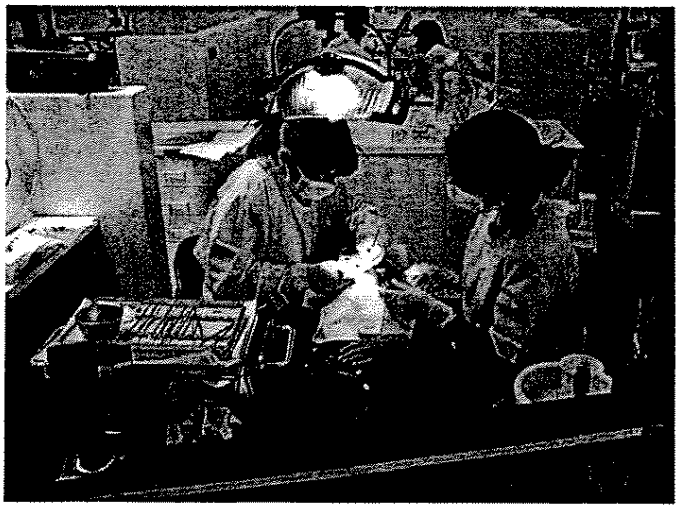
RECEIVED
Human Subjects Division
OCT 31 2003
UW

Casa Pia Children's Amalgam Trial
Report to the Data and Safety Monitoring Board

September 15, 1999

ONLY FILE COPY
Keep with Committee Copy
Send to Reviewer at Renewal
To Records when file is closed
App.# 95-0401-A/C 0



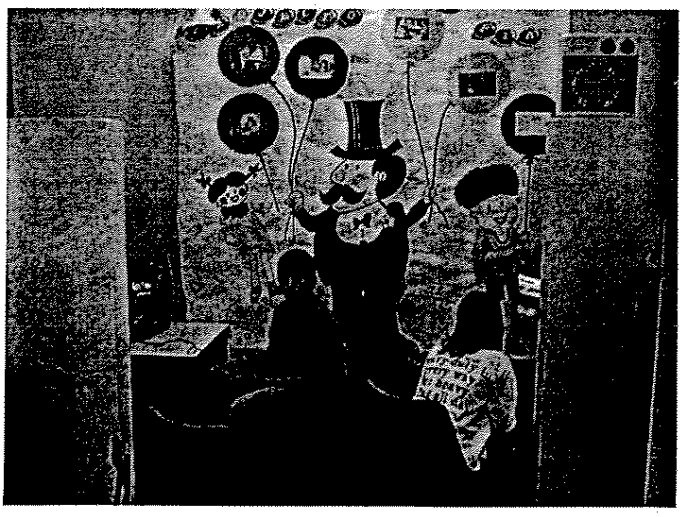


Table of Contents

Introduction 2
0 Changes/Updates Since DSMB May '98 2
0.1 Protocol Changes 2
0.2 Key Staff Changes 3
0.3 Protocol Deviations 3
0.4 Presentations During Last Year 3
1 Screening and Recruitment 5
1.1 Comparison of screened with randomized children 5
1.2 Reasons for ineligibility 6
1.3 Randomization by year and month 6
2 Comparability of Treatment Groups at Baseline 7
2.1 Demographic and health history information 7
2.2 Neurobehavioral 8
2.3 Nerve conduction velocity 10
2.4 Neurological 11
2.5 Oral exam 12
2.6 Laboratory 15
3 Dental Treatment 19
3.1 Treatment time and number of children receiving dental treatment 19
3.2 Number of restorations placed 19
3.3 Mercury exposure 21
4 Follow-up 21
4.1 Number of subjects lost to follow-up and reasons 21
4.2 Number of follow-up visits occurring within visit windows 21
5 First-year Follow-up Results on Primary Outcome Variables 22
5.1 Neurobehavioral 22
5.2 Nerve conduction velocity 25
5.3 Statistical stopping rule 27
5.4 Reassessment of power 28
6 First-year Follow-up Results on Secondary Outcome Variables 30
6.1 Sentinel events by group 30
6.2 Health history 30
6.3 Neurological 30
6.4 Oral exam 31
6.5 Laboratory 33
7 Quality Assurance Monitoring 38
7.1 Timeliness of data entry 38
7.2 Completeness of data 38
7.3 Accuracy of data entry 39
7.4 Examiner reliability 39
8 Appendix A. Additional Tables. 42
9 Appendix B. Additional Figures 45
10 Appendix C. The Statistical Stopping Rule. 54
11 Appendix D. Results of the Bacterial Resistance Study. 55
12 Appendix E. Results of the Dietary Mercury Study. 57
13 Appendix F. Progress on the Prepubertal Periodontitis Study 58
14 Appendix G. Sentinel Health Event Questionnaire 59

Introduction

The Casa Pia Children's Amalgam Trial has continued to progress on schedule during the year since the first DSMB meeting in May of 1998. The initial recruitment goal of 450 children was exceeded, with 507 children being finally enrolled into the trial. One-year follow-up neurobehavioral testing and dental procedures have proceeded with no major problems, delays or interruptions. Due to an unexpectedly high recurrent and new caries rate seen at the first annual dental follow-up examinations, additional personnel were added through a supplemental budget to provide aggressive oral disease prevention activities for the children in the study.

The "Sentinel Event" monitoring system suggested by the DSMB at the 1998 meeting has been implemented, and the results are reported in section 6.1 of this report.

No key personnel have left the project at either site (Lisbon/Seattle) during the last year, although several new individuals have been engaged in the project. This is discussed in section 0.2.

Loss to follow-up in the trial has been minimal, with a 94.9% retention rate of subjects.

There are no important differences seen between the two treatment groups at this point for the neurobehavioral, neurological or dental treatment data. All results are provided in this report according to the outline suggested by Dr. Redford.

0 Changes/Updates Since DSMB May '98

0.1 Protocol Changes

Data on race of the subjects was not initially being recorded, and had to be obtained retrospectively. As of 10/6/98 race was added to the records, and recorded as Asian or Pacific Islander; Black, not of Hispanic Origin; Hispanic; White, not of Hispanic Origin; Other or unknown.

Changes in Neurological Examination Form (Section 2.16.4): 1.) For each field in the examination, the form originally had an "abnormal" check box. If the box was blank, normality was assumed. We added a "normal" box as well, so that one or the other is always indicated. 2.) The instructions for the original form asked the examiner to indicate the presence or absence of "adventitious" movements, but there was no check-box. We have added a check box for this finding. 3.) The original form had a space for the noting of "soft" neurological signs, but no specific check boxes. These have been added. Implemented 11/24/98.

The tracking of "health events" which may occur in the year between annual behavioral or dental visits was considered to be part of the dental record, but not clearly elucidated or tracked. At the suggestion of the DSMB, an annual "Health events" questionnaire is now completed for each child by the parents or guardians for the purpose of tracking any unusual health events which may need follow-up in terms of potential relationship to the study exposures. This questionnaire seeks information about illnesses, medicines taken, doctor-initiated treatments or hospitalizations during the previous year (added to Section 2.10.1 of Manual of Operations). At the recommendation of the DSMB we made particular note of any reported illnesses or episodes from the list below, and follow up with calls and letters to parents/guardians to confirm such reports. Implemented 2/1/99. In addition, the dental hygienists who work with the schools on implementing the prevention program meet with teachers regularly to identify any study

participants who may have experienced any recent illnesses. This serves as an additional mechanism for identification of sentinel health events.

Key Sentinel Events Monitored:

Seizure disorder
Other neurologic disorder (eg, M.S., stroke, A.L.S.)
Tumor (neoplasm)
Metabolic disease (eg, diabetes, renal failure)
Cardiovascular disorder (eg, heart failure, hypertension, C.V.A.)
Pulmonary disorder (eg, asthma)

Additional Treatment Form: The original protocol called for the use of a single form to track additional dental treatment found to be needed at each of the annual follow-up visits. (Section 2.17.2). Because of the wide variation in "anniversary" dates for the subjects (not everyone was enrolled on the same date), it would be difficult to determine what treatment was provided in each study year of the project with the current method. The decision was made to use a new form for each year of the study, uniquely identifying the treatment each subject received in each year of follow-up. Implemented 3/1/99.

Because we found an unusually high recurrent and new caries attack rate among the study subjects, we requested, and were awarded a supplement to provide additional, aggressive oral hygiene instruction, care and follow-up among the study subjects. Two part-time dental hygienists have been assigned to coordinate and provide oral hygiene instruction, fluoride mouthrinse, behavioral modification, and toothbrushing programs among the study subjects. This supplement and activity was initiated 12/1/98.

0.2 Key Staff Changes

Because of the unexpectedly high rate of new and recurrent caries found at the first annual followup dental exams, two part-time dental hygienists have been added to provide additional, aggressive oral hygiene instruction and adjunctive activities. An additional part-time psychometrist has been added for the purpose of speeding up the rescoring of the psychometric tests. This individual does not actually perform testing of the children. Aimee Loar, Graduate Research Assistant in Biostatistics, has joined the staff to assist with data management and biostatistical analysis. Other than these, there have been no changes in key staff since the 1998 DSMB meeting.

0.3 Protocol Deviations

Due to a misunderstanding of protocol, blood samples were collected from 72 subjects in year two of follow-up before this error was discovered and corrected. The protocol calls for blood samples to be collected in the third year of follow-up, but this was misunderstood to be year three of the project. The samples are currently frozen.

There have been no other deviations in protocol since the 1998 DSMB meeting.

0.4 Presentations During Last Year

A major conference of study investigators and staff from both the Seattle and Lisbon sites was held on November 11-12, 1998 in Lisbon for the purpose of discussing progress to date, presenting baseline results within the study team, and coordinating efforts for an initial round of publications based on the

baseline findings. A number of small group meetings were included in the agenda to allow individual units (i.e. neurobehavioral) to focus on study and publication concerns for their particular areas. Participants and the meeting agenda are shown below.

Participants:

Seattle site

Dr. Timothy DeRouen	Principal Investigator
Dr. Michael Martin	Project Director
Dr. Brian Leroux	Biostatistician
Dr. Brenda Townes	Co-Investigator
Dr. James Woods	Co-Investigator
Ms. Gail Rosenbaum	Psychometrist
Dr. Carina Evans	Ancillary Project Director (Dietary Mercury)
Dr. Marilyn Roberts	Ancillary Project Director (Antibiotic Resistance)

Lisbon site

Dr. Jorge Leitão	Co-Investigator
Dr. Mário Bernardo	Lead Dentist
Dr. Alexandre Castro-Caldas	Co-Investigator
Dr. Mamede Carvalho	Neurologist
Dra. Isabel Pavão	Neurologist
Henrique Soares Luis, RDH	Study Records Manager
Lurdes Vaz, RDH	Patient Care Coordinator
Helena Amaral, RDH	Patient Care Coordinator
Helena Nazareth, RN	Nurse Technician
Goretti Feirrera	Psychometrist
Pedro Rodriques	Psychometrist
Dra. Paula Marques	Ancillary Project Director (Juvenile Periodontitis)

Agenda

Wednesday, November 11, 1998

Morning:

Introduction and Welcome	Dr. Timothy DeRouen, Dr. Jorge Leitão
Baseline Neurologic Findings	Dr. Castro Caldas, Dra. Isabel Pavão
Baseline Nerve Conduction Findings	Dr. Castro Caldas, Dr. Mamede Carvalho
Baseline Neurobehavioral Findings	Dr. Brenda Townes, Ms. Gail Rosenbaum
Baseline Oral Findings	Dr. Mario Bernardo, Dr. Michael Martin
12 Month Follow-up Findings	Dr. Michael Martin, Dr. Mario Bernardo

Lunch

Afternoon:

Room A: Group Meetings for Dental Group

Room B: Group Meetings for Neurological/neurobehavioral Group

Thursday, November 12, 1998

Morning:

Baseline Urinary Analysis Results	Dr. James Woods
Dietary Exposure / Blood Analyses	Dr. Carina Evans
Antibiotic Resistance	Dr. Marilyn Roberts

Pre-pubertal Periodontitis Dr. Paula Marques
Lunch
Afternoon:
Room A: Group Meetings for Dental Group
Room B: Group Meetings for Neurological/neurobehavioral Group
Summation and Wrap-up Dr. Timothy DeRouen, Dr. Jorge Leitão

In addition to the November meeting, site visits to Lisbon for routine oversight of progress and quality control were made by Dr. Martin in April and September of 1999, and by Dr. DeRouen in June, 1999.

Several presentations have been made at major research meetings since the 1998 DSMB meeting:

Dr. Timothy DeRouen (P.I.) presented a poster "The Casa Pia Study of the Health Effects of Dental Amalgam" at the IADR meeting in Nice, France, in June, 1998;

Dr. DeRouen also presented a talk/poster discussion with the same topic and updated data at the Fifth International Conference on Mercury as a Global Pollutant in Rio de Janiero, Brazil on May 26, 1999.

Dr. Mario Bernardo (Co-investigator, Lisbon) presented a poster "Oral epidemiology of a population of Portuguese schoolchildren" at the 1999 International Association for Dental Research (IADR) meeting in Vancouver, B.C. on 3/11/99.

