the outline analyzed was taken from one flat plane section of the volume, and there is no formal representation of its "shape.") To the extent that the finding of a correlation between callosal anatomy and CVLT is held in common between the two studies, it suggests particular attention be paid to alcohol effects on the brain regions crucial to the behaviors tapped by the CVLT task. Of the seven scores appearing in Table 2A, three relate to semantic clustering, a working memory task often associated with frontal lobe integrity, and four relate to simple learning or recall, which is associated with perisylvian and temporal lobe integrity.

Clinical Implications

The Institute of Medicine report (Stratton *et al.*, 1996) proposed replacement of present diagnostic systems for fetal alcohol effects by a fivefold categorization: FAS with confirmed exposure, FAS without confirmed exposure, partial FAS with confirmed exposure, alcohol-related birth defects (ARBD), and alcohol-related neurodevelopmental disorder (ARND). The proposed diagnoses of partial FAS and ARND are intended to apply to patients who show particular behavioral or cognitive abnormalities or delays (e.g., learning difficulties or poor metacognition) that cannot be explained by environment or family background alone. However, the effects of environment or family background are quite difficult to assess. We propose instead that the behavioral or cognitive abnormalities scrutinized for this diagnostic purpose be just those found to be linked, by an analysis such as ours, to the crucial intervening variable for any behavioral teratology study, viz., the trace of prenatal brain damage.

In Harris's language, as excerpted in our Introduction, we have produced strongly suggestive evidence that callosal shape is a crucial aspect of the "neurological information" against which information on behavior and the mind is to be integrated for this class of patients. This might, indeed, supply the localized "lesion" required if quantitative aspects of the image are to abet the diagnostic process. For the brain image to become ancillary to diagnosis in this way, the findings in Figs. 3 through 6 need to be converted to practical decision rules. One discrete shape feature that emerges from a closer study of Fig. 6, for instance, is the difference in shape of the region of genu as a whole. The cases that proved extreme on NLV2 tended to be extreme on CSW2 as well, which entails a vertex of curvature interior to the arch at genu that is strikingly too close to or too far from the line connecting rostrum to the front of the callosal curve. The development of this line of investigation into a formal decision rule for detecting that neuroanatomical "lesion" is in preparation.

While there seems to be only one such focus for the telling "lesion" in this analysis, there seem to be two (the two subclusters of Fig. 5c or Table 2) for the corresponding behavioral profile. It is this pair of possible subdiagnoses that, if confirmed, will lead to the more important follow-on hypotheses combining structural imaging, functional imaging, and behavioral assessment. Callosal shape variability may actually be expressing a range of different, ad-hoc developmental accommodations to the direct impact of alcohol neurotoxicity per se. In that event, the distinction here between our "EF-related" and "motor" types—that is, between "thick" and "thin" callosa—will need to be be confirmed by structural or functional studies of the specific brain systems distant from the callosum itself, the systems that link through the arcs of the callosum entailed in the patterns of Fig. 6. To the extent that there are two types of fetal alcohol damage, such studies need to use neurobehavioral profile as a covariate of shape difference; studies averaging over the two clusters in Fig. 7 will lose power just as studies of schizophrenia lose power when they average over the clinical subtypes so well-known there.

How good a detection protocol might we hope for? At the right in Fig. 7 is a schematic of a plausible underlying model for the observed distribution on NLV2: a mixture of four subpopulations. Assume that the population is half fetal-alcohol-affected and half unaffected and that the affected group is divided evenly in thirds over the three component distributions of "EF-deficit," "motor deficit," and "other" alcohol effects. At the top, at mean 1.25, is the motor type of fetal alcohol patient; at the bottom, at mean −1.25, the EF-deficit type. In the middle, at mean 0, are the rest of the fetal alcohol cases in this sample, those who were not extreme on NLV2 (dotted curve), together with all the normals (tall solid curve). The jackknifed quantities reported earlier are consistent with the common standard deviation of about 0.50 for all three of these distributions that is drawn. Consider, for instance, the "detection rule" that assigns values of NLV2 above 0.75 (+1.5 normal standard deviation) to the motor type of fetal alcohol deficit and those below −0.75 to the EF type and calls everybody else unexposed. Given the sample proportions here, we would detect all but 16% of the two extreme types of fetal alcohol deficit, while misclassifying only 14% of the unexposed. Unfortunately, a full 86% of the nonextreme type of cases here, those commingling with the normals in Fig. 7, would likewise be classified as unexposed. The resulting "diagnosis" would have specificity 86% but sensitivity only 60%.

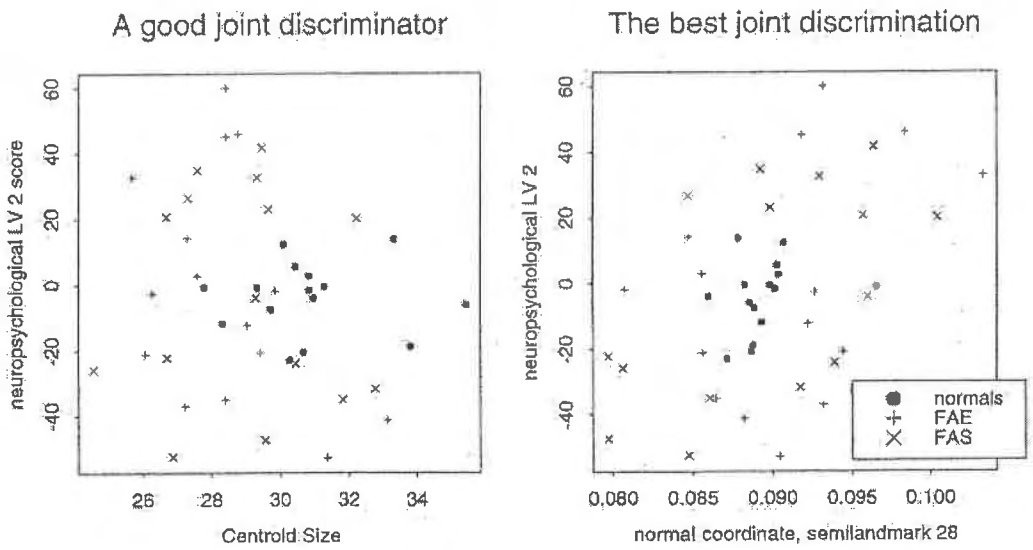


FIG. 8. Better detection rules combining morphometric and neuropsychological information. (Left) A promising combination of morphometric and neurobehavioral scores, centroid size by NLV2, permits the detection of half the exposed cases overlooked by the neurobehavioral categorization alone (see text). (Right) A very lucky discriminator with 100% sensitivity and 93% specificity for exposure: the interior and boundary of the normal cluster (solid circles) here. The horizontal axis is the perpendicular coordinate from semilandmark 28, Fig. 4, and the vertical the neuropsychological score from Fig. 7. The cluster in the middle of the plot incorporates 14 of the 15 unexposed subjects without a single exposed subject intruding.

We can do better by bringing back in information from callosal shape. Even though CSW2 is enormously correlated with the neurobehavioral profile to which we are attending, other dimensions of callosal form need not be. On the left in Fig. 8, for instance, we have scattered the simplest callosal measure of all, centroid size (in effect, its length), against this same bipolar neurobehavioral discriminator and declared all forms abnormal that have either an extreme NLV2 score or else a centroid size less than the minimum of the centroid sizes for the unexposed subsample, regardless of NLV2 score. The number of false negatives immediately drops by half, leaving only four classification errors (three +'s and one × inside the cluster of normals) instead of the eight when we attended to the vertical axis alone. In effect, small size detects a handful of the abnormal callosa that are otherwise normal in neuropsychological profile even as that profile casts suspicion on a great number of subjects having normal callosal length. (An analogous scatter using callosal area in place of centroid size shows no such improvement.)

The shape coordinate incorporated in Fig. 4 turns out to be another such uncorrelated shape variable, one showing quite a bit more effect of exposure than does centroid size. The possible improvement such a focus might afford for differential diagnosis is suggested in the scatter at right in Fig. 8, where this coordinate is combined with the same NLV2 score. All the exposed, along with just one unexposed subject, lie in a ring outside the central cluster of 14 unexposed. The discrimination rule that says "diagnose everything outside this normal cluster as alcohol-damaged" thus nominally has 93% specificity, there being only one "false positive," but 100% sensitivity as well, unconfounded by IQ.

Of course this nearly perfect separation cannot be expected to persevere into additional samples. The "hole" in the middle of the distribution for the 30 patients is surely a fortuitous accident that cannot be expected to recur. Also, the shape coordinate here is not a local measure of the isthmus outline on which it lies; its quantification required the detailed digitizing of the entire 40-point polygon. If we could find a simple shape descriptor uncorrelated with NLV2 that has a variance ratio of 4:1 between the groups (instead of the very helpful ratio of 6:7:1 in the lucky example here), even absent any mean differences by exposure group a further thresholding of the shape measure at ±1.5 of the standard deviation for the unexposed would reduce the false-negative count by nearly half, raising sensitivity to 78% (versus 60% in Fig. 7) while dropping specificity only to 71% (versus 78% in Fig. 7). (The discrimination in Fig. 8a seems to be running at a somewhat higher level, owing to the nongaussianity of the apparent joint distribution here, but we cannot model these deviations effectively in a sample this small.) A supplemental morphometric descriptor like this need not have derived from callosal outline alone, but could involve quantification of the same MR image from any part of the brain supplying the white-matter pathways here. Separately, the profile-relevant parts (Table 2) of the full battery in Table 1 might be winnowed into a simpler net psychometric protocol.

In any event, explicit crossvalidation of these and other detection rules will require additional samples, not reuse of this one. Three additional waves of 45 subjects each (adult females, adolescent males, and adolescent females) are available for these callosal data, but the neurobehavioral battery is not ready yet. When it is, we will carefully replicate this entire structural–behavioral analysis. For the anatomical data alone, the callosal outline form here, the information content relevant to diagnosis appears to be of equivalent amplitude between adult males and adult females (Bookstein *et al.,* 2001), but the pertinent discriminators arise from different subarcs of the callosum. We do not know yet how specific these discriminators may prove against syndromes owing to other causes than prenatal alcohol exposure (Coles *et al.,* 1997; Jacobson, 1998), such as Down syndrome, Asperger's disorder, or fragile-X. In respect of unselected populations, however, our earlier experience with checklists specific to FAS/FAE behavioral eccentricities (Streissguth *et al.,* 1998) is cause for optimism.

Historically, it was the facial pattern of FAS that brought the reality of alcohol teratogenesis to our attention. If our finding is valid, the face does not tap the dimensions that calibrate the effect of prenatal alcohol exposure on brain and behavior, at least for the type of exposed subjects studied here, who are diagnosed with FAS or FAE, but are not frankly mentally retarded. It is the brain image, not the portrait, that is the more informative for these patients. Syndromology aside, then, our finding forces us to face a terrible antinomy having severe societal consequences. Current systems of social service in the United States are keyed to the dysmorphological diagnosis of FAS or else to mental retardation (Sampson *et al.,* 2000). If major dimensions of the neuropsychological damage consequent upon prenatal exposure, however detected, are independent of both those criteria, then it behooves us to pursue alternative diagnostic protocols equally specific to alcohol damage but enormously more sensitive. The findings reported here have approached that most humane end via careful examination of covariation between brain structure and behavior over the full range of neuropsychological alcohol effects. In particular, the specific executive function deficit suggested here as pertaining to some but not all of these patients comprises the aspect of behavioral deficit that, in practice, induces the greatest range of secondary disabilities (Streissguth *et al.,* 1996; Connor *et al.,* 2000). If this profile is indeed predictable by brain imaging at earlier ages, the implications for clinic and community are as substantial as those for developmental neuroscience.

ACKNOWLEDGMENTS

We thank David Haynor, M.D., for clinical readings of all the MRI scans; psychometrist Michael Hampton; Julia Kogan, for her heroic efforts at assembling a balanced patient sample; and Sterling K. Clarren, M.D., who provided most of the diagnoses within the exposed group. Our digitizing used the dig3 module of Bill Green's Edgewarp morphometrics package, which is available for Silicon Graphics workstations and PCs running Linux by anonymous FTP from ftp://brainmap.med.umich.edu/pub/edgewarp3.2. Other graphics were produced in Splus, a commercial statistics package. Functions for duplicating the analyses here are available from the senior author. Four anonymous reviewers of an earlier version helped greatly with the flow of the argument. The research on which this paper is based was supported by USPHS Grants AA-11037 to A.P.S. and GM-37251 to F.L.B.

REFERENCES

Note: References listed here that are not cited in the main text are cited in Table 1 as sources for elements of the behavioral measurement battery.

Aboitiz, F., Scheibel, A. B., and Zaidel, E. 1992. Morphometry of the Sylvian fissure and the corpus callosum, with emphasis on sex differences. *Brain* **115:** 1521–1541.

Andresen, P. R., Bookstein, F. L., Conradsen, K., Ersbøll, B. Marsh, J., and Kreiborg, S. 2000. Surface-bounded growth modeling applied to human mandibles. *IEEE Trans. Med. Imag.* **19:** 1053–1063.

Archibald, S. L., Fennema-Notestine, C., Gamst, A., Riley, E. P., Mattson, S. N., and Jernigan, T. L. 2001. Brain dysmorphology in individuals with severe prenatal alcohol exposure. *Dev. Med. Child Neurol.* **43:** 148–154.

Army Individual Test Battery. 1944. *Manual of Directions and Scoring.* War Department, Adjutant General's Office, Washington, DC.

Arnett, P. A., Rao, S. M., Bernardin, L., Grafman, J., Yetkin, F. Z., and Lobeck L. 1994. Relationship between frontal lobe lesions and Wisconsin card sorting test performance in patients with multiple sclerosis. *Neurology* **44:** 420–425.

Bachman, J. C. 1961. Specificity vs. generality in learning and performing two large muscle motor tasks. *Res. Q.,* **32**(1): 3–11.

Baloh, R. W., Jacobson, K. M., Beykirch, K., and Honrubia, V. 1998. Static and dynamic posturography in patients with vestibular and cerebellar lesions. *Arch. Neurol.* **55:** 649–654.

Barr, H. M., Streissguth, A. P., Darby, B. L., and Sampson, P. D. 1990. Prenatal exposure to alcohol, caffeine, tobacco, and aspirin: Effects on fine and gross motor performance in four-year-old children. *Dev. Psychol.* **26:** 339–348.

Benton, A. L. 1968. Differential behavioral effects in frontal lobe disease. *Neuropsychologia* **6:** 53–60.

Bookstein, F. L. 1991. *Morphometric Tools for Landmark Data: Geometry and Biology.* Cambridge Univ. Press, New York.

Bookstein, F. L. 1996a. Biometrics, biomathematics, and the morphometric synthesis. *Bull. Math. Biol.* **58:** 313–365.

Bookstein, F. L. 1996b. Combining the tools of geometric morphometrics. In *Advances in Morphometrics* (L. F. Marcus, M. Corti, A. Loy, G. J. P. Naylor, D. E. Slice, Eds.), NATO ASI Series A: Life Sciences, Vol. 284, pp. 131–151. Plenum, New York.

Bookstein, F. L. 1996c. Morphometrics. *Math Horizons* **February:** 28–31.

Bookstein, F. L. 1997a. Shape and the information in medical images: A decade of the morphometric synthesis. *Comput. Vis. Image Understand.* **66:** 97–118.

Bookstein, F. L. 1997b. Landmark methods for forms without landmarks: Localizing group differences in outline shape. *Med. Image Anal.* **1:** 225–243.

Bookstein, F. L. 1998. A hundred years of morphometrics. *Acta Zool.* **44:** 7–59.

Bookstein, F. L. 1999. Linear methods for nonlinear maps: Procrustes fits, thin-plate splines, and the biometric analysis of shape variability. In *Brain Warping* (A. W. Toga, Ed.), pp. 157–182. Academic Press, San Diego.

Bookstein, F. L., Streissguth A. P., Sampson, P. D., and Barr, H. M. 1996. Exploiting redundant measurement of dose and behavioral outcome: New methods from the behavioral teratology of alcohol. *Dev. Psychol.* **32:** 404–415.

Bookstein, F. L., Green, W. D. K. 1998. *Edgewarp 3D: A Preliminary Manual.* Posted to the internet as ftp://brainmap.med.umich.edu/pub/edgewarp3.1/manual.html, August, 1998.

Bookstein, F. L., Sampson, P. D., Streissguth, A. P., and Connor, P. D. 2001. Geometric morphometrics of corpus callosum and subcortical structures in the fetal-alcohol-affected brain. *Teratology* **64:** 4–32.

Bullmore, E. T., Suckling, O., Rabe-Hesketh, S., Taylor, E., and Brammer, M. J. 1999. Global, voxel, and cluster tests, by theory and permutation for a difference between two groups of structural MR images of the brain. *IEEE Trans. Med. Imag.* **18:** 32–42.

Coles, C. D., Platzman, K. A., Raskind-Hood, C. L., Brown, R. T., Falek, A., and Smith, I. E. 1997. A comparison of children affected by prenatal alcohol exposure and attention deficit, hyperactivity disorder. *Alcoholism: Clin. Exp. Res.* **21:** 150–161.

Connor, P. D., Streissguth, A. P., Sampson, P. D., Bookstein, F. L., and Barr, H. M. 1999. Auditory and visual attention in adults with and without fetal alcohol syndrome and fetal alcohol effects. *Alcoholism: Clin. Exp. Res.* **23:** 1395–1402.

Connor, P. D., Sampson, P. D., Bookstein, F. L., Barr, H. A., and Streissguth, A. P. 2000. Direct and indirect effects of prenatal alcohol damage on executive function. *Dev. Neuropsychol.* **18:** 331–354.

Delis D. C., Kramer, J. H., Kaplan, E., and Ober, B. A. 1987. *California Verbal Learning Test.* Psychological Corp., San Antonio, TX.

Denckla, M. B. 1974. Development of motor co-ordination in normal children. *Dev. Med. Child Neurol.* **16:** 729–741.

Denckla, M. B. 1996. A theory and model of executive function: A neuropsychological perspective. In *Attention, Memory, and Executive Function* (G. R. Lyon and N. A. Krasnegor, Eds.), pp. 263–278. Brookes, Baltimore.

Dryden, I. L., and Mardia K. V. 1998. *Statistical Shape Analysis.* Wiley, Chichester, UK.

Dugbartey, A. T., Rosenbaum, J. G., Sanchez, P. N., and Townes, B. D. 1999. Neuropsychological assessment of executive functions. *Semin. Clin. Neuropsychiatry* **4:** 5–12.

Eliassen, J. C., Baynes, K., and Gazzaniga, M. S. 2000. Anterior and posterior callosal contributions to simultaneous bimanual movements of hands and fingers. *Brain* **123:** 2501–2511.

Eslinger, P. J. 1996. Conceptualizing, describing, and measuring components of executive function: A summary. In *Attention, Memory, and Executive Function* (G. R. Lyon, and N. A. Krasnegor, Eds.), pp. 367–395. Brookes, Baltimore.

Golden, G. J. 1978. *Stroop Color and Word Test.* Stoelting, Chicago.

Good, P. 2000. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses,* 2nd ed. Springer-Verlag, New York.

Harris, J. C. 1998. *Developmental Neuropsychiatry.* Oxford Univ. Press, New York.

Heaton, R. K., Curtiss, G., and Tuttle, K. 1993. *Wisconsin Card Sorting Test: Computer Version-2.* Psychological Assessment Resources, Lutz, FL.

Heindel, W. C., Salmon, D. P., Shults, C. W., Walicke, P. A., and Butters, N. 1989. A comparison of Alzheimer's, Huntington's, and Parkinson's disease patients. *J. Neurosci.* **9:** 582–587.

Iacoboni, M., Zaidel, E. 1995. Channels of the corpus callosum: Evidence from simple reaction times to lateralized flashes in the normal and the split brain. *Brain* **118:** 779–788.

Jacobson, S. W. 1998. Specificity of neurobehavioral outcomes associated with prenatal alcohol exposure. *Alcoholism: Clin. Exp. Res.* **22:** 313–320.

Jastak, S., and Wilkinson, G. S. 1984. *Manual for the Wide Range Achievement Test Revised.* Jastak Associates, Wilmington, DE.

Kent, J. T., and Mardia, K. V. 2001. Shape, procrustes tangent projections and bilateral symmetry. *Biometrika* **88:** 469–485.

Kerns, K. A., Don, A., Mateer, C. A., and Streissguth, A. P. 1997. Cognitive deficits in nonretarded adults with fetal alcohol syndrome. *J. Learn. Disabil.* **30:** 685–693.

Kodituwakku, P. W., Handmaker, N. S., Cutler, S. K., Weathersby, E. K., and Handmaker, S. D. 1995. Specific impairments in self-regulation in children exposed to alcohol prenatally. *Alcoholism: Clin. Exp. Res.* **19:** 1558–1564.

Kopera-Frye, K., Dehaene, S., and Streissguth, A. P. 1996. Impairments of number processing induced by prenatal alcohol exposure. *Neuropsychologia* **34:** 1187–1196.

Mardia, K. V., Bookstein, F. L., and Moreton, I. J. 2000. Statistical assessment of bilateral symmetry of shapes. *Biometrika* **87:** 285–300.

Mardia, K. V., Kent, J., and Bibby J. 1979. *Multivariate Analysis.* Academic Press, London.

Matthews, C. G., and Klove, H. 1978. *Wisconsin Motor Steadiness Battery: Administration Manual for Child Neuropsychology Battery.* University of Wisconsin Medical School, Neuropsychology Laboratory, Madison, WI.

Mattson, S. N., Jernigan, T. L., and Riley, E. P. 1994. MRI and prenatal alcohol exposure: Images provide insight into FAS. *Alcohol Health Res. World* **18:** 49–52.

Mattson, S. N., Riley, E. P., Gramling, L. S., Delis, D. C., and Jones, K. L. 1997. Heavy prenatal alcohol exposure with or without physical features of fetal alcohol syndrome leads to IQ deficits. *J. Pediatr.* **131:** 718–721.

Mattson, S. N., and Riley, E. P. 1997a. Brain anomalies in fetal alcohol syndrome. In *Fetal Alcohol Syndrome: From Mechanisms to Prevention* (E. L. Abel, Ed.). CRC Press, Boca Raton, FL.

Mattson, S. N., and Riley, E. P. 1997b. Neurobehavioral and neuroanatomical effects of heavy prenatal exposure to alcohol. In *The Challenge of Fetal Alcohol Syndrome: Overcoming Secondary Disabilities* (A. P. Streissguth and J. Kanter, Eds.), pp. 3–14. Univ. of Washington Press, Seattle.

Mattson, S. N., Riley, E. P., Gramling, L. S., Delis, D. C., and Jones, K. L. 1998. Neuropsychological comparison of alcohol-exposed children with or without physical features of fetal alcohol syndrome. *Neuropsychology* **12:** 146–153.

Mattson, S. N., Goodman, A. M., Caine, C., Delis, D. C., and Riley, E. P. 1999. Executive functioning in children with heavy prenatal alcohol exposure. *Alcoholism: Clin. Exp. Res.* **23:** 1808–1815.

McIntosh, A. R., Bookstein, F. L., Haxby, J., and Grady, C. 1996. Multivariate analysis of functional brain images using partial least squares. *NeuroImage* **3:** 143–157.

Meyers, J. E., and Meyers, K. R. 1995. *Rey Complex Figure Test and Recognition Trial: Professional Manual.* Psychological Assessment Resources, Odessa, FL.

Milner, B. 1965. Visually-guided maze learning in man: Effects of bilateral hippocampal, bilateral frontal, and unilateral cerebral lesions. *Neuropsychologia* **3:** 317–338.

Mirsky, A. F., Anthony, B. J., Duncan, C. C., Ahearn, M. B., and Kellam, S. G. 1991. Analysis of the elements of attention: A neuropsychological approach. *Neuropsychol. Rev.* **2:** 109–145.

Olson, H. C., Feldman, J., Streissguth, A. P., Sampson, P. D., and Bookstein, F. L. 1998. Neuropsychological deficits in adolescents with fetal alcohol syndrome: Clinical findings. *Alcoholism: Clin. Exp. Res.* **22:** 1998–2012.

Pellegrino, J. W., Hunt, E., Abate, R., and Farr, S. 1987. A computer-based battery for the assessment of static and dynamic spatial reasoning abilities. *Behav. Res. Meth. Inst. Comp.* **19:** 231–236.

Peterson, L. R., and Peterson, M. J. 1959. Short-term retention of individual verbal items. *J. Exp. Psychol.* **58:** 193–198.

Reitan, R. M., and Davison, L. A. 1974. *Clinical Neuropsychology: Current Status and Applications.* Winston/Wiley, New York.

Reitan, R. M., and Wolfson, D. 1985. *The Halstead-Reitan Neuropsychological Test Battery.* Neuropsychology Press, Tucson.

Reyment, R. A., and Jöreskog, K. G. 1993. *Applied Factor Analysis in the Natural Sciences.* Cambridge Univ. Press, Cambridge.

Riley, E. P., Mattson, S. N., Sowell, E. R., Jernigan, T. L., Sobel, D. F., and Jones, K. L. 1995. Abnormalities of the corpus callosum in children prenatally exposed to alcohol. *Alcoholism: Clin. Exp. Res.* **19:** 1198–1202.

Romano, J. P. 1990. On the behavior of randomization tests without a group invariance assumption. *J. Am. Stat. Assoc.* **55:** 686–692.

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D., and Beck, L. N. 1956. A continuous performance test of brain damage. *J. Consult. Psychiatry* **20:** 343–350.

Ruff, R. M., Evans, R., and Marshall, L. F. 1986. Impaired verbal and figural fluency after head injury. *Arch. Clin. Neuropsychol.* **1:** 87–101.

Sampson, P. D., Streissguth, A. P., Bookstein, F. L., Little, R. E., Clarren, S. K., Dehaene, P., Hanson, J. W., and Graham, J. M. 1997a. Incidence of fetal alcohol syndrome and prevalence of alcohol-related neurodevelopmental disorder. *Teratology* **56:** 317–326.

Sampson, P. D., Kerr, F. B., Olson, H. C., Streissguth, A. P., Hunt, E., Barr, H. M., Bookstein, F. L., and Thiede, K. 1997b. The effects of prenatal alcohol exposure on adolescent cognitive performance: A speed–accuracy tradeoff. *Intelligence* **24:** 329–353.

Sampson, P. D., Streissguth, A. P., Bookstein, F. L., and Barr, H. M. 2000. On categorizations in analyses of alcohol teratogenesis. *Environ. Health Perspect.* **108**(Suppl.3): 421–428.

Shallice, T., and Burgess, P. W. 1991. Deficits in strategy application following frontal lobe damage in man. *Brain* **114:** 727–741.

Sohlberg, M. M., and Mateer, C. M. 1989. *Introduction to Cognitive Rehabilitation: Theory and Practice.* Guilford, New York.

Sowell, E. R., Mattson, S. N., Thompson, P. M., Jernigan, T. L., Riley, E. P., and Toga, A. W. 2001a. Mapping callosal morphology and cognitive correlates. *Neurology* **57:** 235–244.

Sowell, E. R., Thompson, P. M., Mattson, S. N., Tessner, K. D., Jernigan, T. L., Riley, E. P., and Toga, A. W. 2001b. Voxel-based morphometric analyses of the brain in children and adolescents prenatally exposed to alcohol. *Cogn. Neurosci. Neuropsychol.* **12:** 515–523.

Stratton, K., Howe, C., and Battaglia, F., (Eds.) 1996. *Fetal Alcohol Syndrome: Diagnosis, Epidemiology, Prevention, and Treatment.* Institute of Medicine, U.S. National Academy of Sciences, National Academy Press, Washington, DC.

Streissguth, A. P., Barr, H. M., Sampson, P. D., Bookstein, F. L., and Darby, B. L. 1989. Neurobehavioral effects of prenatal alcohol (three parts.) *Neurotoxicol. Teratol.* **11:** 461–507.

Streissguth, A. P., Aase, F. M., Clarren, S. K., Randels, S. P., LaDue, R. A., and Smith, D. F. 1991. Fetal alcohol syndrome in adolescents and adults. *J. Am. Med. Assoc.* **265:** 1961–1967.

Streissguth, A. P., Bookstein, F. L., Sampson, P. D., and Barr, H. M. 1993. *The Enduring Effects of Prenatal Alcohol Exposure on Child Development, Birth through Seven Years: A Partial Least Squares Solution.* University of Michigan Press, Ann Arbor.

Streissguth, A. P., Sampson, P. D., Olson, H. C., Bookstein, F. L., Barr, H. M., Scott, M., Feldman, J., and Mirsky, A. F. 1994a. Maternal drinking during pregnancy: Attention and short-term memory in 14-year-old offspring—A longitudinal prospective study. *Alcoholism: Clin. Exp. Res.* **18:** 202–218.

Streissguth, A. P., Barr, H. M., Olson, H. C., Sampson, P. D., Bookstein, F. L., and Burgess, D. M. 1994b. Drinking during pregnancy decreases Word Attack and Arithmetic scores on standardized tests: Adolescent data from a population-based prospective study. *Alcoholism: Clin. Exp. Res.* **18:** 248–254.

Streissguth, A. P., Barr, H. M., Kogan, J., and Bookstein, F. L. 1996. *Understanding the Occurrence of Secondary Disabilities in Clients with Fetal Alcohol Syndrome (FAS) and Fetal Alcohol Effects (FAE),* final report, CDC Grant CCR008515. Fetal Alcohol and Drug Unit, University of Washington School of Medicine, Seattle, WA.

Streissguth, A. P., Barr, H. M., Kogan, J., and Bookstein, F. L. 1997. Secondary disabilities in people with FAS and FAE. In *The Challenge of Fetal Alcohol Syndrome: Overcoming Secondary Disabilities* (A. P. Streissguth and J. Kanter, Eds.), pp. 203–221. University of Washington Press, Seattle.

Streissguth, A. P., Bookstein, F. L., Barr, H. M., Press, S., and Sampson, P. D. 1998. A fetal alcohol behavior scale. *Alcoholism: Clin. Exp. Res.* **22:** 325–333.

Sulik, K. K., Johnson, M. C., and Webb, M. A. 1981. Fetal alcohol syndrome: Embryogenesis in a mouse model. *Science* **214:** 936–938.

Swayze, V. W., II, Johnson, V. P., Hanson, J. W., Piven, J., Sato, Y., Giedd, J. N., Mosnik, D., and Andreasen, N. C. 1997. Magnetic resonance imaging of brain anomalies in fetal alcohol syndrome. *Pediatrics* **99:** 232–240.

Talland, G. A. 1965. *Deranged Memory.* Academic Press, New York.

Thompson, P. M., Giedd, J. N., Woods, R. P., MacDonald, D., Evans, A. C., and Toga, A. W. 2000. Growth patterns in the developing human brain detected by using continuum-mechanical tensor maps. *Nature* **404:** 190–193.

Wechsler, D. 1981. *Wechsler Adult Intelligence Scale—Revised.* Psychological Corp., San Antonio, TX.

Woodcock, R. W. 1987. *Woodcock Reading Mastery Tests Revised.* American Guidance Service, Circle Pines, NM.

TERATOLOGY 64:4–32 (2001)

Geometric Morphometrics of Corpus Callosum and Subcortical Structures in the Fetal-Alcohol-Affected Brain

FRED L. BOOKSTEIN,[1]* PAUL D. SAMPSON,[2] ANN P. STREISSGUTH,[3] AND PAUL D. CONNOR[3]

[1]*Institute of Gerontology, University of Michigan, Ann Arbor, Michigan 48109*
[2]*Department of Statistics, University of Washington, Seattle, Washington 98195*
[3]*Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, Washington 98195*

ABSTRACT

Background: Although experienced clinicians have been diagnosing fetal alcohol syndrome (FAS) for nearly 30 years, the rest of the spectrum of fetal alcohol damage is not being classified effectively. This article describes a quantification of neuroanatomical structure that may supply a useful discriminator of prenatal brain damage from alcohol. It is demonstrated in a data set of adults of both sexes.

Methods: Ninety adults (45 males) were examined by magnetic resonance imaging (MRI). These subjects were group-matched for age and ethnicity across three diagnoses: FAS, fetal alcohol effects (FAE), and normals. All FAS and FAE were heavily alcohol-exposed in utero; normals were not. From T_1-weighted MR brain images, we extracted 3D morphometric representations of shape for 33-landmark point configurations and 40-point outlines of the corpus callosum along its midline (a slightly nonplanar structure).

Results: There are striking differences between exposed and unexposed in the statistical distributions of these two shapes. The differences are better characterized by excess variance in the exposed group than by any change in average landmark or outline shape. For each sex, combining the callosal outline data with the landmark data leads to a powerful quadratic discriminator of exposed from unexposed. The discriminating features include the relationship of brain stem to diencephalon, and localized variabilities of callosal outline shape, but not diagnosis (FAS vs. FAE).

Conclusions: Statistical analysis of brain shape is a powerful new source of information relevant to fetal alcohol spectrum nosology and etiology. Patients with FAS and FAE do not differ in these brain shape features, but both differ from the unexposed. The aspects of brain shape that are especially variable may be entailed in the underlying neuroteratogenetic mechanisms.

Teratology 64:4–32, 2001.

INTRODUCTION

The teratogenic properties of alcohol were suspected when children with unusual faces, growth deficiency, and a variety of abnormalities were observed among the offspring of alcoholic women (Rouquette, '57; Lemoine et al., '68; Jones et al., '73). Jones and Smith ('73) coined the term "fetal alcohol syndrome" (FAS). Soon afterward, additional groups of children with this diagnosis were reported from France (Dehaene et al., '77), Germany (Majewski et al., '76), Sweden (Olegård et al., '79), and elsewhere. By 1978, after more than 250 published case reports (Clarren and Smith, '78), it was clear that FAS was only one identifiable form of an extended range of disorders associated with maternal alcohol abuse. By 1980, the teratogenic properties of alcohol had been clearly established in animal models (cf. Randall, '77), and neurobehavioral consequences of prenatal alcohol exposure were being discovered that were not necessarily associated with morphologic abnormality or even growth deficiency (Martin et al., '77; Ouelette et al., '77; Sander et al., '77; Landesman-Dwyer et al., '78; Streissguth, '78; Streissguth et al., '80a, b). By the mid-1980s, there was a large body of literature from both animal and human data congruent with the principles of teratology as set out by Wilson and Fraser ('77), showing multiple central nervous system (CNS) effects of prenatal alcohol exposure (West, '86) that depend on the dose, timing, and condition of exposure. Related literature (Riley and Voorhees, '86) enumerated teratogens in addition to alcohol for which brain damage was not necessarily accompanied by morphological abnormalities or growth deficiency.

Yet throughout this period, clinical diagnosis remained focused on FAS. The non-FAS range of the spectrum of fetal alcohol damage has been variously

Grant sponsor: National Institutes of Health; Grant number: AA-10836; Grant number: GM-37251.

*Correspondence to Fred L. Bookstein, Institute of Gerontology, University of Michigan, 300 North Ingalls Building, Ann Arbor, MI 48109-2007. E-mail: fred@brainmap.med.umich.edu

Received 10 August 2000; Accepted 30 December 2000

TABLE 1. Age, race, and IQ by sex and diagnosis

	White	Nat. Am.	Black	Mean age	Age range	Mean IQ	IQ range
Males							
Normals	10	3	2	24.2	19.1–36.9	113	85–137
FAE	9	4	2	23.6	18.6–32.4	87	67–107
FAS	9	6	0	23.9	18.5–36.9	84	65–113
Females							
Normals	9	5	1	23.1	19.0–36.2	114	93–136
FAE	11	4	0	24.9	18.0–37.4	83	75–106
FAS	10	4	1	25.1	18.2–35.9	82	66–102

Nat. Am., Native American; FAE, fetal alcohol effects; FAS, fetal alcohol syndrome.

labeled fetal alcohol effects (FAE) (Clarren and Smith, '78; Hanson et al., '78), expanded FAS (Shaywitz et al., '80), alcohol-related birth defects (NIAAA, '83), prenatal exposure to alcohol (Riley et al., '95), partial FAS and alcohol-related neurodevelopmental disorder (ARND) (Stratton et al., '96), or atypical FAS and alcohol encephalopathy (Astley and Clarren, '00). For 20 years, until just recently, there were few clinical protocols for diagnosis within this range in the individual case, despite the overwhelming evidence that alcohol is teratogenic throughout pregnancy (Guerri, '98) at doses and timings of exposure that may not produce observable dysmorphology or growth deficiency. The widespread abuse of alcohol in our society, combined with this persistent nosological confusion regarding "partial manifestations" of the syndrome, have led to major problems identifying and meeting the therapeutic needs of individuals prenatally damaged by alcohol over this extended range of effects.

A quantitative evaluation of brain morphology might improve this diagnostic process. This article is one in a series examining brain morphology and neuropsychological deficit in a balanced sample of 180 subjects equally divided by age (adults and adolescents), sex, and diagnosis (FAS, FAE, and normals for comparison). The present article examines alcohol-related brain damage using data from three-dimensional (3D) analysis of magnetic resonance images (MRI) for the full sample of 90 adult subjects but defers analysis of adolescents and of neuropsychological sequelae at all ages to later manuscripts.

In studies of other severe childhood disorders, such as schizophrenia or autism, subjects are characterized by typical behaviors of unknown etiology. By contrast, in studies of FAS/FAE, patients have all been damaged by a known teratogen, prenatal exposure to ethanol: they represent the spectrum of consequences of a biological process the cause of which is known. Exploiting this knowledge of etiological homogeneity, in recent years several investigators using MRI have reported morphologic abnormalities in patients with FAS/FAE (Mattson et al., '96: diencephalon, cerebellum, and basal ganglia; Riley et al., '95: corpus callosum; Swayze et al., '97: corpus callosum) that arise from the common embryological challenges confronting these patients' brains. The present article shares this thrust, as well as the rich data resources of contemporary MRI, but exploits a considerably more sophisticated analytic strategy for neuroanatomic data.

The methodology we exploit in this study is landmark-based, as discussed in the section, MR Images and Derived Data. In its handling of size differences, the method is demonstrably more powerful than earlier attempts (e.g., Mattson et al., '96) to "adjust" the size of neuroanatomical components for the microencephaly that is often found to characterize those with the diagnosis of FAS or FAE. The principal sample filter applied in the present study is simply the requirement that the subjects be able to negotiate both the MRI session and the 5-hr neurobehavioral battery. (The neurobehavioral findings will be reported and correlated to the neuroanatomical data in subsequent publications.) The resulting study is the first, we believe, of sufficient sample size and richness of neuroanatomical data structure to develop strategies for individual classification and detection. It is to this elusive question that the present work is addressed, toward the resolution of the conundrum that has driven the entire research program of our group: the prognostically and therapeutically valid classification of patients with brain damage from prenatal alcohol exposure over the full range of forms of damage. Only a broad-spectrum nosology can be expected to drive appropriate service delivery protocols.

SAMPLE

We studied 90 Seattle-area subjects aged 18–37 years, comprising 30 unexposed normals and 60 cases. For brevity, we refer to the pool of all 60 cases as "the exposed," although, of course, each was not only exposed to alcohol prenatally, but also affected, as evidenced by their alcohol-related diagnoses. Thirty had been diagnosed as FAS by a dysmorphologist, 30 as FAE. All groups of 30 were divided equally between males and females, and the subgroups of 15 were group-matched by age and, as far as possible, by ethnicity (Table 1). After giving informed consent, all subjects were examined by the identical protocol. Patient ascertainment was from the Seattle FAS Follow-up Database, accrued over nearly three decades from referrals from dysmorphologists. The diagnosis was made by David W. Smith or one of his fellows or trainees (usually Sterling K. Clarren) after a clinical dys-

TABLE 2. Landmark points: names and operational definitions

Abbreviation	Name	Sagittal view	Axial view	Coronal view
Tip4[a]	Tip of 4th ventricle	Posterior tip of sharp angle	Most posterior point of CSF	Thin canal of CSF connecting two lateral "pools"
AC[a]	Anterior commissure	Center of small oval of white matter	Center of cylindrical band of white matter between hemispheres	Center of cylindrical band of white matter between hemispheres
PC[a]	Posterior commissure	Middle of vertical arc of little C-shaped appendage to colliculi	Center of cylindrical band of white matter between hemispheres	Center of cylindrical band of white matter between hemispheres
Genu	Genu–CC	Point directly in front of internal genu	Local axis of symmetry of callosum–CSF boundary	Short segment of gray between white matter and CSF
Spl	Splenium–CC	Posteriormost point on CC	Local axis of symmetry of CC–CSF	Short segment of gray between white matter and CSF
Fr–l, Fr–r	Frontal horn of lateral ventricle L, R	Frontalmost point on teardrop shape boundary between brain–CSF	Frontalmost point of boundary between brain–CSF	Slightly off center of small oval of CSF
Cd–l, CD–r	Caudate L, R	Triple point posterior bottom of horn of ventricle at spring of thin white capsule	Triple point of CSF, caudate, and white matter	Beginning of the appearance of gray matter of caudate
Chi	Optic chiasm	Anterior/inferior limit of the chiasm	Front of the chiasm at the midline	Locus where chiasm disappears
SC–l, SC–r	Superior colliculus L, R	Center of hump on outer boundary	Tip of hump	Tip of hump
IC–l, IC–r	Inferior colliculus L, R	Center of hump on outer boundary	Tip of hump	Center of white oval
PS–l, PS–r	Posterior point of hippocampus L, R	Top of hippocampus below thin strip of white matter below CSF		Cusp of fimbria where turns down to capsule of the hippocampus
AH–l, AH–r	Temporal horn L, R	Anterior limit of CSF	Anterior point of CSF	Center of little spike of CSF lateral to hippocampus
IntG	Interior genu–CC	Point of sharpest curvature on interior border of CC	Local axis of symmetry on boundary of CC	Local axis of symmetry on boundary of CC
Fx–l, Fx–r	Fornix L, R	Center of the spring of fornix emerging from CC	Center of the spring of fornix emerging from CC	Center of the spring of fornix emerging from CC
TpP	Top of pons	Deepest point of crevice of pons	Posteriormost point of wedge of CSF	Center of small part of CSF at back of pons
BtP	Bottom of pons	Deepest point of crevice of pons	Posteriormost point of wedge of CSF	Center of small part of CSF at back of pons
Obex	Obex	Interior corner of shelf of white matter at beginning point of spinal aqueduct	Deepest point of wedge of CSF	Bottom of wedge of CSF
CCS	Central cerebellar sulcus	Hull of cerebellum bisecting the central sulcus	Local midline of cerebellum	Local midline of cerebellum
Pd–l, Pd–r	Cerebellar peduncle L, R	Center of the spring of the peduncle as it emerges from midbrain	Center of peduncle	Center of peduncle
Rost	Rostrum–CC	Sharp corner of CC	Axis of symmetry at narrowest point of CC	Axis of symmetry of last pixels of CC at "drooping moustache" of white matter
FrPo[b]	Frontal pole	Anteriormost point of cortex	Anteriormost point of cortex	First slice in which cortex is visible
OcPo[b]	Occipital pole	Posteriormost point of cortex	Posteriormost point of cortex	Last slice in which cortex is visible
Top[b]	Top of brain	Superiormost aspect of cortex	Last slice in which cortex is visible	Superiormost aspect of cortex
Left[b]	Left side of brain	Last slice in which cortex is visible	Leftmost point of cortex	Leftmost point of cortex
Rght[b]	Right side of brain	Last slice in which cortex is visible	Rightmost point of cortex	Rightmost point of cortex

CC, corpus callosum; CSF, cerebrospinal fluid; L, left; R, right.
[a]Three points are used to reorient to a midsagittal plane.
[b]For outer extremes of the brain, cross-hairs are placed on the posterior commissure with one line running through the anterior commissure in a plane perpendicular to the midsagittal. Landmarks are then placed along cross-hair lines.

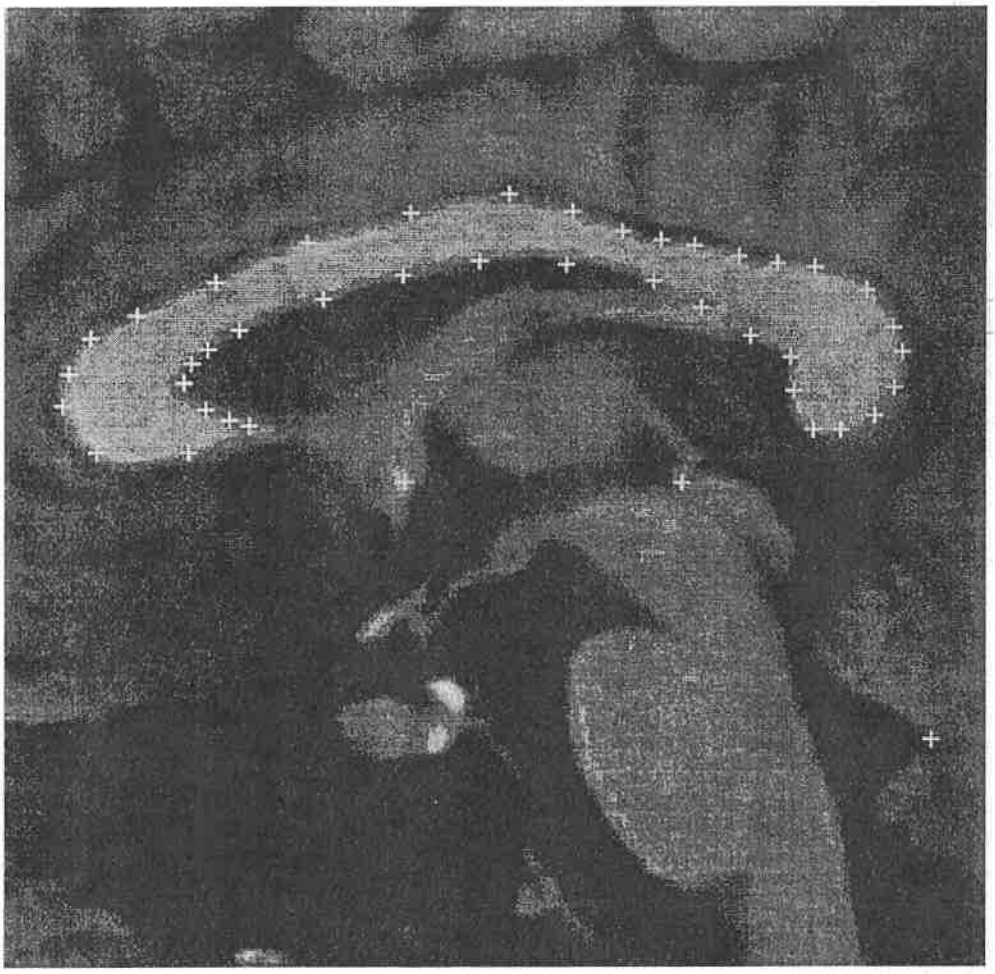


Fig. 1. Aspects of digitizing the corpus callosum in 3D. (**left**) Full outline, one subject, as projected onto a near-parasagittal plane. (The semilandmarks + do not actually lie precisely within this plane or any single plane.) Except for rostrum, all of the points on the outline have been allowed to slip to minimize bending energy with respect to a template form. Other landmarks: anterior commissure, posterior commissure, tip of fourth ventricle, and left and right brain boundaries (not shown) at posterior commissure. (**right**) A typical point of the outline (semilandmark 28, the one used in later figures) for one subject, showing how approximate symmetry is used to determine the point digitized. The section here is perpendicular to the estimated tangent line of the outline in its vicinity. The point digitized is slipped perpendicular to that tangent line until it lies at midgray voxel value on the "midline" located visually by evidence of symmetry, and then is moved along the tangent direction to minimize bending energy of the configuration as a whole.

we ultimately came to use. Twelve are subcortical midline points, 16 come as 8 pairs of bilateral points, and 5 are "extremal landmarks" (Bookstein, '91) at the outer boundary of the cortex with the cranium.

Callosal outlines

There is a small but persuasive literature of alcohol effects on the corpus callosum (e.g., prenatal alcohol exposure seems to be the principal known cause of partial or total callosal agenesis; Riley et al., '95). We therefore determined to represent it more richly than could be managed by its four unpaired landmarks: genu, internal genu, splenium, and rostrum (Table 2). Beginning with the landmark point rostrum already located, we digitized the rest of the callosal midline outline as a 39-point sequence of semilandmarks (landmarks "slipped" along the outline; see below). Tracing was carried out by one of the authors (F.L.B.), who was blind to the diagnostic group. A typical set of digitized locations, totaling 40 points, is at left in Figure 1. These are spaced roughly inverse to curvature on a reference form (the first one digitized). At right is a typical digitizing scene: a perpendicular section through a candidate point along the lower border of isthmus. Each digitized point lies precisely on the "vertical" (axis of symmetry) of an image like this at the apparent boundary of callosum. The word "vertical" is in inverted commas because anatomically it is oriented perpendicular to the callosal outline, and so lies truly vertically only at a few scattered points (top and bottom of the arch, bottom of the splenium). The cross-hairs in this panel indicate how within any such plane, a plane containing

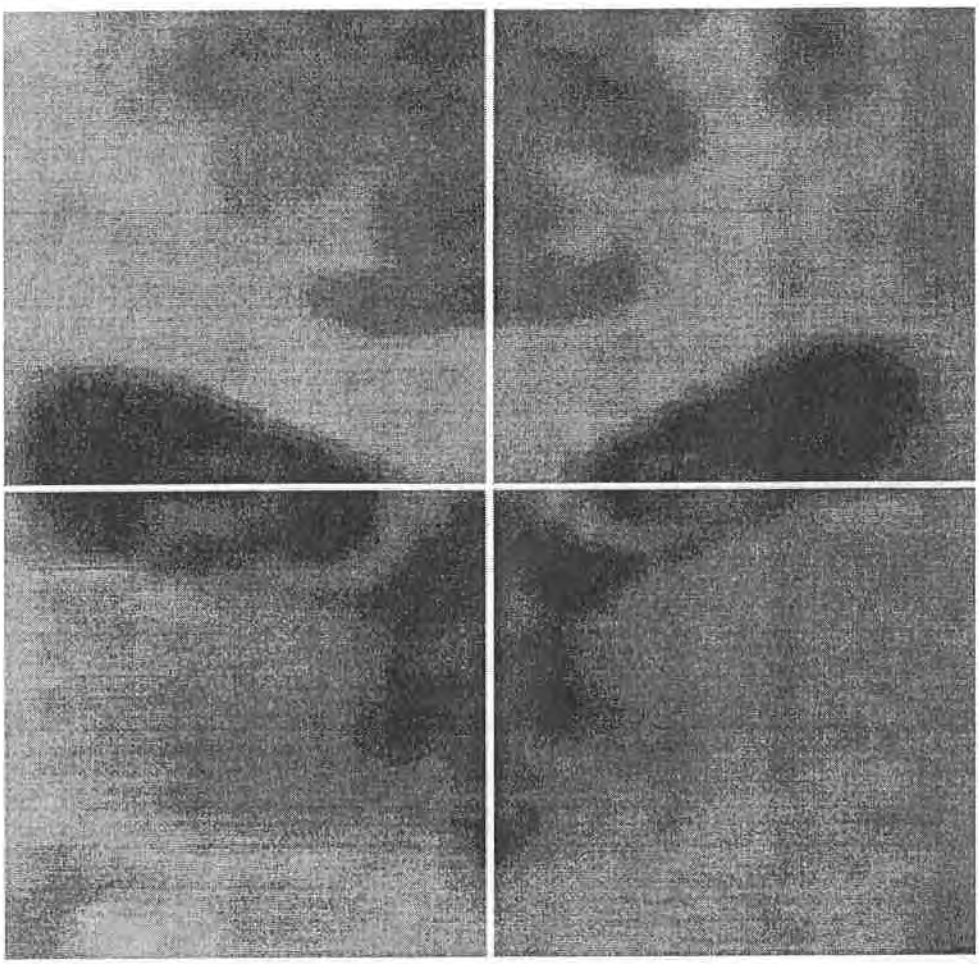


Figure 1. (Continued.)

the normal to the callosal midline somewhere, the particular point we seek is taken by its mediolateral position and its location on the grayscale gradient of the neural tissue. The third coordinate, pointing out of this page, is the coordinate that passes tangentially around the callosum in the left-hand image; it is computed by the sliding algorithm reviewed above, rather than being selected by the digitizing technician.

Even in the normal subgroup, these curves are distinctly nonplanar: they do not lie in any possible "midsagittal plane." For purposes of the statistical analysis to follow, the mediolateral (out-of-plane) coordinate has been suppressed; its distribution shows no differences among the diagnostic subgroups in either mean or variance. Reliability of these outlines over a random subsample of six digitized independently several months apart showed a reproducibility of <0.6 mm in the trajectory of the curve averaged over the whole outline. This is comparable to the standard error of the better landmark points and is considerably smaller than the magnitude of the effects reported in the present analyses (reported in units of Procrustes distance, not mm).

MORPHOMETRIC METHODS

After inspection for errors, the configurations of neuroanatomical landmarks or callosal semilandmarks were analyzed by standard methods of the morphometric synthesis. The entire subcortical region is quantified in a single multivariate analysis that considers not only differences in average size or volume, but also deviations from normal shapes and spatial relations of the different parts of the brain, whether the pattern of damage be found to be gross or localized. The modern statistical toolkit underlying this work has been the subject of a recent textbook (Dryden and Mardia, '98) and several recent review articles (Bookstein, '96, '97a, '98). There is one statistical space for the variation of the shape of the landmark set and another for the variation of the shape of the callosal outline. Thus, in their geometry the landmark data and the callosal outline data have been kept separate. In either of these spaces, the shape of the geometric object in question (a landmark set, or an outline) is represented by one single "observation" per specimen—a rather complicated observation, yes, but one that is treated as a

single algebraic entity for statistical purposes—regardless of the number of points in the original digitized representation. When the spaces are combined (to produce the findings displayed in Fig. 12), it is by statistical, not geometrical, procedures.

Getting from Cartesian coordinates to shape

The construction of a statistical shape space for outline data is a special case of the construction that applies to landmark data. This section provides a quick sketch of the standard method for landmark points and then indicates the nature of the extension to handle smoothly curving forms such as the callosum. The Appendix at the end of this article reviews the standard landmark methods in considerably greater detail.

Shape is the information about landmark configurations that remains unchanged under adjustments of position, orientation, or scale. One straightforward way of representing this information for statistical purposes begins by considering a *shape distance* between any pair of landmark configurations, and then constructing a useful set of geometric coordinates, *shape coordinates,* for which this shape distance is the appropriate "Euclidean" sum of squares. Since the great original paper by Kendall ('84), the distance used for these spaces has always been one or another modification of the following definition of *Procrustes distance.* It is convenient to begin by removing location and size information from each configuration separately, by centering each at its own center of gravity and scaling each to a fixed sum of squares around that center. The scaling factor divided out in this step, called *centroid size,* remains available as a size measure for use at any subsequent stage of analysis. If all forms are standardized in this way, the Procrustes distance between any two is simply the sum of squares of the ordinary Euclidean distances between the matching landmarks of the two landmark configurations when one of them is freely rotated (around the common center of gravity) until this sum of squares is minimized.

Shape averages and shape coordinates

From the shape distance formulation, the rest of the statistical scheme follows very directly. One can define the *average shape* of a set of landmark configurations as the shape from which they have, taken together, the least summed squared Procrustes distance. (This is precisely analogous to the least-squares property of ordinary arithmetic averages: the average of any set of numbers is the value that has the least summed squared difference from the numbers actually being averaged.) After that average has been computed, each original shape of the data set can be superimposed upon it by the rotation described in the preceding paragraph—the rotation that supplies the actual minimum sum of squares serving for its Procrustes distance to the average. The locations at which the original landmarks arrive after this rotation serve us as the shape coordinates of the original landmark configurations with respect to the sample as a whole. For k landmarks, there are $2k$ of these coordinates for 2D data, or $3k$ for three-dimensional data. (But four dimensions of their space (for 2D data) or seven dimensions (for 3D data) necessarily have no variance; instead, they express the constraints on position, orientation, and scale that were imposed during the course of the construction.) These coordinates serve as the set of variables that make possible an analysis of shape by otherwise familiar multivariate procedures. For instance, in Figure 7, each point shown stands for two shape variables—its x-coordinate and its y-coordinate—with group comparisons going forward in terms of the averages, the variances, and the covariances of those variables. (For a general discussion of the relationship between shape coordinates and shape variables, including the most familiar shape variables such as angles or ratios of distances, see Bookstein, '91). All the formulas entailed in this sequence of steps are available in the Dryden textbook and the Bookstein reviews; software is available free of charge, to carry out the computations on most scientific research computer platforms (see the website http://life.bio.sunysb.edu/morph/ maintained by F.J. Rohlf at the State University of New York at Stony Brook). Our analyses were carried out in the Splus statistical system, using functions written by the authors.

Shape coordinates for outline data

A corresponding analysis for the callosal outline data takes into account the indeterminacy of "homologous" points along extended curves like this midline. Bookstein ('97b) suggested that, beginning from any plausible sampling of points along the curves of the sample, a Procrustes average in the sense just reviewed for landmarks alternate with a "sliding" operation that redistributes the semilandmarks of each outline along that outline in such a way as to minimize the "bending energy," quantified in one specific algebraic way, that characterizes the relation of the outline to the full sample average. This approach is preferable to even spacing of points, in permitting coordination of information between top and bottom of the arch, and is preferable to approaches that assign coordinates out of a "center" for the reasons reviewed in Bookstein ('91). For a justification of the relevance of bending energy to this context, see Bookstein ('99). In the callosal data set, our bending energy computation (for sliding) also took account of four conventional landmark points: the commissures, the tip of the fourth ventricle, and the rostrum, the only good point landmark actually located on the callosum. At the convergence of this alternating algorithm, one arrives at a sample average shape and a set of shape coordinates just as before. Points generated by this algorithm are called *semilandmarks.*

Principal components analysis for shape

Once shape coordinates are in hand, a variety of multivariate statistical tools become available that correspond to those that apply to more typical biostatistical data sets. This analysis exploits two of these tools:

principal components analysis and testing statistical significance of group differences in average shape or in the variability of shape.

Whether for landmark data (points) or for semilandmark data (outlines), principal components analysis of shape is carried out by ordinary principal components analysis of the shape coordinates just described. The analysis uses their covariance matrix, not their correlation matrix, in order to preserve the Procrustes geometry through subsequent steps. Within the context of shape analysis, these components are often diagrammed as deformations (warps) of a grand mean (see Fig. 5), and are thus called *relative warps*. (In this article, "deformation" is used only in its mathematical sense, a smooth map from one picture to another.)

For ordinary sets of variables, the first principal component is characterized as the linear composite that has the greatest variance among the set of all possible composites whose coefficients sum in square to 1. Similarly, the first relative warp is the composite shape variable (pattern of joint landmark rearrangement) having the largest variance among all the shape variables of a given Procrustes length. If the concept of the Procrustes length of a variable seems forbidding, there is an exact equivalent that may be more accessible: the Procrustes length of a shape variable is proportional to its variance on a model of "pure digitizing noise," the same small variance in every direction at every landmark. Biological data can often be modeled effectively by this noise distribution, perhaps after systematic factors (e.g., prenatal exposure) have been controlled. The second relative warp is the composite shape variable having greatest variance (per unit Procrustes length) of all those that are uncorrelated with the first relative warp of the sample, and so on. Every subject in the data set has a score (projection) on every relative warp of the data set, and these scores can be scattered to look for patterns or clusters, tested for mean difference between groups, correlated with ostensible causes or effects of shape, and generally treated just like any other set of principal component scores in any other application of multivariate statistics.

Testing statistical hypotheses about shape

Although this morphometric version of principal components analysis is only slightly modified from the standard approach, the way in which one tests hypotheses of mean shape difference changes considerably. In most morphometric data sets, there are more of these shape coordinates (variables) than subjects. Hence our significance tests will usually be permutation tests (Good, '94). In a permutation test, a quantity is selected that captures a scientific question about the relation between two aspects of the data structure. For instance, in studying the association between diagnosis and callosal outline in our 45 males, we might select the Procrustes distance between the average callosal outline shapes of exposed and unexposed as an interesting measure of the scientific signal we are examining. Then the distribution of this quantity is computed over a very large collection of "pseudo datasets" in which diagnosis (FAS vs. FAE vs. normal) is randomly reassigned over the 45 adult male callosal outlines. The significance level of the empirical association between diagnosis and callosal outline, for the chosen quantity (the mean difference), equals the probability that a random permutation of this type results in a Procrustes difference between averages at least as great as the value actually observed. It was Ronald Fisher himself (the "F" of the F-test) who first noted that this is what we actually mean by a statistical significance level, to which any other quantity deriving from textbook formulas, including his own, merely approximates. For data sets large enough to preclude looking at "all possible permutations" (about 340 billion, for our sample of 30 exposed vs. 15 unexposed males), one looks at an adequately large random sample, here, a few hundred to a thousand or so. Permutation tests are easily performed in any of the standard statistical software packages and, for landmark data such as these, are built into Rohlf's TPSregr program, available free of charge for Windows PCs from the Stony Brook, New York, site.

In the landmark data set, the feature underlying the test for mean shift is Procrustes distance. In the outline data set, it is Procrustes distance in the direction normal to the outline. For demonstrations of hypervariability, it is the ratio of the sum of variances of the first three principal components of shape for the exposed to that for the unexposed.

MAIN MORPHOMETRIC COMPARISONS, BY DIAGNOSIS AND SEX

Our morphometric analyses involve comparisons of both kinds of shape by diagnostic group and sex and principal components of both kinds by sex. There are findings of four different types: size differences, means and principal components of landmark shape, means and principal components of callosal outline shape, and the combined pattern of the two shape components. In this section, each of these is recounted separately, leading to a total of six distinct findings. In general, group differences in shape variability far outweigh group differences in average shape or average size, allowing a startlingly powerful discrimination of the exposed from the unexposed by shape alone, whereas the two exposed subgroups (FAS and FAE) differ in shape hardly at all. The reader who does not wish to wade deeply into the details may choose at this point to turn directly to the section, Summary of the Findings, which highlights the six main findings separately from the supporting computations.

Size

A morphometric analysis of either landmarks or semilandmarks should begin by considering centroid size, the scaling factor divided out in the course of producing shape coordinates. Means and standard deviations of this descriptor are presented in Table 3 for

TABLE 3. Size measures by diagnosis and sex*

Group	Landmark CS	Callosal CS	Callosal area
Normal M	248.9 ± 4.8	26.25 ± 1.72	1159 ± 155
FAE M	238.2 ± 9.1	24.22 ± 1.67	1057 ± 216
FAS M	240.5 ± 7.7	24.63 ± 1.99	1103 ± 227
Normal F	241.7 ± 7.3	25.52 ± 1.80	1199 ± 165
FAE F	233.8 ± 6.9	24.73 ± 1.29	1154 ± 181
FAS F	224.6 ± 9.4	23.15 ± 1.10	928 ± 162

CS, centroid size; M, male; F, female; FAE, fetal alcohol effects; FAS, fetal alcohol syndrome.
*Means ± SD in digitizing units of mm, except for area, which is in units of mm^2.

the two different data structures of this study, the landmark configuration and the callosal outline configuration. For the landmark data in the males, both exposed groups differ by about 4% from the normal mean. The females diagnosed with FAE differ from the normal females by about the same amount, but the FAS females were about 8% in deficit. Student's t-tests for these comparisons are all significant at $P < 0.01$. For this landmark configuration, centroid size is approximately proportional to "shoebox size," the diagonal of a rectangular shoebox around the landmark configuration as a whole. For the centroid size of the callosal outline, the difference of 7% between the normal mean and the average for the exposed pool is significant at $P \sim 0.002$ by t-test. For forms that are this long and narrow, centroid size is approximately proportional to the diameter of the callosum itself, the distance from genu to splenium. For the females, the FAS mean is clearly different from both of the others. In terms of callosal area, the conventional alternative to centroid size, the male diagnostic groups likewise show a 7% shortfall for the exposed pool, but none of these differences is significant, as the within-group coefficient of variation for area is far greater than that for length. For the females, mean callosal area in the FAS subgroup is again clearly and significantly different from the other two ($P < 0.002$).

Our shape findings are reviewed in three subsections: landmark shape variation, callosal outline shape variation, and the combination.

Landmark shape variability

Figure 2 illustrates the mean landmark shape configurations according to their three cardinal views (from the front, the top, and the side). The axes are in units of Procrustes coordinates, which are dimensionless. In the top row, the landmarks are named at their grand mean locations; in the bottom row, the sex- and diagnosis-specific means are shown, exaggerated fivefold away from the grand means for visibility. The Procrustes distances among these six means are shown above the diagonal in Table 4, and their significance levels (by permutation test using 500 permutations) below the diagonal. In general, no differences attributable to syndrome are as large as the difference between the male and female normals, and within sex no mean difference between any pair of the three diagnostic groupings is significant. Male and female normals, as well as male and female FAE, are very significantly different, whereas male and female FAS are not too different. However, such contrasts are not the primary interest of this study. Although the groups are known to differ in centroid size, and although centroid size is a covariate of shape variation in general human populations, it is not appropriate to "correct" any of these comparisons for size difference, as brain size is determined by the same prenatal dose that led to the diagnostic assignments of these subjects in the first place.

We learn a great deal more from the examination of relative warps than we could glean from these mean comparisons alone. Figure 3 shows the scatter of the first two relative warps separately by sex. Recall that these are the patterns of relatively greatest shape variation, each one standing for a set of correlated shifts of all the landmarks jointly. Like any other principal component, the pattern they delineate pertains to the pool of all the male subjects, or all the females, and not to any single subject. In either sex, the scatter of the normal sample (filled circles) is much less than that of the + and × symbols standing for the exposed. The original analysis was of the first three relative warps, and tests were carried out using that slightly larger subspace. For the males, the significance level of the extra variance in the exposed subsample, summed over the first three relative warps, is $P \sim 0.01$; for the females, it is $P \sim 0.025$. Most of the excess of variance in the relative warps can be captured by projecting onto the directions shown (separately by sex) in Figure 3. The short axes of the normal subsamples after outliers (one for the males, two for the females) are sequestered. (That is, once those outliers are sequestered, the center of gravity of the remaining control data in each panel is the midpoint of the little segment there, and the best-fitting ellipse to the covariance pattern of these points has the direction indicated as its shortest diameter.) Separately by sex, these patterns of joint landmark rearrangement have considerably more variance in the alcohol-exposed subsample than in the unexposed sample.

Visualization of systematic aspects of 3D shape variation, such as these particular linear combinations, goes forward best via dynamic (tumbling) 3D displays. For publication, one selects a number of still images from these displays. Figure 4 shows the effect of an arbitrary multiple of these changes at all landmarks simultaneously, in the usual three views. In both sexes,

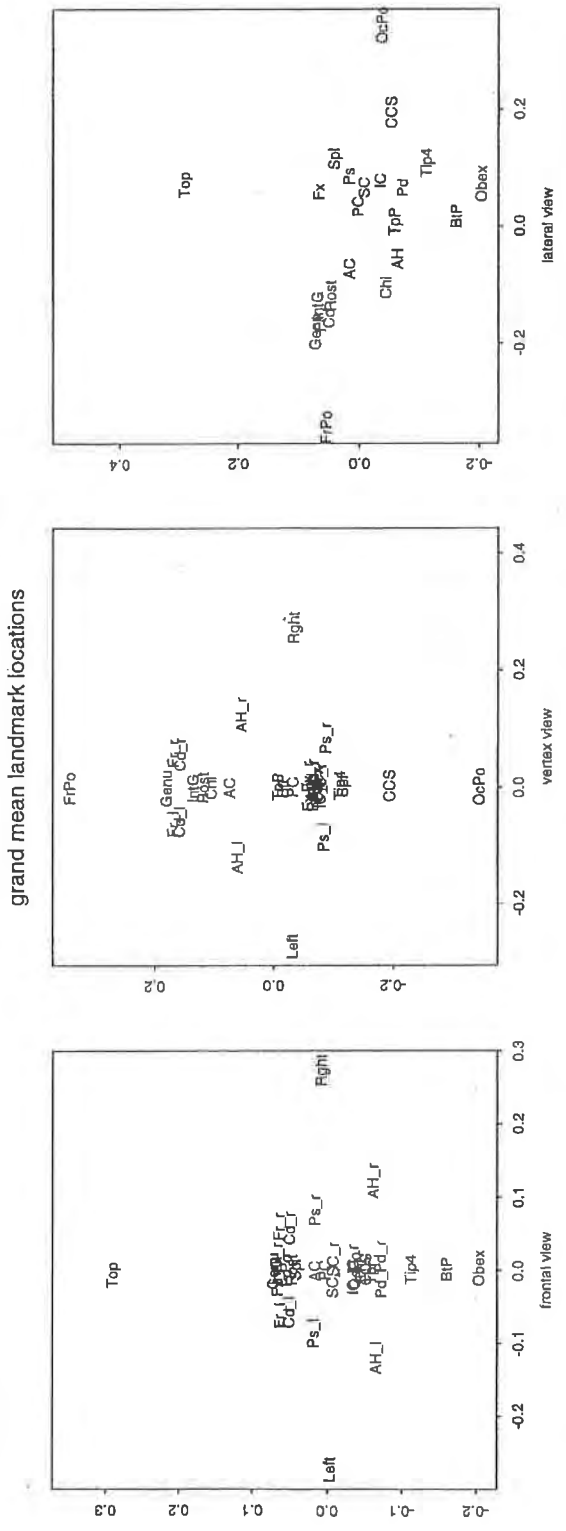


Fig. 2. Averages of Procrustes shape coordinates for the 33-landmark configuration. Upper row, grand means of all 90 cases; lower row, means by diagnostic group and sex as indicated in the legend, represented after fivefold exaggeration of their separation from the grand mean. None of the diagnostic group differences here are statistically significant. Left column, frontal view; center column, vertex view; right column, lateral view. In the lateral view, paired landmarks are represented by their bilateral averages, and left side and right side points are omitted. Axes are in units of Procrustes distance throughout. For full names and operational definitions of all landmarks, see Table 2.

TABLE 4. Procrustes distances between diagnostic groups: landmark point data*

	Male			Female		
	Normal	FAE	FAS	Normal	FAE	FAS
Normal M	—	61	72	138	104	58
FAE M	0.222	—	18	194	129	64
FAS M	0.176	>0.5	—	206	149	71
Normal F	0.002	—	—	—	77	94
FAE F	—	0.002	—	0.052	—	50
FAS F	—	—	0.295	0.323	>0.5	—

M, male; F, female; FAE, fetal alcohol effects; FAS, fetal alcohol syndrome.
*Entries above upper left–lower right diagonal: squared Procrustes distances between group mean shapes, multiplied by 10^5. Below diagonal: significance levels, according to 500 permutations of diagnostic label over landmark configuration.

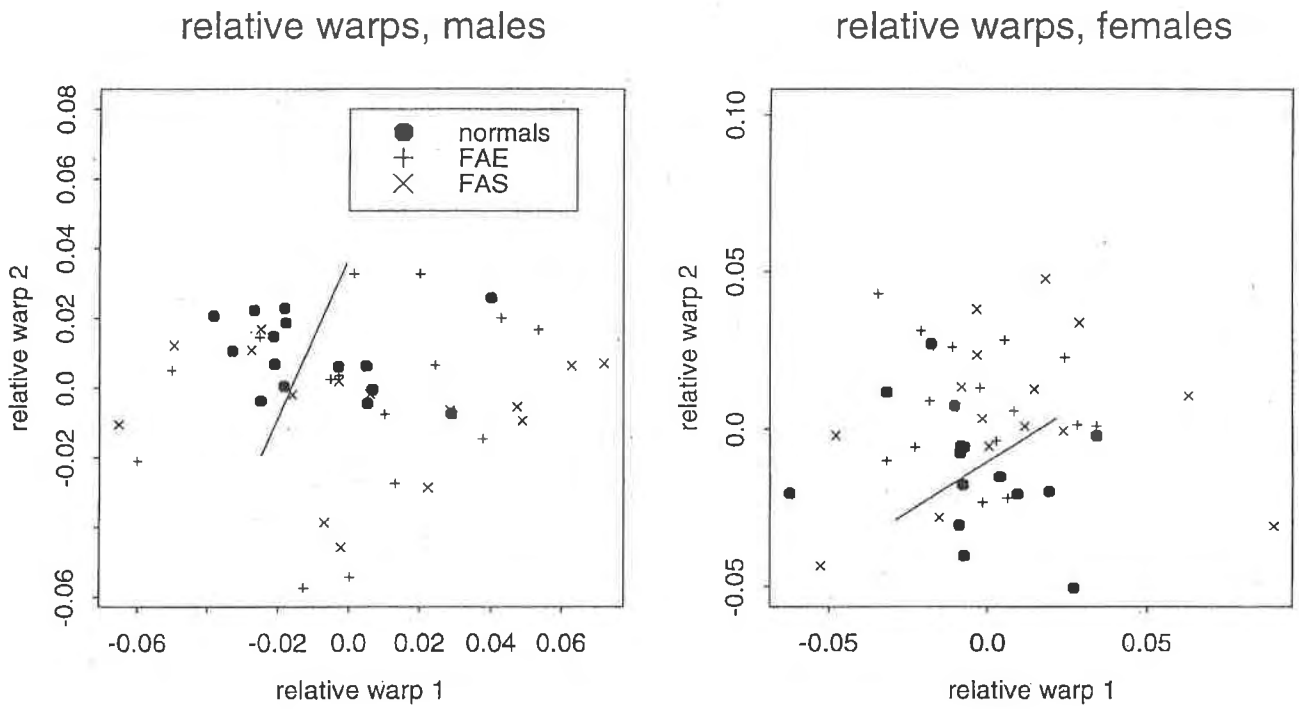


Fig. 3. Relative warps analysis (principal components analysis of shape coordinates) for the landmark data set. **Left,** males; **right,** females. Scatters of relative warp scores 1 versus 2 are in units of Procrustes distance. Short segments in each panel, indicating the dimension of least variance of the normals, are taken along the shorter principal axis for the subscatters of the apparently typical normals (14 males or 13 females). Outliers excluded for this purpose (but included in all statistical testing): for the males, the rightmost normal point; for the females, the leftmost and rightmost normal points.

very little of this correlated hypervariation lies in the mediolateral direction; thus, we can profitably restrict our attention to the lateral view, which nullifies this dimension. In that view, Figure 5 demonstrates the effect of an arbitrary amount of change in the transversal dimensions of Figure 3 by displacement vectors (top) and by splined deformations (bottom). The landmark key is repeated for convenience at lower right. The thin-plate spline, a standard tool of the new morphometrics, depicts a landmark rearrangement as a deformation of the diagram plane in which the landmarks lie. Of all the grids that could be drawn for this purpose, the spline is the smoothest—the one that has the least extent of local bending in one specific algebraic sense. For an extended explanation and justification, see Bookstein ('96, '97a, '98).

Each scene divides into two regions: structures on or adjacent to the pons, and rearrangements in the vicinity of the corpus callosum. In both sexes, the segment connecting fornix (Fx) to bottom of pons (BtP) is rotated strikingly, in accord with the realignment of grid lines from vertical throughout the middle of both grids. The feature in question varies bidirectionally; for example, the rotation of that vertical grid line could have just as

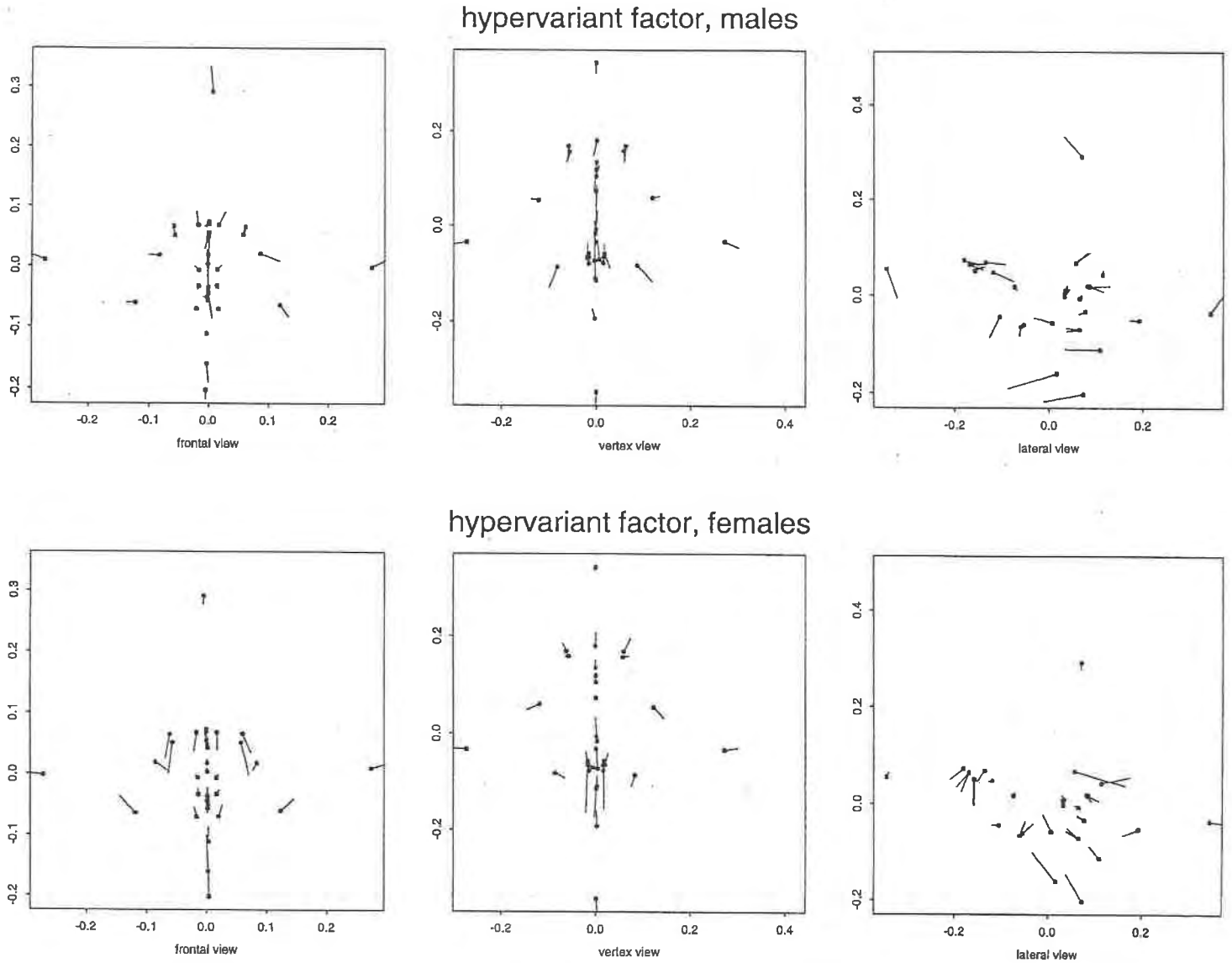


Fig. 4. The segments shown in Fig. 3 correspond to vectors that shift the grand mean landmark shape by correlated changes at all landmarks. **Top row,** an arbitrary multiple of the indicated direction of shape change for the males; **bottom row,** for females. Dots: grand mean configuration. The meaning would be the same if each small segment were rotated 180 degrees around the dots to which it pertains. Most of the information in these vectors pertains to the lateral view.

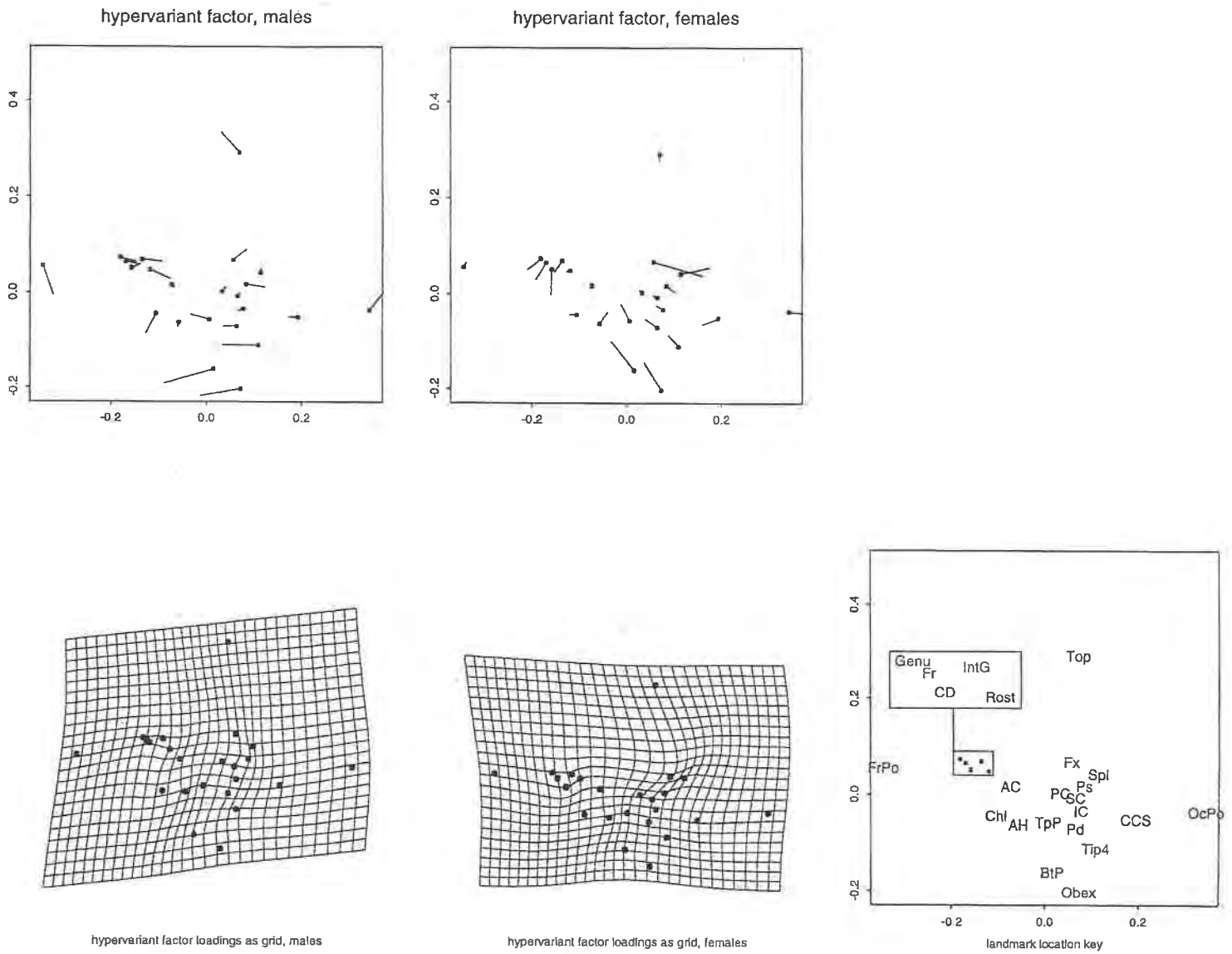


Fig. 5. Landmark hypervariability factors, continued. **Top row,** lateral panels from Fig. 4, with bilateral landmarks averaged and left side and right side landmarks deleted. **Bottom row,** the same shifts diagrammed as thin-plate spline transformation grids. Notice the concentration of the hypervariability at the front of the callosum and along the posterior vertical from splenium through pons.