Dr. Michael Martin (Project Director) also presented a poster "Treatment needs in the Casa Pia trial of dental amalgam" at the 1999 IADR meeting (3/12/99).

Dr. Carina Evens presented "Examination of methylmercury exposure in the 'Children's Amalgam Trial' through dietary information and blood analysis" at the Epidemiological Forum at the 1999 IADR meeting.

1 Screening and Recruitment

Enrollment was completed on October 30, 1998 after 507 subjects had been randomized into one of the two treatment groups (253 in Group A and 254 in Group B)[1].

1.1 Comparison of screened with randomized children

Letters were sent to the parents or guardians of 845 children about participating in the study at the seven participating schools ("Approached" in Table 1.1). Consent to participate in the study was obtained from 647 of these children; the remaining children either refused participation (146, 17%) or did not respond to the request for consent (52, 6%). Of the 647 children who consented to participate and were screened, 131 (20%) were determined to be ineligible and the eligibility status of an additional nine (1%) was undetermined. The randomized children were similar to the ineligible children in terms of age (mean 9.4 years vs. 9.3 years), gender (45.4% female vs. 42.6% female) and race (70.0% White, 28.0% Black, and 2.0% Asian, other, or unknown vs. 76.7% White, 18.6% Black, and 4.7% Asian, other, or unknown).

[1] Treatment groups are labeled A and B in most sections of the report in order to maintain masking of the treatment groups (among study personnel only the data manager knows the identity of the groups). Group identities will be available at the DSMB meeting in a sealed envelope. Different group labels (eg, A and B, 1 and 2) are used for different results because the treatment group identities may be revealed in some data (eg, follow-up urinary mercury) and we must avoid the possibility of linking the treatment groups in these tables with the masked groups in other tables.

1.2 Reasons for ineligibility

Of the 131 subjects declared ineligible, the main reasons for ineligibility were previous exposure to amalgam (54, 41%), no caries on a permanent posterior tooth (38, 29%), and IQ < 67 (32, 24%); in addition, 5 (4%) of these children had urinary mercury > 10 ug/L and 2 (2%) of the children had an excluding health condition (one with diabetes and one with neoplastic disease). No subjects were declared ineligible due to high blood lead level (> 15 ug/dl). Information on the numbers of children screened, consented, randomized, and declared ineligible for each of the seven participating schools is given in Table 1.1.

Table 1.1. Number of children approached, consented, randomized and declared ineligible by school and reasons for ineligibility.

			Randomized		Ineligible					
School	Approached	Consented and Screened	Group A	Group B	a	b	c	d	e	f
Nuno Alvares	206	180	76	77	6	11	6	3	0	0
António Aurélio	7	7	4	2	1	0	0	0	0	0
Pina Manique	150	108	38	41	5	7	10	0	0	0
Maria Pia	186	148	59	56	11	9	10	1	0	2
Santa Clara	137	105	39	37	5	8	14	0	0	0
Santa Catarina	33	24	9	10	1	0	4	0	0	0
N.S. Conceição	126	75	29	30	3	3	10	1	0	0
Total	845	647	254	253	32	38	54	5	0	2

a) IQ < 67
b) No caries on permanent posterior tooth.
c) Previous exposure to amalgam.
d) Urinary Hg>10 ug/l.
e) Blood lead>15 ug/dl.
f) Excluding health condition.

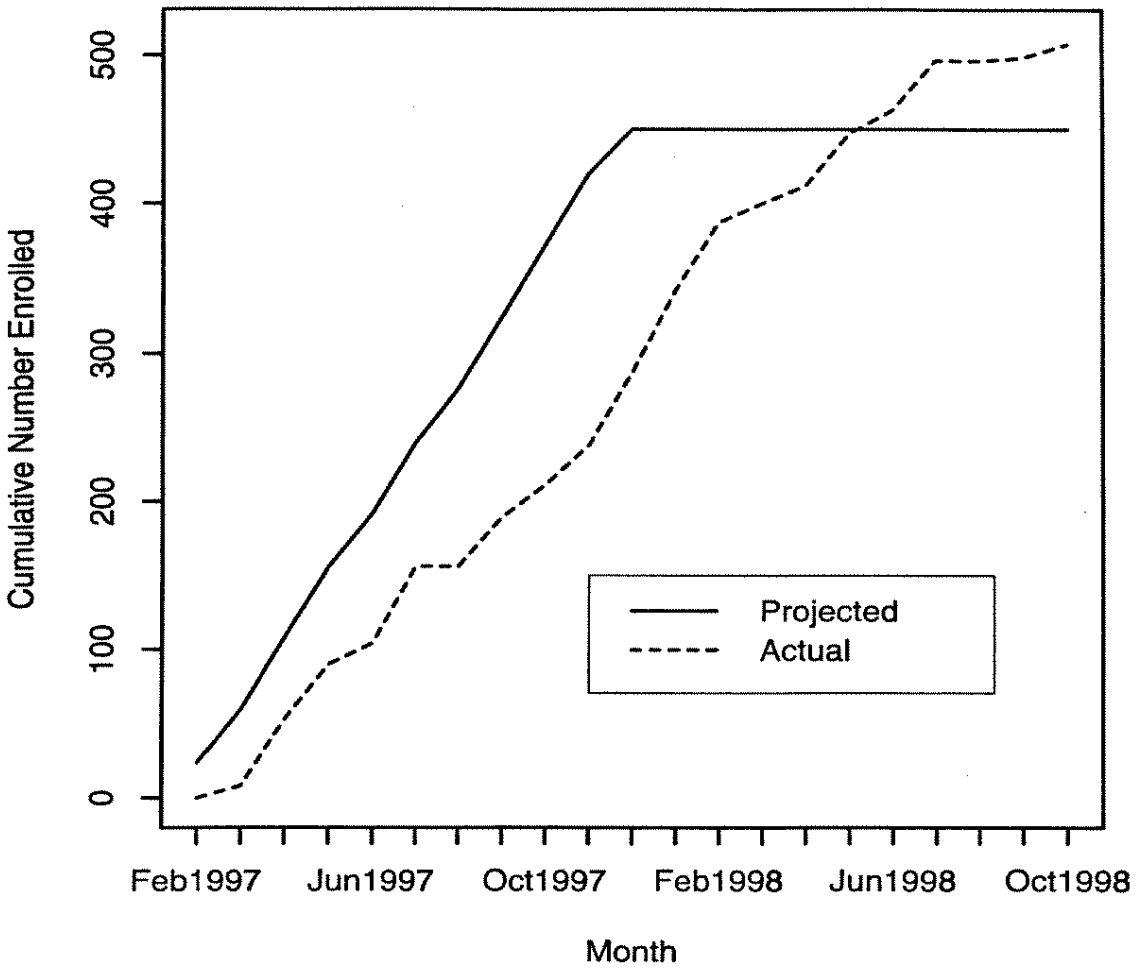
Three protocol errors occurred in which subjects that should have been declared ineligible were randomized (the first two were previously reported):

- 1 subject with abnormal neurological findings was randomized on 4/25/97
- 1 subject with existing amalgam was randomized on 1/7/98
- 1 subject was randomized on 12/23/97 with a blood lead value of 14 ug/dl that was later adjusted by the laboratory to 16ug/dl.

1.3 Randomization by year and month

The number of children randomized for each month is shown in Figure 1.1 along with the projected numbers. Enrollment was delayed at the beginning of the study while the study protocol was being finalized and thereafter occurred at a slightly lower than anticipated rate due to higher ineligibility rates (on average 25 patients were enrolled per month during the recruitment period). However, in the end we were able to enroll more subjects than planned (507 vs. 450) due to the large number of students in the appropriate age range in the schools.

Figure 1.1. Projected and actual number of children randomized by year and month.



2 Comparability of Treatment Groups at Baseline

2.1 Demographic and health history information

The distribution of age, gender, and race for the randomized subjects is given in Table 2.1. Health history information is given in the appendix (Table A.1). The treatment groups were similar at baseline in age, gender and race and on health history items with the exception of small differences for the prevalence of pain, infection, or injury and the percentage of subjects reporting taking vitamin-mineral pills. None of these small differences would be expected to impact the study conclusions.

Table 2.1. Demographic information at baseline by treatment group.

	Group A		Group B	
	Number	Percent	Number	Percent
Year of Birth:				
1986	90	35.6	74	29.1
1987	85	33.6	93	36.6
1988	54	21.3	55	21.7
1989	24	9.5	32	12.6
Sex:				
Male	135	53.4	142	55.9
Female	118	46.6	112	44.1
Race:				
White	176	69.6	179	70.5
Black	75	29.6	67	26.4
Asian or Other	2	0.8	8	3.2

2.2 Neurobehavioral

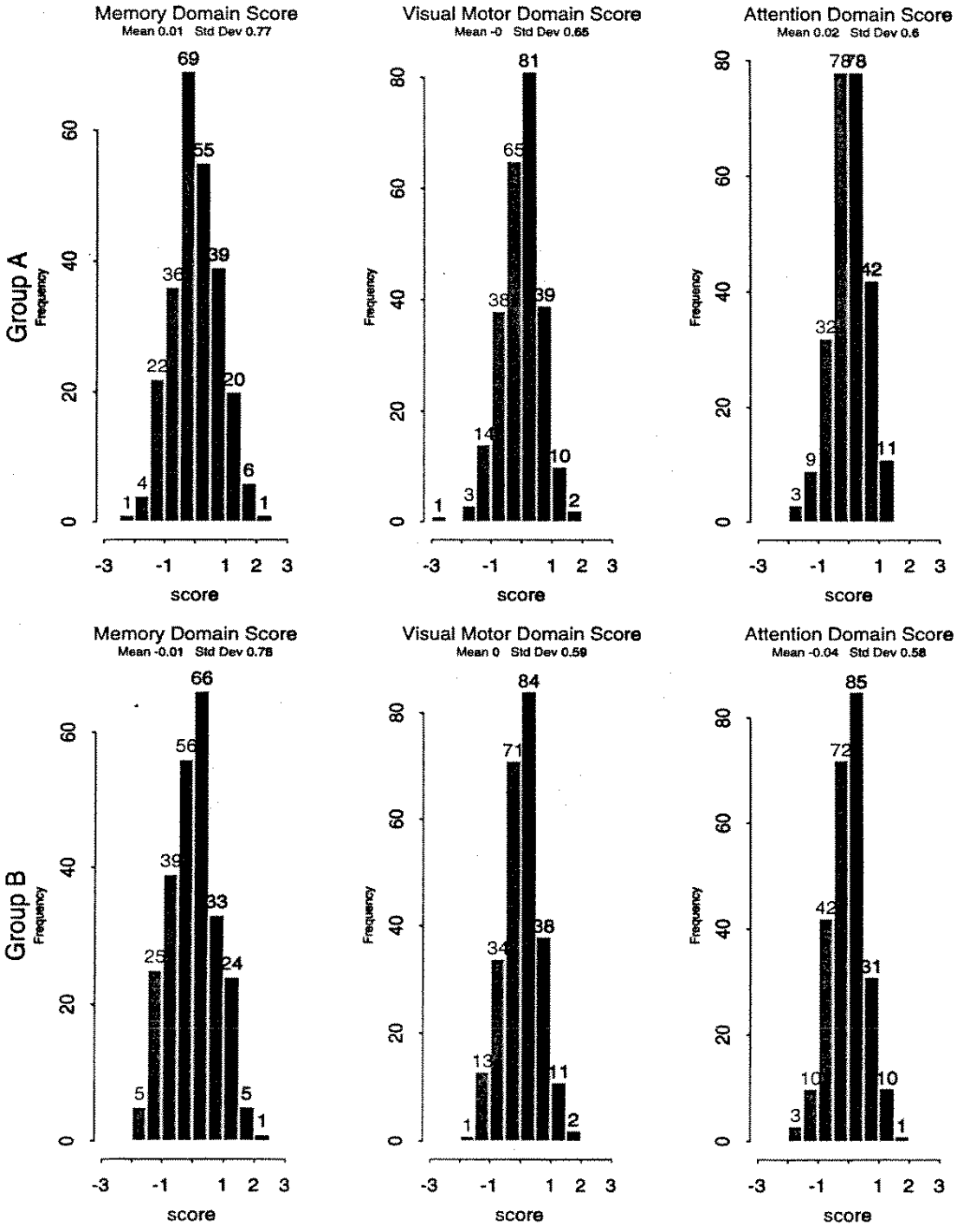
Histograms[2] of the Z-scores[3] at baseline for the three neurobehavioral domains by treatment group are provided in Figure 2.1. These three domain scores are three of the four primary endpoints of the study. Histograms for the individual items used in the calculation of the domain Z-scores are presented in Figure B.1 (Appendix B). The distributions were very similar between the two treatment groups for the domain Z-scores and also for most of the individual test items. IQ test results were also similar in the two groups (Group A: mean 85.2, SD 10.1, range 67-118; Group B: mean 84.9, SD 9.9, range 67-116)[4]. Information on other neurobehavioral items not used in the computation of domain scores will be available at the DSMB meeting but is not included in this report.