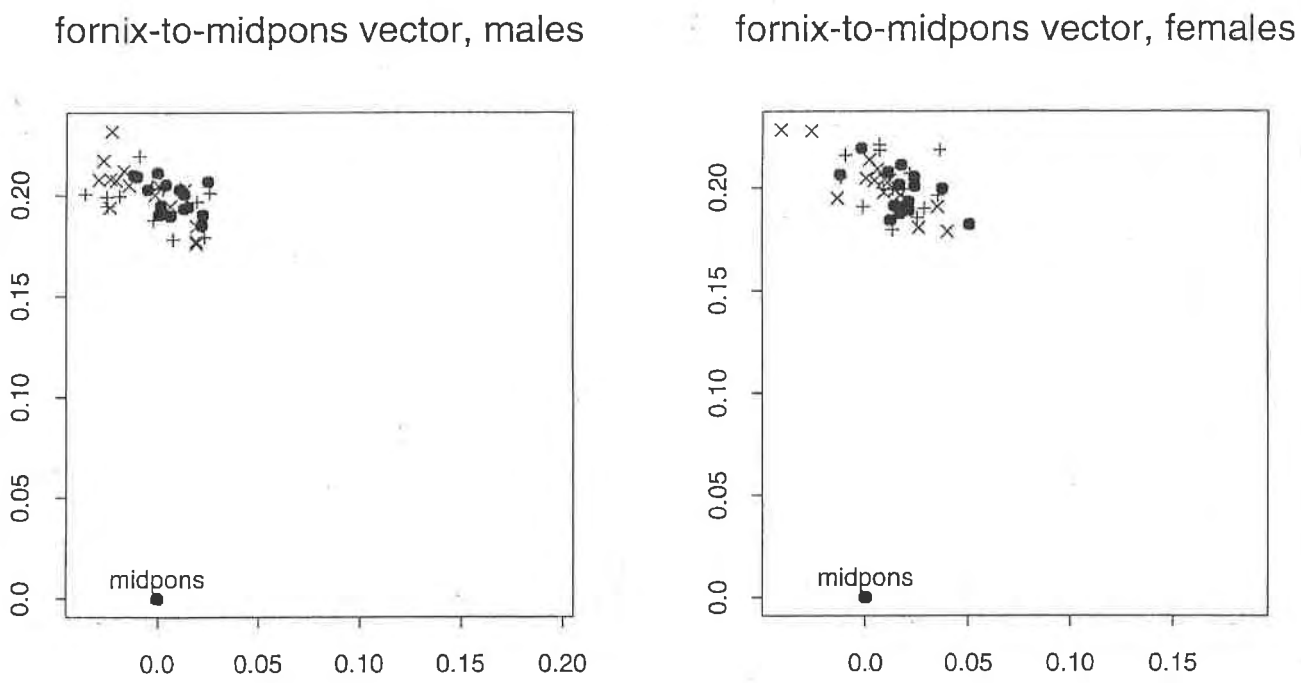


Fig. 6. Interpretation of landmark hypervariance factors by localization to a vector from midpons to midfornix in the lateral view. The vectors are plotted as if all images were registered to a common midpons point, as shown. Legend: ●, normals; +, FAE; ×, FAS. "Midpons" is the centroid of the quadrilateral TpP, BtP, Obex, and Tip4.

well been drawn as counterclockwise, by reversing all the vectors in the top panels. In the males, this factor of hypervariability moves BtP, Obex, Tip4, and TpP strongly anteriorly (respectively posteriorly) with respect to the commissures, whereas all the points of the callosum shift posteriorly (respectively anteriorly) except for Spl, the posteriormost point itself. In the females, this brain stem shift is rather more upward (respectively downward) than forward (respectively backward), the principal shift in the posterior callosum is at the fornices, and the front of the callosum shifts downward (respectively upward) rather than backward (respectively forward) with respect to the commissures. Also in both sexes, the central cerebellar sulcus (CCS) has moved distinctly anterior (respectively posterior) with respect to the occipital pole OcPo. Keep in mind that each of these is to be read as a pattern of coordinated shifts in all landmarks, and that what we are looking at is not the shift of an average shape among the diagnostic groups (none of those was significant), but rather the extra variation apparent in the pool of diagnostic groups by comparison with the normals, separately by sex.

These patterns supply tantalizing hints of localization of the excess variation for confirmation in other samples. For example, as shown in Figure 6, the within-sample variance of the vector from "midpons" (BtP, Obex, Tip4, and TpP) to fornix is 0.00049 in the male exposed pool and 0.00051 in the female, versus 0.00020 in the normal males and 0.00032 in the normal females; this variance ratio is significant at $P \sim .005$ for the males, and at 0.07 for the females, by ordinary F-test. (The variance of a vector is the sum of the variances of its two components separately; this sum is independent of the orientation of the coordinate system in which the components are specified.) The vector may be imagined as a sort of axis for the relationship of brain stem to diencephalon. (In particular, the notion of a "midpons" has no biological reality but is just a convenient way of summarizing the common shift pattern of the four points in its vicinity with respect to the shifts of the two fornix points, which go the other way according to this same pattern.) In this sense, there is considerably more variance in the exposed than in the normals in the angle between brain stem and diencephalon; the hypervariation factors of Figure 4 show this variation along with the displacements of other landmarks, such as those near genu, that happen to be associated with it.

Callosal outline shape variability

The data set of callosal outline shape offers the richer findings here. Figure 7 (left) shows the complete sample scatter of 40 points (39 semilandmarks, together with rostrum) for the complete data set of 90 adults. This variability is commensurate with that of the landmark points analyzed in the previous section. Figure 7 (right) shows the six diagnosis- and sex-specific means. In pairwise mean comparisons (by permutation test), none of the differences among the male diagnostic subgroups is significant, but all pairwise comparisons among averages for the females are nominally signifi-

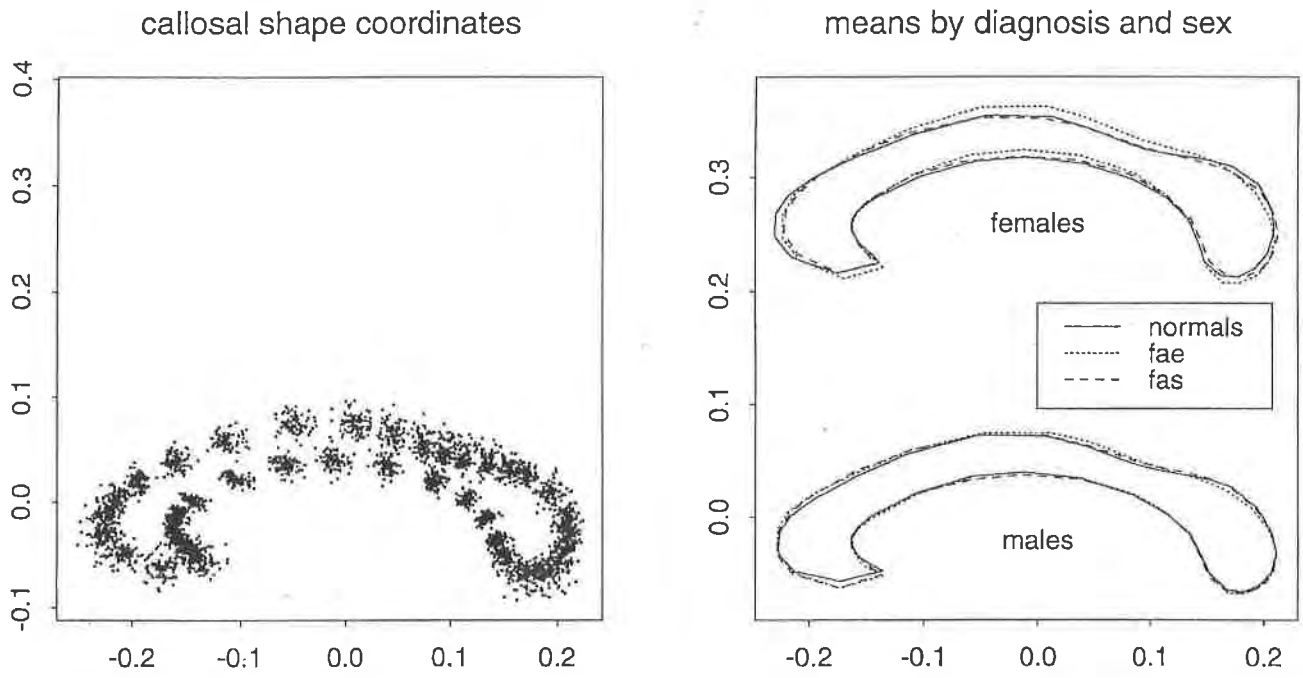


Fig. 7. Procrustes shape coordinates for the projected callosal outlines. **Left,** complete sample scatter of rostrum and 39 semilandmarks for all 90 subjects. **Right,** means by diagnosis and sex.

TABLE 5. Procrustes distances between diagnostic groups: callosal outline data*

	Male			Female		
	Normal	FAE	FAS	Normal	FAE	FAS
Normal M	—	53	34	65	162	63
FAE M	0.425	—	20	71	99	77
FAS M	>0.5	>0.5	—	58	108	57
Normal F	0.126	—	—	—	157	102
FAE F	—	0.120	—	0.006	—	118
FAS F	—	—	0.323	0.040	0.044	—

M, male; F, female; FAE, fetal alcohol effects; FAS, fetal alcohol syndrome.
*Entries above upper left–lower right diagonal: squared Procrustes distances between group mean shapes, multiplied by 10^5. Below diagonal: significance levels, according to 500 permutations of diagnostic label over callosal shape.

cant. Distances and significance levels (by permutation tests of 500 runs) are presented in Table 5. Figure 8 sets out their shape differences pairwise as thin-plate splines, exaggerated threefold for legibility. The normal mean clearly bears a differently shaped splenium from either diagnostic group, and a thicker genu. The isthmus is thinner in the FAS subgroup, with the arch relatively higher in the FAE subgroup. As the top row of Figure 8 suggests, the splenium anomaly for both exposed groups is similar to the representation of the mean shape for normal males as a deformation of that for the normal females.

Even more than the landmarks, our callosal data bespeak a strong excess variability in the exposed subsample. Figure 9 shows the subscatter of its 45 outline shapes for each sex (as in Fig. 7). When the normal subsamples are indicated by connecting the dots, it becomes apparent that the semilandmarks of the syndromal subsample lie well outside the envelope of the outlines for the normal subsamples at several sites. We can test this impression rigorously by applying the same permutation test (variance of relative warps between exposed and unexposed) that was used for the landmark data. Summed over the three callosal shape dimensions of largest Procrustes variance, the net shape variability within the exposed pool is significantly larger than that within the normal pool at $P \sim 0.01$ for the males, $P \sim 0.025$ for the females. Figure 10 localizes this difference by ordinary F-tests (Bartlett's test for difference of variances) at each semilandmark separately. Because variation along the direction of the average curve has already been adjusted out

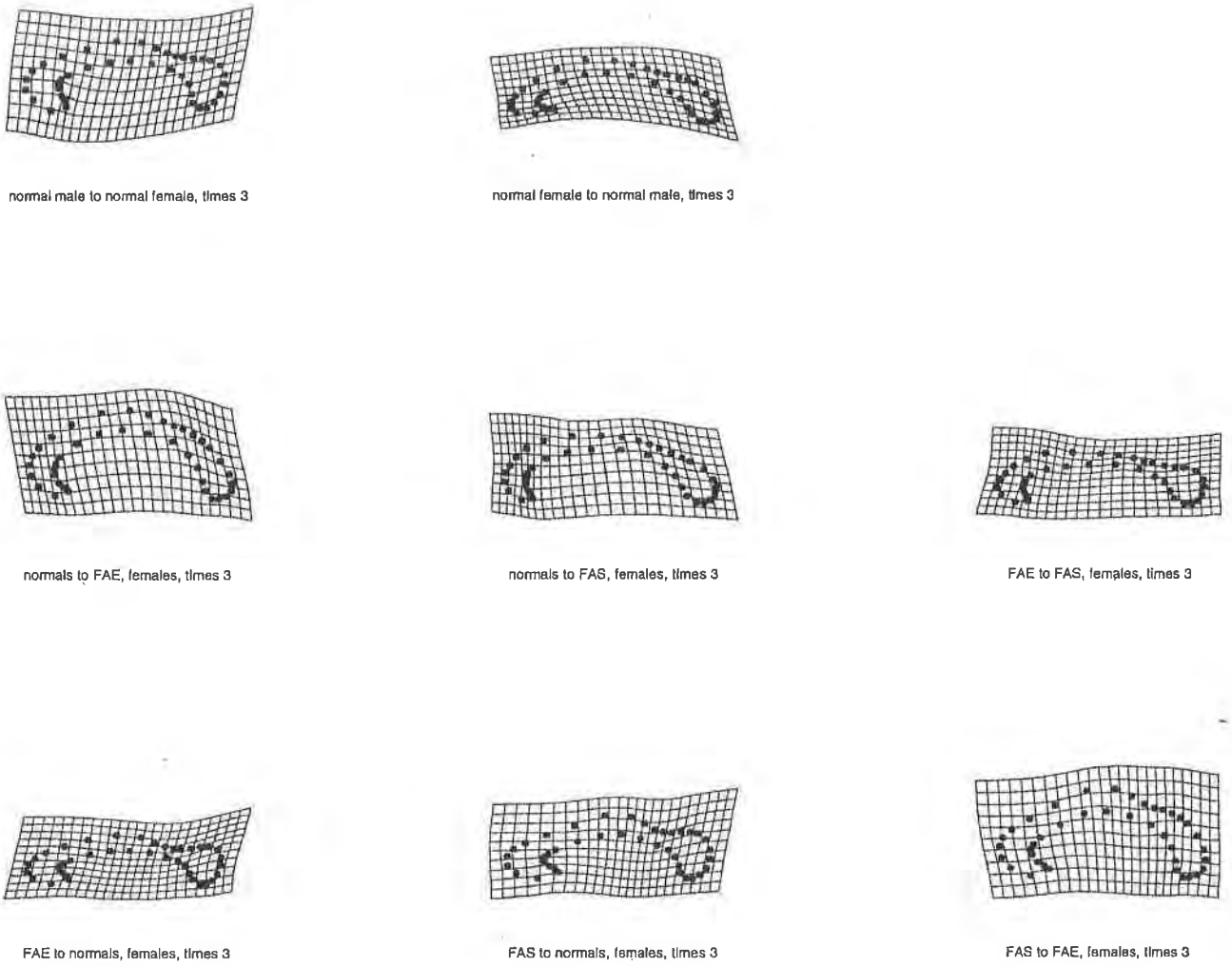


Fig. 8. Thin-plate splines for comparison of normal males and females and also for all pairwise comparisons of the female averages from Fig. 7. Each difference is exaggerated threefold for legibility.

during the course of producing these Procrustes coordinates (see section, MR Images and Derived Data), the test is of variance in one direction only, perpendicular to the average curves of Figure 7. For at least one semilandmark in each sex, the significance of this variance discrepancy is $P < 0.05/39 \sim 0.001$, so that the set of multiple comparisons is significant according to the conventional (Bonferroni-corrected) inference as well. The arcs of greatest discrepancy are quite differently situated in the two sexes: for males, anterocaudal to isthmus; for females, in the arch.

Figure 11 focuses on the best three of these points, for each sex, to illustrate what the F-test is detecting. Point by point, the normals' semilandmark coordinates concentrate themselves about the average outline at least as well as did the relative warp scores for the whole set of landmarks (Fig. 3). We thus harvest a whole new discriminator to augment the pair suggested in Figure 3. For the males, the point of greatest apparent information content, semilandmark 28, is concentrated indeed, and so we take its coordinate perpendicular to the outline as the simplest useful scalar for adducing the hypervariance of the exposed. For the females, the semilandmarks atop the arch show the greatest variance ratios. A summary score that weights upper and lower arcs equally takes the average vertical coordinate of the set of all four.

Combining the shape spaces

Figure 12 combines the findings for landmarks and callosal outlines in one composite display. Separately by sex, the dimension of sharpest increase in variance from the landmark relative warp analysis (Fig. 4) is plotted against the shape coordinate of greatest hypervariance from the outline analysis (Fig. 11). Clearly, the two analyses, by landmark and by callosal outline, are complementary. If, as shown, boxes are taken tightly around the cores of the scatters for the normals, a classification rule emerges that calls subjects alcohol-affected just when they lie outside the boxes. The rule has putative sensitivity and specificity of 0.90 and 0.93 for males, 0.97 and 0.87 for females—enormously

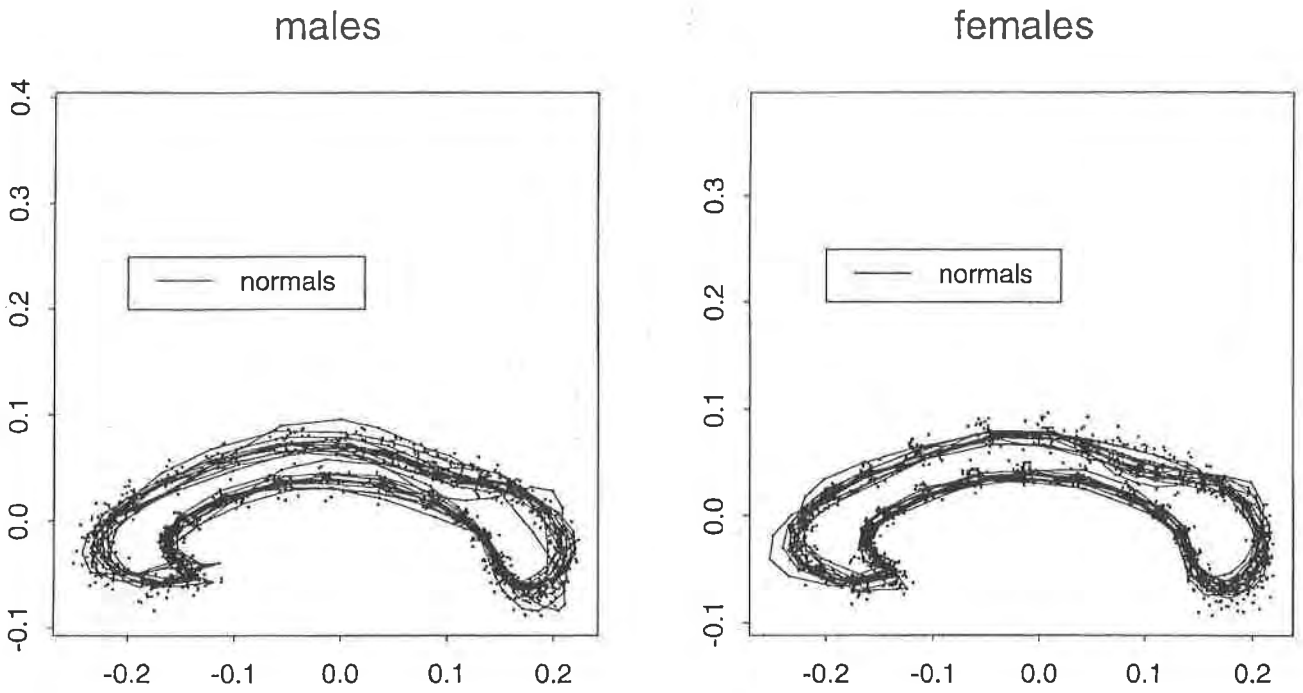


Fig. 9. Procrustes shape coordinates for the outlines shown in Fig. 7, separately by sex, with the outlines of the normal subjects connected. The variation of the exposed shapes outside the range of the normal is clear at several arcs around the circumference. The overall ratio of within-group variances, exposed vs. unexposed, is significant by the permutation test described in the text.

greater than anything hitherto reported from neuroanatomical data, with or without knowledge of size. These classifiers are uncorrelated; thus, each channel of measurement detects some exposed cases as hypervariant that are not detected by the other. But, of course, as the findings are not homologous between the sexes, each needs to be replicated in its own additional sample. Replications using 45 male and 45 female adolescents are in progress.

As the earliest FAS patients were often mentally retarded, it is important to think about IQ as a relevant dimension of these subject populations; nevertheless, because lowered IQ is itself a consequence of the brain damage that is our primary dependent variable, we did not impose any IQ selection criterion during the course of assembling our samples. Figure 13 annotates Figure 12 with the full-scale IQ scores of the alcohol-affected subgroup. The plot is entirely consistent with our emerging awareness that facial stigmata, intellectual deficit, and actual neuroanatomical abnormalities are fairly independent within the relatively high-performing end of the exposed range. There is no apparent pattern of full-scale IQ deficit by position in this plot in either sex, nor is the FAE subgroup "intermediate" in any useful sense between the normal subsample and the subjects diagnosed with the full FAS. As this is an important finding, albeit a negative one, we present it more explicitly in Figure 14, which makes explicit how IQ varies with "net severity" of brain damage, measured as distance from the average of the typical normal subgroups boxed in Figure 12. The line on each graph is a standard scatterplot smoother applied to the exposed subsample only. Clearly, there is no tendency for IQ to decrease with distance from the neuroanatomically normal in this composite shape space. Within the exposed group there is also no association of these shape discriminators with centroid size.

SUMMARY OF THE FINDINGS

The preceding discussion was a lengthy recount of analyses that are likely to be relatively unfamiliar to the reader. It may be helpful to summarize these findings before discussing their implications. We analyzed variations of shape, by sex, among normal subjects and subjects diagnosed with an alcohol-related disorder, either FAS or FAE. Two data structures were extracted from the same MRI: one of 33-point landmark locations and the other of 40-point representations of a callosal midline curve. Our principal findings are the following:

1. For both sexes, and for both shape representations, the exposed group has distinctly more shape variability than the normal group.
2. In the landmark point data, the excess variability seems to involve the brain-stem-to-diencephalon axis.
3. In the callosal outline data, both sexes have more variance among the exposed than among the normals, but the site of particular hypervariation differs by sex. The additional variance seems concentrated under the isthmus in males, but in the height of the arch for females.

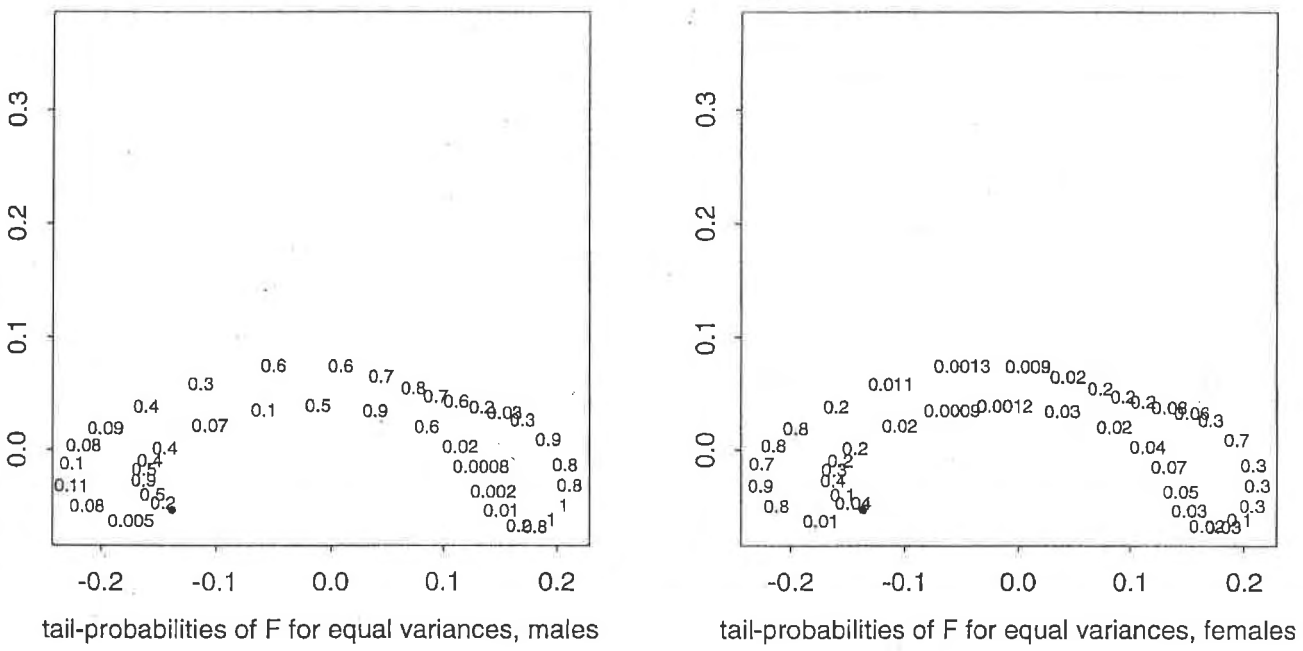


Fig. 10. *P*-values of F-tests for excess variance in the exposed subsample perpendicular to the average outline, semilandmark by semilandmark. For the males, the signal shown concentrates in the isthmus; for the females, in the arch.

4. The combination of these two localizations of variability supports a discrimination between the normals and the alcohol-affected with startlingly high sensitivity and specificity.
5. For both sexes, FAS and FAE each differ from the normal mean, but in neither sex do they differ much from each other. In particular, the discrimination of normals from exposed is not improved by knowledge of subdiagnosis.
6. Nor is the discrimination of normals from exposed improved by knowledge of IQ within this sample population.

DISCUSSION

Our alcohol-exposed subjects had been diagnosed before the onset of the study. Each was assigned either a diagnosis of FAS or what was then called FAE after examination by a dysmorphologist experienced in FAS. We combined these historical diagnostic records with data collected under a new methodology for quantifying brain differences between groups. Two types of data were involved: (1) landmark point configurations, restricted to brain regions where landmarks could easily be located (i.e., subcortical structures); and (2) closed curves in space, required for the analysis of the corpus callosum. The statistical analysis of their variability went forward separately, using closely related algebraic tools, and was then fused at the final stage of the analysis, the discrimination step.

The methodology weaving these data together involves three separate strategic decisions, each more or less unusual within our literature. The principal findings (as reviewed above) are expressed in terms of variances, not mean differences; they refer to shape variables instead of size variables; and the scientific goal they pursue is a discrimination, not a description. Interrelations among these thrusts are built into the methodology of shape coordinate analysis on which we have relied. For instance, the initial decision to examine shape variables more intensively than size variables made possible the important finding of the excess variability among the exposed—the shape features involved are not easily summarized in conventional measures of distance or area within parts of the form. In turn, the strength of the hypervariation underlies the surprisingly effective separation of affected subjects from normals conveyed in Figure 12, the sensitivity and specificity of which are so unexpected. Using either landmark point shape or outline shape, whichever is more practical for the tissue(s) at hand, the new toolkit supports studies of the affected parts of the brain and of the relationship among those parts at the same time and in the same computations. And however widely distributed the data in space, their statistical summary remains a unitary multivariate computation. We highlight the implications of these findings under five headings: the biotheoretical meaning of findings dealing with variability, the corpus callosum in ethanol teratogenesis, the literature of dysregulated brain stem/diencephalon relationships, the growing suspicion that FAS and FAE are the same clinical entities, and some methodological details.

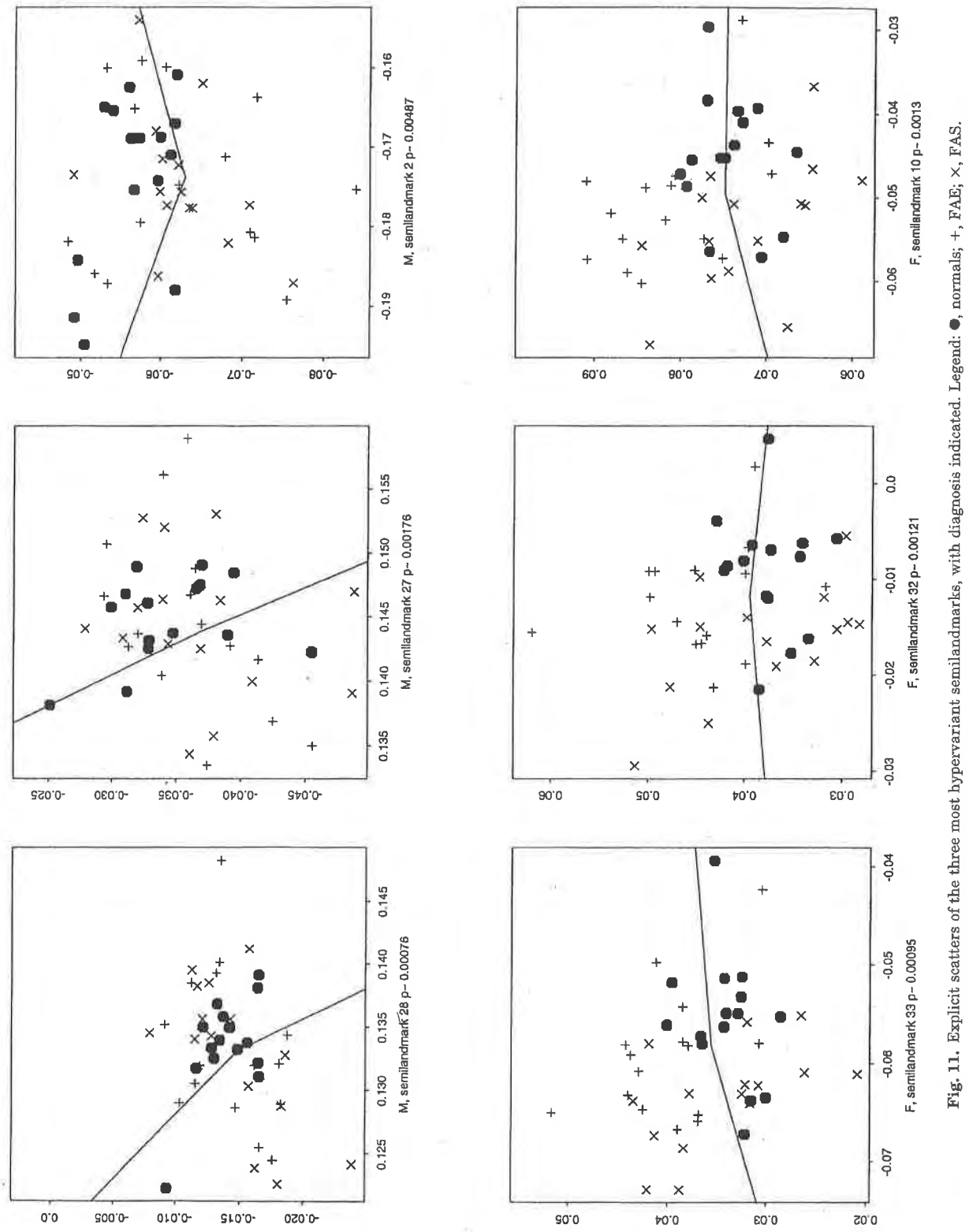


Fig. 11. Explicit scatters of the three most hypervariant semilandmarks, with diagnosis indicated. Legend: ●, normals; +, FAE; ×, FAS.

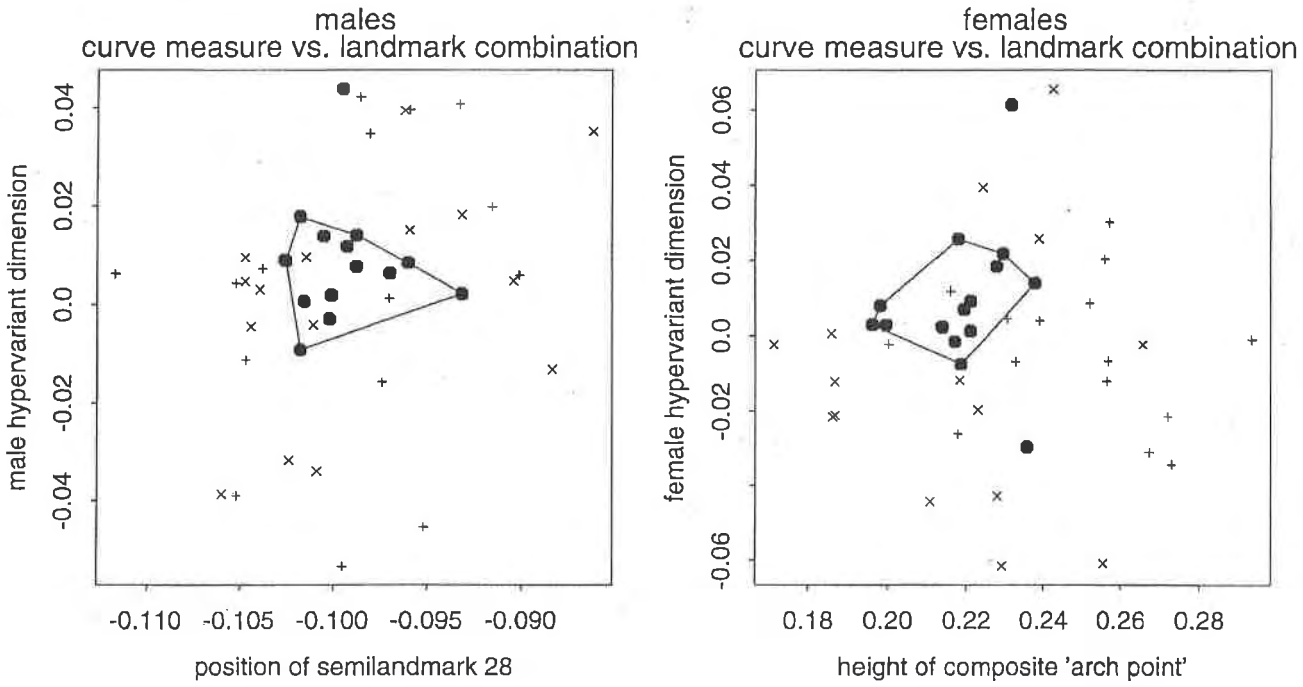


Fig. 12. Scatter of the most hypervariant landmark shape dimension against the selected hypervariant callosal shape coordinate, separately by sex. The great concentration of the normal subgroups (filled circles) is evident. +, FAE; ×, FAS.

Variability as a finding

It is not that the idea of considering differences of variance instead of differences in mean is unfamiliar to the quantitative biologist. It is present in most applied statistics texts under the heading of "verifying the assumptions" (of equal variances) required for conventional analyses of variance. In its essentials the comparison of two variances is a matter of the same F-ratios (variance ratios) that already underlie the usual tests for mean differences. What makes them unusually apposite for teratological applications such as these is their direct application to a specific high-dimensional vector of shape coordinates. In teratology, it is the malformations of shape, not those of size, that are inherently of greatest interest, and so a difference in variances need not be considered a nuisance interfering with inferences about means, but rather can be a powerful finding in its own right.

This finding is in accord with the now rather old literature concerning the origins of normal variation, the denominators of all the variance ratios we have been finding significant in the present study. For a classic overview of this topic, see Chapter 3, "Anatomical Variations—Significance," in Williams ('56). This chapter reviews the literature (through 1956) on quantitative and qualitative variations of structure in animals and humans, concluding, unsurprisingly, that great variation from one normal individual to another can be found in structures in all of the body's physiological systems, and that, in particular, "the brain is extremely variable in every character that has been subject to measurement" (an observation the author attributes originally to Karl Lashley). Williams goes on: "Few studies are available concerning the structural variations in human brain tissue, considering the possible importance of their relation to behavior. Virtually nothing is known about disharmonies of development in the central nervous system except for very gross deficiencies." We know of no review over the intervening 44 years that would substantially alter this summary; the last monograph on methods for the multivariate study of such variations as might arise, Olson and Miller's *Morphological Integration* ('58), also dates from the 1950s. Regarding current methodological fashions, it is worth noting that the index of one popular current reference on psychopathology (Harris, '98) does not include any citations to the use of the term "variability" at all.

All the more startling, therefore, that our study of one abnormal developmental pathway, having the known cause of alcohol teratogenesis, has uncovered two distinct neuroanatomical configurations—the brain stem/diencephalon axis and the quantitative shape of the midline corpus callosum—for which the "disharmonies of development" prove so spatially specific. The finding is a variance ratio; thus, the signal it betokens might pertain either to its denominator (unusual invariance of form in the normally developing embryo) or to its numerator (a disregulation of the normal process in the same embryo that has been alcohol-affected). The former interpretation, the narrowing of range of the most crucial aspects of embryogen-

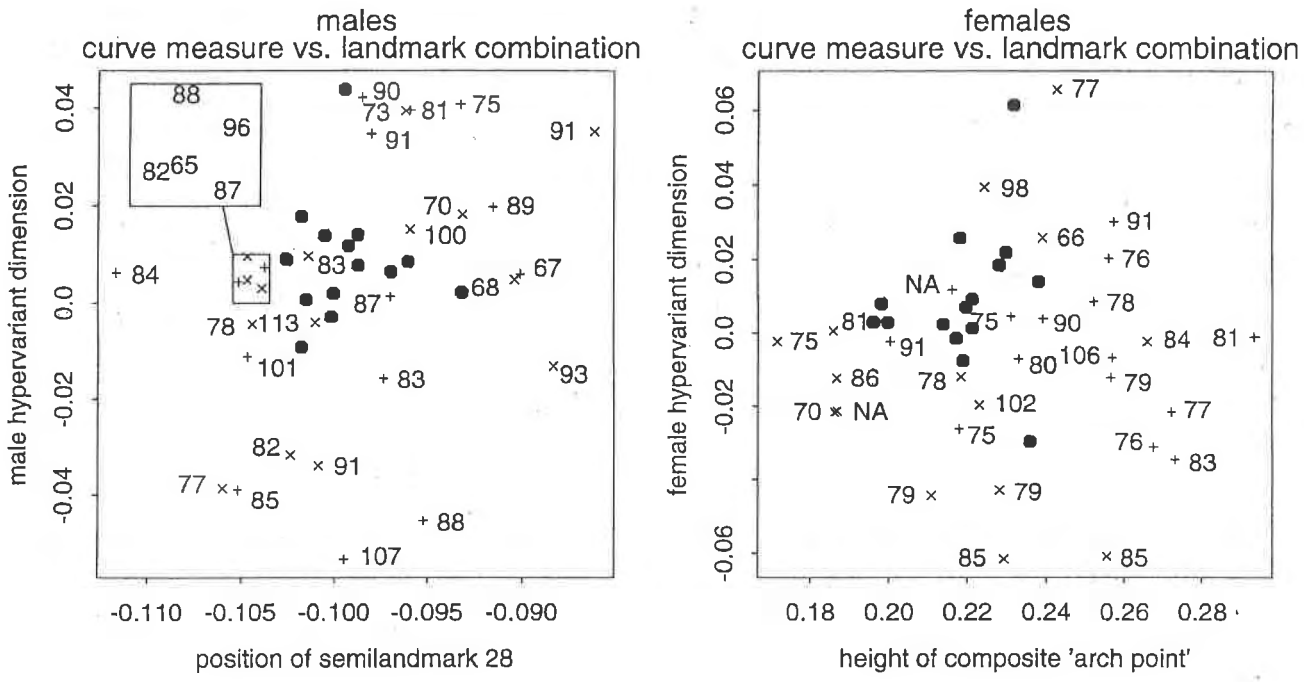


Fig. 13. Enhancement of Fig. 12 by full-scale IQ for the exposed subsamples. NA, score missing (unobtainable).

esis, is the more fruitful theoretically, as it corresponds to Waddington's classic notion of *developmental canalization,* whereby individuals attain a common endpoint of anatomical form despite variation in ontogenetic conditions. The callosum finding has precisely the logical structure of a disrupted canalization. Rhetorically, a "disruption" lacks the implication of a polarity that would otherwise be borne in words such as "deficit"; indeed, in the hypervariation characterizing the present findings, there is no particular direction in which the exposed group has been substantially shifted with respect to the controls—no "direction of deficit," such as would underlie a linear discriminant function; instead, there is a calibration of similarity versus dissimilarity to the normal, a contrast of typical with atypical in all directions of potential shape defect, not just one preferred direction.