[2]Results for continuous variables are presented as histograms; the number of subjects within each interval is printed above the histogram bars and the mean and SD are printed above the figure. Note that data values equal to 0 are located to the left of 0 in these histograms (eg, see Figure 2.3).

[3] Z scores (mean 0, std. dev. 1 across the entire sample) were calculated for each of the key neurobehavioral variables. Then domain scores were obtained by averaging the Z scores within each domain.

[4] IQ scores refer to the entire CTONI—previous results reported were based on only one component of the CTONI. As previously reported, the lower cut-off value for inclusion in the study was changed from 70 to 67.

Figure 2.1. Distribution by group at baseline of neurobehavioral domain Z-scores for memory, visual motor, and attention domains.



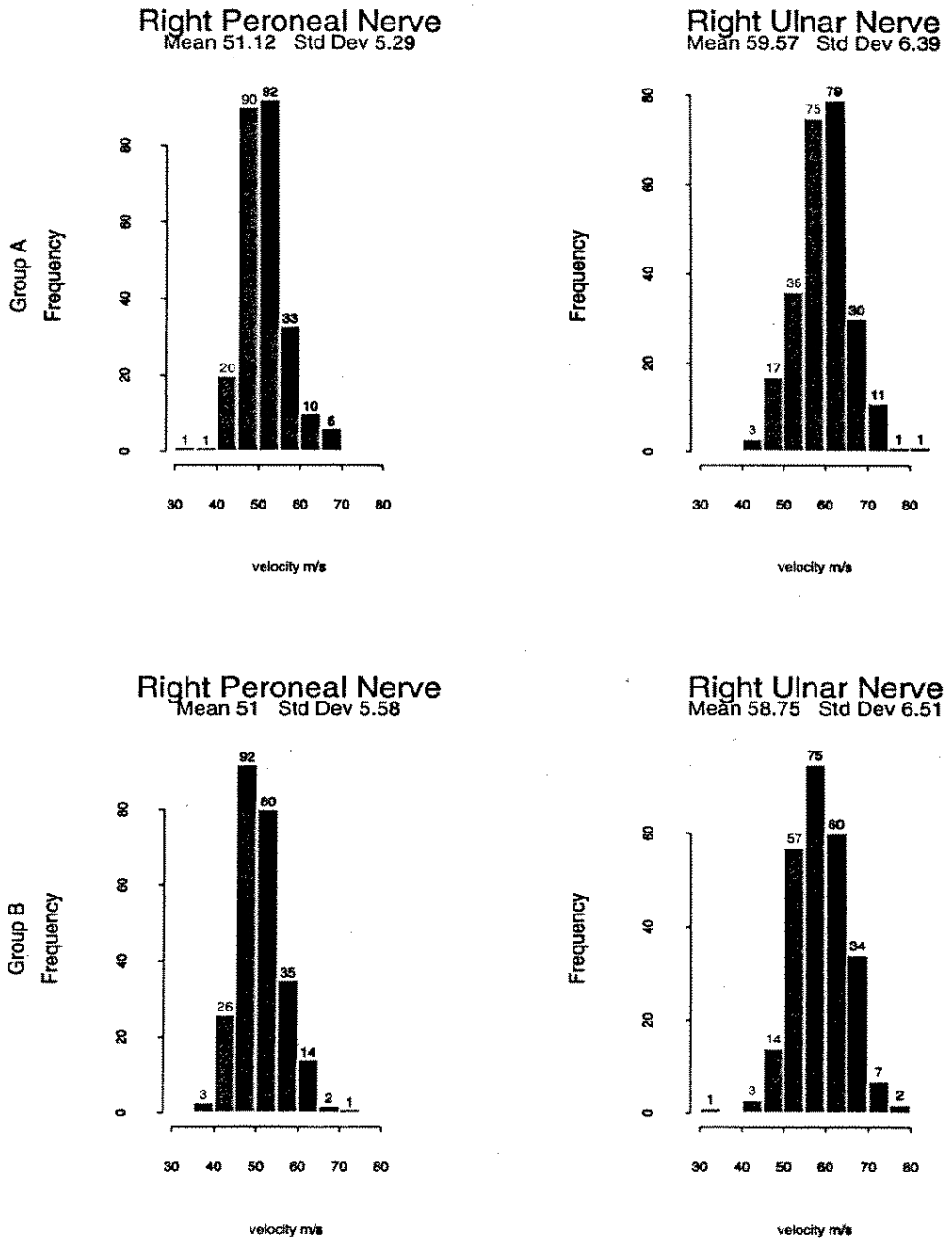
Norms for the key neurobehavioral test items are given in Table A.2 for 9-11 year-old children (note that 85% of enrolled subjects were 9-11 at the time of initial neurobehavioral testing; the mean age was 9.6 years). For most items in the memory and visual motor domains, the average scores for the study subjects were similar to the norms, but for every item in the attention domain, the study subjects had poorer scores than the norms on average. Standard deviations for the study subjects were also similar to the standard deviations from the norms data for most variables.

2.3 Nerve conduction velocity

Histograms of the baseline nerve conduction velocities (right ulnar and peroneal nerves) for each treatment group are provided in Figure 2.2. Additional information on the distances and latencies used to calculate the velocities are given in Table A.3 (Appendix A). The distributions of nerve conduction velocities were very similar in the two treatment groups. The mean values are fairly similar to the published norms we reported previously (Ulnar: mean 67.6, SD 6.0; Peroneal: mean 57.6, SD 7.3). Norms for nerve conduction velocity may be machine dependent.



Figure 2.2. Histograms (with mean, SD) of nerve conduction velocities by treatment group.



2.4 Neurological

The prevalence of neurological abnormalities by treatment group is given in Table 2.2. In addition the prevalence of any abnormality is given in the last line of the table. The prevalence of these abnormalities was low and similar in the two groups. Because of the small numbers of positive cases, the small

differences in prevalences of neurological abnormalities are not statistically significant and also would not be anticipated to have any impact on the analysis of the primary study endpoints.

Table 2.2. Baseline prevalence of neurological abnormalities by treatment group.

	Group A		Group B	
	Number with Abnormality	Percent	Number with Abnormality	Percent
Mental Status	0	0.0	0	0.0
Cranial Nerves	6	2.4	3	1.2
Motor Exam	1	0.4	1	0.4
Sensory	0	0.0	1	0.4
Coordination	2	0.8	1	0.4
Reflexes	3	1.2	0	0.0
Gait/Station	1	0.4	0	0.0
Adventitious Movements[5]	0	0.0	0	0.0
Any Abnormality	9	3.6	6	2.4

2.5 Oral exam

The distributions of key oral health measures at baseline are provided in Figures 2.3 and 2.4. Figure 2.3 shows that there was considerable variation among children in the number of primary teeth (range 0-18) and permanent teeth (range 6-28); however, the overall number of teeth (not shown) was approximately 24 per child in both groups and did not vary much between children (SD=1.8, range 16-30). Subjects in Group A had on average 0.4 more primary teeth and 0.4 fewer permanent teeth than subjects in Group B. Not surprisingly, Group A also had more decayed primary tooth surfaces (9.4 vs 9.1) and fewer decayed permanent tooth surfaces (6.2 vs 6.7); however, the total number of decayed surfaces was similar in the two groups (15.6 vs 15.8). These differences are small and not anticipated to have appreciable impact on the analysis of the primary study endpoints.

[5] Adventitious movements were not recorded in the initial baseline neurological exams; data are available for 385 subjects.

Figure 2.3. Distribution by treatment group of numbers of primary and permanent teeth at baseline.

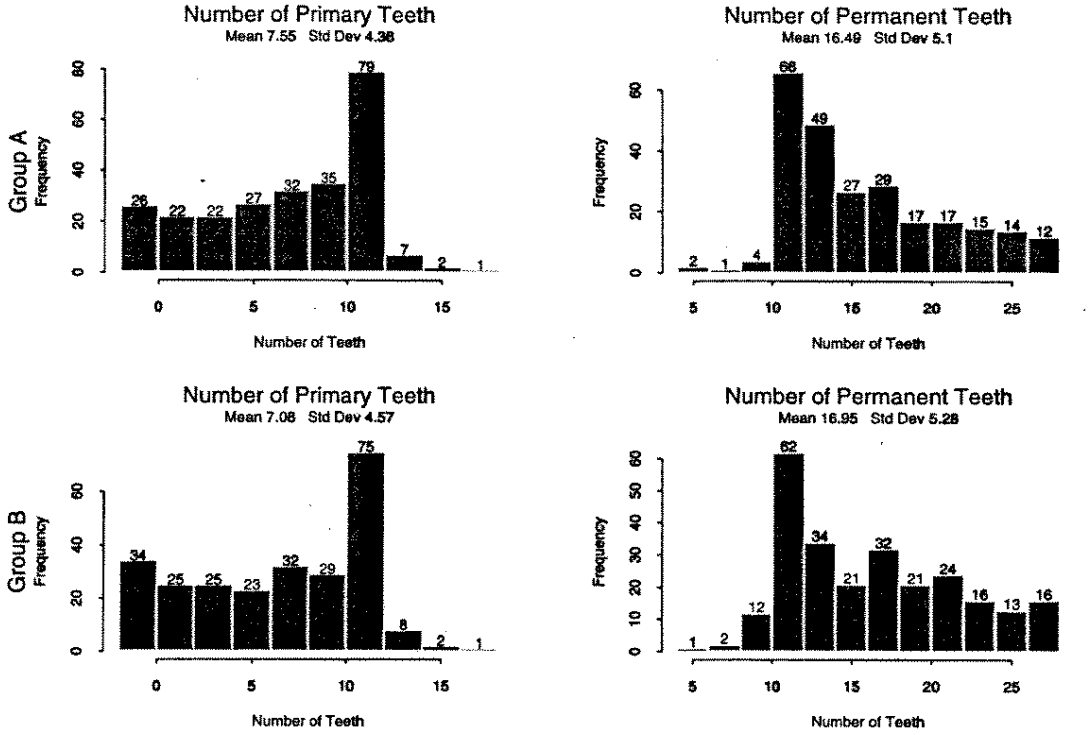
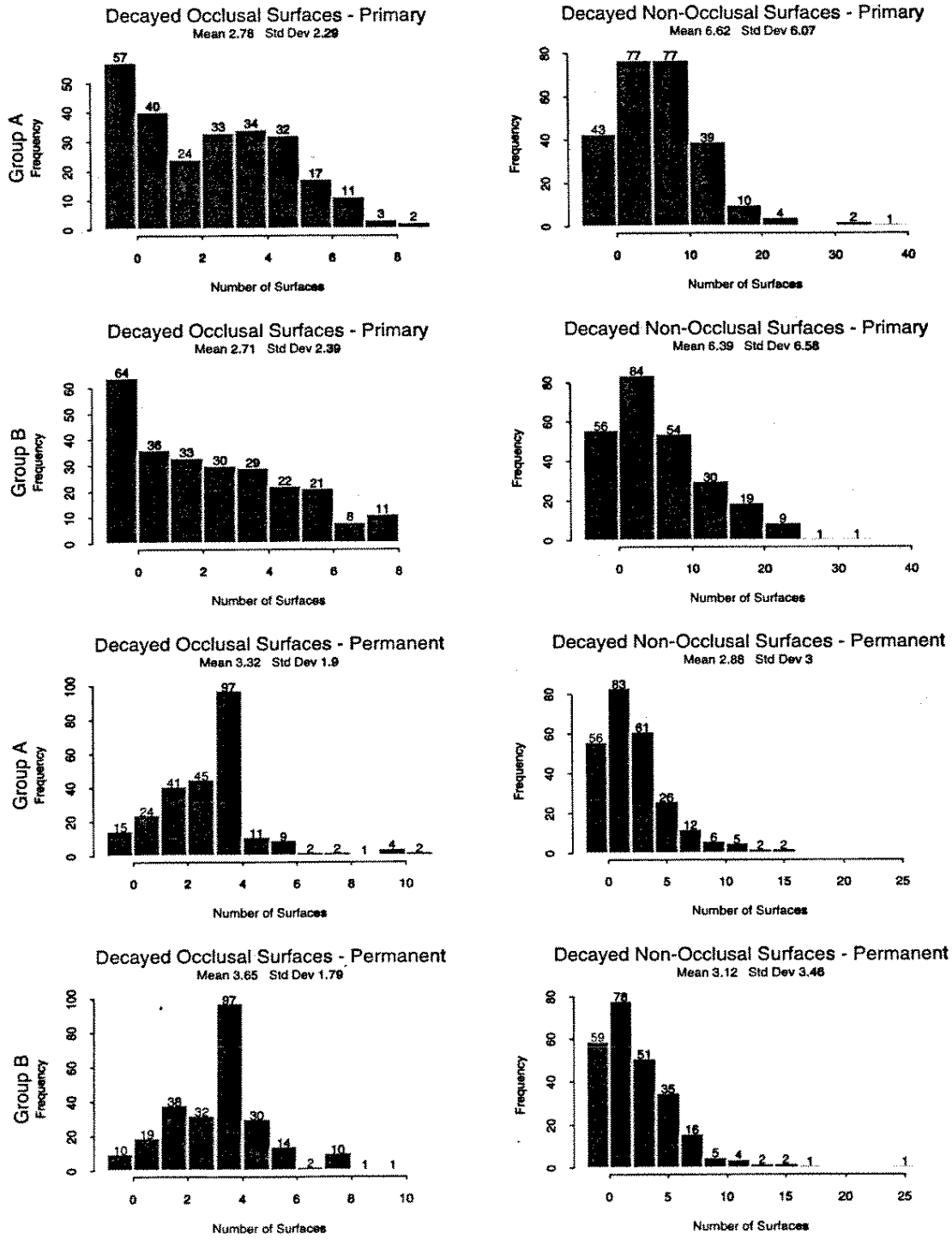
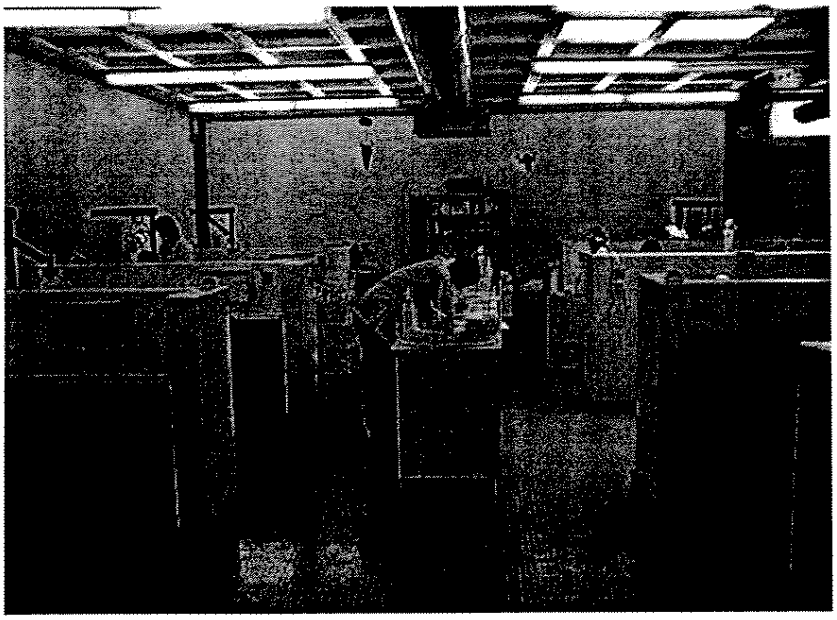


Figure 2.4. Distributions by treatment group of numbers of decayed surfaces at baseline.



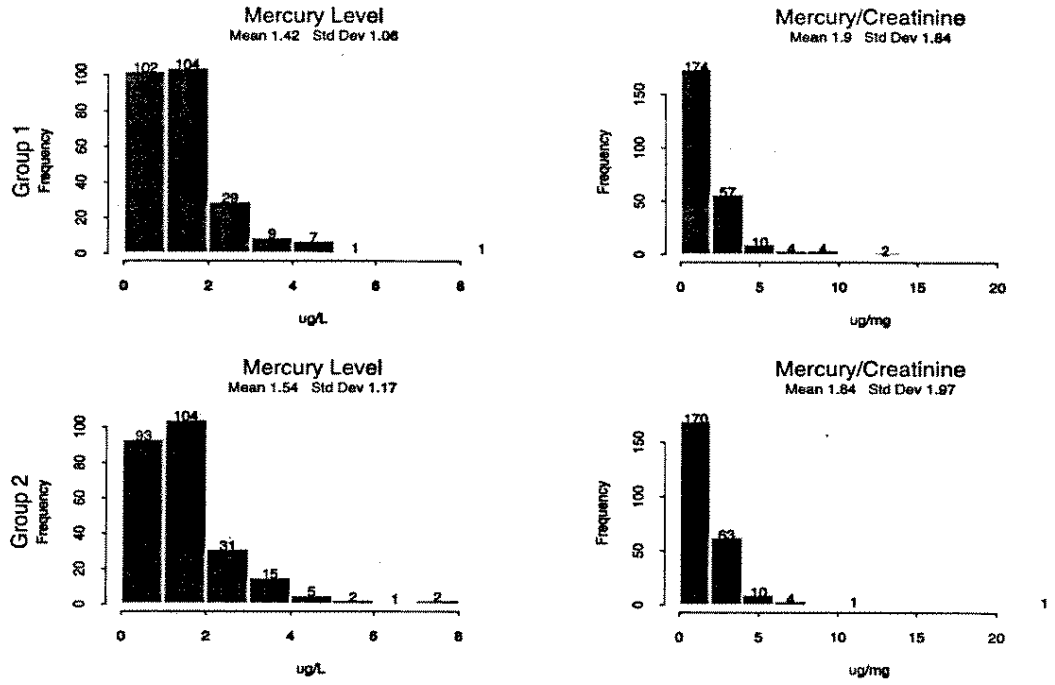
2.6 Laboratory

The distribution of baseline urinary mercury levels is shown in Figure 2.5.[6] The urinary mercury levels were low; among the randomized subjects only 7 out of 507 had levels > 5 ug/L (5 in Group 2 and 2 in Group 1). As reported in Section 1.2, only five subjects had to be excluded because of urinary mercury levels > 10 ug/L. The groups were very similar in the distribution of urinary mercury levels at baseline. There is no "normal" urinary mercury value; the average level in the U.S. population (all ages) is 3.1 ug/L.



[6] The groups are labeled "1" and "2" for laboratory data and one data value has been excluded to create equal size groups and aid in maintaining masking of the treatment groups.

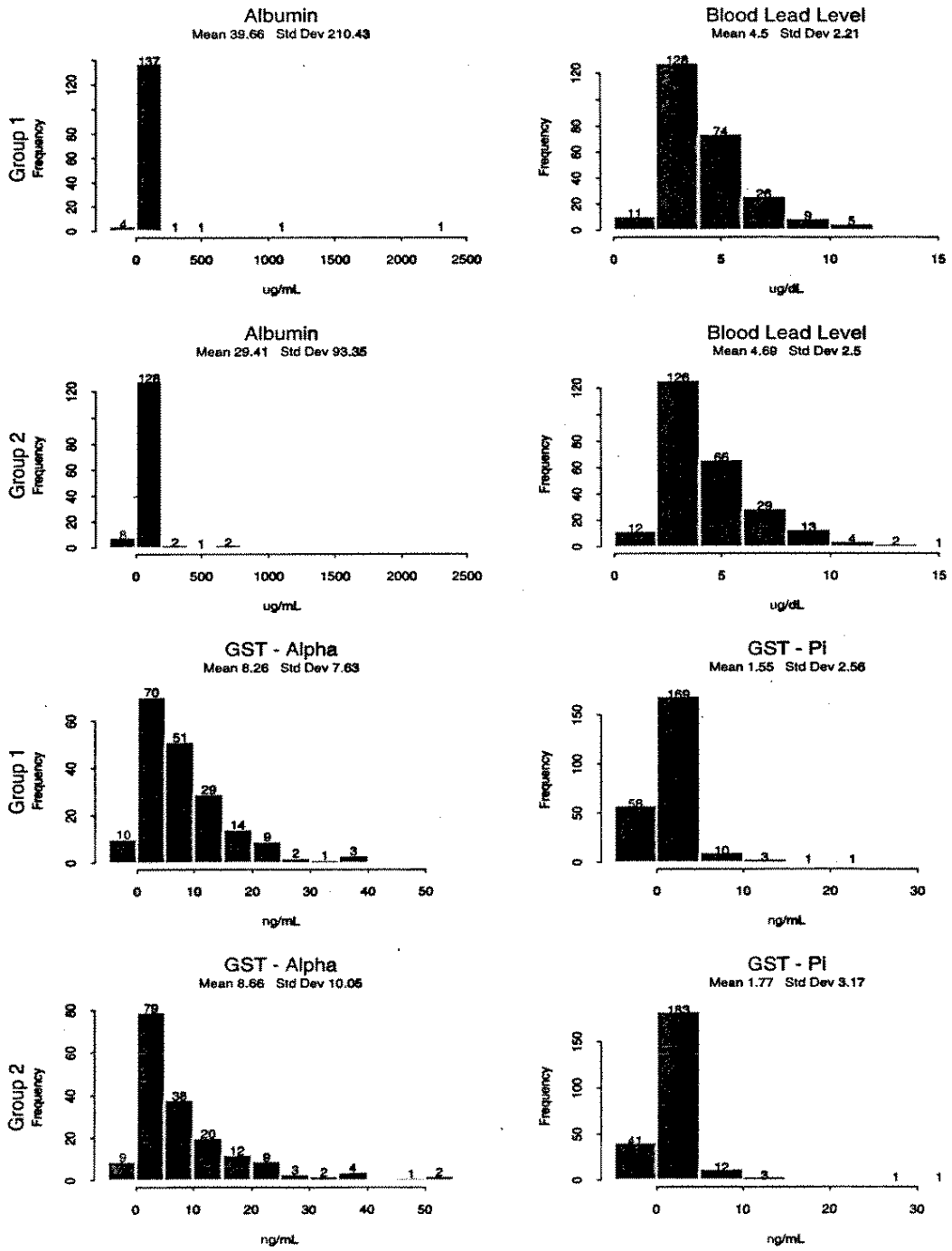
Figure 2.5. The distribution of baseline urinary mercury level and urinary mercury levels adjusted for creatinine by treatment group.



The distributions of other baseline laboratory values are presented in Figure 2.6. The blood lead levels were low and similar in the two groups; only 12 exceeded 10 ug/dl (5 in Group 1 and 7 in Group 2). No subjects had been excluded because of blood lead > 15 ug/dl; however, one subject was randomized before it was learned that his/her corrected lead value was 16 ug/dl (Group 2). Six subjects (3 in each treatment group) had elevated albumin levels (> 500).[7] The distributions of GST-alpha and GST-pi were similar in the two groups.

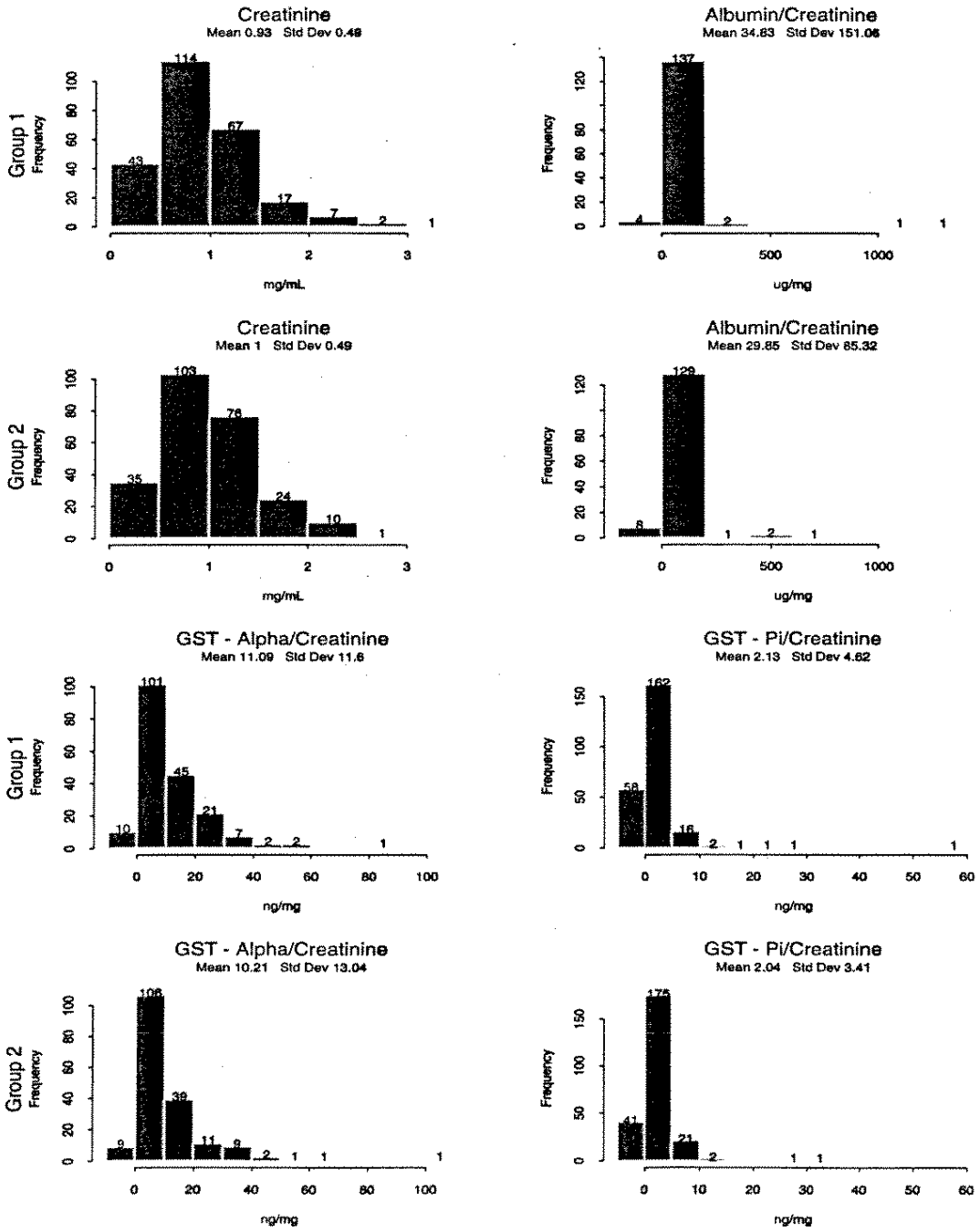
[7] Albumin levels were not measured initially; baseline albumin values are available on only 286 children.