The corpus callosum finding

Development of the human corpus callosum begins on about the 39th postconception day with differentiation of the commissural plate, and callosal fibers differentiate from that plate at about the 74th day, achieving adult morphology by day 115 (Loeser and Alvord, '68). It is therefore reasonable, that damage secondary to prenatal alcohol exposure would result principally from exposure within the first trimester of pregnancy, as is the case for the general "latent brain damage" observed by detailed measurement of behavior in large human samples (Streissguth et al., '93). Indeed, partial or total agenesis of the corpus callosum has been noted before in subjects diagnosed with FAS (Riley et al., '95; Swayze et al., '97)—this observation was the reason we chose to measure callosal outline for the purposes of the present study. We had no cases of diagnosable agenesis in this adult sample. If agenesis itself is a direct consequence of exposure, it is likely to be engendered only at levels of exposure that are higher than those that typify the subjects of the present study (who, the reader will recall, had to be capable of undergoing a 5-hr battery of neurobehavioral tests in addition to the MR acquisition generating the data analyzed in the present study). The callosum typically develops in a rostrocaudal direction (Schaefer et al., '90), and partial agenesis is usually observed posteriorly, but we do not see any concentration of size difference or of variance differences at either end of the callosum (cf. Fig. 7).

Inasmuch as all the patients in this study were diagnosed with alcohol damage, the data set affords no contrasts speaking to any specificity of the findings. Indeed, more serious prenatal insults, such as spina bifida, can entail partial callosal agenesis at considerably greater rates than are found in fetal alcohol populations. But the literature suggests no reason to suspect hypervariance of callosal form in any milder syndrome. We would welcome comparative studies on this theme, especially in other patient groups characterized by attention disorders (cf. Banich, '98).

Although Riley et al. ('95) make no explicit reference to callosal shape variability per se, their Figure 4 includes standard deviations of "proportional area" (areas of five sectors of the callosal outline as a percentage of their total), and thus lets us interpret their data set

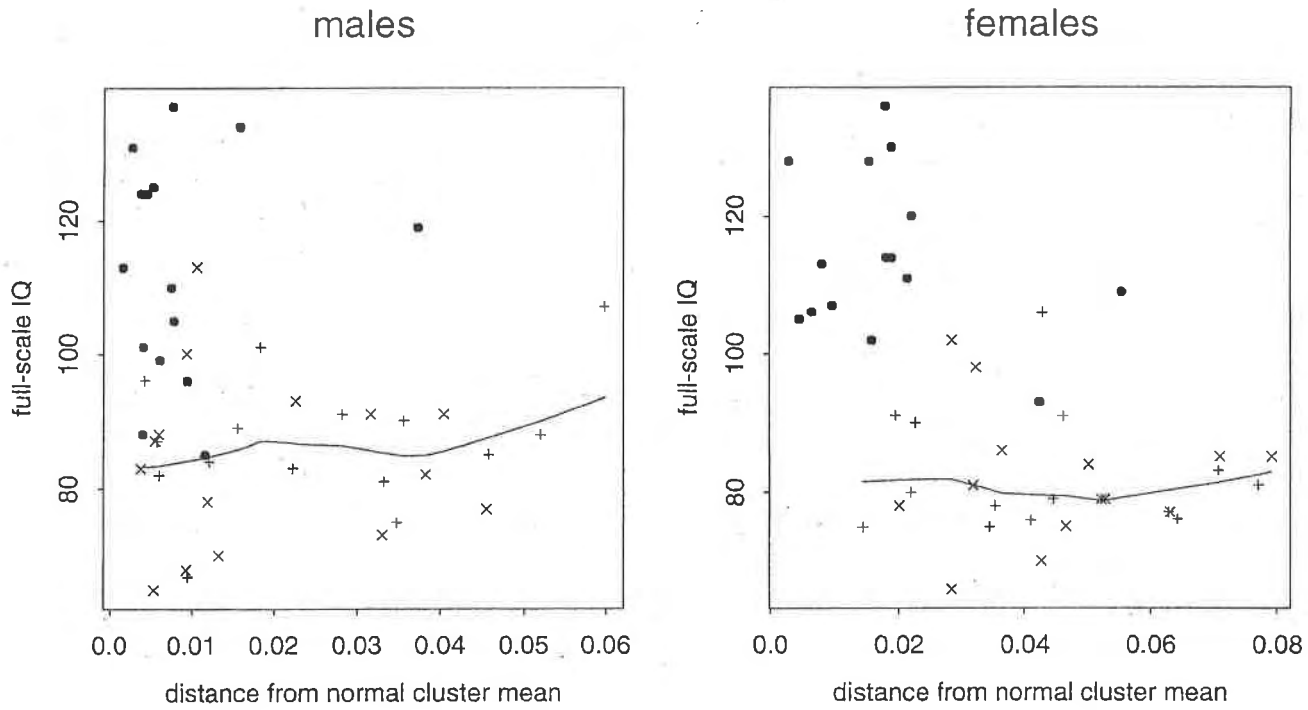


Fig. 14. Within this sample of exposed subjects, none profoundly retarded, atypicality of neuroanatomical form greatly distinguishes them from the normal subgroup but nevertheless is not associated with either the facial features of the full FAS syndrome or the measured full-scale IQ. Broken line, lowess scatterplot smoother (for the exposed subsample only). ●, normals; +, FAE; ×, FAS.

in the light of our present findings. The coefficient of variation of all five of the proportional areas is obviously significantly larger in their exposed subgroup than their unexposed subgroup, "preconfirming" our analysis. Apparently more discrepant is one recent study reporting an increase in callosal size consequent to exposure. Using a binge model in macaques, Miller et al. ('99) found that the rostral half (only) of the exposed callosa increased in size by comparison with unexposed controls. Operationally, the separation between rostral and caudal halves of the callosum is at the divergence of the fornix. This is perhaps an unfortunate strategy for group comparisons in which that point itself may have shifted along the callosal outline. In our human sample, hypervariance is concentrated at that point, especially in females (Fig. 4). Note also that, according to Table 3 of their publication, callosal size is clearly hypervariable in the exposed groups. Hence the Miller finding, based as it is on a total of only 15 animals over three groups, could well be an expression of hypervariance rather than the mean shift reported by those authors.

The brain stem/diencephalon finding

Brain stem anomalies have been found in a number of developmental syndromes, often teratogenetic. In one autopsy of an autistic patient, a shortening of the brain stem was noted, an outcome arising in animal models including Hoxa-I gene knockout and exposure to antimitotics or to thalidomide (Rodier et al., '96, '97). A mouse model of holoprosencephaly (Lanoue et al., '97) shows abnormalities of mid- and hindbrain structure. Retinoic acid early in pregnancy in mice produces Arnold-Chiari malformations (Alles and Sulik, '92), including herniation of the hindbrain, owing to the primary damage to the neural crest and rhombencephalon. Similar brain stem dysmorphology is often observed after alcohol exposure in laboratory animals (Maier et al., '99; Thomas et al., '96). Studies of prenatally alcohol-exposed rats have linked brain stem damage of this sort to some of the behavioral deficits that characterize FAS/FAE in humans: decreased brain stem weight with motor deficits (Thomas et al., '96), or brain stem damage with auditory processing (Church et al., '96)—for the human analogue, see Pettigrew and Hutchinson ('84). But these studies did not pursue the issue of geometric shape as it has been formalized in the analyses presented in this article.

FAS and FAE do not seem to differ neurologically

Studies of neurobehavioral teratology in animals have demonstrated repeatedly that both brain dysmorphogenesis and behavioral dysfunction occur in offspring prenatally exposed to alcohol even in the absence of dysmorphic facial features, limb anomalies, or growth deficiency (Goodlett and West, '91; Riley et al., '90; Means et al., '88). For more than a decade, first qualitative and, recently, quantitative evidence for the absence of clear behavioral distinctions between pa-

tients with FAS and those with FAE has been accruing from several research sites (Streissguth et al., '91; Kodituwakku et al., '95; Aronson, '97; Mattson et al., '97; '98; Mattson et al., '99; Autti-Rämö, '00). Previous efforts to detect brain anomalies in patients with FAS compared with FAE (Clark et al., '00; Mattson et al., '97) have been straitjacketed by their reliance on size and volume assessments, which are irrevocably confounded with IQ levels, and thus with diagnosis. (Alcohol-exposed children who are mentally retarded are far more likely to be diagnosed with FAS than are those who are not; see Sampson et al., '00). The sharp focus on shape variation here more effectively quantitates the intricacies of brain maldevelopment observed qualitatively by teratologists and embryologists for decades. That the diagnosis of FAS cannot be distinguished from that of FAE from solid brain MRI (as quantified in the way we have done in the present study) bears enormous implications for the clinical course of these patients, inasmuch as brain dysmorphology is the center of the prenatal teratogenetic effect of alcohol. Diagnoses that can take these into consideration are thus likely to improve the delivery of appropriate social services, as they will be more closely tailored to the actual neuroteratological basis for prognosis in this class of patients.

Morphometric data

Of the two morphometric data channels combined in this article, that of landmark point locations is the less unfamiliar. The general issue of homology among points of the cortex proper (e.g., points on gyri or sulci, points on the gray–white boundary) is currently the subject of intense debate among many research groups (see, e.g., Toga, '99). Thus far there is no methodology for judging the comparative merits or demerits of suggestions from this class: for instance, there is not yet any general agreement that Talairach's ('88) original clinical suggestion of an AC–PC registration is demonstrably wrong on scientific grounds, let alone agreement on the reasons why it should be considered wrong (cf. Toga, '99). The registration here, using the machinery of Procrustes shape coordinates, arises from a justification that is mathematical, not necessarily empirical: it is the optimal way of visualizing shape variations of landmark configurations in general. In empirical applications, analyses that assign correspondences to points stand or fall on the covariations of the ensuing shapes or sizes with known causes or consequences of form, such as the fetal alcohol spectrum diagnoses in the present study.

In this connection, we had originally included, very tentatively, three other landmarks that did not make it into the final list of 33 (Table 2). The "bottom of the top of the cingulate sulcus"—a midline landmark locating the base of the cingulate sulcus where it turns laterally along the superior aspect of the brain—proved intolerably unreliable across raters. It is typical, we now believe, of landmark structures that are truly curves in space, to be represented as such for statistical purposes, but we restricted the present investigation to just one such curve, the callosum, because of the clear implication of prenatal alcohol in callosal anomalies already noted in the literature.

The posterior analogues of landmarks Fr–r and Fr–l, tips of the frontal horns of the ventricle, would be the two matching tips of the occipital horns. These could indeed be localized reliably on the individual subject, usually by following the narrow crevices of CSF posteriorly to some extent. However, the points thus arrived at appear to be bimodally distributed, with one tentative location about 2 cm posterior to the other. The appearance of the landmark in one or the other of these positions proved not to be associated with diagnosis, sex, or, in many cases, the contralateral position. In fact, this "tip" is a pair of landmarks, one of which was missing on each side for each subject. We omitted these landmarks because, taken individually, they were systematically missing in this way. Otherwise, the reader may have noted (by the absence of a footnote to the contrary in Table 2) that there are no missing landmark data across the full collection of 33 points here for 90 subjects. We would welcome the attempts of others to extend Table 2 to include cortical points that could be characterized on the same principles.

The corpus callosum outline data set here, which follows the midline of that structure as it gently twists in the vicinity of the putative midsagittal plane, is not itself a plane curve. As compared with the more usual technique of callosal visualization, which extracts an outline from one single image plane, the twisting leaves measures of area, both total and sectoral, relatively unaltered, but greatly affects the assessment of variability of actual locations at small scale. As we have seen, the areal finding is weaker than that for curving outline shape as regards mean differences by diagnosis, and offers no equivalent of the techniques for localizing shape hypervariability that constitute the principal signal we have found. (For more on contemporary strategies for callosal morphometrics, see Bookstein, '00.)

ACKNOWLEDGMENTS

The authors have been generously supported by NIH grants AA-10836 (to A.P.S.) for the project as a whole and GM-37251 (to F.L.B.) for the development of analytical methods. Malay Chan digitized thousands of landmark points with extraordinary care. David Haynor, M.D., University of Washington, supplied the clinical reviews of all the subjects' MR images. Julia Kogan and Lisette Austin managed the assembly of a balanced set of six study groups subject to constraints on imaging and testing staff and facilities. Data were digitized in Edgewarp, a flexible package for navigating solid medical images. A manual can be found at ftp://brainmap.med.umich.edu/pub/EWSH3.1. Helen Barr has supported us enthusiastically throughout the research that led to this article, and has helped greatly to clarify several underlying points of exposition.

APPENDIX: MORE ON PROCRUSTES METHODS

This Appendix expands on the terse summary of multivariate morphometric tools sketched under the Morphometric Methods section. Four main notions are explained in detail: Procrustes shape distance, Procrustes shape averages, Procrustes shape coordinates, and relative warps. The presentation is limited to the simplest case, landmark points in two dimensions, but the extensions to 3D data and to semilandmarks (points on curves) are straightforward. Sources that treat all this material at greater length include Bookstein ('96, '97a, '98) and Dryden and Mardia ('98). These particular characterizations of the main technical terms were worked out during the mid-1990s. A useful on-line glossary (Slice et al., '95) expands on their interrelationships, which represent a consensus among a community of several toolbuilders, along with a variety of other terms central to earlier applications of these methods.

Procrustes shape distance

To carry out multivariate analysis of the "shape" of a data set of k-landmark point configurations, it is sufficient to have a distance measure between the two shapes that obeys the usual rules. Suppose we have two landmark configurations, that is to say, two sets X_1, X_2 of k points with coordinates (x_{1i}, y_{1i}), $i = 1, \ldots, k$, for the first form and (x_{2i}, y_{2i}) for the second. If we were talking about "location" rather than shape, a reasonable notion of squared distance between the two would just be the usual Pythagorean sum

$$\sum_{i=1}^{k} [(x_{1i} - x_{2i})^2 + (y_{1i} - y_{2i})^2]$$

of all squared coordinate differences between the positions of corresponding landmarks in the two forms.

We need to adapt this formula so that it gives the same answer whenever either of the two shapes is moved, rotated, or rescaled: then it will be talking about shape, as we want it to, rather than merely about locations in the original digitizing planes. It turns out best if we reformulate the problem in a way that turns out to reduce to just this Euclidean formula under certain conditions. We circumvent the problem of change of position by not allowing position to vary. Each form X_1 or X_2 has to be put down with its coordinates centered at (0, 0)—that is, we subtract $\sum_{i=1}^{k} x_{1i}/k$ from each x_{1i}, and similarly for y_{1i}, x_{2i}, and y_{2i}. Geometrically, the effect of this is just a shift of the origin of coordinates of each form, leaving its shape, as well as its size, alone. Also, we circumvent change of scale by likewise not allowing scale to vary: replacing each set of centered coordinates with a new set chosen so that the sum of their squared distances from the origin (0, 0) of coordinates, which is now also their centroid, is exactly 1. We do this by dividing each centered form by a suitable scale factor, namely, the square root of whatever that sum of squares was before this operation. This factor, called centroid size, is examined by group in Table 3 (see text).

With position and scale both standardized, that leaves rotation. Just as we have repositioned each form independently, and rescaled each form independently, we could imagine having to rotate each form independently to some arbitrary horizontal or vertical. But that is not the way the method actually goes. It turns out a much better idea to state instead that we would like the result of adjusting rotation to give us back the Euclidean sum of squares corresponding to the rotated points. When two sets of points are both rotated by the same angle, the sum of squared distances doesn't change at all, and so what is needed is the relative rotation of one form with respect to the other that leads to the least Euclidean sum of squares out of all possible relative rotations. One version of Procrustes shape distance is then defined as the square root of this sum of squares when the relative rotation is chosen appropriately. Other versions of the definition differ from this one only by inconsequential adjustments, like using the arcsine of a small number in place of the number.

The steps in this computation follow down the rows of Figure A1. At the top are two quadrilaterals of landmarks presumed to arise from real images. Erase whatever outline information goes with these landmarks, but treat them purely as configurations of disconnected points. Then connect each landmark to the centroid of its own form. Its centroid size is the square root of the sum of squares of these lengths. For each form, rescale the sum of squares of the distances shown to unity (second row) by dividing by this centroid size. Next (third row, left) translate one of the forms so that its centroid directly overlies the centroid of the other form. Finally, identify the rotation (third row right) that minimizes the sum of squares of the residual distances between matched landmarks. The squared Procrustes distance between the forms (fourth row, redrawn with their own outlines back in) is (to a very good approximation) the sum of squares of those residuals at this minimum: total area of the circles at lower right, divided by π.

Procrustes average shape

To this point, there is now a formalism for computing a shape distance measure between any two landmark configurations, but not, as yet, any way to add or subtract them, so that we can't define the average of a shape data set as its sum divided by its count, the way we do for vectors—at least, not yet. Turn instead to a different characterization of the same idea: the average shape as the shape about which the specimens of a data set have the least sum of squared Procrustes distances. Although it may seem to you that this definition is somehow both vacuous and circular, in fact it is perfectly rigorous mathematically. Furthermore, for any landmark data set you will ever encounter in practice, the (unique) average shape can be computed by the

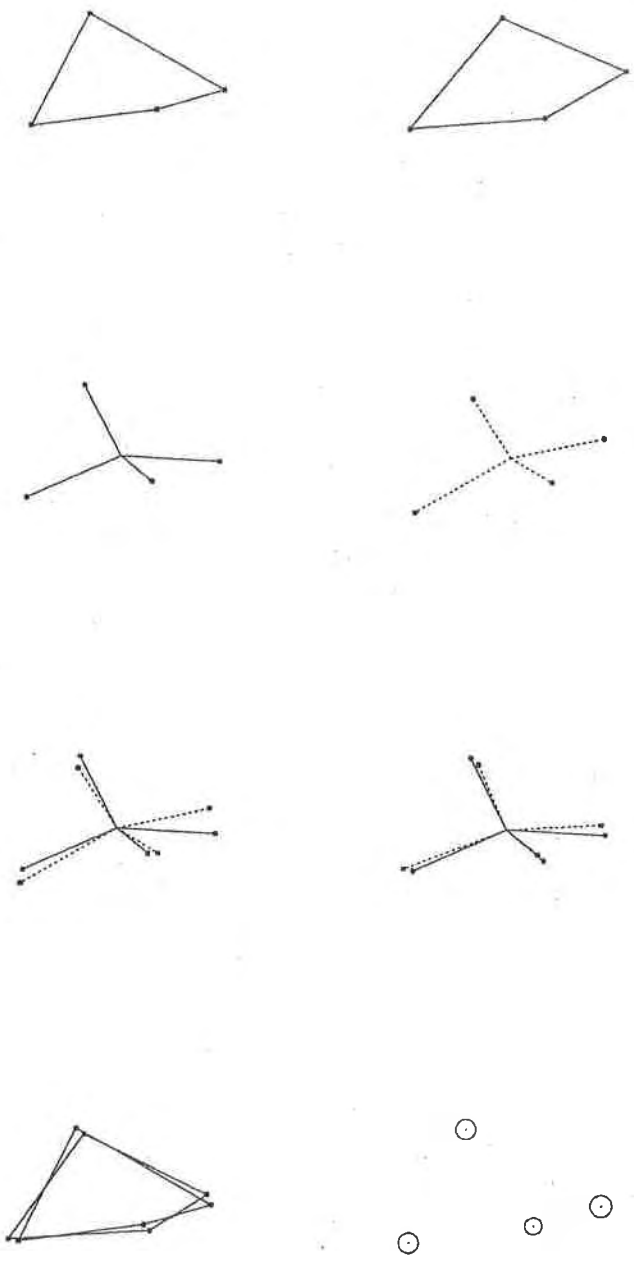


Fig. A1. The Procrustes superposition for a pair of forms. (**top row**) Two forms of four homologous landmarks. (**second row**) Each form is rescaled so that the sum of squares of the distances to the centroid of its four landmarks is 1. This is the sum of squares of the four lines shown. (**third row**) The centroids are superposed, and then one form is rotated over the other so that the sum of squared distances between corresponding landmarks is a minimum. (**fourth row**) With the construction lines erased, the squared Procrustes distance between the pair of forms is that sum of squared distances. It is proportional to the total area of the circles drawn at lower right.

iterative algorithm sketched in Figure A2 for a sample of four four-landmark shapes.

The top row displays some "raw data": four quadrilaterals of similar but not identical shapes. Guess at the value of the average shape—not a wild guess, but something in the vicinity of the real data: for instance, guess that the average shape is the same as the shape of the first specimen. Then (second row) superimpose each of the original forms (including this form) over this guessed average in exactly the posture implied by Figure A1—the position in which that Euclidean sum of squared distances is smallest.

The next guess for the average shape is made up of the averages of the positions of the landmarks, one by one, after they are superimposed over the previous guess in this way. The third row of Figure A2 shows what happens when you update the candidate for average shape and go through the cycle of fits again: you get the same average, to the accuracy of the printer's dots in Figure A2. In nearly every real data set, this straightforward iterative algorithm converges to adequate precision by the end of the second iteration. To keep the figures from getting smaller row after row, it is also customary to rescale each candidate average to centroid size 1 before diving back into the second row, the refitting procedure.

Procrustes shape coordinates

The last panel in Figure A2 shows two concepts: the "Procrustes average shape" we sought, and copies of all the original forms scattered around it, each Procrustes-fitted to their common average. This composite image is the crux of the value of the whole Procrustes toolkit. By definition of Procrustes distance, the sum of squared distances of the shape coordinates of each original shape from those of the Procrustes average shape is its squared Procrustes distance from that average. But also, the sum of squared distances between the final positions of the landmarks of any two of the original forms is also *their* squared Procrustes distance.

For Figure A3, erase everything except the little scatters around the average at lower right in Figure A2, and put a separate coordinate system down centered at each averaged landmark position in turn. We thereby arrive at an exact analogue of one of the two great customary ways of setting up a multivariate statistical analysis, the approach usually called principal coordinates analysis, beginning with sums of squared distances instead of values of variables. Readers familiar with factor analysis will recognize this under the name of R-mode analysis. Beginning with distances, we have arrived (in an essentially unique way) at an equivalent set of $2k$ ordinary variables.

These are the Procrustes shape coordinates, which represent all the information in the shape of the original sets of landmarks for any linear multivariate statistical purpose. Any question about the correlation of shape with its causes or effects can be answered by using this single set of coordinates as a "vector of shape variables" in the corresponding standard multivariate procedure. For instance, to talk about averages of these coordinates by subgroups of the sample (e.g., the three diagnostic groups of this paper), it is sufficient to average their Procrustes shape coordinates, which produced the locations plotted in Figures 2 and 7. Our permutation tests for significance of these differences

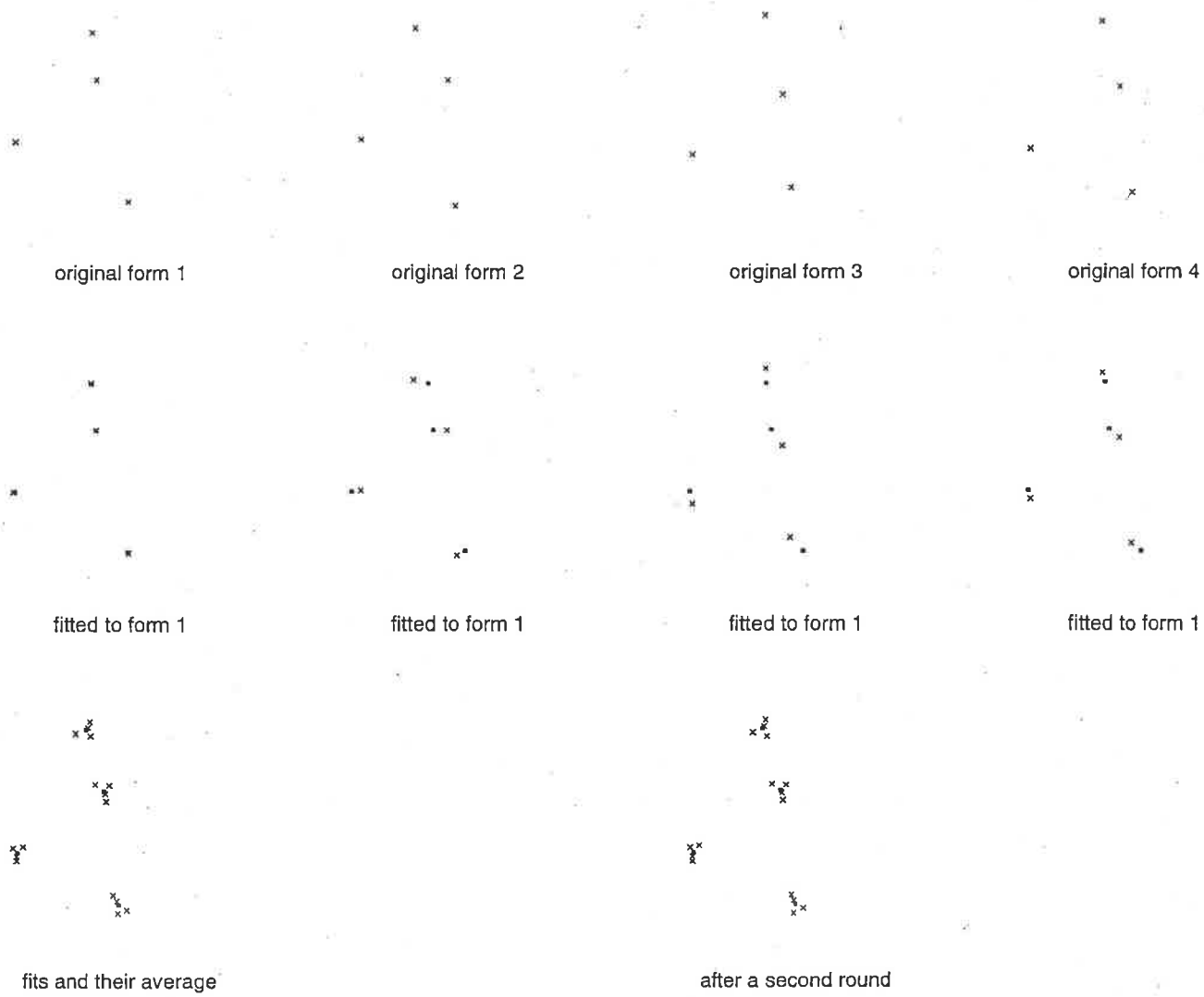


Fig. A2. Procrustes averaging and Procrustes shape coordinates. **Top row:** four forms of four landmarks. **Middle row:** Procrustes fit of each (X's) to an arbitrary starting guess (dots: the first form). **Bottom left:** the next estimate of the average (dots) is the average of the fitted locations from the previous step. **Bottom right:** a second round of fits and averages changes it hardly at all—the algorithm seems to have converged already.

computed exactly analogous averages over pseudogroups rather than real groups. Correlations of shape with its causes or consequences, such as IQ, likewise proceed coordinate by coordinate in this representation, and are tested by permutation procedures using explained and unexplained squared Procrustes distances just as for variables arrived at by direct observation in the ordinary way.

Relative warps

Finally, a useful factor analysis of shapes is one version of ordinary principal components of these same Procrustes shape coordinates. Instead of starting from their correlation matrix, the usual procedure in most branches of applied biometrics, one works with their covariance matrix (an option available under the name "unscaled" in most packages). The reason for this is, at root, a magnificent mathematical elegance underlying the entire Procrustes toolkit. When data arise on a model of wholly random digitizing error around the same true landmark locations—digitizing error that is the same at every landmark and in every direction, what is called isotropic noise—the resulting distribution of Procrustes shape coordinates has extraordinary symmetries, regardless of what that average form was. Specifically, under that strong isotropic assumption (which is not far from applying to many real data sets once systematic factors of form are regressed out), the theoretical distribution of Procrustes shape coordinates necessarily has $2k - 4$ dimensions of exactly the same variance, and a final 4 dimensions of no variance at all, regardless of the average shape.

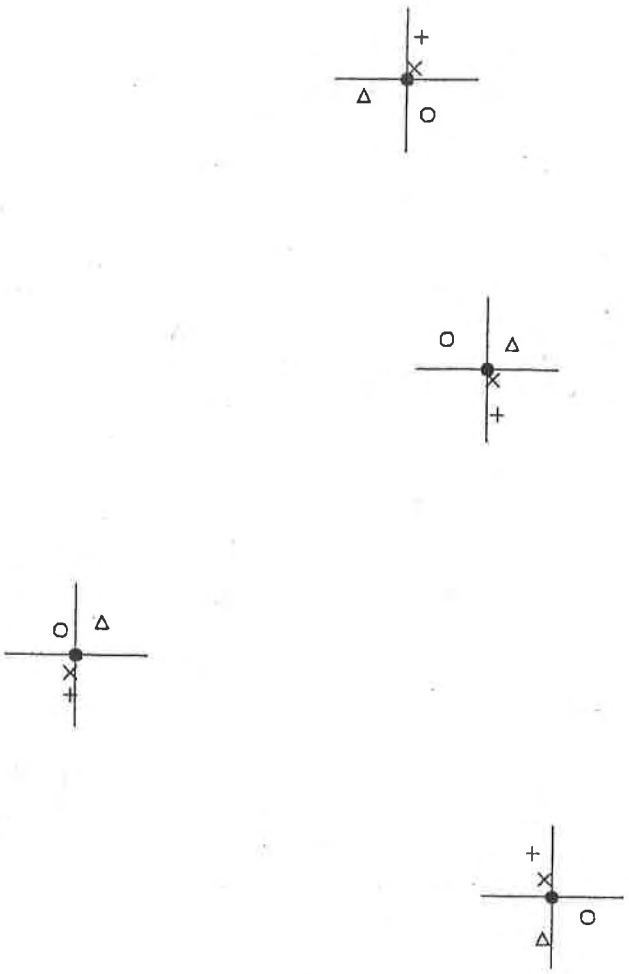


Fig. A3. Procrustes shape coordinates (now indicated in different symbols for each form) are the deviations of fitted landmarks from the average shape (dots) at the convergence of the algorithm in Fig. A2. There are a total of eight of these coordinates: two each for the four landmarks of the data set. The sum of squared differences between the values of the coordinates in any pair of specimens equals the squared Procrustes shape distance between the two specimens as originally digitized; that is, these coordinates serve as a set of principal coordinates for the Procrustes shape distance data (see text).

Because this is important, we say it another way as well. No matter what the average form looks like, if data arise from it by uninformative noise, the probability distribution of all those Procrustes shape coordinates is pretty much proportional to e^{-cPD^2}, where c is a suitable precision-like constant that takes into account the centroid size of the "true picture" as well as the amplitude of digitizing noise, and PD^2 is the squared Procrustes distance of any digitized form from the true average. If a spherical covariance matrix stands for no information, one that is not spherical stands for exactly the kind of information at which a principal components analysis is aimed. Principal components of Procrustes shape coordinates (under the covariance-matrix option) represent precisely the dimensions of shape variability that have the highest variance "per unit Procrustes distance" just as principal components of ordinary lists of variables have the highest variance "for unit sum of squared coefficients," and those dimensions of extra variance help us ordinate data distributions with the greatest efficiency just as do scatterplots of factor scores in most other applications.

You are probably used to seeing such components emerge from packages only in tabular form—columns of coefficients, one for each component, headed by its eigenvalue ("explained variance"). For shape coordinates, the corresponding tables are immediately converted to geometric diagrams showing how the points move away from the average shape, landmark by landmark, in strongly or weakly correlated ways (depending on the magnitude of the analogous eigenvalues). Specialized principal components of this sort, restricted to the covariance matrix of shape coordinates, are called relative warps because these displacements are usually drawn out in turn by images of deformed (warped) Cartesian grids. Each of these graphical styles is exploited in Figure 5.

LITERATURE CITED

Aase JM. 1990. Diagnostic dysmorphology. New York: Plenum Press.

Alles AJ, Sulik KK. 1992. Pathogenesis of retinoid-induced hindbrain malformations in an experimental model. Clin Dysmorphol 1:187–200.

Aronson M. 1997. Children of alcoholic mothers: results from Göteborg, Sweden. In: Streissguth AP, Kanter J, editors. The challenge of fetal alcohol syndrome: Overcoming secondary disabilities. Seattle, WA: University of Washington Press. pp. 15–24.

Astley SJ, Clarren SK. 2000. Diagnosing the full spectrum of fetal alcohol exposed individuals: introducing the 4-Digit Diagnostic Code. Alcohol Alcohol 35:400–410.

Autti-Rämö I. 2000. Twelve-year follow-up of children exposed to alcohol in utero. Dev Med Child Neurol 42:406–411.

Banich MT. 1998. The missing link: The role of interhemispheric interaction in attentional processing. Brain Cognition 36:128–157.

Bookstein FL. 1991. Morphometric tools for landmark data: geometry and biology. New York: Cambridge University Press. 455 p.

Bookstein FL. 1996. Biometrics, biomathematics, and the morphometric synthesis. Bull Math Biol 58:313–365.

Bookstein FL. 1997a. Shape and the information in medical images: a decade of the morphometric synthesis. Computer Vision Image Understand 66:97–118.

Bookstein FL. 1997b. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. Med Image Anal 1:225–243.

Bookstein FL. 1998. A hundred years of morphometrics. Acta Zool Hung 44:7–59.

Bookstein FL. 1999. Linear methods for nonlinear maps: Procrustes fits, thin-plate splines, and the biometric analysis of shape variability. In: A. Toga, editor. Brain warping. San Diego, CA: Academic Press. pp. 157–181.

Bookstein FL. 2000. Morphometrics for callosal shape studies. In: Zaidel E, Iacoboni M, editors. The parallel brain: cognitive neuroscience of the corpus callosum. Cambridge, MA: MIT Press. In press.

Bookstein FL, Green WDK. 1998. Edgewarp 3D: A preliminary manual. Posted to the internet as ftp://brainmap.med.umich.edu/pub/edgewarp3.1/manual.html.

Church MW, Abel EL, Kaltenbach JA, Overbeck GW. 1996. Effects of prenatal alcohol exposure and aging on auditory function in the rat: preliminary results. Alcoholism 20:172–179.

Clark CM, Li D, Conry J, Conry R, Loock C. 2000. Structural and functional brain integrity of Fetal Alcohol Syndrome in nonretarded cases. Pediatrics 105:1096–1099.

Clarren SK, Smith DW. 1978. The fetal alcohol syndrome. N Engl J Med 198:1063–1067.

Dehaene P, Titran M, Samaille-Villette C, Samaille P, Crépin G, Delahousse G, Walbaum R, Fasquelle P. 1977. Fréquence du syndrome d'alcoolisme foetal. Nouv Presse Med 6:1763.

Dehaene P, Walbaum R, Titran M, Samaille-Villette Ch, Samaille PP, Crépin G, Delahousse G, Decocq J, Delcroix M, Caquant F, Querleu D. 1977. La déscendance des mères alcooliques chroniques: À propos de 16 cas d'alcoolisme foetal. Rev Fr Gynecol Obstet 72:491–498.

Dryden IL, Mardia KV. 1998. Statistical shape analysis. New York: John Wiley & Sons.

Evans AC, Dai W, Collins L, Neelin P, Marrett T. 1991. Warping of a computerized 3-D atlas to match brain image volumes for quantitative neuroanatomical and functional analysis. SPIE Proc 1445: 236–247.

Good PI. 1994. Permutation tests. New York: Springer-Verlag.

Goodlett CR, West JR. 1992. Fetal alcohol effects: rat model of alcohol exposure during the brain growth spurt. In: Zagon IS, Slotkin TA, editors. Maternal substance abuse and the developing nervous system. San Diego, CA: Academic Press. pp. 45–75.

Guerri C. 1998. Neuroanatomical and neurophysiological mechanisms involved in central nervous system dysfunctions induced by prenatal alcohol exposure. Alcoholism 22:304–312.

Hanson JW, Streissguth AP, Smith DW. 1978. The effects of moderate alcohol consumption during pregnancy on fetal growth and morphogenesis. J Pediatr 92:457–460.

Harris JC. 1998. Developmental neuropsychiatry. New York: Oxford University Press.

Jones KL. 1986. Fetal alcohol syndrome. Pediatr Rev 8:122–126.

Jones KL, Smith DW. 1973. Recognition of the fetal alcohol syndrome in early infancy. Lancet 2:999–1001.

Jones KL, Smith DW. 1975. The fetal alcohol syndrome. Teratology 12:1–10.

Jones KL, Smith DW, Ulleland CN, Streissguth AP. 1973. Pattern of malformation in offspring of chronic alcoholic mothers. Lancet 1:1267–1271.

Kendall DG. 1984. Shape-manifolds, procrustean metrics, and complex projective spaces. Bull Lond Math Soc 16:81–121.

Kodituwakku PW, Handmaker NS, Cutler SK, Weathersby EK, Handmaker SD. 1995. Specific impairments in self-regulation in children exposed to alcohol prenatally. Alcoholism 19:1558–1564.

Landesman-Dwyer S, Keller LS, Streissguth AP. 1978. Naturalistic observations of newborns: effects of maternal alcohol intake. Alcoholism 2:171–177.

Lanoue L, Dehart DB, Hindsale ME, Maeda N, Tint GS, Sulik KK. 1997. Limb, genital, CNS, and facial malformations result from gene/environment-induced cholesterol deficiency: further evidence for a link to sonic hedgehog. Am J Med Genet 73:24–31.

Lemoine P, Harousseau H, Borteyru J-P, Menuet J-C. 1968. Les enfants de parents alcooliques: anomalies observées, à propos de 127 cas. Ouest Med (Paris) 21:476–482.

Loeser JD, Alvord EC. 1968. Agenesis of the corpus callosum. Brain 91:553–570.

Maier SE, Miller JA, West JR. 1999. Prenatal binge-like alcohol exposure in the rat results in region-specific deficits in brain growth. Neurotoxicol Teratol 21:285–291.

Majewski F, Bierich JR, Löser H, Michaelis R, Leiber B, Bettecken F. 1976. Zur Klinik und Pathogenese der Alkohol-Embryopathie: Bericht über 68 Fälle. Munch Med Wochenschr 118:1635–1642.

Martin JC, Martin DC, Lund CA, Streissguth AP. 1977. Maternal alcohol ingestion and cigarette smoking and their effects on newborn conditioning. Alcoholism 1:243–247.

Mattson SN, Riley EP, Sowell ER, Jernigan DF, Sobel DF, Jones KL. 1996. A decrease in size of the basal ganglia in children with fetal alcohol syndrome. Alcoholism 20:1088–1093.

Mattson SN, Riley EP, Gramling L, Delis DC, Jones KL. 1997. Heavy prenatal alcohol exposure with or without physical features of fetal alcohol syndrome leads to IQ deficits. J Pediatr 131:718–721.

Mattson SN, Riley EP, Gramling L, Delis DC, Jones KL. 1998. Neuropsychological comparison of alcohol-exposed children with or without physical features of fetal alcohol syndrome. Neuropsychology 12:146–153.

Mattson SN, Goodman AM, Canine C, Delis DC, Riley EP. 1999. Executive functioning in children with heavy prenatal alcohol exposure. Alcoholism 23:1808–1815.

Means LW, Burnette MA, Pennington SN. 1988. The effect of embryonic ethanol exposure on detour learning in the chick. Alcohol 5:305–308.

Miller MW, Astley SJ, Clarren SK. 1999. Number of axons in the corpus callosum of the mature macaca nemestrina: increases caused by prenatal exposure to alcohol. J Comp Neurol 412:123–131.

NIAAA (National Institute on Alcohol Abuse and Alcoholism). 1983. Fifth Special Report to the U.S. Congress on Alcohol and Health. U.S. Department of Health and Human Services, Public Health Service, ADAMHA (Alcohol, Drug Abuse, and Mental Health Administration), NIAAA.

Olegård R, Sabel KG, Aronsson M, Sandin B, Johansson PR, Carlsson C, Kyllerman M, Iversen K, Hrbek A. 1979. Effects on the child of alcohol abuse during pregnancy. Acta Paediatr Scand (Stock) 275(suppl):112–121.

Olson EC, Miller RL. 1958. Morphological integration. Chicago, IL: University of Chicago Press.

Ouellette EM, Rosett HL, Rosman NP, Weiner L. 1977. Adverse effects on offspring of maternal alcohol abuse during pregnancy. N Engl J Med 297:528–530.

Pettigrew AG, Hutchinson I. 1984. Effects of alcohol on functional development of the auditory pathway in the brainstem of infants and chick embryos. Ciba Found Symp 105:26–46.

Randall CL. 1977. Teratogenic effects of in utero ethanol exposure. In: Blum K, Bord D, Hamilton M, editors. Alcohol and opiates: neurochemical and behavioral mechanisms. New York: Academic Press. pp 91–107.

Riley EP. 1990. The long-term behavioral effects of prenatal alcohol exposure in rats. Alcoholism 14:670–673.

Riley EP, Vorhees CV. 1986. Handbook of behavioral teratology. New York: Plenum Press.