Figure 2.6. The distribution by treatment group of additional baseline laboratory measures: albumin, blood lead, GST-alpha, and GST-pi.



Creatinine levels were also similar in the two groups, and adjustment of laboratory values by creatinine level did not alter the group comparisons for albumin, GST-alpha and GST-pi (Figure 2.7).

Figure 2.7. The distribution by treatment group of additional laboratory measures: creatinine and creatinine-adjusted values of albumin, GST-alpha, and GST-pi.



3 Dental Treatment

3.1 Treatment time and number of children receiving dental treatment

All initial dental treatment is complete. Of the 507 randomized children, all but five have received dental treatment as part of the study (of these five, two have ceased participation in the study, one could not be located, and two others did not have any restorations placed for other reasons). The amount of time from the date of randomization to the date of the last dental treatment visit was typically about 12 weeks for both groups (medians equal to 12.0 vs. 11.7 weeks). The interquartile ranges (25^{th} - 75^{th} percentile) of the treatment time were 7.1 - 19.0 weeks and 7.0 - 19.9 weeks, respectively. Thus the typical treatment time was very similar in the two groups.

3.2 Number of restorations placed

The distribution by treatment group of the number of restorations placed by surface (occlusal vs non-occlusal), tooth type (primary vs permanent) and treatment group is shown in Figure 3.1. Subjects in Group A had more surfaces restored on primary teeth than subjects in Group B (mean of 3.94 vs. 3.26), which can be explained by the fact that these subjects had more decayed primary tooth surfaces (9.4 vs 9.1, Figure 2.4). Subjects in Group A had fewer surfaces restored on permanent teeth (mean of 6.46 vs 6.93), which can be explained by a smaller number of decayed surfaces on permanent teeth (Figure 2.4, 6.2 vs 6.7). Overall the total number of restored surfaces was similar in the two groups (10.4 vs 10.2). Note that the number of surfaces restored on primary teeth was considerably less than the number of decayed primary tooth surfaces found on the initial oral exam, because of teeth being extracted or exfoliating before completion of dental treatment. On the other hand, there were slightly more surfaces restored on permanent teeth than decayed surfaces recorded at the oral exam (additional decay may have occurred between the oral exam and the end of dental treatment). Only three surfaces on primary anterior teeth were restored (all in Group B).

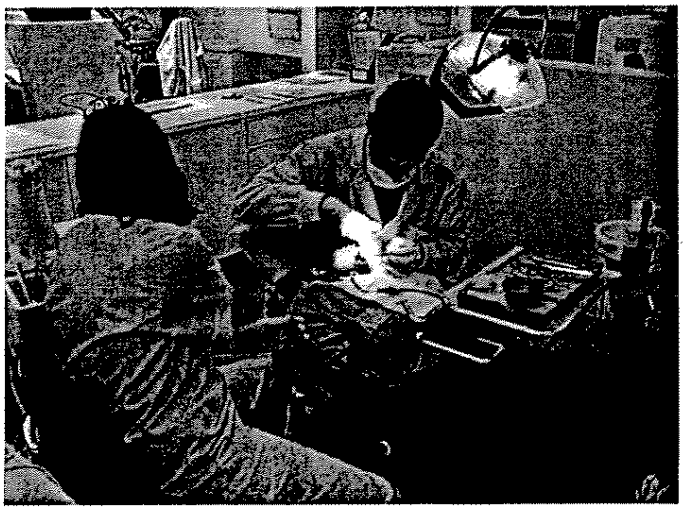
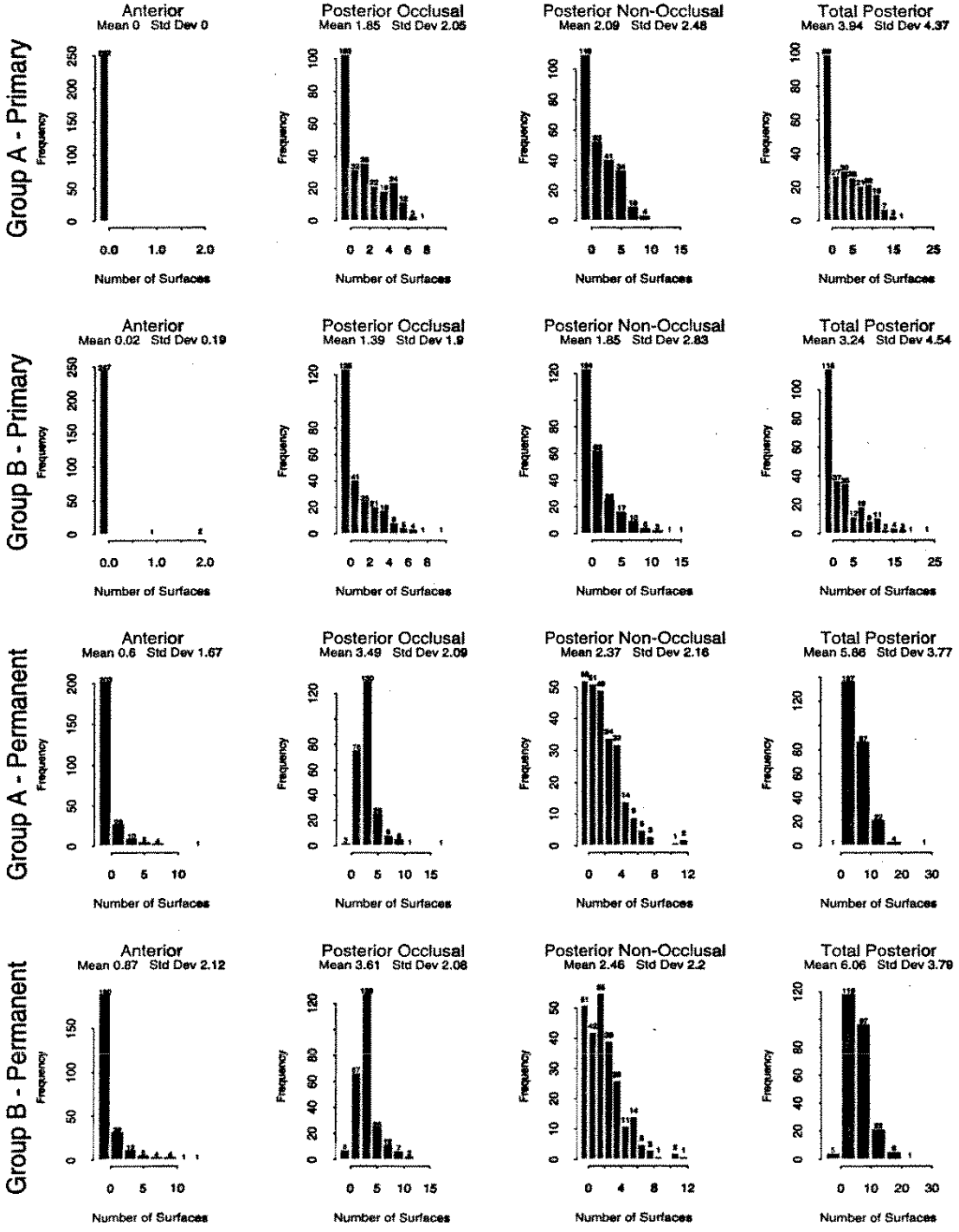


Figure 3.1. The distribution of numbers of restored surfaces by treatment group.



3.3 Mercury exposure

In Figure 3.1, the type of restoration material is not specified in order to maintain the masking of treatment groups. As an indication of the exposure to mercury received by subjects in the amalgam group, an average of 9.7 surfaces per child were restored with amalgam. In this group, 99.5% of restored posterior surfaces used amalgam, but only seven anterior teeth (nine surfaces) were restored with amalgam. Children in the amalgam group have received an additional 1.52 amalgam surfaces on average in follow-up treatment, and 99.2% of the restored surfaces on posterior teeth used amalgam.

One subject in the composite group mistakenly received amalgam restorations on 7 teeth (10 surfaces); this child also received composite restorations on 2 teeth (3 surfaces) and had 3 teeth extracted. (This protocol deviation was previously reported.)

Only one restoration was replaced within the first year after randomization (a one-surface restoration in a posterior primary tooth in a composite group subject). During follow-up treatment, a total of 58 restorations have been replaced (24 in one group, 34 in the other group). Although there were a large number of extractions performed during initial treatment (419 and 389 per group), only 5 extractions have been performed during follow-up treatment (all in one group).

4 Follow-up

4.1 Number of subjects lost to follow-up and reasons

First-year follow-up data on at least one of the primary endpoints has been obtained on 481 subjects (94.9% of the 507 randomized). For 2 of these, only nerve conduction velocity data has been obtained, and for 17 subjects, only neurobehavioral data has been obtained. Efforts continue to get complete data on all 19 of these subjects. Of the 26 children for whom we have no first-year follow-up data on the primary endpoints, a letter has been received requesting that the child be removed from the study for 10; no further attempts at follow-up will be made for these children. For the remaining 16 children, follow-up data was not obtained because the child has moved a long distance from Lisbon or could not be located for testing; however, attempts are being made to maintain contact with these children and obtain follow-up data in future whenever possible. (Second-year follow-up data has already been obtained on 5 of these 16 children.) Collection and analysis of urine samples from first-year follow-up visits has been completed for 475 subjects.

Second year follow-up neurobehavioral and nerve conduction velocity testing has been completed on 159 and 82 subjects, respectively. Urine samples have been collected and analyzed for 146 subjects.

4.2 Number of follow-up visits occurring within visit windows

In monitoring the timing of follow-up visits, particular attention is paid to the visits for neurobehavioral testing because these provide three of the primary outcome variables and the results could potentially be influenced by the timing of the visit. The target for these visits is two weeks on either side of the anniversary date. The majority of these visits (298, 62%) have occurred within this window; an additional 121 visits (25%) have occurred within 6 weeks of the anniversary, and 59 visits (12%) have occurred more than 6 weeks away. In some cases, children were seen late because they were unavailable at the scheduled time but were found to be available on return to the school at a later time. In other cases, children were seen before their anniversary because it was suspected that they would not be available

around the time of their anniversary date. The timing of follow-up visits is similar in the two treatment groups (mean time between baseline and follow-up visits of 52.5 weeks vs. 53.5 weeks). In future follow-up we may expect some children to have sporadic follow-up information; if there are differences between treatment groups, this will be handled by adjusting for time since baseline in statistical analyses.

The first-year follow-up visits for the nerve conduction velocity testing, neurological exams, and oral exams tended to occur at slightly more than one year from baseline. The reason is that these visits occurred first during enrollment, whereas the neurobehavioral testing was conducted later after eligibility status was determined. At follow-up, all visits were scheduled close together near the neurobehavioral testing anniversary date. The groups were similar in the timing of these follow-up visits (mean number of weeks between visits of 60.3 vs. 59.9 for nerve conduction velocity, 54.5 vs. 54.0 for neurological exams, and 59.6 vs. 59.2 for oral exams).

5 First-year Follow-up Results on Primary Outcome Variables

5.1 Neurobehavioral

Histograms of the follow-up results on the neurobehavioral domain scores by treatment group are provided in Figure 5.1. The histograms for individual items used in the calculation of the domain Z-scores are provided in Figure B.2 (Appendix B). The distributions of follow-up neurobehavioral scores are similar in the two groups.

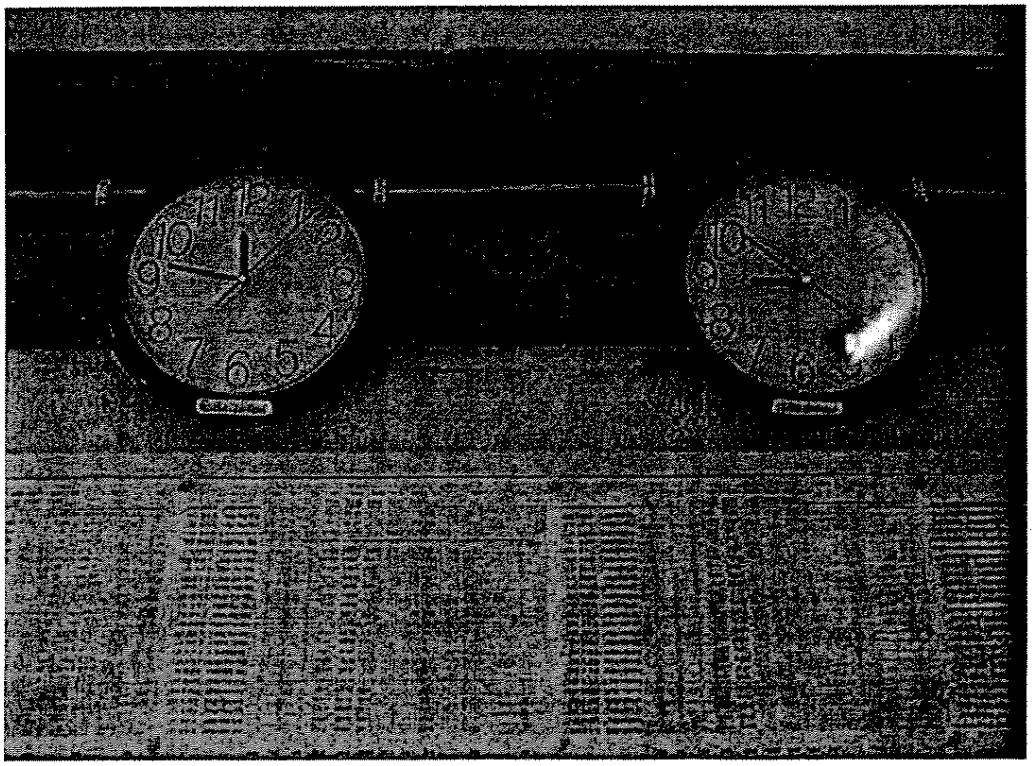
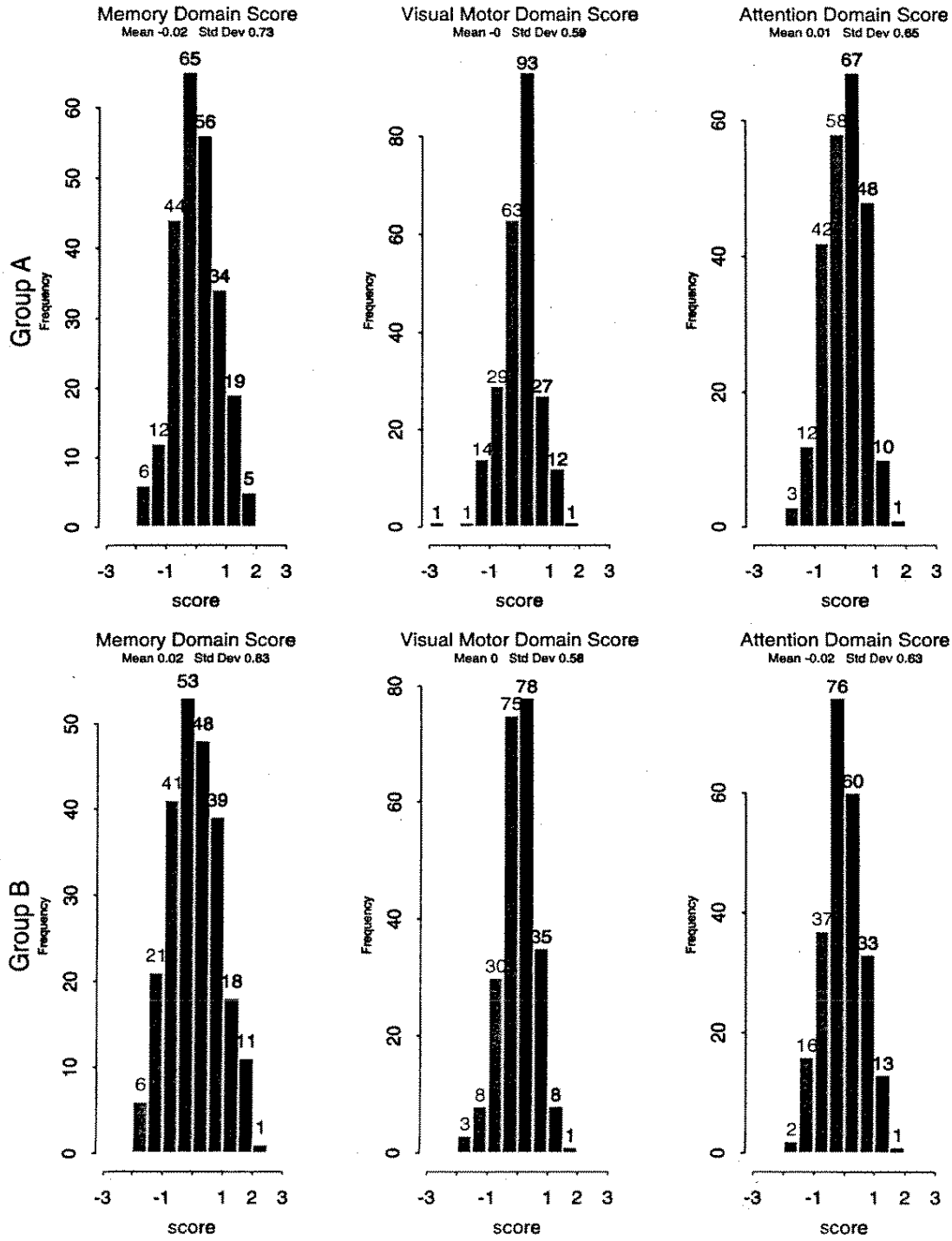
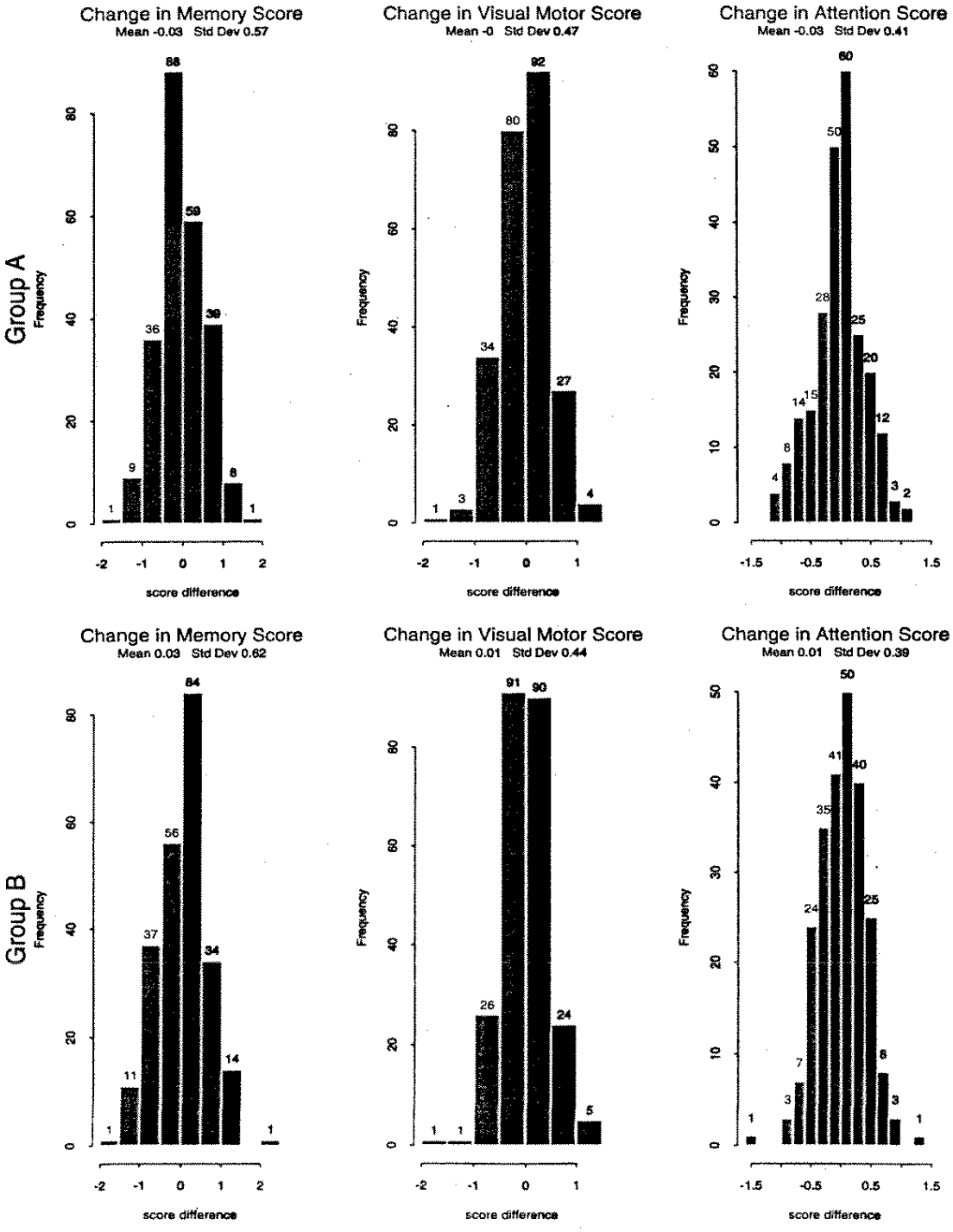


Figure 5.1. The distribution of follow-up domain Z-scores for memory, visual motor, and attention domains, by group.



Histograms of the changes from baseline (follow-up value minus baseline value) for the neurobehavioral domain scores are provided in Figure 5.2. The average change was close to 0 for each variable. Histograms of the changes in individual items used in computation of domain scores are given in Figure B.2 (Appendix B).

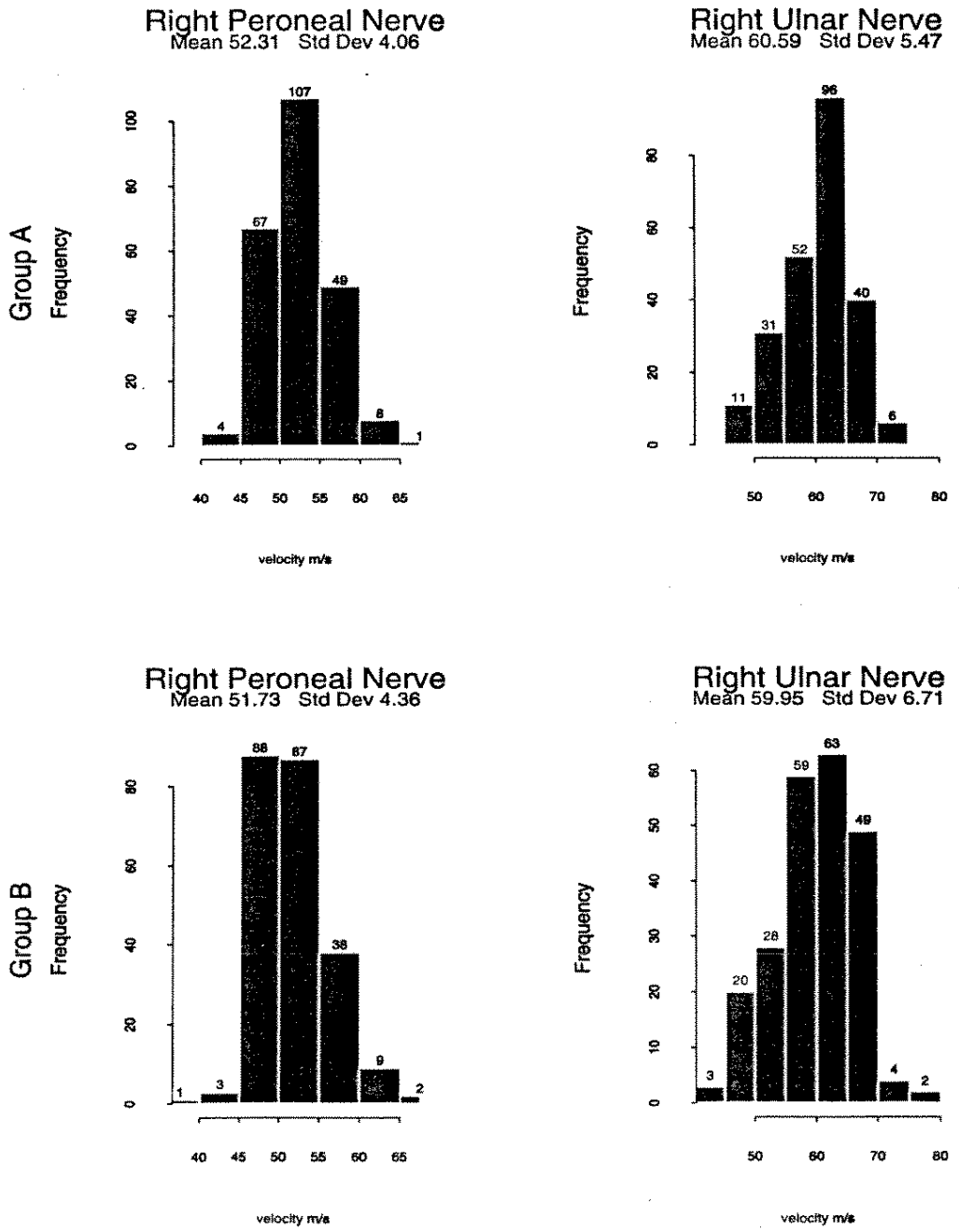
Figure 5.2. The distribution of changes from baseline in domain Z-scores for memory, visual motor, and attention domains, by group.