Riley EP, Mattson SN, Sowell ER, Jernigan TL, Sobel DF, Jones KL. 1995. Abnormalities of the corpus callosum in children prenatally exposed to alcohol. Alcoholism 19:1198–1202.

Rodier PM, Ingram JL, Tisdale B, Croog VJ. 1997. Linking etiologies in humans and animal models: studies of autism. Reprod Toxicol 11:417–422.

Rodier PM, Ingram JL, Tisdale B, Nelson S, Romano J. 1996. Embryological origin for autism: developmental anomalies of the cranial nerve motor nuclei. J Comp Neurol 370:247–261.

Roebuck TM, Mattson SN, Riley EP. 1998. A review of the neuroanatomical findings in children with fetal alcohol syndrome or prenatal exposure to alcohol. Alcoholism 22:339–344.

Rouquette J. 1957. Influence de l'intoxication alcoolique parentele sur le développement physique et psychique des jeunes enfants. Thesis University of Paris, Paris, France.

Sampson PD, Streissguth A, Bookstein FL, Barr HM. 2000. On categorizations in analyses of alcohol teratogenesis. Environ Health Persp 108(suppl 3):421–428.

Sander LW, Snyder PA, Rosett HL, Lee A, Gould JB, Ouellette E. 1977. Effects of alcohol intake during pregnancy on newborn state regulation: a progress report. Alcoholism 1:233–241.

Schaefer GB, Shuman RM, Wilson DA, Saleeb S, Domek DB, Johnson SF, Bodensteiner JB. 1990. Partial agenesis of the anterior corpus callosum: correlation between appearance, imaging, and neuropathology. Pediatr Neurol 7:39–44.

Shaywitz SE, Cohen DJ, Shaywitz BA. 1980. Behavior and learning difficulties in children of normal intelligence born to alcoholic mothers. J Pediatr 96:978–982.

Slice D, Bookstein FL, Marcus LF, Rohlf FJ. 1995. A glossary for geometric morphometrics. Electronic publication: http://life.bio.sunysb.edu/morph/gloss1.html.

Smith DW. 1976, 1983. Recognizable patterns of human malformation; genetic, embryologic, and clinical aspects, 2nd, 3rd ed. Philadelphia: WB Saunders.

Stratton KR, Howe CJ, Battaglia FC, editors. 1996. Fetal alcohol syndrome: diagnosis, epidemiology, prevention, and treatment. Washington, DC: Institute of Medicine, National Academy Press.

Streissguth AP. 1978. Fetal alcohol syndrome: an epidemiologic perspective. Am J Epidemiol 107:467–478.

Streissguth AP, O'Malley K. 2000. Neuropsychiatric implications and long term consequences of fetal alcohol spectrum disorders. Semin Clin Neuropsychiatry 5:177–190.

Streissguth AP, Barr HM, Martin DC, Herman CS. 1980a. Effects of maternal alcohol, nicotine and caffeine use during pregnancy on infant mental and motor development at 8 months. Alcoholism 4:152–164.

Streissguth AP, Landesman-Dwyer S, Martin JC, Smith DW. 1980b. Teratogenic effects of alcohol in humans and laboratory animals. Science 209:353–361.

Streissguth AP, Aase JM, Clarren SK, Randels SP, LaDue RA, Smith DF. 1991. Fetal alcohol syndrome in adolescents and adults. JAMA 265:1961–1967.

Streissguth AP, Bookstein FL, Sampson PD, Barr HM. 1993. The enduring effects of prenatal alcohol exposure on child development, birth through seven years. Ann Arbor, MI: University of Michigan Press.

Streissguth AP, Barr HM, Kogan JA, Bookstein FL. 1996. Understanding the occurrence of secondary disabilities in clients with Fetal Alcohol Syndrome and Fetal Alcohol Effects. Final report, CDC grant CCR008515. Fetal Alcohol and Drug Unit, University of Washington School of Medicine.

Swayze VW II, Johnson VP, Hanson JW, Piven J, Sato Y, Giedd JN, Mosnik D, Andreasen NC. 1997. Magnetic resonance imaging of brain anomalies in fetal alcohol syndrome. Pediatrics 99:232–240.

Talairach J, Tournoux P. 1988. Co-planar stereotactic atlas of the human brain: 3-dimensional proportional system—an approach to cerebral imaging. Stuttgart: Thieme Verlag.

Thomas JD, Wasserman EA, West JR, Goodlett CR. 1996. Behavioral deficits induced by bingelike exposure to alcohol in neonatal rats: importance of developmental timing and number of episodes. Dev Psychobiol 29:433–452.

Toga AW, editor. Brain warping. New York: Academic Press.

West JR, editor. 1986. Alcohol and brain development. New York: Oxford University Press.

Williams RJ. 1956. Biochemical individuality. New York: John Wiley & Sons.

Wilson JG, Fraser FC, editors. 1977. Handbook of teratology. Vol. 1: General principles, etiology. New York: Plenum Press.

THE ANATOMICAL RECORD (NEW ANAT.) 269:162–174, 2002

FEATURE ARTICLE

Midline Corpus Callosum Is a Neuroanatomical Focus of Fetal Alcohol Damage

FRED L. BOOKSTEIN,* PAUL D. SAMPSON, PAUL D. CONNOR, AND ANN P. STREISSGUTH

Prenatal exposure to high levels of alcohol often induces birth defects that combine morphological stigmata with neurological or neuropsychological deficits. But it has proved problematic to diagnose these syndromes in adolescents and adults, in whom the morphological signs are absent or attenuated, the behavioral deficits nonspecific, and the exposure history often difficult to reconstruct. Localizing the associated brain abnormalities might circumvent most of these difficulties. To this end, three-dimensional (3D) locations were recorded for 67 homologous points on or near the corpus callosum in magnetic resonance (MR) brain images from 60 adolescents and adults who were normal, 60 diagnosed with fetal alcohol syndrome, and 60 diagnosed with fetal alcohol effects. We combined the standard statistical approach to this type of geometric data, Procrustes analysis, with a multivariate strategy focusing on differences in variability. In this data set, the shape of the corpus callosum and its vicinity proves systematically much more variable in the alcohol-affected brains than in those of the normal subjects. From this excess variability follows a promising classification rule, having both high sensitivity (100 out of 117) and high specificity (49 out of 60) in this sample. The discrimination uses four landmark points and two summary scores of callosal outline shape. The information from the corpus callosum and vicinity, as viewed in MR brain images of full-grown subjects, may serve as a permanent record of the prenatal effects of alcohol, even in patients who are first suspected of these syndromes relatively late in life or who lack the facial signs of prenatal alcohol damage. The statistical pattern underlying the callosal diagnosis also leads to speculations on mechanisms of the prenatal damage. *Anat Rec (New Anat) 269:162–174, 2002.* **© 2002 Wiley-Liss, Inc.**

KEY WORDS: fetal alcohol syndrome; FAS; fetal alcohol effects; FAE; corpus callosum; hypervariance; Procrustes analysis; symmetry curves; neuroscience; birth defects; child health; neurodevelopment; magnetic resonance imaging; MRI

INTRODUCTION

Fetal alcohol syndrome (FAS) is a relatively frequent diagnosis based in observations of growth deficiencies of prenatal origin, morphological anomalies, and neurological or neuropsychological symptoms together with a history of prenatal exposure to unusually high levels of alcohol. The brain is the organ most affected by prenatal alcohol exposure (West et al., 1994), with effects discernible in many regions by conventional statistical analysis of sectoral areas or segmented volumes and their disproportions (Roebuck et al., 1998). Yet contemporary diagnostic systems for prenatal alcohol damage do not use neuroanatomical data in any quantitative way (e.g., Astley and Clarren, 2000). Also, many persons suffering from the associated brain damage lack the external facial anomalies required for the FAS diagnosis. Typically, these patients have been labeled with less precise descriptions, such as "fetal alcohol effects" (FAE), but there is little agreement on any nosology within this extended domain.

For example, a recent report from the Institute of Medicine (Stratton et al., 1996) proposes a definition of "alcohol-related neurodevelopmental disorder" as applying to patients exposed to high levels of ethanol in

Dr. Bookstein is Honorarprofessor of Biomathematics in the Institute of Anthropology, University of Vienna, and Senior Distinguished Research Scientist in the Institute of Gerontology, University of Michigan. For more than 20 years he has been a major force in developing the data acquisition software and data analysis methods that are used in this study, and for nearly that long he has collaborated with Drs. Sampson and Streissguth on morphometric and psychometric studies of children and adults heavily exposed to prenatal alcohol. Dr. Sampson is Research Professor in the Department of Statistics, University of Washington (UW), where he is Director of the Statistical Consulting Program and Assistant Director of the National Research Center for Statistics and the Environment. As Senior Statistician at the Fetal Alcohol and Drug Unit (FADU), Department of Psychiatry and Behavioral Sciences, UW, he is responsible for the development and application of multivariate and morphometric methods of data analysis in this and other human samples. Dr. Connor is a neuropsychologist with the FADU and Acting Instructor in the Dept. of Psychiatry and Behavioral Sciences at UW. For the past six years he has been studying the neuropsychologic and neuroanatomic consequences of prenatal alcohol exposure in clinical and subclinical populations. Dr. Streissguth, Director of the FADU and Professor in the Department of Psychiatry and Behavioral Sciences at UW, has since 1973 worked with patients with fetal alcohol syndrome and other prenatal alcohol effects. She has carried out epidemiologic studies of long-term consequences of prenatal alcohol exposure on psychosocial development and the central nervous system and is interested in developing clinically relevant methods of detecting prenatal alcohol brain damage in patients of all ages.

This research was supported by USPHS grants AA-10836 and GM-37251.

*Correspondence to: Fred L. Bookstein, Institute of Gerontology, University of Michigan, 300 North Ingalls Building, Ann Arbor, MI 48109-2007. Fax: 734-936-2116; E-mail: fred@brainmap.med.umich.edu.

DOI 10.1002/ar.10110
Published online in Wiley InterScience (www.interscience.wiley.com).

utero who show "evidence of CNS neurodevelopmental abnormalities" and/or "evidence of a complex pattern of behavior or cognitive abnormalities" that "cannot be explained by familial background or environment alone." (All quotes are from Stratton et al., 1996, Table 4-1.) In other words, the alcohol-related diagnosis is a residual category for damage that cannot be attributed to these socioeconomic causes. But as the effects of familial background and environment are actually extremely difficult to assess in any biometrically rigorous manner, it might be more productive to improve the measurement of their brains than to reach out to the additional measures of families or postnatal environments that would otherwise be necessary.

The best neuroanatomical quantification would be a short suite of measurements characterizing the fact of prenatal alcohol brain damage reliably enough to differentiate affected from unaffected persons even should other associated signs be ambiguous. Such a protocol would be very helpful inasmuch as estimates of the fraction of the Western adult population affected by fetal alcohol damage range up to 1%, whereas the number actually diagnosed is quite a bit lower (Sampson et al., 1997). Although there are reports in the clinical literature about anomalies of neuroanatomy such as partial or total callosal agenesis in certain extreme fetal alcohol cases (Riley et al., 1995; Swayze et al., 1997), they have not hitherto suggested quantifications that would permit assessment of similar but subtler effects in the more typically affected individuals whose brains appear grossly and qualitatively normal.

This report introduces a new quantification of brain form specialized for service as a detection tool over the broad range of people with brain damage from prenatal exposure to alcohol. We have verified its effectiveness in a large sample of adolescents and adults, the most difficult age for diagnosing these effects. By late adolescence, the prenatal growth deficiency entailed in the fetal alcohol damage often has been reversed, and the telling facial characteristics, if they were ever present, may have been masked by ordinary growth variations (Lemoine and Lemoine, 1992; Löser et al., 1999). Although this study used patients diagnosed in the conventional manner, the method we suggest does not refer to facial features at all, but only specific measurable features of the MR brain image.

METHODS FOR GEOMETRIC NEUROANATOMICAL DATA

As regular readers of this journal are aware, the toolkit of computational neuroanatomy currently overflows with powerful new image-based techniques. *(Editor's note: see the special issue of* The New Anatomist *on biomedical imaging, described in Lester and Olds (2001) and available free online at* www.wiley.com/anatomy/bioimage.) This surge in novel methods seems driven equally by sudden improvements in the basic medical physics of imaging devices and by the sudden emergence of functional neuroimaging as the liveliest contemporary branch of cognitive neuroscience. One pinnacle of work in the new tradition, as reported by Toga and Thompson (2001), involves extensive mathematical modeling of the geometric relationships among solid brain images from large samples of normal or nearly normal individuals. In most applications of this sort, geometric relationships are encoded in very complex descriptive structures, such as detailed displacement fields, that are displayed on powerful workstations, scanned for local extremes of signal, and published as compelling images in glorious full color.

Contemporary diagnostic systems for prenatal alcohol damage do not use neuroanatomical data in any quantitative way.

This study is not in that tradition. Although the methods we will rely on are new and important, they are not computationally intensive in the image-processing phase. Instead, we rely on an older neuroanatomical approach, the careful manual tracing of structures that are known in advance to be informative for particular comparative applications. Computer software aids this tracing task but does not guide it with any sort of algebraic or geometric intelligence. The purpose of the tracing is to supply data for the rather more intelligent tools that carry out modern shape measurements of the structures of interest. For this purpose, a massive initial reduction in information content of the data set is not only desirable but actually essential. Although each subject's image is approximately eight million bytes of gray-level information, the tracing protocol we chose reduces it to the three Cartesian coordinates of a mere 67 points per case. Characterizations of these 67 points, furthermore, were selected in advance of looking at any of the data at hand; they arose variously from the literature of the disease under study, which continually mentioned but failed to quantify abnormalities of callosal form, or from general tenets of good morphometric practice when confronting an extended heterogeneous structure. In a series of steps involving expert judgments by both the teratologist and the statistician at every stage, the statistical analysis will subsequently distill those 201 dimensions down to a mere 4. It is these statistical methods, not the image analyses, that embody the scientific intelligence of these new neuroanatomical tools.

Subjects

We studied 180 Seattle-area subjects aged 14 to 37 years: 60 normal subjects unexposed to high prenatal doses of alcohol, 60 subjects diagnosed with fetal alcohol syndrome, and 60 with fetal alcohol effects. A diagnosis of FAS entailed evidence of a compromised central nervous system (CNS), growth deficiency of prenatal origin, and the uniquely characteristic facial stigmata subsequent to heavy fetal alcohol exposure: short palpebral fissures, flat philtrum, thin upper vermilion, and flat midface. The related diagnosis of FAE was typically applied to patients with CNS compromise and a history of exposure without the full set of physical findings. Each of the three main subject groupings—unex-

BOX 1. SYMMETRY CURVES

Many point-like structures appear on each side of the head—the colliculi, the head of the caudate nucleus—and others, such as the commissures or the tip of the fourth ventricle, can be taken as loci where obviously paired structures fuse to cross an apparent midplane with respect to which they are symmetrical. The apparent symmetry extends well beyond these point-like components to include at least one important curving form, the outline of the corpus callosum as we have digitized it here (Figures A and B). Any smooth curve in space, whether or not it is one of these callosal outline curves, has a tangent line at every point. Everywhere on these callosal midlines curves, there is a plane through that tangent line, a local symmetry plane, with respect to which the local volume is (approximately) symmetric, left against right. Call the direction perpendicular to this plane, the direction of left–right symmetry, the symmetry axis. Then if that local volume really is symmetric, we should be able to see the same symmetry in any plane through the symmetry axis. In other words, each point of this symmetry curve lies on the (approximate) line of symmetry of the image cut by any plane through the symmetry axis at that point. One of these cut planes includes the tangent to the curve itself, and another the actual (geometric) normal to the curve. In Figure 2 (see text), the symmetry plane itself is shown at upper left for the point under the crosshairs; the horizontal crosshair lies along the tangent line there. The (putative) symmetry axis, which is the normal to this plane, lies horizontally in both of the confirming images. At upper right, the image vertical is the tangent line itself. Notice how the white matter *just* touches the section plane at the crosshairs; everywhere else the section is outside the callosum. At lower left is the plane at 90 degrees, the customary normal plane to the curve (in other words, the normal section of the callosum through the vertical crosshair of the upper-left panel). Again, the image is nearly symmetric with respect to its vertical.

At lower right in Figure 2 (see text) is the most commonly encountered special case: where the symmetry curve of the callosum is not on the image boundary itself, but is instead effaced by juxtaposition with some other tissue of equivalent gray level, the point is located by extrapolation. The panel here shows a point of the callosal midline that is more or less hidden at the spring of fornix. (Its location in sagittal view is noted in connection with Box Figure B.) Rostrum is also a special case, as the outline there has no tangent line, or, rather, has two different tangents. The direction used in the visualization is the bisector of that pair of tangents.

The complete representation of any subject's midline callosal curve for our analysis is the set of 39 of these ordinary points, together with Rostrum. Although the spacing of the points is gently uneven (it will be standardized by the bending algorithm to be introduced presently), the statistical analysis to follow is almost invariant to changes in the spacing of these points. As Box Figure A shows, although the 40-point polygon has the familiar "shape of the callosal arch" in sagittal view, it is not in fact a plane curve. Other specimens, especially among the exposed subjects, can appear a good deal more warped out-of-plane than the normal adolescent male here. The same set of 40 points is displayed in Figure B superimposed over a nearby parasagittal image plane. The digitized outline does not lie on any single such section but typically wanders inside or outside of the apparent curve corresponding to the three-dimensional surface shape in its vicinity. The point displayed at lower right in Figure 2 appears here above posterior commissure and slightly anterior to it; in this view, it is difficult to distinguish from fornix.

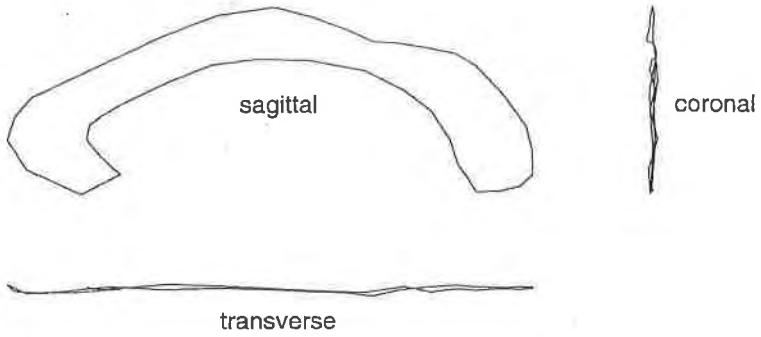


Figure A. Callosal midline curves are not planar. The three panels here show the curve for this subject in three precisely perpendicular planes. Our morphometric statistical analysis of the curve was restricted to the two-dimensional sagittal view here.

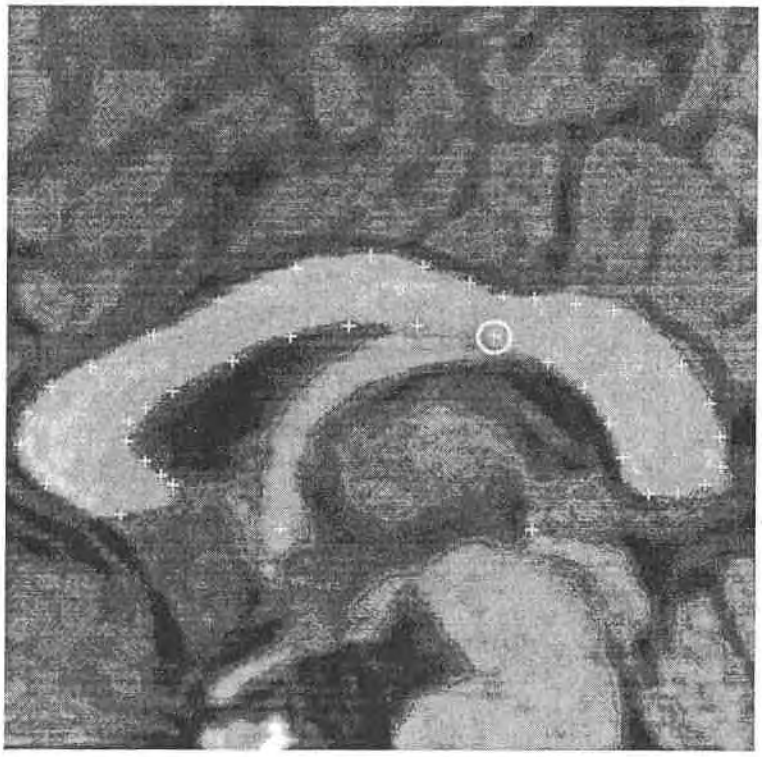


Figure B. The full set of 40 semilandmarks for the same subject is shown here superimposed over a nearby parasagittal plane. The callosal outline used in the statistical analysis is not the outline of the apparent callosal arch in this or any other planar image section. Digitized points not on the callosum: anterior and posterior commissures. The semilandmark at the crosshair in the lower right panel of Figure 2 (see text) is the point seeming to lie mid-fornix above and slightly anterior to posterior commissure here (circled).

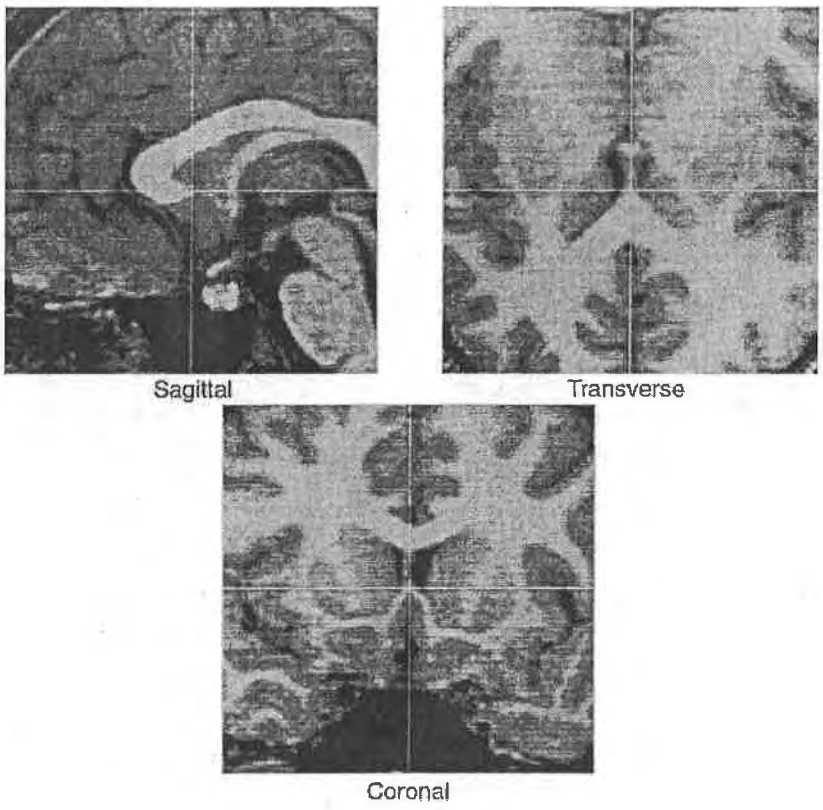


Figure 1. Digitizing Rostrum for 1 of our 15 normal adolescent males. View orientations as labeled. The anteroposterior axis, horizontal in the sagittal view and vertical in the transverse image, is the conventional anterior commissure–posterior commissure line, but those points need not lie on the plane here. The axis that is vertical in sagittal and coronal views is the direction perpendicular to anteroposterior in the plane of symmetry of the callosum through Rostrum, and the axis running left–right in transverse and coronal views is the symmetry axis at Rostrum, the normal to this plane of symmetry (see text). Transverse view: posterior toward top of page. Coronal view: top of head toward top of page. In this coordinate system, Rostrum is the sharp corner of the sagittal view on the plane of symmetry and having the posteriormost point of midgray voxel content. Similar sets of three perpendicular views implement and confirm the location of the other landmarks described in Table 1 and all the other landmarks of the full three-dimensional data set (Bookstein et al., 2001).

posed, FAS, FAE—was represented by 15 adolescent males (aged 14 through 17), 15 adult males, 15 adolescent females, and 15 adult females. For brevity, we refer to the pool of all 120 diagnosed subjects as "the exposed," although of course each of them was affected as well as exposed.

In studies of this kind, informed consent is typically obtained, once the procedures of a study are explained, variously from the subject or, for young or unemancipated subjects, from their custodial parent or guardian. Such consent was obtained for all the subjects in this study under the oversight of the cognizant University of Washington institutional review board.

Patient ascertainment was from the Seattle FAS Follow-up Database, accrued over nearly three decades from referrals from or to dysmorphologists. The patients were diagnosed FAS or FAE, often prefixed by "possible" or "probable," according to the clinical guidelines of the time (Streissguth et al., 1991). Normal subjects were recruited from employees and their children at local health care facilities and educational institutions so as to approximately match the age and ethnic composition of the exposed groups. All potential subjects/patients were excluded who had AIDS, were taking neurotoxic medications, were legally blind, wore dental braces, had had psychological testing within the past year, or did not have English as their primary language. Additionally, potential normal subjects were excluded who had alcohol or drug problems, neurological problems, birth defects involving the brain, who reported hearing voices or seeing visions, who had a bachelor's degree or higher education, or whose biological mothers had a history of alcohol or drug problems or had a history of binge drinking around the time of their pregnancy with the subject. There was no IQ criterion for acceptance into the study. WAIS-R IQ scores averaged 109 for the unexposed subjects, 83 for those diagnosed FAE, and 80 for those diagnosed FAS; a total of 19 IQ scores, all from exposed subjects, fell below 70. Complete morphometric data were obtained for 177 of the 180 subjects (omitted from analyses: two adolescent females with hydrocephalus and one with partial agenesis of the corpus callosum; all three were in the exposed group).

Images

T1-weighted sagittal SPGR images were acquired over 12 min in a GE 1.5 T Signa scanner at the University of Washington. TE was 8 msec, TR 29 msec, flip angle 45 degrees. The resulting images contained $256^2 \times 124$ voxels of size $0.86^2 \times 1.50$ mm^3.

Data Acquisition

From these images, two data sets were extracted by using the *Edgewarp* program package (Bookstein and Green, 2002; Bookstein et al., 2001). All digitizing was carried out blind to subject characteristics except age group and gender. One data set consists of 28 discrete landmark point locations on many different subcortical structures; the other, a series of 39 arbitrarily spaced points on one specific curve upon a single structure. We describe these data sets separately.

Locations of 12 unpaired (midsagittal) points and 8 bilaterally paired points were digitized in a coordinate system aligned with the plane through anterior and posterior commissures and tip of fourth ventricle. A landmark is digitized when it is located authoritatively in two orthogonal sections of the solid image and confirmed in a third orthogonal image. For example, the landmark point Rostrum, which will also lie on the midline callosal outline, is characterized as the

TABLE 1. Names and operational definitions of landmarks near the crease in Figure 5

Name	Sagittal view	Axial view	Coronal view
Genu (CC)	Point directly in front of interior genu	Local axis of symmetry of callosum–CSF boundary	Short segment of gray between white matter and CSF
Caudate (L, R)	Triple point posterior bottom of horn of ventricle at spring of thin white capsule	Triple point of CSF, caudate and white matter	Beginning of the appearance of gray matter of caudate
Frontal horn (L, R)	Most frontal point on teardrop shape' boundary between brain and CSF	Most frontal point of boundary between brain and CSF	Slightly off center of small oval of CSF
Interior genu (CC)	Point of sharpest curvature on interior border of CC	Local axis of symmetry on boundary of CC	tiny lacuna touching CSF within a large sector of CC tissue
Splenium (CC)	Posteriormost point on CC	Local axis of symmetry on boundary of CC	Tiny segment of grey touching CC within a large sector of CSF
Rostrum (CC)	Sharp corner of CC	Axis of symmetry at narrowest point of CC	Axis of symmetry of last pixels of CC

*CC, corpus callosum; CSF, cerebrospinal fluid; L, R, left and right. Anterior and posterior directions accord with the anterior commissure–posterior commissure line. Modified from Bookstein et al., 2001, Table 2.

tip of the apparent retrograde corner of the callosum in one plane, the extreme of midgray image contents on an anteroposterior line of bilateral symmetry in a second plane, and a triple point of intersecting thin curves in a third plane. These three planes are shown in Figure 1 for one of our normal subjects, an adolescent male.

The point Rostrum demonstrated here is, properly speaking, the only fully acceptable point-landmark on the entire callosal outline: the only one characterized by details of local image geometry. A point almost as reliable, Interior Genu (IntG), is characterized by the curving of the local callosal outline in its vicinity—IntG is the sharpest corner of the inner boundary of the anterior arch where callosum abuts the third ventricle in the midline. Two other points that will appear in the analysis are not proper landmark locations at all, but quasi-landmarks or "Type III" landmarks in the sense of Bookstein (1991). In the coordinate system aligned with the commissures, Genu is the point of the midline outline directly in front of Interior Genu, and Splenium is the most posterior point of the same outline. Operational definitions of these four points, along with two others arising in the sequel, are collected in Table 1, extracted from the complete roster of 28 definitions in Table 2 of Bookstein et al. (2001).

Considerable additional information was extracted from the same images by digitizing an additional 39 points, *semilandmarks*, all the way around the outline of the corpus callosum in its midline. Semilandmarks are not new to readers of this journal (see Bookstein et al., 1999), and when derived from two-dimensional images, they have appeared in earlier studies of the callosal outline itself (Bookstein, 1997b). But the careful way in which we have digitized them for the analysis here is new. The callosal midline curves we are digitizing need not, and in fact do not, lie in any single geometric plane within the original three-dimensional data volumes of these brain images. Nor do they arise, like sulcal tracings, as actual curves in the three-dimensional space of the brain. Instead, they are an abstraction exploiting the near-symmetry of brain form in our species: a *symmetry curve* (see Box 1) standing for a locus of symmetry upon a larger 3D surface in each specimen. Because Rostrum is actually a point on this symmetry curve, you have already seen an example of how symmetry is used in the course of the digitizing (Figure 1). A more typical point of the curve for the same subject is demonstrated in Figure 2, again by way of a "side view" of the symmetry plane, together with two confirming images.

Morphometric Strategy

Each of the morphometric data sets was converted to *Procrustes shape coordinates*, representations of the relative positions of the points from which variations of overall size, position, and orientation have been removed a priori (Bookstein, 1997a, 1999; Dryden and Mardia, 1998) for the purpose of a multivariate analysis of shape. Each of the data sets required a special tool for this purpose.

For the landmark data set, the concern was for symmetry. Each of the 28-landmark shapes was averaged with its own mirror image (Mardia et al., 2000). There results a completely symmetric form in which unpaired landmarks, such as the commissures, lie precisely on one plane, whereas the paired landmarks (like Head of Caudate) become perfectly symmetrical with respect to that plane. The paired points were then projected onto the same midplane, the "synthetic midplane" whence anteroposterior and craniocaudal coordinates were sent

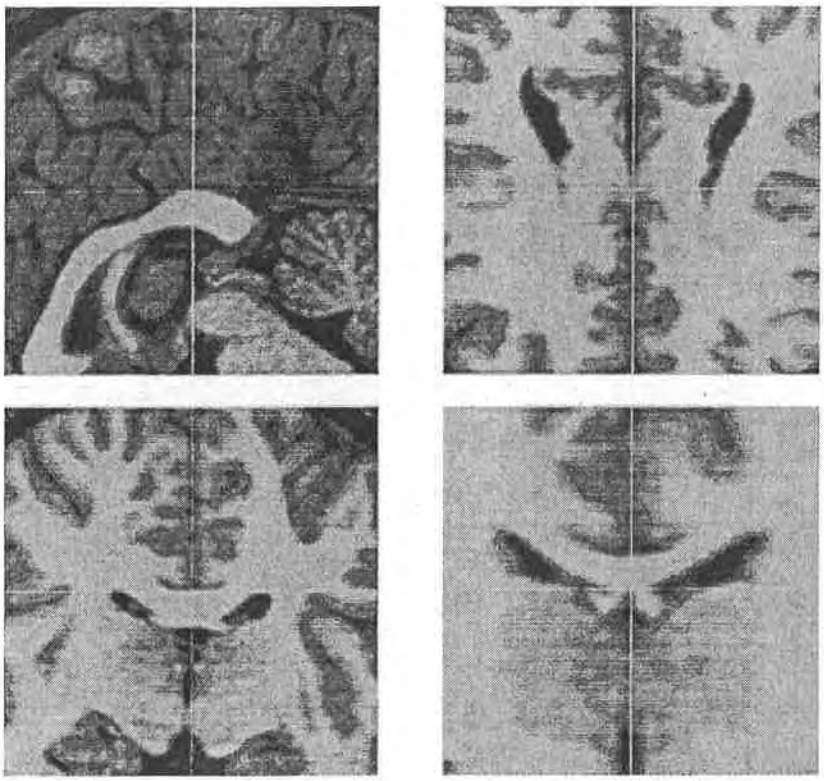


Figure 2. Tracing callosal midline curves: typical and exceptional points. The subject is the same male as in Figure 1. A typical point lies on the apparent boundary of the callosum in some symmetry plane (upper left). A coordinate system is taken along the callosum's tangent line there (the horizontal crosshair). The symmetry axis is a direction perpendicular to the tangent such that every plane through that direction looks approximately symmetric as an image. The upper right panel shows such a plane through the symmetry axis (horizontal) and the tangent to the curve (vertical); the lower left panel, a plane through the same symmetry axis perpendicular to the curve. In principle, these axes lie in different directions at every point of this curve. Lower right panel, the normal view at an exceptional point (see text and Box 1 Figure B).

for Procrustes analysis by means and scatters of the shape coordinate pairs shown in Figure 3.

For the callosal outline data set, the concern was to standardize positions of the semilandmarks along their curves to optimize the power of the statistical analyses to follow. For this, we relaxed individual semilandmarks along their outlines by the method of bending energy (Bookstein, 1997b). In this algorithm, specimen by specimen, the semilandmarks are slid simultaneously along the outline to minimize the net bending energy of the transformations (modeled as a thin-plate spline) by which each is produced from the grand Procrustes average. In an earlier application (Bookstein et al., 1999), the semilandmarks arose from two different 2D curves, the inner and outer faces of the frontal bone in the midsagittal plane; in the present application, they arise from two widely separated arcs of the same 3D curve. In both applications, the sliding proceeds simultaneously both at relatively large scale, by which spacing of the points is adjusted "from end to end," and at relatively small scale according to which nearby points on different arcs are corrected to consistent relative positions throughout the sample. The relaxation minimizing bending energy took account, in addition, of three landmark points not on the callosum: anterior and posterior commissures and tip of fourth ventricle. Rostrum, although itself one of these outline points, was not allowed to slide.

After relaxation, the set of 177 40-point polygons was converted to shape coordinates by 3D Procrustes analysis and then projected onto its own plane of best fit to yield a 40-point polygonal outline for each subject. Those sets of 40 points were then sent for Procrustes analysis of the shape coordinates shown in Figure 4.

Multivariate Statistical Tactics

Separately for males and females, average shape in the unexposed subsample was tested against that in the FAS subsample, the FAE subsample, and the full-exposed pool by permutation test (Good, 2000) of net squared Procrustes distance. For the landmark data set, this was by the ordinary formula: the sum of squared distances between the landmarks of one subject and the landmarks of the grand average or a diagnosis-specific average. For the outlines, the distances squared and summed were taken only in the direction perpendicular to the grand average outline shape. As the reader may have already suspected from the right-hand sides of Figure 3 or Figure 4, no meaningful mean differences were found among the diagnostic groups (FAS, FAE, unexposed) in either sex. In separate computations not displayed here, the coordinates omitted from Figures 3 and 4, i.e., mediolateral coordinates of the landmarks and semilandmarks, likewise showed no mean effects of diagnosis either point by point or combined.

Principal components of shape coordinates, also called relative warps, are principal components of Procrustes shape coordinates computed from their covariance matrix (not their correlation matrix). These principal components have been published previously for the adult subsample of this study (Bookstein et al., 2001). Both for the landmark data set and for the curve data set, the plots of components 1 through 3 show hints of excess variance within the exposed subsamples. There continue to be no mean differences distinguishing the exposed from the unexposed or the subgroup diagnosed FAS from those diagnosed FAE.

Relative warps are the shape variables that, over the combined sample of both exposed and unexposed, have the largest variance as a multiple of a certain technical quantity, their "Procrustes length." For this study, we strongly enhanced the legibility of

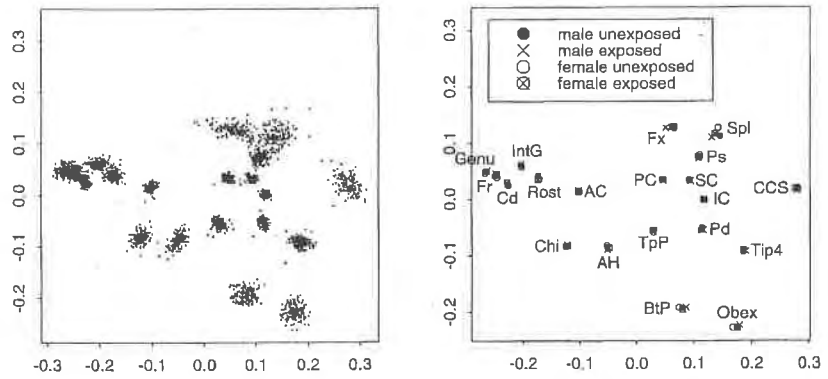


Figure 3. Scatters of Procrustes shape coordinates for 20 landmarks in the synthetic midplane and their averages by sex and exposure. (Left) 12 midline landmarks and 8 bilateral pairs, symmetrized as in the text, for 177 subjects. (Right) Means by sex and exposure status (N = 30, 60, 30, 57), with landmark legend. Fr, frontal horn of the lateral ventricle; Cd, head of the caudate; IntG, interior genu; Rost, rostrum of the corpus callosum; AC, anterior commissure; Chi, optic chiasm; AH, anterior limit of the temporal horn; PC, posterior commissure, TpP: top of pons at crevice in midline. BtP: bottom of pons at corner in midline. Fx: spring of fornix. Ps: tip of pes of hippocampus. Spl: splenium of corpus callosum. CCS: central cerebellar sulcus. Tip4: tip of fourth ventricle. SC: superior colliculus. IC: inferior colliculus. Pd: spring of cerebellar peduncles. Landmarks Fr, Cd, AH, Fx, Ps, SC, IC, and Pd are bilateral.

those hints of extra syndromal variance by *relative eigenanalysis*, the computation instead of the shape variables that have the greatest variance in the exposed subgroup *as a proportion of their variance in the unexposed*. This approach, although novel, is not too technical; we provide the details in the Appendix. Any interesting linear combination of shape coordinates, whether mean difference, relative warp, relative eigenvector, or any other derived multivariate descriptor, specifies a simultaneous shift in position of all the shape coordinates (all the projected landmark locations or all the slid semilandmarks) simultaneously. We visualize these deformations of the sample average configuration by the method of thin-plate splines (Bookstein, 1997a).

FINDINGS

The deformation grids in Figure 5 pertains to the landmark data. Table 1 describes the landmarks. For these 20-landmark shapes, this is the shape dimension that maximizes the ratio of variation within the exposed females (29 FAS, 28 FAE) to variation within the unexposed 30 females. (The dimensionality d described in the Appendix was set at 11.) The grid at left diagrams a particularly legible multiple of this shape factor, the one for which the mesh first appears to "fold" (the method of creases; Bookstein, 2000). (Taken in the opposite sense, Figure 5, right, the deformation involves localized vertical extension in the region of genu. The signal here is statistically just the same, but the figure is much less powerful—the crease method takes advantage of a trick of human visual processing.) Evidently, the information that drives this geometrical pattern emphasizes a localized vertical compression in the vicinity of the five anteriormost landmarks: Genu, Interior Genu, Rostrum, and the frontal horn of the ventricle and the head of the caudate as projected onto the midplane (Figure 3). The emergence of this factor from a relative eigenanalysis means that the shape factor shown—relative vertical extension or compression of this part of the image—is more variable in the exposed female subjects, as a multiple of its variance in the unexposed, than any other shape variable that might arise in the space of the first 11 principal components of the full female data set. No other part of the grid shows as severe a deformation as does the anterior callosum. The deformation is very clearly localized, meaning that exposed form is very much more variable than unexposed form just in this region delimited by three midline landmarks on the genu of corpus callosum, together with two nearby neighbors.

A careful exploration of the geometry, involving more relative eigenanalyses that need not be reported here, leads to the simpler quantification in Figure 6, which applies to both sexes at once. The diagram specifies a two-point shape coordinate construction (Bookstein, 1991) involving four of the five landmarks entangled in the crease of the grid in Figure 5. The baseline is the segment from Genu to Rostrum, and Interior Genu and Head of Caudate are the moving points. Four shape coordinates are sufficient to capture the shape of this quadrilateral, as indicated in the figure caption. Of the four, the two representing distance perpendicular to the baseline are aligned with the direction in which the crease of Figure 5 has crumpled and, hence, prove particularly informative. Figure 7 presents their scatter coded by diagnostic

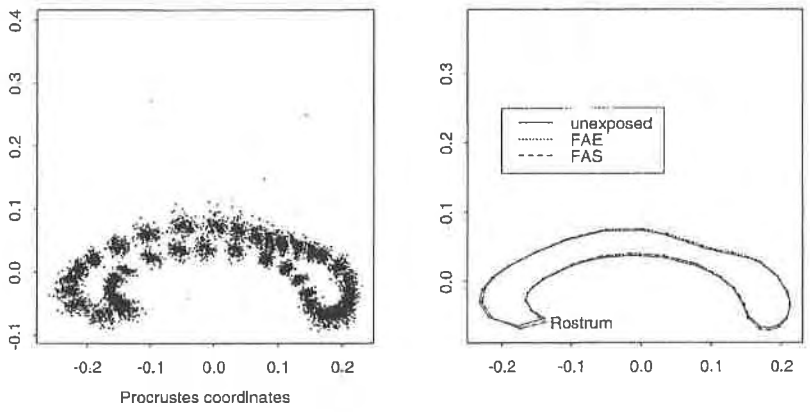


Figure 4. Scatters and means of Procrustes shape coordinates for the outline data. (Left) Scatter of Procrustes shape coordinates for Rostrum and 39 semilandmarks. (Right) Means by diagnosis (N = 60, 60, 57). There are evidently no meaningful mean differences here.

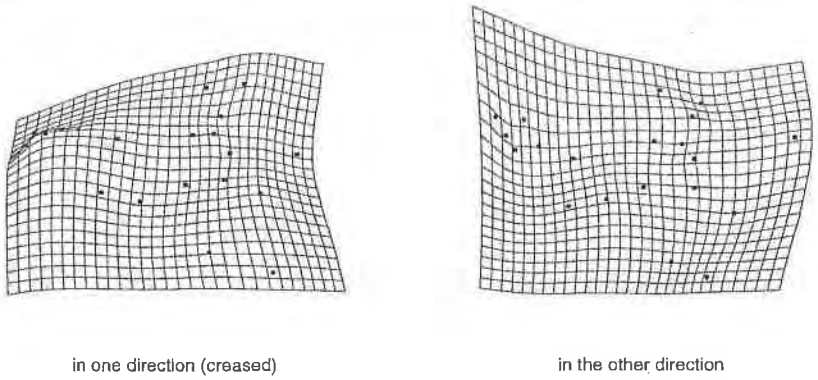


Figure 5. The first relative eigenvector for landmark shape variability, exposed versus unexposed, in females, by using 11 principal components of landmark shape. It is drawn as a thin-plate spline. (Left) Extrapolation in one direction until the graph has "folded," a pattern particularly easy to read. Five landmarks appear to contribute to this shape feature (see Figure 3 for placement and Table 1 for description): Genu, Interior Genu, Rostrum, averaged Caudate, and averaged Frontal Horn. (Right) Grid for the opposite transformation.

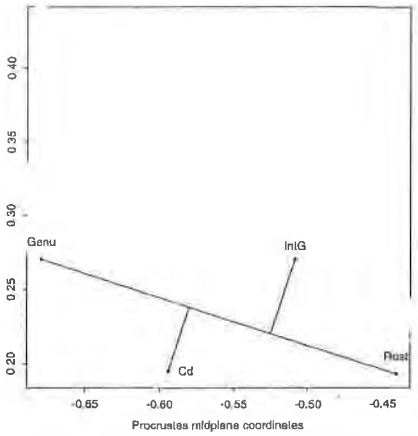


Figure 6. Shape coordinate pairs of Interior Genu (IntG) and Caudate (Cd) to a baseline from Rostrum (Rost) to Genu. To produce the shape coordinates for IntG, measure its perpendicular distance above the Genu–Rostrum line, and also the distance of the foot of that perpendicular from Genu; then divide each of these distances by the distance from Genu to Rostrum. The second pair of shape coordinates are produced analogously by using Cd in place of IntG. For the operational definitions of these four landmarks, see Table 1.

group. There is apparent clustering of the normal subjects so tight that only 6 of the 60 fall outside a conveniently drawn hull. By comparison, 73 of the 117 exposed fall outside the same hull. The practical meaning of the separation is as follows: "normally" (i.e., in the unexposed), IntG lies above the Rostrum–Genu line by approximately 20% of the length of that line, and projected Caudate (Cd) lies below that line by approximately the same fraction. Variation of IntG by more than 0.08 or so, or of Cd by more than 0.15 or so, or any other combination indicated by the polygon on the figure, suggests (in this sample) a diagnosis of fetal alcohol damage. (This language does not suggest that Rostrum and Genu are placed "normally" and IntG or Cd "abnormally" in the exposed subjects or anybody else outside the hull; the deviations of these ratios can arise from derangement or dysmorphology at any one or more of the four points involved.) Note that although the computations leading to the discovery of this pattern were highly multivariate, the actual discrimination in Figure 7 is an elementary one, involving only two ratios of distances arising from three measures among four landmark points.

From this suggestive tableau, we turn to the other data resource of 177 callosal outline shapes, the Procrustes analysis of which was shown in Figure 4. Again, although the average outline shapes do not substantially differ across the diagnostic groups, there is a strong signal of hypervariability from the conventional principal component analysis that is enhanced by relative eigenanalysis over a range of dimensionalities d. It is convenient to proceed using $d = 11$ dimensions of principal components, the same strategy as in Figure 5 (even though these are of course a completely different set of 11 dimensions pertinent to a completely different data set). For the outline data, it proved worth preserving two of these interesting dimensions instead of just one.

Figure 8 shows the first two relative eigenvectors, again as deformation grids deforming the average shape by arbitrary positive and negative multiples. These grids do not simplify into discrete shape coordinates as Figure 6 managed to do for the landmark data. Instead, they refer to phenomena at a gratifyingly large scale. The exposed callosum seems to be systematically hypervariable in arch shape and in thickness along the arch. The pattern of relative eigenvector 1 (Figure 8a, b) appears to be a contrast of relative thickness in the front of the callosum with respect to that at the back. There is also entailed an opening of the angle of the arch and an anteroposterior variability in the position of rostrum and interior genu with respect to genu itself. The second of these most highly hypervariable patterns, Figure 8c, combines the vertical "crushing" near genu that we have already seen in the landmark data set (Figure 5) with a relative thickening of the entire remaining arch of the callosum. Or the pattern could as well be characterized by the opposite deformation, vertical extension at genu along with thinning of the rest of the arch (Figure 8d).

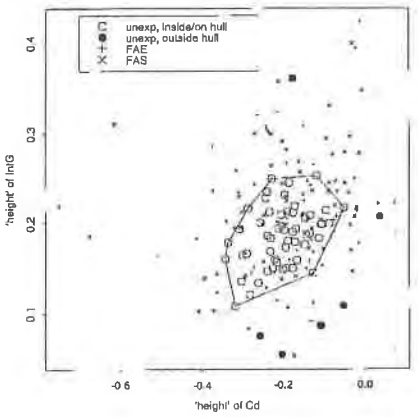


Figure 7. Full sample scatter (all 177 landmark sets) of the two vertical shape coordinates indicated in Figure 6. The unexposed subjects, regardless of age group or sex, cluster quite tightly, with exposed subjects varying from this central tendency in every direction. The polygon representing the classification rule was drawn by hand to pass through a suitable selection of subextreme normal subjects. The plotting symbol used for the unexposed in Figure 9 arises from position in this scatter as indicated. FAE, fetal alcohol effects; FAS, fetal alcohol syndrome; IntG, Interior Genu; CD, Caudate.

BOX 2. SIGNIFICANCE TESTS

As the relative eigenvector shown in Figure 5 (see text) was computed intentionally to maximize the variance of the exposed females as a ratio to that of the unexposed, one cannot test that ratio (which happens to be 8.1) for significance by arithmetic on those variances. Instead, we *crossvalidated* the analysis. The same formula that had the maximum variance ratio for the females can be applied to form scores (linear combinations of the shape coordinates) for the independent pool of 90 male subjects, and the exposed males turned out to have 4.3 times as much variance as the unexposed in this entirely separate subsample. That variance ratio of 4.3 can be tested as an F test with 59 and 29 degrees of freedom ($F_{59,29}$); the tabled significance level is $P \sim 0.00003$. As this was the best of a series involving between 6 and 20 principal components in either sex, a conservative correction of this tail probability would multiply it by 30, yielding a true significance level of approximately 0.001 for the finding of excess variation in the exposed. Over the full data set of 177 subjects, the net ratio of variances on this score, exposed 117 versus unexposed 60, is 5.89. Exposed cases are, on the average, almost 2.5 times as far from the grand average as are the unexposed. There seem to be no differences in this score by age group, sex, or diagnosis within the exposed subsample.

For the relative eigenanalysis of callosal outline semilandmarks (Figure 8 in text), our findings did not vary much by dimension or by sex, and so we pursue the matter further by pooling the sexes. A suitable permutation test exploits the sum of the first two eigenvalues (corresponding to the two relative eigenvector scores shown in the next figure) against the distribution generated when the diagnostic labels are randomly permuted over the 177 cases. The observed eigenvalues are greater than any found over 1,000 random permutations of the diagnostic label, so that the corresponding significance level may be estimated, again, at $P \sim 0.001$.

The scores corresponding to these two relative eigenvectors are scattered in Figure 9. On the horizontal axis (relative eigenvector 1), the exposed have 3.57 times as much variance as the unexposed; on the vertical axis (relative eigenvector 2), 3.01 times as much. As in Figure 7, FAS and FAE subjects, intermixed, obviously fall disproportionately outside a hull of concentration for the shapes of the unexposed. We have plotted the unexposed by using the symbols from Figure 7 to show that, of those six "outlying" unexposed subjects, three lie inside this hull for midline callosal shape, two lie on the hull, and one lies outside a second time. To follow the exposed subjects in both figures, it is convenient to produce a third scatter, Figure 10, from which the unexposed subsample is omitted. The legend indicates the diagnosis (FAS or FAE) and the position with respect to both decision boundaries, that in Figure 7 and that in Figure 9. A total of 17 FAS or FAE printed in solid symbols here evade detection in either discrimination.

Combining the information in Figures 7, 9, and 10, we see that a decision rule assigning a subject to the "exposed" subgroup if he or she lies outside the normal cluster in *either* Figure 7 or Figure 9 detects 100 of our 117 exposed subjects while making only 11 misclassifications of the unexposed. In the conventional terminology, this is a *sensitivity* of 100/117 = 85% and a *specificity* of 49/60 = 82%. The 17 false-negative classification errors (alcohol-affected subjects inside both of the convex hulls, Figures 7 and 9) are evenly distributed over diagnosis and sex: 3 FAE males, 4 FAS males, 5 FAE females, and 5 FAS females. Box 2 describes the statistical significance of these findings.

BIOLOGICAL AND BIOMETRIC CONCERNS

The logic of the discrimination here, which contrasts a fairly well-regulated form in the unexposed with a widely varying collection of dysmorphies in the exposed, may be unfamiliar, but the idea of considering differences in variance irrespective of averages is not wholly unfamiliar to the quantitative biologist. Its logical contradiction, the equality of variances, is mentioned in many biometry texts as one of the assumptions that need to be checked if inferences coming from the usual analysis-of-variance computations are to be scientifically valid. In two-group studies, checking this assumption involves the same F-tests already mustered in the course of testing for mean differences. What makes the F-tests quite unexpected in the present context is the application to a real hypothesis about directly observable variances. It is malformations of shape, not size, that are of the greatest interest in neuroanatomy; therefore, a difference in variances of shape is not a nuisance interfering with inferences about means but a powerful finding in its own right.

This finding is in accord with the now rather old literature concerning the origins of normal variation, the denominators of all those variance-ratios we were maximizing. For a charming overview of this topic, see the chapter on "Anatomical variations—significance" in Williams (1956). The chapter reviews the literature through publication date on quantitative and qualitative variation of structure in animals and humans, concluding, unsurprisingly, that great variation from one normal individual to another is found in structures in all of the body's systems, and that, in particular, the brain is "extremely variable in every character that has been subject to measurement." The matter of this variability is only rarely considered in contemporary anatomy atlases. Anson's (1950, 1961) great atlas, with its dozens of plates of normal variation of form and of topology, is long out of print, as is even the much more recent compendium by Bergman et al. (1988).

Williams (1956) goes on: "Virtually nothing is known about disharmonies of development in the central nervous system except for very gross deficiencies." We know of no review over the past half-century that would alter this summary. The latest monograph on associated multivariate methods, Olson and Miller's *Morphological Integration* (1958), is limited to simple morphological measures. (This is not

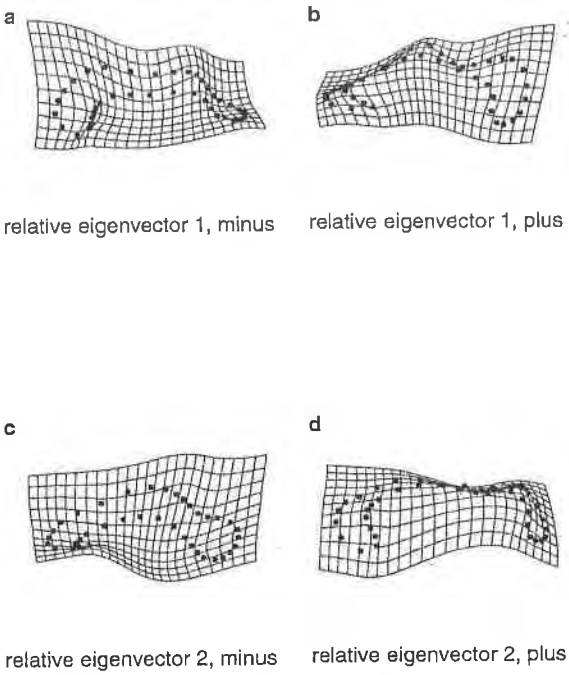


Figure 8. The patterns of greatest variability of callosal outline shape, exposed versus unexposed, for the full sample of 177. Diagrams of these two relative eigenvectors, in both directions, are drawn by the method of thin-plate splines. The multiples of the relative eigenvectors rendered here exaggerate the actual sample range of the next figure to show more clearly how changes of curve position are correlated around the outline.

their fault; shape coordinates were not to be invented for another 25 years.)

All the more startling, then, that our study of one single abnormal developmental pathway—environmentally induced ethanol neuroteratogenesis—has uncovered so clearly focused a measurement protocol: a target organ, and a method of representing it, for which the "disharmonies of development" prove so spatially specific. Because the finding is a variance-ratio, the mechanism here might be entailed in either a large numerator (lots of variance in the exposed) or a small denominator (unexpectedly tight regulation of form in the normal embryo). The latter interpretation is the more fruitful theoretically, as it corresponds to Waddington's (1966) classic notion of a *chreod*, arrival at a common developmental end point in the face of variation in ontogenetic conditions or timings. Logically, our finding regarding the corpus callosum has the form of a disrupted chreod. The developing callosum that was exposed to ethanol seems not to be able to access the information about strict form-constraints that apply to the normal version. The response of these embryonic neural tissues to ethanol toxicity is in a variety of patterns, corresponding, perhaps, to the variability of dose or timing or to the variety of behavioral deficit profiles seen in people with this disorder.

Although the corpus callosum itself consists mainly of white matter, the mechanisms governing its primordia may be shared with those of the overlying cerebral cortex. Rakic and Kornack (2001), summarizing an extensive literature, note that cortical neurons are not generated within the cortex itself, but rather near the surface of the cerebral ventricle (which includes, of course, the underside of the callosal arc we traced). The callosum begins to be laid down at approximately the 39th postconception day (Loeser and Alvord, 1968), in the middle of the time of greatest neuroembryonic sensitivity to alcohol damage, and alcohol exposure is known to derail normal neuronal migrations (Ikonomidou et al., 2000). The callosum, thus, might serve as a quantification of the earliest stage of teratogenetic processes that, when the cortex is affected, account for the decreased

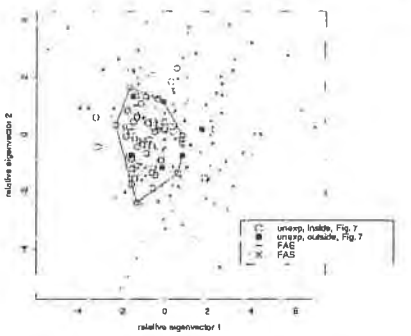


Figure 9. Full sample scatter (all 177 callosal outlines) of scores for the first pair of relative eigenvectors shown in Figure 8. Again the normal subsample clusters tightly except for six unexposed outliers outside a suggestive hand-drawn hull. The coding of diagnostic subgroups and outliers of the unexposed cluster is as in Figure 7. Of the six subjects having an outlying genu shape as measured by shape coordinates, five lie inside or on the hull here. FAE, fetal alcohol effects; FAS, fetal alcohol syndrome.

cerebral size and diverse behavioral deficits typical of this group of patients.

Algebraically, these are *quadratic* discriminations, detections of abnormality in any direction out of the central (normative) tendency, not just shift of central tendency in some direction along some unidimensional score. Then we should avoid polarized language, such as references to "deficits," in discussing these shape phenomena. In the hypervariation of Figures 7 or 9, there is no single direction in which the exposed have been sub-

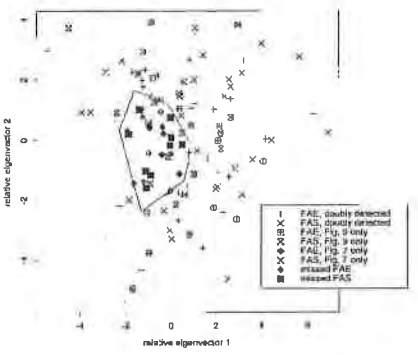


Figure 10. Accuracy of the combined detection rule for the exposed subjects. The plotting symbol encodes every combination of diagnosis, detection in Figure 7, and detection in Figure 9. Of 117 exposed subjects with callosal outline data, the 17 plotted in solid symbols here lie inside the hulls of both Figure 7 and Figure 9 and, thus, are the false negatives of the classification rule proposed in the text. FAE, fetal alcohol effects; FAS, fetal alcohol syndrome.

stantially shifted in mean with respect to the unexposed—no "direction of deficit." There is instead a rough but effective assessment of similarity or dissimilarity to the normal, a contrast of typical with atypical in all directions. The distinction here between modes of developmental explanation drawing on different spatial metaphors is not unique to neuroanatomy but applies throughout theoretical biology as a whole (cf. Weiss, 1973).

CLINICAL IMPLICATIONS

The difference between the diagnoses of FAS and FAE that this study spans lies primarily in the presence or absence of facial stigmata. Those features presumably *were* present at the time that the FAS subgroup of the present cohort were diagnosed, or else the diagnosis of FAS would not have been made. And because both growth deficiency and the particular facial stigmata of FAS attenuate with time, adults who were diagnosed with FAS as children might not have been diagnosed so had they first been encountered as adults (Streissguth et al., 1985, 1991; Spohr et al., 1993). Indeed, a substantial number of people with fetal alcohol damage are first suspected of the diagnosis only late in adolescence or as adults (Streissguth et al., 1996), when it is much less likely that they could be called FAS. This subgroup of too-long-delayed FAS diagnoses combines with another subgroup, the patients in our FAE subgroup, who, in many earlier diagnostic schemes, would never have been diagnosed with fetal alcohol damage at all.

In our data set, the FAE and FAS subgroups do not differ in any salient feature of brain form. This finding is consistent with earlier observations that the groups substantially overlap in the behavioral domain as well (Mattson et al., 1998; Connor et al., 2000; Bookstein et al., 2002). Criteria for access to medical and social support services that are predicated on details of the diagnosis (Sampson et al., 2000) thus would seem to be on an especially shaky clinical foundation in this age range. To rule in or rule out organic neurological causes for neuropsychological deficits not noted until so late in development, it would be important to have a specific protocol that manifests a strong signal even when other signs may have faded. The callosal shape signal in Figures 7 and 9 is just such a diagnostic protocol. (Compare the computationally more intensive protocols reported in Archibald et al., 2001, or Sowell et al., 2001, where the principal concern was with description of mean differences, not with diagnosis per se.).

The net power of this discrimination is remarkably high in view of the limited data resources it taps. The geometric data pay no attention to facial features, head size, IQ deficits, behavioral abnormalities, or any of the other signs conventionally used for establishing fetal alcohol spectrum diagnoses, and the information that proved useful derives from a small subvolume of the full brain image. Although it would certainly be appropriate to explore whether quantitative callosal anomalies are entailed in other syndromes overlapping with the adult neuropsychological profile of fetal alcohol—specifically, mental retardation—nevertheless our results argue strongly that the MR brain image should be incorporated in the standard clinical protocol for suspected fetal alcohol damage in adolescents and adults, and should be explored for equivalent discriminatory power at earlier ages. (Callosal shape abnormalities seen in adults could well be detectable at birth, or even earlier.) The unexpectedly high density of information pertinent to fetal alcohol damage in the MR image processed in the careful quantitative way we have demonstrated here ought to be explicitly used in producing these diagnoses. Other environmental or genetic factors that increase the risk of gross anomaly, whether in the brain or elsewhere, might also benefit from this careful new style of quantification, and other diagnostic tasks might similarly be converted from qualitative to quantitative by analyses like these. We would then enrich our knowledge of the "disharmonies of development" (Williams, 1956) beyond the gross deficiencies into a new domain of real clinical subtlety.

Our results argue strongly that the MR brain image should be incorporated in the standard clinical protocol for suspected fetal alcohol damage in adolescents and adults, and should be explored for equivalent discriminatory power at earlier ages.

ACKNOWLEDGMENTS

We thank William D. K. Green, for custom features of *Edgewarp* that eased our digitizing task; Malay Chan, for locating 5,040 landmark points; Julia Kogan, for coordinating recruitment of 180 hard-to-trace subjects; and Helen Barr, for management of the multiple databases involved here. A.S. and F.B. received funding from the USPHS. An earlier version of this manuscript was presented at the 2001 Annual Meeting of the Teratology Association in Montreal, Quebec.

APPENDIX: RELATIVE EIGENANALYSIS

Although the technique of multivariate analysis of variance (MANOVA) is familiar by now to a wide range of applied biologists, the closely related technique of relative eigenanalysis, from which the results of this study spring, is less familiar. MANOVA helps to analyze differences among two or more groups by constructing axes of greatest mean differences (in a sum-of-squares sense) combining two or more variables. To this end, two partial covariance matrices are built: one, Σ_{within}, for within-group covariance; and the other, $\Sigma_{between}$, for the covariances of the group means. An algebraic analysis is then carried out to find directions X (that is, linear combinations of the variables involved in the covariance matrix) for which the between-group variance $X'\Sigma_{between}X$ is largest as a multiple of the within-group variance $X'\Sigma_{within}X$.

This algebraic analysis is called a *relative eigenanalysis*, and the resulting X is called the "first canonical variate." The second canonical variate is the linear combination with the largest such ratio among the set of all those perpendicular to the first, and so on, in analogy to ordinary principal components. Actually, the relative eigenvectors X arrive all at once in an ordered series by reduction of the problem to a conventional eigenextraction (Mardia et al., 1979).

In the present context, where there seem to be no meaningful mean differences between the exposed and the unexposed, the MANOVA computation per se makes little sense, but a closely related one turns out to be quite helpful: the relative eigenanalysis of the within-*exposed* covariance matrix Σ^d_{exp} with respect to the within-*unexposed* covariance matrix Σ^d_{unexp}. The resulting dimensions are the shape variables having the greatest variance in the exposed subsample as a multiple of the shape variance in the unexposed subsample, and so may serve to "concentrate" or "localize" our understanding of that excess variability.

There is a superscript d (^{d}) in the preceding formulas that requires some explanation. There are only 30 unexposed subjects of either sex, but more than 30 shape coordinates in either of the data sets. The relative eigenanalysis thus requires a restriction of dimension of the "smaller" matrix, the denominator of the ratio we're maximizing. We restrict our attention to subspaces of the first d principal components of the combined samples, exposed with unexposed. That is, for each dimensionality d we work in the subspace of the d most informative dimensions of sample shape variation as a whole. Thus the matrices Σ^d are $d \times d$, where d ranges from 6 to 20.

LITERATURE CITED

Anson BJ. 1950, 1961. An atlas of human anatomy. Philadelphia: Saunders.

Archibald SL, Fennema-Notestine C, Gamst A, Riley EP, Mattson SN, Jernigan TL. 2001. Brain dysmorphology in individuals with severe prenatal alcohol exposure. Dev Med Child Neurol 43:148–154.

Astley SJ, Clarren SK. 2000. Diagnosing the full spectrum of fetal alcohol exposure in individuals: Introducing the 4-digit diagnostic code. Alcohol Alcohol 35:400–410.

Bergman RA, Thompson SA, Afifi AK, Saadeh FA. 1988. Compendium of human anatomic variation. Baltimore: Urban and Schwarzenberg.

Bookstein FL. 1991. Morphometric tools for landmark data: Geometry and biology. New York: Cambridge University Press.

Bookstein FL. 1997a. Shape and the information in medical images: A decade of the morphometric synthesis. Comput Vision Image Understand 66:97–118.

Bookstein FL. 1997b. Landmark methods for forms without landmarks: Localizing group differences in outline shape. Med Image Analysis 1:225–243.

Bookstein FL. 1999. Linear methods for nonlinear maps: Procrustes fits, thin-plate splines, and the biometric analysis of shape variability. In: Toga A, editor. Brain warping. San Diego: Academic Press. p 157–181.

Bookstein FL. 2000. Creases as local features of deformation grids. Med Image Analysis 4:93–110.

Bookstein FL, Green WDK. 2002. User's manual, EWSH3,19. ftp://brainmap.med.umich.edu/pub/ewsh3,19.man.html

Bookstein FL, Schäfer K, Prossinger H, Seidler H, Fieder M, Stringer C, Weber GW, Arsuaga J-L, Slice DE, Rohlf FJ, Recheis W, Mariam AJ, Marcus LF. 1999. Comparing frontal cranial profiles in archaic and modern Homo by morphometric analysis. Anat Rec (New Anat) 257:217–224.

Bookstein FL, Sampson PD, Streissguth AP, Connor PD. 2001. Geometric morphometrics of corpus callosum and subcortical structures in the fetal-alcohol-affected brain. Teratology 64:4–32.

Bookstein FL, Streissguth AP, Sampson PD, Connor PD, Barr HM. 2002. Corpus callosum shape hypervariation covaries with neuropsychological deficits in adult males with heavy fetal alcohol exposure. Neuroimage 15:233–251.

Connor PD, Sampson PD, Bookstein FL, Barr HM, Streissguth AP. 2000. Direct and indirect effects of prenatal alcohol damage on executive function. Dev Neuropsychol 18:331–354.

Dryden IL, Mardia KV. 1998. Statistical shape analysis. New York: Wiley.

Good P. 2000. Permutation tests. 2nd ed. New York: Springer-Verlag.

Ikonomidou C, Birtigau P, Ishimanu MJ, Wozniak DF, Koch C, et al. 2000. Ethanol-induced apoptotic neurodegeneration and fetal alcohol syndrome. Science 287:1056–1059.

Lemoine P, Lemoine Ph. 1992. Avenir des infants de meres alcooliques (étude de 105 cas retrouvés à l'âge adulte) et quelques constatations d'interet prophylactique. Ann Pediatr 39:226–235.

Lester DS, Olds, JL. 2001. Biomedical imaging: 2001 and beyond. Anat Rec (New Anat) 265:37–53.

Loeser FJ, Alvord EC. 1968. Agenesis of the corpus callosum. Brain 39:553–570.

Löser H, Bierstedt T, Blum A. 1999. Fetal alcohol syndrome in adults: long-term observations on 52 patients. Dtsch Med Wochenschr 124:412–418.

Mardia KV, Kent JT, Bibby JM. 1979. Multivariate analysis. London: Academic Press.

Mardia KV, Bookstein FL, Moreton IL. 2000. Statistical assessment of bilateral symmetry of shapes. Biometrika 87:285–300.

Mattson SN, Riley EP, Gramling LS, Delis DC, Jones KL. 1998. Neuropsychological comparison of alcohol-exposed children with or without physical features of fetal alcohol syndrome. Neuropsychology 12: 146–153.

Olson ED, Miller RL. Morphological integration. 1958. Chicago: University of Chicago Press.

Rakic P, Kornack DR. 2001. Neocortical expansion and elaboration during primate evolution. In: Falk D, Gibson K, editors. Evolutionary anatomy of the primate cerebral cortex. New York: Cambridge University Press. p 30–56.

Riley EP, Mattson SN, Sowell ER, Jernigan TL, Sobel DF, Jones KL. 1995. Abnormalities of the corpus callosum in children prenatally exposed to alcohol. Alcohol Clin Exp Res 19:1198–1202.

Roebuck TM, Mattson SN, Riley EP. 1998. A review of the neuroanatomical findings in children with fetal alcohol syndrome or prenatal exposure to alcohol. Alcohol Clin Exp Res 22:339–344.

Sampson PD, Streissguth AP, Bookstein FL, Little RE, Clarren SK, Dehaene P, Hanson J, Graham J. 1997. Incidence of fetal alcohol syndrome and prevalence of alcohol-related neurodevelopmental disorder. Teratology 56:317–326.

Sampson PD, Bookstein FL, Streissguth AP, Barr HM. 2000. On categorizations in analyses of alcohol teratogenesis. Environ Health Perspect 108(Suppl 3):421–428.

Sowell ER, Thompson PM, Mattson SN, Tessner KD, Jernigan TL, Riley EP, Toga AW. 2001. Voxel-based morphometric analyses of the brain in children and adolescents prenatally exposed to alcohol. Cogn Neurosci Neuropsychol 12:515–523.

Spohr HL, Willms J, Steinhausen HC. 1993. Prenatal alcohol exposure and long-term developmental consequences. Lancet 341:907–910.

Stratton KR, Howe CJ, Battaglia FC, editors. 1996. Fetal alcohol syndrome: Diagnosis, epidemiology, prevention, and treatment. Washington, DC: National Academy Press.

Streissguth AP, Clarren SK, Jones KL. 1985. Natural history of the fetal alcohol syndrome: A 10-year follow-up of eleven patients. Lancet 2:85–91.

Streissguth AP, Aase JM, Clarren SK, Randels SP, LaDue RA, Smith DF. 1991. Fetal alcohol syndrome in adolescents and adults. JAMA 265:1961–1967.

Streissguth AP, Barr HM, Kogan J, Bookstein FL. 1996. Understanding the occurrence of secondary disabilities in clients with fetal alcohol syndrome and fetal alcohol effects. Seattle, WA: Fetal Alcohol and Drug Unit, University of Washington School of Medicine.

Swayze VW II, Johnson VP, Hanson JW, Piven J, Sato Y, Giedd JN, Mosnik D, Andreasen NC. 1997. Magnetic resonance imaging of brain anomalies in fetal alcohol syndrome. Pediatrics 99:232–240.

Toga AW, Thompson PM. Maps of the brain. 2001. Anat Rec (New Anat) 265:37–53.

Waddington CH. 1966. Principles of development and differentiation. London: Macmillan.

Weiss PA. 1973. The science of life. New York: Futura.

West JR, Chen NJA, Pantazis J. 1994. Fetal alcohol syndrome: The vulnerability of the developing brain and possible mechanisms of damage. Metab Brain Dis 9:291–322.

Williams RJ. 1956. Biochemical individuality. New York: John Wiley.

Exhibit 15

Attachment C

BYRON LEWIS BLACK, Petitioner

No. 3:00-0764

vs. Judge Campbell

RICKY BELL, Warden, Respondent

DECLARATION OF STEPHEN GREENSPAN, Ph.D.

Declarant, Dr. Stephen Greenspan, states:

Background and Focus of My Evaluation

I was retained by attorneys Kelley Henry and Michael Passino of the Office of the Federal Public Defender in Nashville to perform various tasks in order to render an opinion concerning the validity of the claim of their client, Byron Lewis Black, to have mental retardation (MR) and, thus, to be exempt from execution in light of the 2002 US Supreme Court ruling in Atkins v. Virginia. I am being compensated at the rate of $200 per hour, plus travel expenses, for my services in this case.

Byron Black is an African-American male who at the present time is within a week or two of his 52nd birthday. He is under a sentence of death for three homicides committed in 1988, when he was 32 years of age. In 2004, a hearing was held before Tennessee Circuit Court judge Walter C. Kurtz to determine whether Mr. Black was exempt from execution under Atkins as well as van Tran v. State (Tennessee, 2001). On May 5, 2004, Judge Kurtz ruled that Mr. Black did not have MR. It is my understanding that my role is to render an opinion, based on my review of documents as well as new data collected by me, concerning whether or not I believe the earlier conclusion (namely that Mr. Black does not have MR) was justified.

The main basis for Judge Kurtz's conclusion, as I understand it, was that Mr. Black did not appear to meet the third—"Developmental Criterion"—prong of the legal definition of MR. This prong requires that "significant deficits in intellectual functioning" (the first prong) and "deficits in adaptive functioning" (the second prong) need to have been present and

noted before the age of 18. With respect to the period before age 18, Judge Kurtz was unconvinced that Mr. Black met either the intellectual or adaptive functioning criteria. With respect to Mr. Black's status as an adult, Judge Kurtz stated that while it appeared that Mr. Black did meet the intellectual functioning prong, he was unconvinced that he met the adaptive functioning prong as an adult.

The main focus of my evaluation is on whether I believe that Mr. Black did or did not meet the intellectual and adaptive functioning criteria during the developmental period. In addition, I will render an opinion as to whether or not Mr. Black meets the adaptive functioning criterion as an adult.

<u>My Qualifications</u>

In the past four years, I have been qualified as an expert on MR and related cognitive disorders in four or five capital proceedings in the states of Arizona, California and Colorado. In addition, I have previously been qualified as an expert on MR in family court proceedings in New Jersey and Connecticut. I am a licensed psychologist in the state of Nebraska and was previously licensed in the state of Tennessee (current status: inactive). In addition to testifying in several so-called "Atkins" proceedings, I have been a consultant (and submitted declarations) in numerous other cases. Although my work thus far has always been at the request of attorneys representing defendants, I have found that a claim of mental retardation was unjustified in approximately half of the cases in which I actually examined a defendant (in contrast to other cases, in which my role was limited to educating the court about the nature of mental retardation and/ or opined about the adequacy of reports by other experts.)

I am a Clinical Professor of Psychiatry at the University of Colorado Health Sciences Center, and Emeritus (retired) Professor of Educational Psychology at the University of Connecticut. I received a Ph.D. in Developmental Psychology from the University of Rochester, and was a Postdoctoral Fellow in Mental Retardation and Developmental Disabilities at the University of California at Los Angeles' Neuro-psychiatric Institute. Before moving to Connecticut, I held academic appointments at the University of Nebraska and at George Peabody College of Vanderbilt University.

I have been elected "Fellow" (a designation given only to the most qualified members) by the Mental Retardation division of the American Psychological Association and by the American Association on Mental Retardation. I was also elected to a term as President of the Academy on Mental Retardation, which is the most prestigious research organization in the field. I have published extensively on MR, with particular emphasis on "adaptive behavior." I am a leading scholar in the MR field, as seen in the most recent diagnostic manual of the American Association on Mental Retardation (AAMR), AM. ASS'N ON MENTAL RETARDATION, MENTAL RETARDATION: DEFINITION, CLASSIFICATION AND SYSTEMS OF SUPPORTS (10th Edition, 2002) (hereinafter "the 2002 AAMR Manual"), which cited at least twelve publications by me, more than that of any other authority. My book WHAT IS MENTAL RETARDATION, co-edited with H. Switzky (AAMR; 2003; rev. ed. 2006) has, in a short time, become one of the most-quoted reference works in the field of mental retardation and has been described by Yale professor Edward Zigler as "the best book ever written about the definition and diagnosis of mental retardation." In 2008, AAMR recognized my contributions to the field by granting me its highest honor, the Gunnar and Rosemary Dybwad Award for Humanitarianism.

<u>Materials Examined and Activities Performed</u>

Expert reports or declarations examined:

- **Expert disclosure of Eric Engim, PhD dated July 2, 2003**
- **Declaration of Ruben Gur, PhD dated November 15, 2001**
- **Declaration of Daniel Grant, EdD, dated November 16, 2001**
- **Psychological Evaluation by Patti van Eys, PhD, dated March 28, 2001**
- **Report by Albert Globus, MD, dated November 14, 2001**
- **Report by Susan Vaught, PhD, dated May 2003**

Affidavits and Interviews from lay witnesses examined:

- **Affidavit of Arlita Black Swanson (sister), dated January 11, 2003**
- **Affidavit of Freda Black Whitney (sister), dated January 11, 2003**
- **Affidavit of Lynette Childs Black (sister), dated January 15, 2003**

- **Affidavit of Finis Black (uncle),, undated copy**
- **Affidavit of Alberta Black Crawford (sister), dated January 13, 2003**
- **Affidavit of Melba Black Corley (sister), dated January 11, 2003**
- **Affidavit of Mary Craighead (Elementary School Administrator) dated May 8, 2003**
- **Notes of Interviews with most of the above**
- **Notes of interview with Julia Mai Black (mother)**
- **Notes of interview with Renee Granberry, MD (cousin)**
- **Notes of interview with Richard Corley (co-worker and supervisor)**
- **Notes of interview with Rossi Turner (childhood friend)**
- **Notes of interview with Bart Tucker (high school counselor)**
- **Notes of interview with Karen Greer (sister)**

Other Documents examined:

- **Elementary and Secondary School grade reports for Byron Black**
- **Memorandum and order by Judge Walter C. Kurtz, dated may 5, 2004**
- **Independent Living Scale manual and record form (faxed from Dr. Grant)**

Activities Performed:

- **In-person Interview with Al Harris (former high school football coach)**
- **Phone interview with Mary Black (aunt by marriage)**
- **In-person interview and Vineland adaptive behavior assessment with Rossi Turner**
- **In-person joint interview and Vineland adaptive behavior assessment with Melba Black Corley and Freda Black Whitney**
- **In-person interview and assessment of Byron Black**
- **Phone interview with Dr. Daniel Grant (regarding the Independent Living Scale)**

Criteria To Use in Diagnosing Mental Retardation

As described in my widely-cited book WHAT IS MENTAL RETARDATION? (American Association on Mental Retardation, 2006), MR is not always an easy diagnosis to make, especially with individuals in the range of mild MR, where virtually all Atkins applicants are likely to be found. In this brief discussion, I shall discuss the three prongs to be used in diagnosing MR, emphasizing both the letter and the spirit of these prongs.

Virtually all legal definitions of MR used in the US are derived from either or both of the diagnostic manuals published by the American Association on Mental Retardation (AAMR, recently renamed the American Association on Intellectual and Developmental Disabilities) and the American Psychiatric Association, through its "Diagnostic and Statistical Manual" (DSM). The AAMR diagnostic manual has gone through several revisions, with the most recent being the tenth edition (AAMR-10), published in 2002. DSM has also gone through several revisions, with the most recent being the text-revised fourth edition (DSM-4TR), published in 2000. Starting with DSM-3 (1980), the definition of MR contained in each version of DSM has been derived entirely, except for minor wording changes, from the most current AAMR manual. Thus, the definition of MR contained in the 2000 DSM-4TR is derived from the 1992 AAMR-9, while it is highly likely that the definition of MR in the forthcoming DSM-5 will be nearly identical to the definition of MR contained in the 2002 AAMR-10. Therefore any differences in the definitions of MR in DSM and AAMR manuals reflect the fact that the most recent DSM manual pre-dates the most recent AAMR manual, and does not reflect substantive or philosophical differences between the two organizations.

The definitions of MR in the AAMR and DSM manuals contain two parts: a conceptual (abstract) definition, followed by an operational (concrete) definition. While the operational definitions of MR have changed somewhat over the years, the conceptual definitions have remained essentially unchanged since they were first formulated by AAMR over 45 years ago, in the fifth edition of its manual, published in 1961.

The conceptual definition of MR, as reflected in both AAMR and DSM manuals, and in statutes and court opinions in Tennessee and most other states, has three parts: (a) deficits in intellectual functioning, (b)

concurrent deficits in adaptive functioning (also known as adaptive behavior), and (c) evidence of the disorder before the onset of adulthood. As stated above, these conceptual criteria have remained essentially unchanged in various AAMR and DSM editions.

One difference between DSM 4-TR and AAMR-10 is that DSM 4-TR emphasizes "significantly subaverage intellectual functioning" and "concurrent deficits or impairments in present adaptive functioning" while AAMR-10 emphasizes "significant limitations in intellectual functioning and in adaptive behavior".

The Tennessee statute (TCA-39-13-203) defining MR in criminal cases is aligned more closely with DSM 4-TR, in that it emphasizes "deficits" in adaptive functioning rather than "significant deficits". Specifically, the statute reads: "…Mental Retardation means significant subaverage general intellectual functioning …, deficits in adaptive behavior … [and it] must have been manifested during the developmental period…"

This difference between "deficits" and "significant deficits" is more than a semantic distinction, in that it has implications for the operational definition that follows. The difference is that AAMR-10 applies the same criterion (approximately two standard deviations below the mean, or the second percentile of the population) for both intelligence and adaptive behavior, while DSM 4-TR applies the two standard deviation criterion only for intellectual functioning but does not specify any statistical criterion for meeting the second prong of the definition. Thus, "significant deficit" implies a more stringent criterion (typically set at the second percentile of the population) while "deficit" or "impairment" implies a much less stringent criterion, which if it is specified (not the case with DSM 4-TR or the Tennessee statute) is typically set at approximately one standard deviation below the mean (a standard score of 85, which indicates a percentile rank of about the 16th percent of the population).