5.2 Nerve conduction velocity

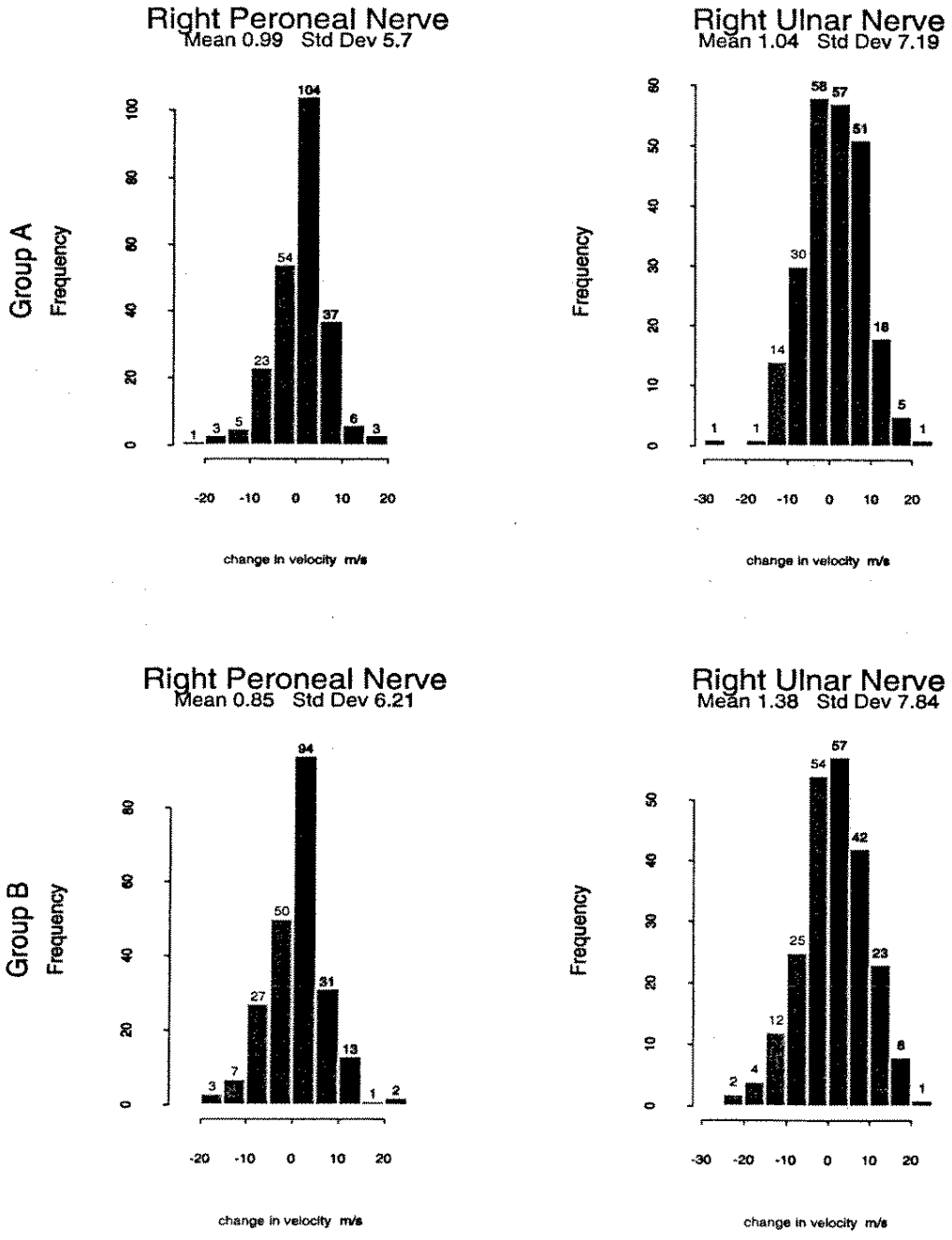
Histograms of the follow-up results on nerve conduction velocities by treatment group are provided in Figure 5.3. Additional information on the items used to calculate the velocities is provided in Table A.4 (Appendix A). The distributions of follow-up nerve conduction velocities are similar in the two groups.

Figure 5.3. The distribution of follow-up nerve conduction velocities by group.



Histograms of the changes from baseline in nerve conduction velocities are provided in Figure 5.4. The average change was approximately equal to one for each nerve in both groups; 95% confidence intervals for the mean change for the peroneal nerve are 0.27 – 1.72 (Group A) and 0.05 – 1.66 (Group B) and for the ulnar nerve are 0.13 – 1.96 (Group A) and 0.36 – 2.40 (Group B).

Figure 5.4. The distribution of changes from baseline in nerve conduction velocities by group.



5.3 Statistical stopping rule

The statistical stopping rule was presented in the manual of operations (see also Appendix C for a summary). The analysis is performed using the four primary endpoints (memory, visual motor, and attention domain scores from the neurobehavioral testing, and a combined nerve conduction velocity score). Two separate tests are performed: an O'Brien type test that will detect small group differences that are similar across the four endpoints and a Hotelling T^2 test that will detect inconsistent group differences, e.g., a difference in just one of the variables. The overall significance level of the test accounts for the overlap between these two tests (i.e., it is possible that both tests reject the null hypothesis).

The primary statistical analysis was based on the intent to treat principle, using imputed values for missing observations, and with no adjustment for covariates. Imputation of a small number of missing values in the baseline data was performed using linear regression with age, gender, dominant hand, and examiner (1 or 2) as predictors; imputation of follow-up data was done using linear regression with baseline values as predictors. A complete-case analysis was also performed (i.e., using only the 462 subjects with complete data on the four primary endpoints). A case-deletion approach was used to determine if any single subject had undue influence on the results; in this approach, each subject is excluded from the analysis one at a time, and the results in each case are compared with the results obtained using the entire sample.

The O'Brien and Hotelling tests depend on the pooled sample correlation matrix of the four endpoints, which is given in the following table:

	Memory	Visual Motor	Attention	Nerve Cond. Velocity
Memory	1.00	0.32	0.39	0.15
Visual Motor	0.32	1.00	0.53	0.13
Attention	0.39	0.53	1.00	0.09
Nerve Cond. Velocity	0.15	0.13	0.09	1.00

This correlation matrix is quite similar to that obtained from a pilot study and used in our original power calculations. The three neurobehavioral domain scores are moderately positively correlated (correlation values 0.32 to 0.53) and the nerve conduction velocity score has only a small positive correlation with the neurobehavioral scores (correlations 0.09 to 0.15).

The results of the test components for the intent to treat analysis are given in the table below along with the required significance levels for an overall significant test result (see Appendix C). The O'Brien test statistic is a two-sample t-statistic computed using a derived variable that weights the four endpoints according to the inverse of the pooled correlation matrix. A 1-sided p-value is reported for this test. The required significance levels for this test component depend on which treatment group has the worse outcome (a higher significance level is allowed for amalgam so that we will have higher power to detect adverse effects of amalgam). Note that if the O'Brien statistic p-value was intermediate between the two required significance levels, the treatment groups would have to be unmasked in order to determine if the required significance level was achieved; in this event, we would provide the treatment identities to the DSMB only in order to maintain masking of study personnel. Differences in either direction are treated the same by Hotelling's test; therefore, a 2-sided p-value is reported for this test. The selection of

significance levels for the entire follow-up period was described in the manual of operations; a summary is provided in Appendix C.

Table 5.1. Results of intent-to-treat analysis using imputed data at first-year follow-up.

	Test Statistic	P value[8]	Required Significance Levels	
			Hypothesis: Amalgam Harmful	Hypothesis: Composite Harmful
O'Brien	T=0.70	0.24	0.0075	0.0025
Hotelling's T^2	F=1.02	0.39	0.0036	

As can be seen from the above table, the two test components are not close to achieving statistical significance. Thus, there is no evidence for a difference between treatment groups in mean values of the four primary endpoints at the first-year follow-up. Results for the complete-case analysis were very similar for both the O'Brien test (T=0.83, P=0.20) and Hotelling test (F=1.01, P=0.40). The case-deletion analysis found no subjects to be influential for the O'Brien test (T=0.57 to 0.85, P=0.20 to 0.28) and the Hotelling test (F=0.91 to 1.15, P=0.33 to 0.46). No adjustment for covariates was done because the treatment groups were balanced on all important covariates.

5.4 Reassessment of power

Our original power calculations assumed effect sizes for each of the primary endpoints that were expressed as fractions of a standard deviation. Therefore, the power calculations did not make any assumptions on the values of the standard deviations. However, the power calculations did assume the values of the correlations between endpoints and also assumed that we would have five years of follow-up data on 400 subjects. Our assumptions on the correlation matrix were based partly on data from a pilot study and were quite close to the correlations actually observed.

Regarding the sample size, we enrolled 507 subjects and currently have complete follow-up on primary endpoints (neurobehavioral and nerve conduction velocity) on 462 subjects (91.1%). We anticipate that the rate of loss to follow-up will be lower in subsequent years because many of the subjects at risk for dropping out have already done so and the remaining subjects are likely to be more stable. Also, we will have follow-up data in future years on some subjects for whom we did not get first-year follow-up data (we have recovered five such subjects already in 2nd year follow-up). Therefore, the anticipated number of subjects with data at years 2, 3, 4, and 5 is 450, 438, 426, and 410, respectively. Under these assumptions, we have re-done the power calculations; the original power values presented in the manual of operations and the revised power figures are presented in Tables 5.2 and 5.3. These results show that the study is on track to have sufficient power, although the power is anticipated to be slightly lower than in the original calculation.

[8] The O'Brien p-value is 1-sided, the Hotelling p-value is 2-sided.

Table 5.2. Power to detect various sizes and patterns of adverse amalgam effects. The values given are the probabilities of rejection of the null hypothesis for the three test components in at least one interim analysis. The total sample size at years 1 to 5 is 462,450,438,426,410 divided equally between the two groups.

(a) Linear trend effect close to uniform across domains. Cumulative effect sizes (SD units): MEMORY: 0.3, ATTENT: 0.3, MOTOR: 0.3, NEURO: 0.15.

Test Component	Original Power Calculation	Revised
Amalgam	0.957	0.920
Composite	0.000	0.000
Hotelling	0.016	0.017
Total	0.973	0.937

(b) Linear trend effect in a single outcome variable. Cumulative effect sizes (SD units): MEMORY: 0, ATTENT: 0, MOTOR: 0, NEURO: 0.5.

Test Component	Original Power Calculation	Revised
Amalgam	0.452	0.468
Composite	0.000	0.000
Hotelling	0.532	0.480
Total	0.984	0.948

Table 5.3. Power to detect various sizes and patterns of adverse composite material effects. The values given are the probabilities of rejection of the null hypothesis for the three test components in at least one interim analysis. The total sample size at years 1 to 5 is 462,450,438,426,410 divided equally between the two groups.

(a) Linear trend effect close to uniform across domains. Cumulative effect sizes (SD units): MEMORY: 0.3, ATTENT: 0.3, MOTOR: 0.3, NEURO: 0.15.

Test Component	Original Power Calculation	Revised
Amalgam	0.000	0.000
Composite	0.903	0.829
Hotelling	0.041	0.040
Total	0.944	0.869

(b) Linear trend effect in a single outcome variable. Cumulative effect sizes for revised procedure (SD units): MEMORY: 0, ATTENT: 0, MOTOR: 0, NEURO: 0.5.

Test Component	Original Power Calculation	Revised
Amalgam	0.000	0.000
Composite	0.284	0.287
Hotelling	0.693	0.650
Total	0.977	0.937

First-year Follow-up Results on Secondary Outcome Variables

6.1 Sentinel events by group

Sentinel events have occurred for two children enrolled in the study: one child suffered a stroke and one has been diagnosed with epilepsy. All three of the neurobehavioral domain scores decreased at first-year follow-up relative to baseline for each of these two children; however, none of the decreases were greater than 1.25. The child with epilepsy had a decrease of 1.25 for the memory domain z score, and all of the other decreases observed for either child were less than 0.61.. In addition to monitoring of sentinel health events, we have also monitored changes in key endpoints. We identified 12 subjects with a decrease of 1.25 or more in one of the neurobehavioral domain Z-scores (6 subjects in each treatment group). We also identified 7 children with increases of 5 ug/L or more in urinary mercury (6 in the amalgam group and 1 in the composite group). One of these children with a large change in urinary mercury level also was one of the 12 with a large decrease in one of the neurobehavioral Z-scores. However, there were no other overlaps between these two groups of children. We have also monitored the status of the child with baseline blood lead of 16 ug/dL: this child did not have large decreases in neurobehavioral Z-scores.

6.2 Health history

At the suggestion of the DSMB, an annual health history questionnaire is completed for each child by the parents or guardians for the purpose of monitoring hospitalizations, illnesses, medications, and allergies. Note that any serious health events are reported as sentinel events in Section 6.1. The results from the first follow-up are provided in Table 6.1. The treatment groups were similar at follow-up on all items.