The operational criteria for diagnosing MR, and the complications involved in applying them in this particular case, are discussed briefly in the following three sub-sections and in the Findings section that follows those.

(1) <u>The Intellectual Criterion</u>. MR is a disorder whose core impairment is in the area of intelligence. This construct is typically measured through one's performance on an individually-administered test of intelligence which results in a full-scale IQ score that locates one's functioning in relation to the mean for the general population. IQ tests are constructed so that the population mean is set at a score of 100, with a standard deviation (an index of statistical variability) of 15. The ceiling for MR is currently established as "approximately two standard deviations below the population mean". The term "approximately" refers mainly to the fact that no test is fully reliable and one should take various factors into account when interpreting a test number. The main thing to take into account is the fact that test scores vary approximately five points around one's "true score". As two standard deviations (2 x 15) equals 30 points, the upper IQ level for meeting the intellectual criterion for MR is 75 (100 minus 30 plus 5 [the reliability index]). In addition, one should take into account factors such as practice effect (possible learning from taking a second test too soon), changes in and adequacy of test norms, and possible malingering.

One of the factors to take into consideration when interpreting IQ scores is what has been termed the "Flynn effect". This term refers to the fact that the overall population has been gaining in performance on IQ tests at a rate of 3 points per decade (0.3 points per year), and this finding is taken into account by test developers when they develop new test editions every few years, in that the norms are toughened. Because a diagnosis of MR could be affected significantly depending on when in a test's cycle a person is tested, the Flynn effect has been used to adjust Full Scale IQ scores using the following formula: (a) subtract the year of the of the test's publication (or, ideally, when the norms were compiled, which typically is two years earlier) from the year a test was administered; (b) multiply this figure by 0.3; (c) subtract this figure from the person's obtained IQ score, with the resulting number being the Flynn-adjusted score.

Thus if someone was tested in 1990 on a test normed in 1978 and received an IQ score of 78, one would multiply 12 (1990-1978) by 0.3, with the resulting number being 3.6. Subtracting 4 points (the rounded sum) from 78, one would receive an adjusted IQ score of 74. A discussion of the Flynn effect in diagnosing MR is contained in a paper by me (Stephen Greenspan, Spring 2006. Issues in the use of the Flynn Effect to adjust IQ scores when

diagnosing MR, which appeared recently in PSYCHOLOGY IN MENTAL RETARDATION AND DEVELOPMENTAL DISABILITIES, which is the official publication of the mental retardation Division of the American Psychological Association. As indicated in that paper, the Flynn effect adjustment formula when diagnosing MR has been accepted as a legitimate practice by state and Federal trial courts (e.g., Walker v. True, 399 F.3d 315, 322-32, 4th Cir. 2005). It is also beginning to be recognized in various appellate courts. As example, on February 28, 2007 the U.S. Navy-Marine Corps Court of Criminal Appeals stated: "In determining whether an offender meets this definition [of MR], standardized IQ scores scaled by the SEM and the Flynn effect will be considered" (web: NMCCA, code 07).

To summarize, the phrase "approximately two standard deviations below the population mean on a standardized test of intelligence" means that one should not rely rigidly on an IQ score number, but should take into account the adequacy of the test, the nature and meaning of the norms, the context in which the test was administered, ethnic and linguistic factors, etc. This is the main use for "clinical judgment" in diagnosing MR. As noted in the book CLINICAL JUDGMENT (AAMR, 2006) by Robert Schalock and Ruth Luckasson (two of the main authors of AAMR-10), clinical judgment in diagnosing MR is <u>not</u> a matter of relying on intuition or gut feeling (which can be misleading, especially in unqualified clinicians) but rather involves using test scores in a thoughtful and scientifically valid manner. A rigid reliance on a test score, without such thoughtfulness, can and often does result in "false positives" (wrongly concluding someone has MR when he does not) or "false negatives" (wrongly concluding someone does not have MR when he does".)

Although a clinician diagnosing MR should not rely on gut feeling (which can vary from clinician to clinician), the notion of clinical judgment (which is relied on heavily in reaching any diagnosis in the human services, not just MR) requires the clinician to interview and have some personal contact, however brief, with the person he or she is diagnosing. This is a matter of basic professional ethics and practice. In the 2004 state court MR hearing both of the two prosecution psychologists testified that they did not believe Mr. Black to have MR, in spite of their never having interviewed or even laid eyes on him. To me, such a "paper diagnosis" lacks credibility and serves to undermine the validity of their findings.

Because in the past, clinicians often relied rigidly and mindlessly on an IQ number, and particularly failed to rake into account the five-point standard error of test scores, AAMR-10 operationally defined approximately two standard deviations below the mean as "a score below 70-75". This indicates that clinicians or agencies making a determination of MR solely on whether a score is below or above 70 are not engaging in acceptable practice. Raising the ceiling from 70 into 70-75 also reflected a policy decision that past manuals, in their concern to eliminate false positives had defined the MR class too narrowly and some loosening of the criteria needed to be undertaken to avoid the now-widespread problem of false negatives.

DSM 4-TR (which preceded AAMR-10) does not use the 70-75 formula. However, it is stated quite clearly that one should take into account standard error of the test and not just rely rigidly on the obtained score. In addition, both AAMR-10 and DSM 4-TR indicate that there are circumstances where reliance on a single "full-scale" IQ score can be misleading. Specifically, it is well-known that individuals with known brain damage syndromes present a mixed pattern of intellectual competence and incompetence, and summarizing across to obtain a single score can serve to obscure the true nature and extent of an individual's impairment. In such circumstances, one must be especially careful to go beyond just full-scale IQ and look at other (sometimes more qualitative) sources of data where these are available and useful.

Finally, the emphasis in both AAMR-10 and DSM 4-TR is on use of individualized and adequately standardized measures, and not on group administered and/ or brief screening instruments. There are only a few such individualized instruments suitable for diagnosing MR, such as the Wechsler scales (WAIS-3), the Stanford-Binet (SB-5), the Woodcock Johnson cognitive battery, etc. Group measures are not acceptable for ruling MR in or out for several reasons, the two most important being: (a) their much weaker reliability and validity, and (b) lack of information about the circumstances of administration (e.g., the possibility that someone may have received help, not been paying attention, etc).

(2) <u>The Adaptive Behavior Criterion</u>. For over the past 45 years, it has no longer been considered adequate to rely solely on IQ scores in determining whether one has or does not have MR. This is because IQ test scores, particularly in the "mild" level of impairment, do not always translate to other settings, and a diagnosis of MR should indicate a fairly global impairment affecting many areas of functioning. Thus, to qualify for a diagnosis of MR, one should show significant deficits in both IQ and "adaptive behavior". The current conceptualization of adaptive behavior relies on a "tripartite model" of intelligence and adaptive functioning that I developed over 25 years ago, and uses my work as the basis. This model has three parts: (a) "conceptual" adaptive skills (understanding academic processes); (b) "practical" adaptive skills (understanding physical processes) and (c) "social" adaptive skills (understanding people and social processes). In determining if someone meets the Adaptive Behavior criterion, it is necessary to show significant deficits in only one of these three areas (AAMR-10). Sources of data can come, preferably, from formal test scores on rating instruments (such as the Vineland or ABAS) administered to informants, supplemented sometimes by formal test scores on individually administered measures (such as the Street Smarts Survival Questionnaire), and from qualitative information gathered from affidavits, records, and observation by an evaluator.

The 2002 AAMR manual specified that the most important source of information regarding whether an individual meets the adaptive behavior criterion is whether one falls approximately two standard deviations (i.e., a standard score below the 70-75 range) on a standardized rating measure of adaptive behavior such as the Vineland. Two pathways to meeting the AAMR's adaptive behavior criterion were offered: (a) a standard score below 70-75 on an overall (composite) score, or (b) a standard score below 70-75 on at least one of the three adaptive skill areas of conceptual adaptive skills, practical adaptive skills or social adaptive skills.

In establishing the possibility of being above 70-75 in one or even two of the three adaptive skill areas (or having good scores on particular items within sub-average adaptive skill areas), the AAMR wished to emphasize that having mild MR is not incompatible with being able to do many things, such as drive a car, hold a job, be married, have relatively normal language and (even) commit crimes that may require some degree of planning and volition.

In its Users Guide, which is a supplement to the 2002 Manual and written by the same authors, the AAMR indicates that in high stakes assessments, such as an Atkins hearing, the use of retrospective ratings of adaptive behavior is often necessary, and is justified in such cases. In such retrospective ratings, raters are asked to rate an individual not as he is today but as he was at the time when the rater knew him best, living in the community. Retrospective ratings are needed because the current setting (e.g., Death Row) does not provide opportunities to assess success or failure in more typical roles (e.g., worker) or tasks (e.g., operating appliances or dealing with neighbors). Also, MR is a disability that can best be understand as a need for supports in fulfilling such community roles and tasks. Another reason for retrospective assessment of adaptive behavior is because such assessments may not have been carried out during the Developmental period and retrospective assessment helps to establish if the individual had significant impairments during that period.

As already mentioned, one operational difference between AAMR-10 and DSM 4-TR, in terms of adaptive behavior/ functioning, is that DSM uses the words "limitations" and "deficits", implying either no statistical cutting score or, at most, a minus one SD (standard score of 85) criterion. AAMR-10, on the other hand, uses the words "significant deficits", implying minus two SDs (standard score below 870-75), although as mentioned, this can be accomplished either in terms of an overall adaptive composite (quotient) of 70-75 or less, or such a score in only one of the three domains of "social", "practical" or "conceptual" adaptive skills.

In DSM 4-TR, the criterion for adaptive functioning (the term this manual prefers, but which means the same thing as adaptive behavior) is defined as deficits in at least two out of eleven functional areas: communication, self-care, home living, social/ interpersonal skills, use of community resources, self-direction, functional academic skills, work, leisure, health and safety. This list is derived from AAMR-9 (1992), which was published eight years before DSM 4-TR. In AAMR-9, the adaptive behavior criterion was established as deficits in 2 out of 10 adaptive skill areas (health and safety were combined into one area) or deficits in overall composite adaptive quotient. In AAMR-10, these ten (11 in DSM 4-TR) skill areas were collapsed into the three adaptive behavior domains (social, practical, conceptual) mentioned above.

In the Tennessee statute (TCA-39-13-203), the adaptive behavior criterion (which is described simply as "deficits in adaptive behavior"), is stated globally and is not broken down into component skills or domains (unlike DSM 4-TR's 11 skills and AAMR-10's 3 domains). Because of that globality, and also because the standard is "deficits" rather than "significant deficits", the Tennessee definition appears to offer considerable flexibility (including the use of non-statistical data) in determining whether or not someone meets the adaptive behavior criterion.

(3) <u>The Developmental Criterion</u>. MR is a term indicating that an individual has serious intellectual impairments which first manifested during what is termed the "developmental period". The developmental period is defined as anytime between birth and 18 (some interpret this as before the end of one's 18th year). The purpose of this criterion is to rule out those who were normal in childhood but whose impairments first manifested in adulthood, such as through a motor vehicle accident. Information about whether one meets the developmental criterion can come from a variety of sources, such as medical or school records and testimony by teachers, family members and peers.

One of the controversies in interpretation of the developmental criterion involves whether or not the individual must have been eligible for a diagnosis of MR before the age of 18. This appears to have been the standard used by Judge Kurtz, but it my respectful view that he was mistaken in making that interpretation. If one takes that tack, then one can use the absence of any IQ score, or adaptive behavior score, before the age of 18 as evidence that would rule out a current diagnosis of MR. In my view, this is an incorrect, and overly rigid, interpretation of the developmental criterion.

A more appropriate, and flexible, interpretation of the developmental criterion is that when a person qualifies as having MR as an adult, one should be able to show that there were precursors or indicators that developed or were evident during the childhood or adolescent period. In other words, a diagnosis of MR would be inappropriate if a child was of average or above average intellectual and adaptive functioning prior to 18 but suddenly showed a steep decline, perhaps because of some injury that developed during adulthood. Outcome-based evidence, such as a child being retained in elementary school (which occurred in this case) and very

low academic achievement (also true in this case) can also be used as evidence that the developmental criterion has been met.

A related issue has to do with evidence of organic (i.e., biological) etiology, such as diagnosed brain damage that is most likely attributable to a developmental process that started early in life. To establish mild MR (which is the sub-category most relevant in this case), one does not have to have evidence of a known etiology, and such evidence is typically lacking. However, such evidence-when it exists—can by itself be used to satisfy the developmental criterion. A good example of this is if there is brain imaging evidence that is highly suggestive of neurological abnormalities indicative of Fetal Alcohol Spectrum Disorder (a major known cause of mild MR). Where such evidence exists (as it does in this case), this could also be used to buttress the conclusion that the third prong for a diagnosis of MR has been met.

<u>My Findings Regarding Whether Byron Black Has MR</u>

It is my conclusion that Byron Black qualifies for a diagnosis of mild MR. My reasons flow from my finding that he meets all three of the definitional prongs. These are discussed under each of the prongs below.

(a)<u>Intellectual Functioning Prong</u>. In adulthood, it is clear that Mr. Black meets the intellectual functioning prong of a diagnosis of MR. In November 2001, Dr. Daniel Grant obtained a full-scale IQ on the Stanford-Binet (SB-4) of 57. On the C-TONI, the best non-verbal IQ test which correlates highly with full-scale IQ, Dr. Grant obtained an IQ score of 64. In October 1993, Dr. Gillian Blair obtained a WAIS-R full-scale IQ score of 73, which is under the 70-75 ceiling. The WAIS-R was normed in 1979 and was, thus, 14 years obsolescent in 1993. A Flynn adjustment would reduce this IQ score by 4 points (0.3 for each year of norm obsolescence), bringing it to 69. In 1997, Dr. Pamela Auble also used the WAIS-R and obtained a full-sale IQ score of 76, which would be reduced another 6 points (for the 18 years of norm obsolescence). In March, 2001, Dr. Patti van Eys administered the more current WAIS-3 and obtained a full-scale IQ of 69, which is under the 70-75 cutting score, and very much in line with the Flynn-corrected scores for the outdated WAIS-R.

Thus, the overwhelming consensus among all of these individualized IQ administrations is that Mr. Black meets the first intellectual functioning) prong for a diagnosis of MR as an adult.

Individualized IQ data for Mr. Black as a child is lacking, for the simple reason that he left high school in the very same year that the federal statute (PL-94-142) that mandated special education was enacted. During the time that Mr. Black was in elementary school, the assumption was that a child would be socially promoted if he was well-behaved (which by all accounts, Mr. Black was), regardless of how little he learned (see Affidavit by Mary Craighead, an administrator at Mr. Black's elementary school). Just the same, Mr. Black was retained in the second grade, even given that tendency to overlook such learning difficulties. Undoubtedly, an individualized IQ test would have been administered had Mr. Black been born ten years later. The absence of such IQ data makes it impossible to know whether he would have qualified for a diagnosis of MR during that period.

Mr. Black's relatively good report cards in elementary school are incongruent with the fact that he was retained and also with his marginal or failing grades in High School. The mystery is cleared up when reading the statements by his fifth and sixth grade teachers (noted in point #17 in the declaration by Dr. Grant). They stated that "I would never allow a student to get a bad grade" (6th grade teacher) and "teachers were liberal in their grading" and a B would be the equivalent of a D at a later time (5th grade teacher). Furthermore, administrator Mary Craighead indicated in her affidavit that the emphasis back then was on helping low-achieving African-American children to feel good about themselves and to experience success in all of their endeavors.

This attitude likely also explains why Mr. Black obtained relatively high scores on group administered IQ tests, as it is very possible, indeed likely, that these tests (which even state experts testified are not appropriate for diagnosing MR) were administered in a non-standard manner that could even have involved teacher assistance.

Even so, it should be noted that the IQ criterion for diagnosing MR was minus 1 SD (full-sale score of 85), during the years 1961 to 1973, and that the 85 that Mr. Black obtained on the Otis-Lennon group IQ test could, thus, have qualified him at that time.

Dr. Grant correctly noted that the best evidence that Mr. Black would have met the MR intellectual functioning criterion in the Developmental period was his very low performance (standard scores of 71 and 67) on the Differential Abilities Test (DAT). Although not specifically termed an IQ test, the DAT correlates very highly with IQ and in the absence of an IQ test can be used as a substitute. Furthermore, Mr. Black's mostly failing grades in High School (where the overprotective stance of his elementary school no loner applied) is probably a better indicator of the depth of his intellectual limitations. Those limitations carry over today into his very low achievement standard score (72) as an adult on the WRAT-III and the Nelson-Denny reading test.

In short, Mr. Black gave clear evidence of intellectual limitations in the developmental period, and there is continuity rather than discontinuity linking his intellectual limitations today and his intellectual limitations as a child.

(b)<u>Adaptive Functioning Prong</u>. The main focus of my evaluation of Byron Black was on his level of adaptive functioning. That is because he appears, as summarized above, to meet the intellectual criterion, but questions were raised by Judge Kurtz regarding whether he met the adaptive functioning criterion either currently, or more specifically, prior to the age of 18.

Adaptive Behavior is most typically evaluated through a rating instrument, such as the ABAS-2 or the Vineland-2 (the two instruments which, along with the SIB, are most widely used in Atkins cases). Using a rating instrument to evaluate the adaptive functioning of someone who has been in prison, especially death row, for a number of years is difficult, if not impossible, for a number of reasons. These reasons include the difficulty in finding raters but more importantly, the absence of opportunities to perform many of the behaviors (such as cooking or using public transportation) that are items on such instruments. Furthermore, the whole purpose underlying the development of these instruments is to assess the supports needed to live successfully in the community, and to face the kinds of challenges and ambiguities one would find in the community. Obviously, death row is a setting that provides few such challenges and ambiguities.

A common mistake that is often made when evaluating the adaptive functioning of someone in prison is to look at his level of adjustment, such as through the presence or absence of discipline write-ups. Some experts, usually those testifying for the state, will look at a defendant who is not a discipline problem and conclude that he could not have MR. The problem with such a conclusion is that adjustment in prison is typically a matter of whether or not one has a cooperative versus hostile personality, and being a cooperative and pleasant person in no way rules out MR. In fact, it is likely the case that people with mild MR, assuming they do not also have mental illness, will tend to be more apt to go along with rules and orders, in part because such a tendency generally served them well in covering up their limitations in work, school and other settings in the community. Furthermore, there are relatively few choices one has to make on death row, and the rules are few, clear and unambiguous. So it is fair to say that people with mild MR are likely to adjust better in a highly structured setting such as death row, and such adjustment in no way can be used to infer how impaired one's adaptive functioning would be in the community.

For these reasons, to assess one's level of current adaptive functioning in prison, one would most likely have to rely on the few "direct" measures of adaptive functioning, such as the "Independent Living Scales" (ILS) used by Dr. Grant, or the "Street Survival Skills Questionnaire" (SSSQ) used by me. Both measures are direct in the sense that one presents everyday problems to a subject (such as filling out a bank deposit slip, or figuring out a paycheck) and seeing whether the subject passes such items. Both the ILS and the SSSQ are mainly measures of the "Practical Adaptive Skills" domain of adaptive functioning, and they have population norms.

Dr. Grant stated in his report that Mr. Black received a standard score in the 70-75 range on three of the five ILS sub-scales that, together, give information about the adaptive behavior domain of "Practical Adaptive Skills". These sub-scales are labeled "managing money" (standard score of 73), "managing home and transportation"(standard score of 73), and "health and safety" (standard score of 72). He was in the normal range on two other ILS sub-scales that, in my view, are unrelated to MR: memory and "social". The reasons why the social sub-scale on the ILS is not diagnostically relevant are two-fold: (a) it mainly taps happiness/ agreeableness which I have already noted is not indicative one way or the other of MR, and (b) it involves solely self-report (rather than problem-

solving) and self-report is notoriously unreliable as a source of diagnostic information in people with MR (who almost universally inflate their description of themselves in order to appear competent (this well-established phenomenon is termed "the cloak of competence". See the classic book of the same name by UCLA Professor Robert Edgerton).

As an independent validation of Dr. Grant's ILS data, I administered the SSSQ, another direct measure of adaptive behavior that mainly taps Practical Adaptive Skills. This test has over 200 items in which a subject is presented with an object or process and then picks the correct one out of four pictures that depicts the object or process. Mr. Grant obtained an overall standardized score (78) which is highly congruent with the 73, 73 and 72 standard scores obtained by Dr. Grant on three relevant sub-scales and certainly meets the "deficit" or "impairment" (minus one SD) standard implicit in DSM 4-TR and in TCA-39-13-203. Also, I found that Mr. Black was below the minus 2 SD standard on three of the nine SSSQ sub-scales and below the minus one SD standard on a fourth.

Before testing Mr. Black on the SSSQ, I administered the Dot Counting Test, which is one of the most used and respected measures of possible malingering on cognitive tasks. This test shows pictures with dots and the task is to count them correctly and in a short period of time. Mr. Black made zero mistakes, and this fact plus the very short average time per picture gave very strong indication that he approached the testing situation in a fully attentive and effortful manner. Thus, I concluded that the SSSQ scores were highly valid and lacked any indication of malingering.

Qualitative data suggesting Mr. Black met the adaptive behavior criterion in adulthood (but prior to conviction in this case) are that he never lived independently (lived with parents, even after marriage), never had a check book, never cooked, never washed his clothes, never did anything suggestive of adult status other than holding a job (which most adults with mild MR do) and driving a car (which many individuals with mild MR do, as suggested in the AAMR criterion of significant impairment in only one out of three domains). Another indication of Mr. Black's impaired adaptive status came from my interview with his high school football coach, Al Harris, who indicated that in over 30 years as a coach, Mr. Black stood out as especially slow. He indicated that although Byron had good physical skills, he could generally not be used on offense for the reason that he could

not learn the plays and was used on offense only when a highly simplified playbook was developed for his use.

Because lack of evidence of adaptive incompetence before the age of 18 appeared to be a major issue in Judge Kurtz's ruing, I conducted a retrospective assessment of Mr. Black's adaptive functioning, using the age 17 years-six months as the target age. I used the most widely-used and respected adaptive behavior rating instrument, the Vineland-2. This instrument is published by Pearson Assessment, the publisher of the most widely respected intelligence test, the Wechsler Scales, and is the publisher that adheres to the highest standards for test development.

The Vineland-2 is filled out by an examiner after each interview with one or more informants. I conducted two such interviews, one with a boyhood friend, Rossi Turner, who knew Mr. Black until he left Nashville to go to school outside the state, and a joint interview with two sisters: Melba Black Corley (older sister) and Freda Black Whitney (younger sister). In the latter interview, I asked for consensus between the two sisters before scoring each item and generally such consensus was obtained. I should note that all three informants hold responsible professional jobs and appear to be people of average or above average intelligence. All three of them indicated they knew Mr. Black very well during the age period (17-6) being rated.

The Vineland-2 labels its domains somewhat differently than does AAMR-10, but they are generally equivalent. The three domains on the Vineland-2 are: "Communication" (which taps basically what AAMR-10 calls "Practical Adaptive Skills"; "Daily Living Skills"(which taps what AAMR-10 calls "Practical Adaptive Skills") and "Socialization" (which taps what AAMR-10 calls "Social Adaptive Skills"). In addition, one sums across all of the items on the scale to obtain a Composite (overall) adaptive quotient.

The standard scores obtained on the Vineland-2 were as follows:
On Communication (Conceptual Adaptive Skills), Mr. Black received a standard score of 75 on the Vineland based on interview with the sisters, while he obtained an identical score on the Vineland based on interview with Mr. Turner.

On Daily Living (Practical Adaptive Skills), Mr. Black received a standard score of 76 on the Vineland based on interview with the sisters, while he obtained a standard score of 71on the Vineland based on interview with Mr. Turner.

On Socialization (Social Adaptive Skills) Mr. Black received a standard score of 63 on the Vineland based on interview with the sisters, while he obtained a standard score of 67 on the Vineland based on interview with Mr. Turner.

On overall Composite Adaptive Behavior, Mr. Black received a standard score of 70 on the Vineland based on interview with the sisters, while he obtained an identical standard score of 70 on the Vineland based on interview with Mr. Turner.

In short, Mr. Black met the AAMR-10 criterion of significant (minus two SD) deficit on adaptive behavior on both sets of Vineland ratings, and he also met the AAMR criterion of significant (70-75 or below) on one out of three domains. Using the somewhat less stringent standards embedded in DSM 4-TR and the Tennessee statute, his qualification is even more clear-cut.

(c)<u>Developmental Prong.</u> As indicted earlier, this prong can be interpreted as either meaning that one must show evidence that could cause a diagnosis of MR to be met prior to 18 (Judge Kurtz's apparent interpretation) or rather only evidence that adult impairments can be traced to indicators of failure, low functioning or causation evident prior to 18 (my interpretation).

Using the looser interpretation, there is no doubt in my mind that Mr. Black satisfies this prong. Although he attended an elementary school considered the most disadvantaged and low-functioning in the district (as reflected in its being chosen for a special Ford Foundation program), Mr. Black was made to repeat second grade, which is a clear indication that he was considered to be very "slow" even in that much slower than average setting. There is also very clear evidence from standardized achievement scores that Mr. Black functioned intellectually at a very low level.

Finally, very powerful evidence that Mr. Black meets the developmental criterion can be found in the very clear-cut evidence obtained by Dr. Gur of structural damage to his brain (abnormal corpus colussum, or mid-brain, seen in MRI image) suggestive of Fetal Alcohol Spectrum Disorder).

Using the more stringent approach to the Developmental criterion apparently used by Judge Kurtz, I believe Mr. Black also meets the developmental criterion, defined in TCA-39-13-203 as "the MR must have been manifested during the developmental period, or by eighteen (18) years if age". The main evidence that could be pointed to as suggesting that Mr. Black was of normal intelligence were the group IQ scores, but these are unreliable tests that cannot be substituted for individualized tests which were not routinely administered (because special education had not yet been federally mandated). Furthermore, the atmosphere at that time was one of helping children such as Byron Black to have feelings of success and it is possible, indeed likely, that he was given assistance with those tests. The Differential Aptitude Test given in 9th grade, and which showed scores under the 70-75 ceiling, along with mostly failing grades in High School are much stronger evidence of the extent of Mr. Black's limitations during the period before he turned 18.

Conclusion

It is my professional opinion, to a high degree of psychological certainty, that Byron Lewis Black meets all three criteria for a diagnosis of mild MR, whether using DSM 4-TR, AAMR-10 or TCA 39-13-203.

FURTHER DECLARANT SAITH NOT.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

Dated: March 13, 2008

Stephen Greenspan, Ph.D.

Attachment D

DECLARATION OF MARC J. TASSÉ, PhD, FAAIDD

I, Marc J. Tassé, declare under penalty of perjury and the laws of the United States, the following to be true to the best of my information and belief:

1. My name is Marc J. Tassé, Ph.D., FAAIDD and I am a licensed psychologist in North Carolina (NC #2613). I completed my Ph.D. in research-clinical psychology at the Université du Québec à Montréal. My doctoral dissertation focused on the study of adaptive behavior assessment in individuals with mental retardation. Following my Ph.D., I completed a post-doctoral fellowship in mental retardation and developmental disabilities at The Ohio State University Nisonger Center, University Center for Excellence in Developmental Disabilities Education, Research, and Service. I am also a "Fellow" of the American Association on Intellectual and Developmental Disabilities.

   I am an Associate Professor in the Department of Child and Family Studies at the University of South Florida (USF). I am also the Associate Director of the USF Florida Center for Inclusive Communities (FCIC). The USF FCIC is a federally funded University Center for Excellence in Developmental Disabilities. Our Mission is three-fold: (1) provide training to undergraduate, graduate and post-graduate students in the field of mental retardation and related developmental disabilities (MR/DD), (2) offer services and state-wide technical assistance to individuals with MR/DD across the age span and to agencies providing supports and services to these individuals, and (3) conduct research in the field of MR/DD.

   I've worked with individuals with mental retardation for the past 20 years. I have provided direct clinical services as well as supervised graduate and post-graduate psychology students in providing direct services to individuals with MR/DD. I've been involved in hundreds of psychological assessments and eligibility/diagnostic evaluations of mental retardation involving children, adolescents, and adults. I have worked extensively over the past 20 years directly with individuals with mental retardation of all ages. I have provided consultative services and technical assistance to families, service providers, and state MR/DD agencies. Over the past 10 years, I have also been involved in providing individual therapy to adolescents and adults with mental retardation and co-occurring psychiatric disorders or complex behavior problems.

   In the past (i.e., 1985 to 1993), I also worked as a behavior specialist (Douglas Hospital; Montreal, Canada), providing behavior programming and developing intervention plans for children and adults with mental retardation and co-occurring behavior problems or psychiatric disorders.

   In addition to my clinical work, I actively conduct research in the field of mental retardation. I have published over 65 book chapters, peer-reviewed journal articles, and monographs in the area of mental retardation or developmental disabilities. I have given over 100 presentations, workshops, or seminars at local, state/provincial, national, and international scientific/professional meetings in the field of mental retardation.

I am a co-author on the American Association on Intellectual and Developmental Disabilities (AAIDD; formerly known as the American Association on Mental Retardation) 2002[1] Manual that defines mental retardation and the recently published AAIDD User's Guide (Schalock et al., 2007)[2]. I have also worked on the development of standardized tests in the field of mental retardation. One such assessment instrument was the *Supports Intensity Scale* (SIS). The *SIS* is a standardized measure of individual support needs for adolescents and adults with mental retardation. I have also worked on the development and refinement of the Quebec Adaptive Behavior Scale, as well as other standardized assessment instruments in the area of measuring problem behavior and psychopathology in individuals with mental retardation. I currently Chair the American Association on Intellectual and Developmental Disabilities' *ad hoc* committee on the development of the Diagnostic Adaptive Behavior Scale (DABS). The DABS has been in development for approximately three years and should result in a standardized test of adaptive behavior that will focus on diagnosing the presence of "significant adaptive behavior deficits" for the purpose of diagnosing mental retardation. I was recently awarded the "Service" award by the American Association on Intellectual and Developmental Disabilities for my work with individuals with mental retardation and complex behavior support needs.

I am an active member of the following professional associations:

- American Association on Intellectual and Developmental Disabilities (Fellow)
- American Psychological Association [member of Divisions: 5 (Assessment), 33 (I&DD), 41 (Psychology & Law Society)]
- International Association for Behavior Analysis
- National Association for the Dually Diagnosed (MR/MI)
- North Carolina Psychology Board of Psychologists (License #2613)

I am an *ad hoc* reviewer for the following professional journals:

- American Journal on Mental Retardation
- Intellectual and Developmental Disabilities
- International Clinical Psychopharmacology
- Journal of Autism and Developmental Disorders
- Journal of Intellectual Disability Research
- Research in Developmental Disabilities
- Revue francophone de la déficience intellectuelle

2. I was asked by Attorneys Kelley Henry and Michael Passino, on behalf of their client Mr. Byron Black (D.O.B.: 3/23/1956), to do the following:

[1] Luckasson, R., Borthwick-Duffy, S., Buntinx, W. H. E., Coulter, D. L., Craig, E. M., Reeve, A., Schalock, R. L., Snell, M. E., Spitalnik, D. M., Spreat, S., & Tassé, M. J. (2002). *Mental retardation: Definition, classification, and system of supports.* Washington, DC: American Association on Mental Retardation.

[2] Schalock, R. L., Buntinx, W. H. E., Borthwick-Duffy, S., Luckasson, R., Snell, M. E., Tassé, M. J., & Wehmeyer, M. L. (2007). *User's Guide Mental Retardation: Definition, Classification, and Systems of Supports, 10th Edition. Applications for Clinicians, Educators, Disability Program Managers, and Policy Makers.* Washington, DC: American Association on Intellectual and Developmental Disabilities.

a. Discuss the nature and common characteristics of mental retardation (MR) and the criteria and methods used in making a diagnosis of MR.

b. Review available reports by other experts in this case and evaluate their adequacy in relation to the criteria and methods discussed in (a).

c. Make recommendations to the attorneys regarding what additional assessment information might be needed to further establish the presence or absence of a diagnosis of mental retardation in this case.

d. Read the Memorandum and Order written by Judge Walter C. Kurtz of the Fifth Circuit Court for Davidson County, Tennessee on May 5th, 2004. Provide comments on aspects related to the diagnosis of mental retardation contained in this Order that might shed additional light in this case.

3. In undertaking the tasks described above, I examined the following relevant case materials relating to Mr. Byron Black:
   - Psychological/Psychiatric Evaluation/Opinion: Ms. Jaros and Drs. Anchor, Auble, Blair, van Eys, Vaught, Grant, Engum, Gur, Bernet.
   - Declaration of Dr. Globus
   - Deposition of Dr. Gur
   - Declaration of Dr. Greenspan
   - Social History and Life Time Line
   - Judge Kurtz's Memorandum and Order in the Fifth Circuit Court for Davidson County, TN (5/5/2004)
   - Post-conviction Hearing Transcripts 1989
   - Post-conviction Hearing Transcripts 2004

4. DEFINITION OF MENTAL RETARDATION

Van Tran v. State determined the mental retardation definition to be applied in Tennessee. Van Tran v. State defined mental retardation as follows: ***"significantly subaverage general intellectual functioning as evidenced by a functional intelligence quotient (I.Q.) of seventy (70) or below; (2) deficits in adaptive behavior; and (3) mental retardation manifested during the developmental period or by eighteen (18) years of age."***

The definition of mental retardation found in the Tennessee Code is consistent with the definitions endorsed by the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR; American Psychiatric Association, 2000)[3] and the American Association on Intellectual and Developmental Disabilities (AAIDD; Luckasson et al., 2002).

The **DSM-IV-TR** defines mental retardation as follows: (a) significantly subaverage intellectual functioning: an IQ of approximately 70 or below on an individually administered IQ test; (b) concurrent deficits or impairments in present adaptive functioning in at least two of the

[3] American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders (4th Edition, Text Revision; DSM-IV-TR)*. Washington, DC: Author.

following areas: communication, self-care, home living, social/interpersonal skills, use of community resources, self-direction, functional academic skills, work, leisure, health, and safety; and (c) onset is before age 18 years.

The **American Association on Intellectual and Developmental Disabilities'** (AAIDD; formerly known as the American Association on Mental Retardation) defines mental retardation as: "***a disability characterized by significant limitations both in intellectual functioning and in adaptive behavior as expressed in conceptual, social, and practical adaptive skills. Mental retardation originates before age 18.***" The AAIDD operationally defined "significant limitations" to be at least two standard deviations below the population mean (i.e., typically a standard score of 70 when the mean = 100 and the standard deviation = 15). The adaptive behavior prong of this definition is met if the individual has significant limitations in (1) conceptual, practical, or social skills or (2) the overall composite (e.g., full-scale) score of adaptive behavior.

Intellectual Functioning

The assessment of intellectual functioning is a task that requires specialized professional training. For the purpose of diagnosing mental retardation, AAIDD stipulates that IQ assessment data should be obtained and interpreted by an examiner experienced with people who have mental retardation and who is qualified in terms of professional and state regulations as well as publisher's guidelines for conducting thorough and valid evaluations of intellectual functioning.

The determination that an individual's intellectual functioning is "significantly" sub-average fulfills the first requirement for being diagnosed with mental retardation. "Significant sub-average intellectual functioning" is defined as a performance that is represented by a full-scale IQ score of approximately 70 or less, while considering all sources of test error. A standard score or intelligence quotient of "70" represents a population-referenced performance that is two standard deviations below the population mean (i.e., population average score = 100, standard deviation = 15). Significant deficits in intellectual functioning are best determined using an individually administered standardized test of intelligence. The full scale or composite IQ is generally regarded as the best estimate of an individual's general intellectual functioning (Luckasson et al., 2002).

Assessment of intellectual functioning must be done using an individually administered comprehensive standardized test of intelligence. The results obtained from group administered tests of intelligence or abbreviated measures of intellectual functioning lack the sufficient reliability and psychometric robustness to be used for the purpose of making a diagnosis of mental retardation. These instruments serve a screening purpose but should not be relied upon when making or refuting a diagnosis of mental retardation.

The Wechsler Adult Intelligence Scale – Third Edition, when used in accordance to best practice, is considered by many as the gold standard for measuring an adult individual's intellectual functioning. Other well accepted individually administered full-scale measures of intellectual functioning for adults include: Stanford-Binet Intelligence Scale-Fifth Edition, Woodcock-Johnson III Test of Cognitive Abilities, and Kaufman Adolescent and Adult Intelligence Test.

Established practice in intellectual assessment informs us that there are several important factors to consider when interpreting the IQ score. The IQ score obtained on any standardized IQ test is an estimate of the individual's "true" intelligence. This estimate is not without error. In addition to the standard error of measurement of the test used, it is important to consider the Flynn effect and possible practice effect when interpreting IQ results (see AAIDD's User's Guide).

The AAIDD User's Guide proposed a number of guidelines to ensure proper assessment of intellectual functioning for the purpose of diagnosing mental retardation. Chief among these elements are the following:

- "*intellectual functioning is best understood as being composed of a general factor ('g')* [i.e., full-scale IQ score].
- *appropriate standardized measures should reflect the individual's social, linguistic, and cultural background and that proper adaptations must be made for any motor or sensory limitations.*
- *psychometric instruments that assess intelligence perform best when used with people who score within two to three standard deviations of the mean and that extreme scores are more subject to measurement error.*
- *assessment of intellectual functioning through the reliance on intelligence tests is fraught with the potential for misuse if consideration is not given to possible errors in measurement.*" (Schalock et al., 2007; page 12).

Sources of Error for the Test Administered

The AAIDD and DSM-IV-TR agree on the importance of taking into consideration all factors contributing error to the obtained IQ test results when interpreting someone's intellectual functioning for the purpose of making a diagnosis of mental retardation. The AAIDD (Luckasson et al., 2002) stipulated the following: *"Although far from perfect, intellectual functioning is still best represented by IQ scores when obtained from appropriate assessment instruments. The criterion for diagnosis is approximately two standard deviations below the mean, considering the standard error of measurement for the specific assessment instruments used and the instrument's strengths and weaknesses."* (page 14). Furthermore, according to the DSM-IV-TR (American Psychiatric Association, 2000), **the IQ prong of mental retardation is met if an individual's full-scale IQ score falls between 70 – 75 (roughly accounting for a 95% confidence interval resulting from standard error of measurement on most IQ tests) or lower (DSM-IV-TR; see pages 41 – 42).** In addition to the standard error of measurement, sources of error surrounding the obtained IQ score may include error that is attributable to the Flynn effect and/or practice effect, and thus the interpretation of the results should account for these factors (see Schalock et al., 2007).

Flynn Effect

The "Flynn effect" is a well-established scientific fact that IQ scores on standardized tests for the American population have been steadily increasing for more than 70 years. Dr. James R. Flynn is a well-respected researcher who studied this rise in IQ scores. Flynn's research uncovered that IQ scores have been increasing from one generation to the next in the United States, as well as in all other developed countries for which we have IQ data. This increase in IQ scores over time was dubbed the "Flynn effect" by Hernstein and Murray, the authors of the book *The Bell Curve*. Some have advanced plausible explanations for this increase in IQ scores that have included: improved nutrition, trend towards smaller families, better education, etc. The

only **theoretical** aspect to the Flynn effect is the "why." The causal factors driving this trend have not yet been scientifically established. Most likely, it is an interaction of multiple factors.

Flynn reported a greater increase in the Wechsler Performance IQ, which is more heavily loaded on fluid abilities, than on the Wechsler Verbal IQs. According to Flynn's research, the average gain in global IQ scores since 1932 is approximately 0.3 points per year. Because of this, IQ tests need to be renormed periodically to recalibrate the scores. In cases where a test with aging norms is used, a correction for the obsolescence of the norms is warranted (e.g., 0.3 points per year since norms were compiled). I will use the WAIS-III to illustrate this point. The population mean on the WAIS-III was set at 100 when it was originally normed in 1995 (test published in 1997). Hence, if the WAIS-III was used to assess an individual's IQ in 2005, the individual's score should be corrected downward as follows: 0.3 points x 10 = 3 points ("10" being the number of years elapsed since the norming of the WAIS-III). After taking the Flynn effect into consideration it is still necessary to account for the test's standard error of measurement when interpreting an individual's test results.

The AAIDD *User's Guide* (Schalock et al., 2007) emphasizes the importance of considering the Flynn effect when interpreting an individual's IQ score in making a diagnosis of mental retardation.

The so-called "Flynn effect" is NOT a theory. It is a well-established scientific fact that the US population is gaining an average of 3 full-scale IQ points per decade. The Flynn effect has been consistently documented over the past 60-plus years. There is NO published scientific evidence currently existing that casts any doubt over it relevance with respect to ongoing IQ gains in the American population. In fact, a recent study published in the *American Psychologist* (a top-rated peer-reviewed scientific journal published by the American Psychological Association), reported on data supporting the effects of the Flynn effect specifically on individuals with mental retardation (see Kanaya, Scullin, & Ceci, 2003[4]). The passage of time since an IQ test was normed is directly related to that test's obsolescence. More time has passed since the norming of an IQ test the greater will be the artificial inflation of the obtained IQ scores on that test. This obsolescence of the test's norms contributes to the error that surrounds the obtained IQ score and we must take this source of error into account when interpreting an individual's obtained IQ score.

National standards are crucial in any field to ensure a uniform and consistent application of best practice. National standards are based on a foundation of empirical knowledge, science, and peer-review and are meant to serve as a guide for proper practice in that respective field. Professional practice should be consistent with established national guidelines, when such standards are available. The AAIDD User's Guide published by the former American Association on Mental Retardation (Schalock et al., 2007) represents the accepted national standard on the proper diagnosis of mental retardation. These national standards clearly indicate that when trying to establish a diagnosis of mental retardation, with respect to the assessment of general intellectual functioning, it is necessary to correct any obtained IQ score for all sources of error associated with the test used. These professional guidelines specifically mention correcting for the obsolescence of a test's norms (i.e., "Flynn effect").

[4] Kanaya, T., Scullin, M. H., & Ceci, S. J. (2003). The Flynn effect and U.S. policies: The impact of rising IQ scores on American society via mental retardation diagnoses. *American Psychologist, 58,* 778 – 790.

Adaptive Behavior

Van Tran defines adaptive behavior as referring to "*how effectively individuals cope with common life demands and how well they meet the standards of personal independence expected of someone in their particular age group, socio-cultural background, and community setting.*" In the AAIDD 2002 manual, adaptive behavior is defined as an individual's conceptual, social, and practical adaptive skills (see Luckasson et al., 2002). The AAIDD recommended that significant limitations in adaptive behavior be established through the use of standardized measures that have been normed on the general population. These three adaptive skills domains are defined as follows:

Conceptual Skills: defined by communication skills, functional academics, and self-direction.

Social Skills: defined by such abilities as interpersonal skills, social responsibility, following rules, and self-esteem. Higher order social skills have also been identified to include such elements as gullibility, naiveté, and avoiding victimization.

Practical Skills: consist of basic personal care skills such as hygiene, domestic skills, health and safety as well as work skills.

The AAIDD specified: "*The examination of adaptive skills must be documented within the context of community environments typical of the individual's age peers and culture*" (page 78). Hence, assessing an individual's adaptive behavior in an institutional context is inappropriate for the purpose of determining if an individual has mental retardation. Assessing if someone is well adapted in an institutional setting (e.g., a prison) might be useful for determining if additional structure is needed or for planning interventions to facilitate integration, but has no relevance in determining how an individual's adaptive functioning compares to the general population for the purpose of establishing a diagnosis of mental retardation.

Another important aspect of adaptive behavior assessment is the measure of the individual's "typical performance" and not best or assumed ability (Luckasson et al., 2002). Thus, when assessing the individual's adaptive behavior, we assess what the person **typically does** and not what he/she can do or could do. This is a critical distinction with the assessment of intellectual functioning, where we assess best or maximal performance.

The AAIDD 2002 definition reminded us of an important understanding about mental retardation. Namely, that within an individual with mental retardation, significant impairments often co-exist with strengths. Individuals with mild mental retardation are capable of doing many things. Most of these individuals will have strengths and areas of competence that might surprise many laypersons or even professionals who have limited experience in working with individuals with mild mental retardation. In the process of diagnosing mental retardation, the finding of significant limitations in conceptual, social, or practical adaptive skills is not outweighed by the presence of some ability on the individual's part. These discrete abilities are not uncommon in individuals with mild mental retardation and should not be viewed as discounting a diagnosis of mental retardation.

Age of Onset and Etiology

With respect to the possible cause of mental retardation, more than 40% of all cases of mild mental retardation are of undetermined etiology. The cause of mental retardation is often likely related to a combination of risk factors. These might include, but are not limited to, pre-natal maternal malnutrition, in uterine insult or trauma, genetic disorders, fetal alcohol spectrum disorder, pre-natal and post-natal exposure to toxins, childhood malnutrition, neglect, abuse, and/or impoverished and under-stimulating home environment.

There are several hundreds of disorders associated with mental retardation. Genetic disorders, such as Down syndrome, which have a well known phenotype (including almond shaped eyes, short stature, round face, etc) is more often associated with moderate to profound level of mental retardation. Again, the cause for more than 40% of cases of mild mental retardation remains unknown. AAIDD has listed numerous risk factors that might explain mental retardation, these risk factors may be of prenatal origin, perinatal, and/or postnatal (see table below).

Mental Retardation is a functional diagnosis, based on evidence regarding someone's functioning in academic and real-world settings. As such, knowledge of the cause of someone's mental retardation is not necessary in order to make a diagnosis, and in the majority of cases (especially of mild MR) one cannot say for certain what caused the condition. Nevertheless, knowledge of a possible or likely cause is a valuable thing to have, especially in establishing whether someone meets the developmental criterion. In the case of mild MR, especially in individuals from impoverished and disadvantaged backgrounds, it is often the case that environmental deprivation and parental under-stimulation in infancy and early childhood are contributing risk factors. However, one can be from such a background and still have contributing biological factors such as pre-maturity, low birth weight, prenatal infection or malnutrition, mother's alcohol consumption during pregnancy, birth trauma, chromosomal syndromes, etc. The key in diagnosing individuals from disadvantaged backgrounds is to see if an individual is viewed within his own family and community as unusually impaired, even when compared to other individuals from the same background. It also helps in making a diagnosis if one can also point to biological risk factors, such as severe head injuries or maternal alcohol consumption during pregnancy, even though evidence of a known cause is not necessary to make a diagnosis of mental retardation.

Table 1. Table of Risk Factors for Mental Retardation (see Luckasson et al., 2002; page 127)

	Biomedical	Social	Behavioral	Educational
Prenatal	Chromosomal Dx Single-gene Dx Syndromes Cerebral dysgensis Maternal illnesses Parental age	Poverty Maternal malnutrition Domestic violence Lack of access to prenatal care	Parental drug use Parental alcohol use Parental smoking Parental immaturity	Parental cognitive disability without supports Lack of preparation for parenthood
Perinatal	Prematurity Birth injury Nenatal Dx	Lack of access to birth care	Parental rejection of caretaking Parental abandon-ment of child	Lack of medical referral for inter-vention services at discharge

Postnatal	Traumatic brain injury Malnutrition Meningoencephalitis Seizure Dx Degenerative Dx	Impaired child-caregiver Lack of adequate stimulation Family poverty Chronic illness in the family Institutionalization	Child abuse and neglect Domestic violence Inadequate safety measures Social deprivation Difficult child behaviors	Impaired parenting Delayed diagnosis Inadequate early intervention services Inadequate special-education services Inadequate family Support

Dx = Disorders

5. MYTHS AND MISCONCEPTIONS REGARDING MENTAL RETARDATION

For most people with mental retardation, there is not a "mentally retarded" look. There are no distinctive features or personality types to mental retardation. It is important to remember the sage words of Ruth Luckasson (1990): "*Ninety percent of persons with mental retardation don't drool, don't stumble, aren't mute. They have significantly impaired intellectual ability, but often don't have any physical stigmata that indicate mental retardation. They won't 'look' a certain way.*" It is dangerously naïve to think that one can "tell" if someone is mentally retarded, or not mentally retarded, by looking or talking to them. Less than 10% of all cases of mental retardation are attributable to a condition such as Down syndrome. The vast majority (approximately 80%) of individuals with mental retardation function in the mild range of intellectual and adaptive behavior deficits.

The DSM-IV-TR notes: "*No specific personality and behavioral features are uniquely associated with mental retardation. Some individuals with mental retardation are passive, placid, and dependent, whereas others can be aggressive and impulsive*" (see page 44 – 45). Additionally, mental retardation can co-exist with any number of other psychiatric disorders or personality traits. The DSM-IV-TR is quite explicit on page 47 when it states: "*The diagnostic criteria for mental retardation do not include an exclusion criterion; therefore, the diagnosis should be made whenever the diagnostic criteria are met, regardless of and in addition to the presence of another disorder.*" Thus, for example, an individual may have both mental retardation and conduct disorder as a child or mental retardation and antisocial personality disorder as an adult. The presence of a co-existing mental disorder should not summarily be used to deny the individual's functioning if it meets criteria for a diagnosis of mental retardation.

6. CLINICAL JUDGMENT

The American Association on Intellectual and Developmental Disabilities (Luckasson et al., 2002) has recognized the important role of the professional's experience and knowledge of mental retardation and individuals with this condition, in diagnosing mental retardation. The AAIDD has defined clinical judgment as it relates to diagnosing mental retardation as follows:

> *"Clinical judgment is a special type of judgment rooted in a high level of clinical expertise and experience; it emerges directly from extensive data. It is based on the clinician's explicit training, direct experience with people who have mental retardation, and familiarity with the person and the person's environments"* (page 95).

AAIDD further clarified clinical judgment by stating:

> "... *[clinical judgment] should be viewed as a tool of clinicians with training and expertise in mental retardation and ongoing experiences with – and observations of – people with mental retardation and their families*" (page 95).

The professional must use his or her clinical judgment throughout the diagnostic process. The experience and clinical judgment in mental retardation informs the professional to take well-established phenomena such as Flynn effect, practice effect, and cloak of competence into consideration when evaluating the data used in making a diagnosis of mental retardation (see AAIDD User's Guide; Schalock et al., 2007).

When diagnosing other mental health disorders such as schizophrenia, clinical judgment plays a central role. In such a process, the clinician weighs various bits of evidence and then judges if an individual fits the behavioral criteria for a particular disorder. In the case of MR, however, the role of clinical judgment has very little room to operate, and is used mainly to see if test scores can be depended on reliably. There are two reasons for this: (a) many psychologists and psychiatrists have little or no training or experience in this area, and their clinical judgment about MR may be untrustworthy; and (b) because people with mild MR can have areas of relatively normal functioning, and not express obvious signs of sub-normality, clinical judgment can be very misleading, especially when it is used to rule out a diagnosis of MR. Thus, while clinical judgment has a role in diagnosing MR, it does not play as prominent a role as in other disorders (in which test scores have little or no diagnostic role) and clinical judgment should not be used as an independent diagnostic criterion separate from its use in commenting on and interpreting IQ and adaptive behavior test scores.

7. REVIEW OF EXPERT REPORTS REGARDING MENTAL RETARDATION

The records indicate that Mr. Black was never administered an individual standardized test of intellectual functioning prior to his incarceration. All IQ scores reported in his school records were obtained from group administered tests of intelligence. These measures are not well normed nor possess the psychometric properties necessary to be used in diagnostic decision-making. For this reason, these results cannot be relied upon to confirm or refute prong 1 of a diagnosis of mental retardation.

Since his incarceration, Mr. Black has been evaluated on several occasions using individually administered tests of intellectual functioning. In this section I focus my comments on the psychological evaluations and reports that centered on the question of mental retardation.

Kenneth Anchor, Ph.D. Psychological Evaluation dated 1/17/1989 – Mr. Black was 32 years old.

Dr. Anchor interviewed and conducted some individual assessments with Mr. Black. Dr. Anchor administered the Shipley-Hartford Institute of Living Scale – Revised Norms and obtained an IQ score of 76. It should be noted that the Shipley-Hartford Institute of Living Scale is a short self-answered paper-pencil questionnaire that provides an abbreviated estimate of intellectual functioning and should not be relied upon for the purpose of confirming or refuting a diagnosis of mental retardation (see AAIDD; Luckasson et al., 2002).

Gillian Blair, Ph.D. Psychological Report dated 10/7/1993 – Mr. Black was 37 years old.

Dr. Blair administered the WAIS-R during an evaluation conducted at the Riverbend Maximum Security Institution. During this evaluation, Mr. Black obtained the following scores on the WAIS-R: VIQ = 73, PIQ = 75, FSIQ = 73. Dr. Blair also administered to Mr. Black a series of other tests that measured memory and personality (e.g., Rorchach, MMPI-2, PAI, Sentence completion test, WMS-R); however, she did not attempt to assess his adaptive behavior.

Pamela Auble, Ph.D. Psychological Report dated 3/5/1997 – Mr. Black was almost 41 years old.

Dr. Auble administered a battery of tests of personality, malingering, attention, memory, and intellectual functioning. Dr. Auble administered the WAIS-R (an individually administered test of intellectual functioning) to Mr. Black and obtained the following scores: VIQ = 76, PIQ = 77, FSIQ = 76. There was no assessment attempted of Mr. Black's academic skills or adaptive behavior.

Patti van Eys, Ph.D. Psychological Report dated 3/28/2001 – Mr. Black was 45 years old.

Dr. van Eys was retained to assess Mr. Black's intellectual functioning. Dr. van Eys administered the WAIS-III on which Mr. Black obtained a VIQ = 67, PIQ = 79, FSIQ = 69. No other assessment instruments were completed at this time.

Daniel H. Grant, Ph.D. Affidavit of Testing Conducted on 10/15 & 10/16/2001 – Mr. Black was 45 years old.

Dr. Grant administered a battery of assessment instruments to Mr. Black at Riverbend Maximum Security Institution. During this psychological evaluation, Dr. Grant assessed Mr. Black using the Stanford-Binet Intelligence Scale – Fourth Edition (SB-FE), Wide Range Achievement Test – 3rd Edition (WRAT-3), Nelson-Denny Reading Comprehension Test, among other tests.

Mr. Black's academic skills as measured on the WRAT-3 and Nelson-Denny Reading Comprehension Test yielded grade-equivalents of 4th grade for both arithmetic and reading comprehension. His performance on the SB-FE yielded the following scores: Verbal Reasoning = 56, Abstract Reasoning = 76, Quantitative Reasoning = 61, Short-term Memory = 56, and Composite Score = 57. The SB-FE Composite Score is comparable to the WAIS-III FSIQ. It should be noted, however, that the mean and standard deviation on the SB-FE are 100 and 16, respectively. Thus, a Composite Score = 68 would represent a score that is 2 standard deviations below the population mean.

Dr. Grant also administered the CTONI, a test of non-verbal intelligence. I will not review Mr. Black's results on this instrument since it is a narrow band test of intelligence and not as reliable as the SB:FE and should be used only when more robust and global measures cannot be used, according to AAIDD 2002 (Luckasson et al., 2002), which was not the case here.

Susan R. Vaught, Ph.D. Review of Existing Psychological Evaluation Data and Professional Opinion Regarding the Question of Mental Retardation dated May 2003 – Mr. Black was 45 years old.

Dr. Vaught was asked to conduct a file review of Mr. Black's previous psychological evaluations and extensive records. Following this review of previously administered intellectual evaluations, Dr. Vaught concluded that Mr. Black met prong 1 of the diagnostic criteria for mental retardation.

It would appear that Dr. Vaught never met with, nor interviewed, Mr. Black or anyone else who may have had knowledge about his adaptive behavior or developmental/social history. Dr. Vaught's conclusions regarding Mr. Black's adaptive behavior appear to be based entirely on a paper review. There is no evidence in Dr. Vaught's report either that she requested any specific or additional standardized testing be done to assist her in reaching her clinical opinion in this matter. It should be noted that Dr. Vaught relied on the AAIDD (Luckasson et al., 2002) Manual in making her determination of prong 2 "deficits in adaptive behavior"; however, AAIDD (2002) clearly specifies that "*for the diagnosis of mental retardation, significant limitations in adaptive behavior should be established through the use of standardized measures normed on the general population, including people with disabilities and people without disabilities. On these standardized measures, significant limitations in adaptive behavior are operationally defined as performance that is at least two standard deviations below the mean of either (a) one of the following three types of adaptive skills: conceptual, social, or practical, or (b) an overall score on a standardized measure of conceptual, social, and practical skills*" (see Luckasson et al., p. 76).

Eric S. Engum, Ph.D., J.D. Review of Existing Psychological Evaluation Data and Professional Opinion Regarding the Question of Mild Mental Retardation dated 7/2/2003 – Mr. Black was 45 years old.

Dr. Engum was asked to review the data from existing psychological evaluations and case records and opine regarding whether or not Mr. Black has mental retardation. Dr. Engum neither assessed nor interviewed Mr. Black before formulating his clinical opinion and completing his written report. Dr. Engum reviewed Dr. van Eys' psychological evaluation and asserted that Mr. Black had to be malingering during Dr. van Eys' administration of the WAIS-III because he obtained a scaled score of 4 on Digit Span and scaled score of 2 on Arithmetic. Dr. Engum's inference is solely based upon the fact that Mr. Black's scaled scores on these two subtests on the WAIS-III administration done in 2001 by Dr. van Eys were lower that Mr. Black's scores obtained on the previously administered WAIS-R in 1997 by Dr. Auble. First, one must be very cautious comparing results on different versions of an intelligence test. In 1997 Mr. Black was administered the WAIS-R and in 2001 he was administered the WAIS-III. These are entirely different versions of the WAIS and research has shown that individuals obtain consistently lower IQ scores when tested on a more recent version of the same IQ test (see above – the Flynn effect). This difference in scaled scores should not be assumed to be an indication of malingering on Mr. Black's part.

I disagree with Dr. Engum's assertion that one cannot or should not correct obtained IQ scores for error of measurement. Research over the past several decades has clearly shown that IQ scores are rising and that an individual score artificially higher on a test with aging norms than he would on a test with more recent norms (see Table 1 & Flynn effect above). This is in fact recommended by

Mr. Byron's Previous Results on IQ Testing

Flynn effect: IQ inflation = 0.3/year

TEST USED	YEAR NORMED	YEAR ADMIN.	# YEARS ELAPSED		IQ SCORES OBTAINED	IQ INFLATION	IQ SCORES CORRECTED FOR FLYNN EFFECT	TEST STANDARD ERROR OF MEAUREMENT IQ < 70 – 75 PRONG 1 MET?
WAIS-R	1979	1993	14	VIQ	73			
		Age: 37 y.o.		PIQ	75			
Dr. Blair				FSIQ	**73**	4.2	**69**	**YES**
WAIS-R	1979	1997	18	VIQ	76			
		Age: 41 y.o.		PIQ	77			
Dr. Auble				FSIQ	**76**	5.4	**71**	**YES**
WAIS-III	1995	2001	6	VIQ	67			
		Age: 45 y.o.		PIQ	79			
Dr. van Eys				FSIQ	**69**	1.8	**67**	**YES**
SB-FE	1986	2001	15	VR	56			
		Age: 45 y.o.		AR	76			
				QR	61			
				Mem	56			
Dr. Grant				Comp	**57**	4.5	**53**	**YES**

the AAIDD when interpreting IQ results for the purpose of making a diagnosis of mental retardation. It should be noted that when Mr. Black was administered the WAIS-R in 1993 by Dr. Blair, the WAIS-R had been normed almost 15 years earlier, thus resulting in an inflation of approximately 4 points on the WAIS-R Full Scale IQ. This is a significant source of discrepancy between the measured IQ (obtained on the WAIS-R) and the individual's true IQ.

I respectfully disagree with Dr. Engum's conclusion that there is no evidence indicating that Mr. Black has significant subaverage intellectual functioning. Table 1 clearly indicates that Mr. Black meets prong 1 of the definition of mental retardation.

8. After reviewing the existing psychological evaluations and reports available, I recommended to Mr. Black's attorneys that they hire a professional to conduct a thorough assessment of Mr. Black's adaptive behavior. This adaptive behavior assessment should be conducted by a professional experienced in the area of mental retardation and adaptive behavior assessment. Since Mr. Black has been incarcerated for numerous years and that a contemporary assessment of his current adaptive behavior is impossible, the best available method would be to interview relatives and other individuals who knew him well prior to his incarceration and possibly prior to age 18 years. Retrospective assessment of adaptive behavior is recommended in such cases by the AAIDD Guidelines for diagnosing mental retardation. I thought that this assessment would yield definitive information regarding prong 2 and contribute valuable clinical information regarding whether or not Mr. Black has mental retardation.

9. RECENT COMPREHENSIVE ASSESSMENT OF MR. BLACK'S ADAPTIVE BEHAVIOR

Stephen Greenspan, Ph.D., a nationally-recognized and respected expert in the field of mental retardation, conducted a comprehensive adaptive behavior assessment using multiple sources of information including: the Vineland Adaptive Behavior Scales – 2nd Edition (a comprehensive standardized assessment of adaptive behavior), a review of existing records, a review of existing affidavits from relatives and other individuals who know Mr. Black.

Dr. Greenspan followed the guidelines put forth by the AAIDD (Schalock et al., 2007) in conducting his retrospective adaptive behavior assessment. Dr. Greenspan interviewed three different individuals in order to complete the VABS-2. A retrospective assessment is sometimes the best method available of assessing the individual's adaptive behavior. Again, adaptive behavior must be assessed in relation to community living. Using a retrospective assessment of adaptive behavior is in some circumstances the only adequate means of assessing adaptive behavior since all existing diagnostic systems, including Van Tran, define adaptive behavior as: "[adaptive behavior] *refers to how effectively individuals cope with common life demands and how well they meet the standards of personal independence expected of someone in their particular age group, socio-cultural background, and community setting.*" Hence, this refers to how the individual copes and adapts to society's expectations in the community, not prison.

Dr. Greenspan also asked these individuals to recall and assess Mr. Black's adaptive behavior prior to his 18th birthday. The advantage of conducting a retrospective assessment in this manner is that it also allows a determination if the age of onset (prong 3) criterion was met.

Based on Dr. Greenspan's evaluation of Mr. Black's adaptive behavior, Mr. Black presents significant deficits in social adaptive skills as well as significant deficits in his overall adaptive behavior (VABS-2 Composite Score = 70), thus meeting AAIDD (Luckasson et al., 2002) and Tennessee Code Annotated section 39-13-203's prong 2 criterion for mental retardation.

10. COMMENTS ON JUDGE KURTZ'S CONCLUSIONS REGARDING MENTAL RETARDATION

Mental retardation is a developmental disability, with its origin during the developmental period. Again, although it originates during the developmental period, it is not always correctly identified and diagnosed during this developmental period. Mental retardation is a chronic and life-long condition from which one seldom out grows. Conversely, one does not acquire mental retardation in adulthood. Mental retardation is a functional definition, which has no pre-set cause or etiology that must be present to be diagnosed. Similarly, there are no co-existing conditions that preclude making a diagnosis of mental retardation. Hence, if an individual functions with significant impairments in intellectual and adaptive functioning and it can be reasonably assumed to have originated during the developmental period a diagnosis of mental retardation is warranted.

There was no reliable individualized assessment of Mr. Black's intellectual functioning conducted during his school years. One should not assume that because a child was not referred for testing or special education that the child in question was not struggling in school. Clearly Mr. Black struggled in school, doing poorly in reading and having been retained in second grade.

There appears to be compelling evidence that Mr. Black's current intellectual functioning is significantly subaverage. Most experts agree that Mr. Black meets prong 1 of the definition of mental retardation. Dr. Greenspan's recent comprehensive evaluation of Mr. Black's adaptive behavior provides strong evidence indicating that Mr. Black has significant limitations in adaptive behavior and that these deficits were manifested prior to age 18 years.

As per any diagnostic system as well as the Tennessee statute 39-13-203, prong 3 refers only to documenting that the onset of significant subaverage intellectual functioning and deficits in adaptive behavior were manifested prior to age 18. No diagnostic system requires that a definitive diagnosis of mental retardation be made before the individual reaches the age of 18 years. An initial diagnosis of mental retardation can be made at any age, as long as the manifestation of prongs 1 and 2 can be documented during the developmental period or in other words, before the individual turns 18 years old.

I declare under penalty of perjury and the laws of the United States that the foregoing is a true and correct statement.

Signed on this 17th day of March, 2008.

Marc J. Tassé, PhD, FAAIDD

Attachment E

DECLARATION OF EMILY OLSON-GAULT, ESQ.

1. I, the undersigned declarant, Emily Olson-Gault, am over eighteen years of age and competent to testify to the statements contained in this Declaration. I am an attorney licensed to practice law in the State of New York and the U.S. Supreme Court. I am the Director and Chief Counsel of the American Bar Association Death Penalty Representation Project, which is based in Washington, D.C.

2. The American Bar Association ("ABA") created the Death Penalty Representation Project (the "Project") in 1986 to address the lack of qualified counsel available to those facing a death sentence.

3. The ABA and the Project do not take a position on the death penalty itself. The Project is committed to ensuring that basic constitutional protections have been provided to all individuals who are charged with a capital crime or sentenced to death. To this end, the Project promotes policies and procedures that will guarantee that all those facing execution are represented at every stage of the proceedings.

4. The ABA has promulgated Guidelines that govern the appointment and performance of defense counsel in death penalty cases. The Project is sometimes asked to provide affidavits in death penalty cases about the ABA Guidelines and standards for representation, and it does so not to advantage a particular litigant but to ensure that basic constitutional protections and due process have been provided to all individuals under a death sentence.

5. The ABA Guidelines for the Appointment and Performance of Defense Counsel in Death Penalty Cases (hereinafter "ABA Guidelines" or "Guidelines"), first adopted in 1989, were revised and updated in 2003 so that they would accurately reflect current death penalty law and practice. 31 HOFSTRA L. REV. 913 (2003), *available at* http://ambar.org/2003guidelines. The Death Penalty Representation Project led the effort to revise and update the Guidelines. The ABA House of Delegates approved the revised ABA Guidelines in February 2003.

6. After the revision of the ABA Guidelines in 2003, the Project and other organizations developed the Supplementary Guidelines for the Mitigation Function of Defense Teams in Death Penalty Cases (hereinafter "Supplementary Guidelines") to address the urgent need to help defense counsel understand how to supervise the development of mitigation evidence and direct a key member of the defense team, the mitigation specialist. The Supplementary Guidelines are a complementary extension of the ABA Guidelines. They serve to spell out important features of the existing standards of practice that enable mitigation specialists and defense attorneys to work together effectively to uncover and develop evidence that humanizes the client. *See* Supplementary Guidelines for the Mitigation Function of Defense Teams in Death Penalty Cases, 36 HOFSTRA L. REV. 679 (2008).

7. The ABA Guidelines and Supplementary Guidelines have been cited favorably in nearly 400 state and federal capital appellate decisions, including the United States Supreme Court.

See, e.g., *Padilla v. Kentucky*, 559 U.S. 356, 366-67 (2010) ("We long have recognized that '[p]revailing norms of practice as reflected in American Bar Association standards and the like ... are guides to determining what is reasonable ...'" (citing *Bobby v. Van Hook*, 558 U.S. 4 (2009) (per curiam)); *Florida v. Nixon*, 543 U.S. 175, 191, 191 n.6 (2004); *Wiggins v. Smith*, 539 U.S. 510, 524 (2003); *Williams v. Taylor*, 529 U.S. 362, 396 (2000)). *See also* ABA Death Penalty Representation Project, *List of Cases Citing the Guidelines* (Mar. 27, 2020), available at https://www.americanbar.org/content/dam/aba/administrative/death_penalty_representation/allcites.pdf.

8. The Guidelines have been adopted in substantive part or officially acknowledged as an accurate description of the standard of care for defense representation in death penalty cases by organizations such as the State Bar of Texas, the Department for Public Advocacy for the Commonwealth of Kentucky, the Idaho Public Defender Commission, the Georgia Public Defender Standards Council, and numerous others. The ABA Guidelines have also been adopted in substantive part by state legislative action or court rule in Louisiana, Nevada, and Arizona. *See* ABA Death Penalty Representation Project, *Implementation Fact Sheet* (Jul. 2018), available at https://www.americanbar.org/content/dam/aba/administrative/death_penalty_representation/ImplementationFactSheetJul2018.pdf.

9. The ABA Guidelines did not themselves create the national standard of care for capital representation; rather they simply codified long-standing norms of capital defense practice in the United States. *See Hamblin v. Mitchell*, 354 F.3d 482, 487 (6th Cir. 2003) ("the [ABA Guidelines] merely represent a codification of longstanding, common-sense principles of representation understood by diligent, competent counsel in death penalty cases.").

10. They are intended to provide guidance to judges and capital defense counsel regarding the skills and training death penalty counsel must possess when representing a person charged with a capital crime or sentenced to death.

11. The professional norms encapsulated by the ABA Guidelines govern every stage of capital proceedings. *See* Guideline 1.1(B) ("These Guidelines apply from the moment the client is taken into custody and extend to all stages of every case in which the jurisdiction may be entitled to seek the death penalty, including initial and ongoing investigation, pretrial proceedings, trial, postconviction review, clemency proceedings and any connected litigation."). In a capital case, "every stage" includes litigation in the context of an issued execution warrant, *see* Guideline 10.15.1, "Duties of Post-Conviction Counsel," Commentary, at 1081 n.335, as well as "advocacy outside the confines of the capital case itself," such as systemic and administrative challenges. Guideline 1.1, Commentary, at 923-24.

12. Underlying much of the ABA Guidelines is the recognition that defending capital cases requires extraordinary time and effort at every stage of a capital proceeding, including post-conviction, habeas corpus, and once an execution warrant has issued. *See* Guideline 1.1, Commentary ("'Every task ordinarily performed in the representation of a criminal defendant is more difficult and time-consuming when the defendant is facing execution.' . . . Due to the extraordinary and irrevocable nature of the penalty, at every stage of the

proceedings counsel must make 'extraordinary efforts on behalf of the accused'" (quoting, first, Douglas W. Vick, *Poorhouse Justice: Underfunded Indigent Defense Services and Arbitrary Death Sentences*, 43 BUFF. L. REV. 329, 357-58 (1995) and, second, ABA Standards for Criminal Justice: Defense Function, Standard 4-1.2(C), (3d ed. 1993)). The need for time and resources to prepare the defense is due in part to the tremendous amount of investigation that the capital team must complete to adequately represent a person under a death sentence.

13. In my position as Director of the Project, I am in frequent contact with capital defenders and pro bono attorneys to discuss and assist with issues related to capital representation.

14. During the month of March 2020, I have spoken with capital defenders and pro bono attorneys all over the United States as they attempt to cope with the unprecedented situation created by the COVID-19 global pandemic. My understanding from these conversations is that most capital defense teams are unable to conduct the large majority of the investigation and expert work required in capital representation (*see* ¶¶18-31, *infra*). This is due to restrictions set in place by state and local governments, as well as departments of corrections and institutional defender offices and law firms, out of a concern for public health and the welfare of employees. As a result, the already extremely limited time available to capital teams has been truncated significantly because of health concerns related to COVID-19.

15. Time is a scarce resource in all capital representation, and never more so than at the post-conviction or habeas corpus stages, or when an execution warrant has issued. *See* Guideline 1.1, Commentary ("Post-judgment proceedings demand a high degree of technical proficiency, and the skills essential to effective representation differ in significant ways from those necessary to succeed at trial. In addition, death penalty cases at the post-conviction stage may be subject to rules that provide less time for preparation than is available in noncapital cases. Substantive pleadings may have to be prepared simultaneously with, or even be delayed for, pleadings to stay the client's execution."); Guideline 10.15.1, Commentary, n.335 ("When a capital case enters a phase of being 'under warrant'—i.e., when a death warrant has been signed—time commitments for counsel increase, "due in large part to the necessary duplication of effort in the preparation of several petitions which might have to be filed simultaneously in different courts.'").

16. When the already limited time is further truncated, whether by operation of the legal system or by something wholly external like a natural disaster, counsel will not have adequate time to prepare their case and this, in turn, jeopardizes due process and fairness in capital cases. *See* Guideline 6.1, Commentary ("Regardless of the context, no system that involves burdening attorneys with more cases than they can reasonably handle can provide high quality legal representation. In the capital context, no such system is acceptable.").

17. The Guidelines' description of the nature of investigation required at every stage of a capital proceeding provides insight into the extraordinary need for time in all capital proceedings.

18. Capital defense counsel and their teams *at every stage of a capital proceeding* must conduct thorough, independent investigations related to both the guilt and penalty phases of the trial. *See* ABA Guideline 10.7(A); Guideline 10.15.1(E)(4) (post-conviction counsel must "continue an aggressive investigation of all aspects of the case"). Because "the trial record is unlikely to provide either a complete or accurate picture of the facts and issues in the case," post-conviction investigation must be "thorough" and "independent." Guideline 10.15.1, Commentary, at 1085-86.

19. Additionally, investigation in post-conviction proceedings poses specific challenges and unique obligations, as it requires a reinvestigation of "the facts underlying the conviction and sentence," a re-investigation of the client to put together a more thorough and up-to-date social history, and additional investigation into "trial counsel's performance, judicial bias, or prosecutorial misconduct." Guideline 10.15.1, Commentary, at 1086. The ABA Guidelines make clear that this obligation – like the other requirements of capital defense counsel – is "on-going" and contemplates all phases of litigation subsequent to the trial. Guideline 10.15.1(E).

20. As part of the requisite investigation, capital cases also require comprehensive, multi-generational psychosocial history construction based on "exhaustive investigation." ABA Guideline 4.1, "Defense Team and Supporting Services," Commentary, at 959. These histories must extend back at least three generations in the defendant's family. *See also* Supplementary Guideline 10.11(E)(2)(a).

21. The areas for investigation include (1) medical history, including "hospitalizations, mental and physical illness or injury, alcohol and drug use, pre-natal and birth trauma, malnutrition, developmental delays, and neurological damage;" (2) family and social history; (3) educational history; (4) military service; (5) employment and training history; and (6) prior juvenile and adult correctional experience. Guideline 10.7, Commentary, at 1022-23. See also Supplementary Guideline 10.11(B) (listing the same areas enumerated by the ABA Guidelines and further adding "multi-generational family history, genetic disorders and vulnerabilities, as well as multi-generational patterns of behavior; . . . religious, gender, sexual orientation, ethnic, racial, cultural and community influences; socio-economic, historical, and political factors.").

22. The areas for investigation listed in ABA Guideline 10.7 are not intended to be exhaustive, and the Guidelines explicitly contemplate additional investigation for other legal issues: "Additional investigation may be required to provide evidentiary support for other legal issues in the case . . . Whether within the criminal case or outside it, counsel has a duty to pursue appropriate remedies if the investigation reveals that such conditions exist." Guideline 10.7, Commentary, at 1027. *See also* Supplementary Guideline 10.11(B) ("The defense team must conduct an ongoing, exhaustive and independent investigation of every aspect of the client's character, history, record and any circumstances of the offense, or other factors, which may provide a basis for a sentence less than death.").

23. The ABA Guidelines outline a dual-track approach to conducting this investigation, requiring both witness interviews and records collection. *See* Guideline 10.7, Commentary,

at 1024 ("It is necessary to locate and interview the client's family members (who may suffer from some of the same impairments as the client), and virtually everyone else who knew the client and his family, including neighbors, teachers, clergy, case workers, doctors, correctional, probation, or parole officers and others" and "[r]ecords—from courts, government agencies, the military, employers, etc. . . . should be requested concerning not only the client, but also his parents, grandparents, siblings, cousins, and children.").

24. Simply locating a single source for this information is often insufficient. The Guidelines recognize that "[t]he collection of corroborating information from multiple sources—*a time-consuming task*—is important wherever possible to ensure the reliability and thus the persuasiveness of the evidence." Guideline 10.7, Commentary (emphasis added).

25. The Guidelines also make clear that in-person interviews with the client, witnesses, and family members are at the core of any adequate investigation: "Team members must conduct *in-person, face-to-face, one-on-one* interviews with the client, the client's family, and other witnesses who are familiar with the client's life, history, or family history or who would support a sentence less than death. Multiple interviews will be necessary to establish trust, elicit sensitive information and conduct a thorough and reliable life-history investigation." Supplementary Guideline 10.11(C) (emphasis added). *See also* ABA Guideline 10.5, Commentary, at 1008 ("Even if counsel manages to ask the right questions, a client will not—with good reason—trust a lawyer who visits only a few times before trial, does not send or reply to correspondence in a timely manner, or refuses to take telephone calls."). Remote technology options such as video conferencing and phone calls do not provide an adequate alternative for capital defenders, mitigation specialists, experts, or investigators.

26. This time-intensive, in-person contact is essential for establishing a relationship of trust with the client, client's family, and other witnesses, which is indispensable to effective representation. *See* Supplementary Guideline 10.11(C) ("Team members must endeavor to establish the rapport with the client and witnesses that will be necessary to provide the client with a defense in accordance with constitutional guarantees relevant to a capital sentencing proceeding."); ABA Guideline 10.5 and Commentary ("Client contact must be ongoing, and *include sufficient time spent at the prison to develop a rapport between attorney and client*.") (emphasis added); ABA Guideline 10.15.1(E)(1)-(2) (Post-conviction counsel must "maintain close contact with the client" and "monitor the client's mental, physical, and emotional condition [].").

27. In-person visits with the client and client's family are also a crucial tool in dictating the defense team's choices regarding necessary mental health screening and experts, which are especially important given the near-ubiquity of mental health issues in capital cases. Defense counsel's observations of the client's mental state are a necessary piece of the puzzle, as are the observations of a member of the defense team specifically trained to screen for disorders and recommend follow-up investigation and appropriate experts. *See* ABA Guideline 4.1, Commentary, at 956. In many cases, the results of such observation render a "psychologist or other mental health expert [] a needed member of the defense team." Guideline 10.4, Commentary, at 1004.

28. The mental health services provided as a result of observations and screenings are also oftentimes necessary to ensure that a client is "cognitively and emotionally competent to make sound decisions concerning his case." Guideline 4.1, Commentary, at 959. Moving capital proceedings forward while in-person visits cannot take place prevents counsel from ensuring that the client is competent.

29. Additionally, expert evaluations of the client are time-consuming, particularly for issues related to intellectual disability and mental illness or competency. *See* Guideline 4.1, Commentary ("Creating a competent and reliable mental health evaluation consistent with prevailing standards of practice is a time-consuming and expensive process."). In order to ensure the heightened reliability of such evaluations, a thorough investigation must first be conducted. *Id.* ("Counsel must compile extensive historical data, as well as obtain a thorough physical and neurological examination. Diagnostic studies, neuropsychological testing, appropriate brain scans, blood tests or genetic studies, and consultation with additional mental health specialists may also be necessary."); *see also* Guideline 4.1, Commentary, at 959 (the mitigation specialist "provides social history information to experts to enable them to conduct competent and reliable evaluations"). Judgment calls regarding which expert evaluations are recommended are necessarily the product of in-person visits between the client and the defense team, *see* ¶¶27-28, *supra*, and are informed by a comprehensive and thorough investigation.

30. The investigation is also a necessary precursor to making key strategic decisions and preparing pleadings. *See* ABA Guideline 10.7, Commentary, at 1021 ("Counsel cannot responsibly advise a client about the merits of different courses of action, the client cannot make informed decisions, and counsel cannot be sure of the client's competency to make such decisions, unless counsel has first conducted a thorough investigation with respect to both phases of the case.").

31. In addition to the above-described duties and decisions, cases with an active execution warrant require additional urgent, time-consuming tasks that cannot be completed prior to the issuance of the warrant, such as the preparation of petitions for executive clemency, seeking stays of execution, and arranging for expert evaluations regarding the clients' competency to be executed. See Guideline 1.1, Commentary at 937 ("Recent advances in the use of DNA technologies, combined with restrictions on the availability of post-conviction review, have elevated the important role that clemency has played as the "fail-safe" of the criminal justice system, and *increased the demands on counsel*") (emphasis added); Guideline 10.15.1(B) ("If an execution date is set, post-conviction counsel should immediately take all appropriate steps to secure a stay of execution and pursue those efforts through all available fora."); Guideline 10.15.2(B)-(D) (detailing duties of clemency counsel, including duty to "conduct an investigation in accordance with Guideline 10.7").

32. The norms of practice reflected in the ABA Guidelines are not aspirational. *See* Guideline 1.1, "Objective and Scope of Guidelines," Commentary, at 920. They represent the minimum requirements for adequate representation. If counsel lacks adequate time to prepare their case, or if defense counsel, mental health experts, investigators and mitigation

specialists are unable to conduct in-person meetings and interviews to discharge the duties outlined in ¶¶18-31 above, fundamental fairness and accuracy are put at risk.

33. As the Guidelines have recognized, there is an indispensable need for "effective representation on appeal, in state and federal post-conviction proceedings, and in applications for executive clemency. Because each of those proceedings has a unique role to play in the capital process, because both legal and social norms commonly evolve over the course of a case, and because of 'the general tendency of evidence of innocence to emerge only at a relatively late stage in capital proceedings,' jurisdictions that retain capital punishment must provide representation in accordance with the standards of these Guidelines, as outlined in Subsection B, 'at all stages of the case.'" Guideline 1.1, Commentary at 929-30. The Guidelines "recognize[] the simple truth that any other course has weighty costs—to be paid in money and delay if cases are reversed at later stages or in injustice if they are not." *Id.* at 930.

34. The ABA Guidelines are the most authoritative and up-to-date articulation of the investigative and other responsibilities of capital defense counsel. The American Bar Association believes that meeting these responsibilities is essential to ensuring justice in capital cases.

I hereby swear under penalty of perjury that the above and foregoing is a true and correct statement.

Dated this 3rd day of April, 2020.

Emily Olson-Gault