Table 6.1. Follow-up health history information by treatment group.

	Group A		Group B	
	Number with Event	Percent	Number with Event	Percent
Hospitalized in the last year?	23	9.1	20	7.9
Have any disease in the last year?	52	20.6	44	17.3
Taking any medications in the last year?	88	34.8	81	31.9
Taking any medications today?	33	13.0	26	10.2
Had any allergies in the last year?	28	11.1	25	9.8

6.3 Neurological

The prevalence of neurological abnormalities by treatment group is given in Table 6.2. In addition the prevalence of any abnormality is given in the last line of the table. Prevalences of abnormalities remain low and similar in the two treatment groups.

Table 6.2. Prevalence of neurological abnormalities at follow-up by treatment group.

	Group A		Group B	
	Number with Abnormality	Percent	Number with Abnormality	Percent
Mental Status	2	0.8	3	1.3
Cranial Nerves	4	1.7	6	2.6
Motor Exam	0	0.0	1	0.4
Sensory	0	0.0	0	0.0
Coordination	0	0.0	1	0.4
Reflexes	3	1.3	2	0.9
Gait/Station	0	0.0	0	0.0
Adventitious Movement	3	1.3	1	0.4
Any Abnormality	11	4.7	11	4.7

6.4 Oral exam

Information on oral health measures is provided as histograms by treatment group in Figures 6.1 and 6.2. There are fewer than 4 primary teeth and more than 20 permanent teeth per subject on average in each group. The groups are very similar in the distributions of numbers of teeth. Some new decay was found at the follow-up oral exam (approximately 0.7 decayed surfaces per child on primary teeth and 1 decayed surface per child on permanent teeth). The amount of new decay was similar in the two groups.

Figure 6.1. Distribution by treatment group of numbers of primary and permanent teeth at follow-up oral exam.

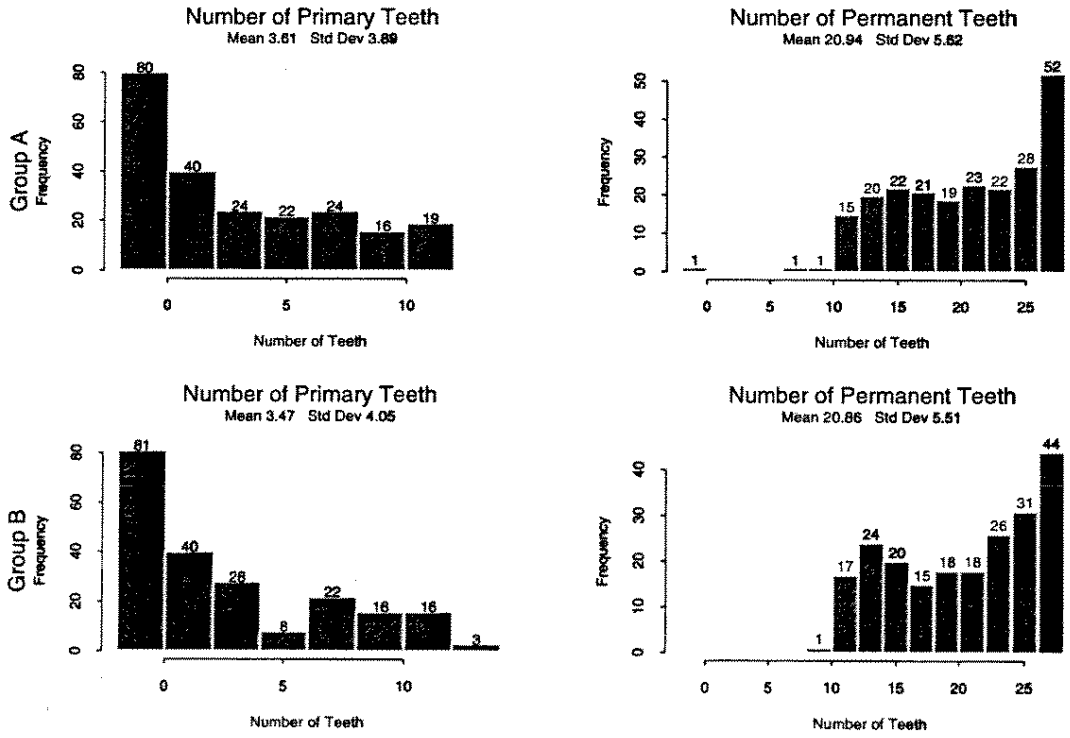
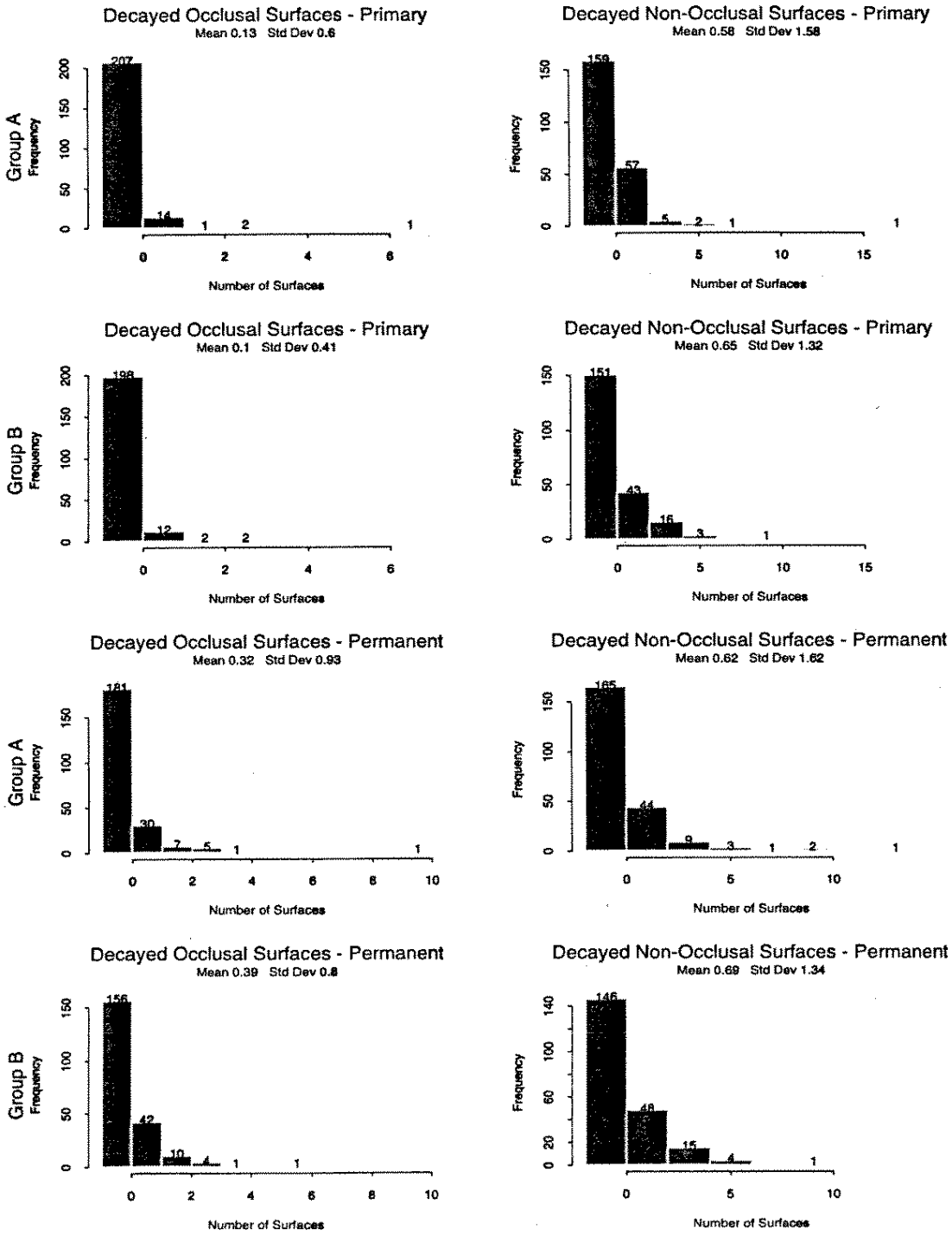


Figure 6.2. Distribution by treatment group of numbers of decayed surfaces at follow-up oral exam.

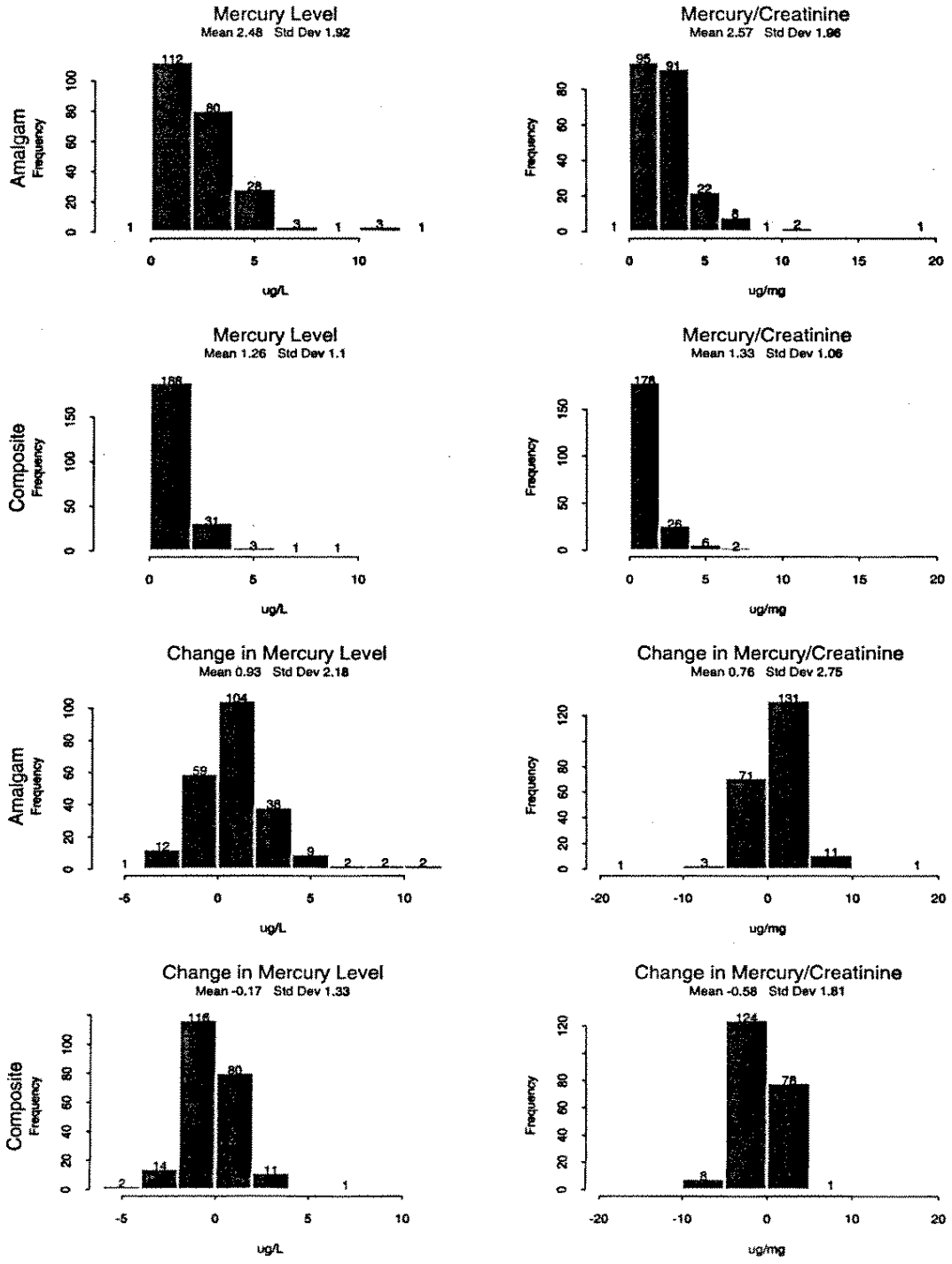


6.5 Laboratory

Distributions of follow-up urinary mercury levels and changes from baseline to follow-up are presented in Figure 6.3 for the Amalgam and Composite groups. There was an increase of 0.93 ug/L on average in the amalgam group (95% confidence interval for mean change: 0.65 – 1.21). There was no evidence for a mean change in the composite group (sample mean = -0.17, 95% confidence interval: -0.344 – 0.004).



Figure 6.3. The distribution of follow-up urinary mercury levels and changes from baseline to follow-up in urinary mercury levels (unadjusted and adjusted for creatinine) for the Amalgam and Composite groups.



Distributions of other laboratory data are presented in Figure 6.4. One subject in the composite group had an elevated albumin level (> 500 ug/ml). The distributions of GST-alpha and GST-pi were similar in the two groups although mean levels were slightly higher in the amalgam group. (Note that blood lead is not measured at first-year follow-up.)

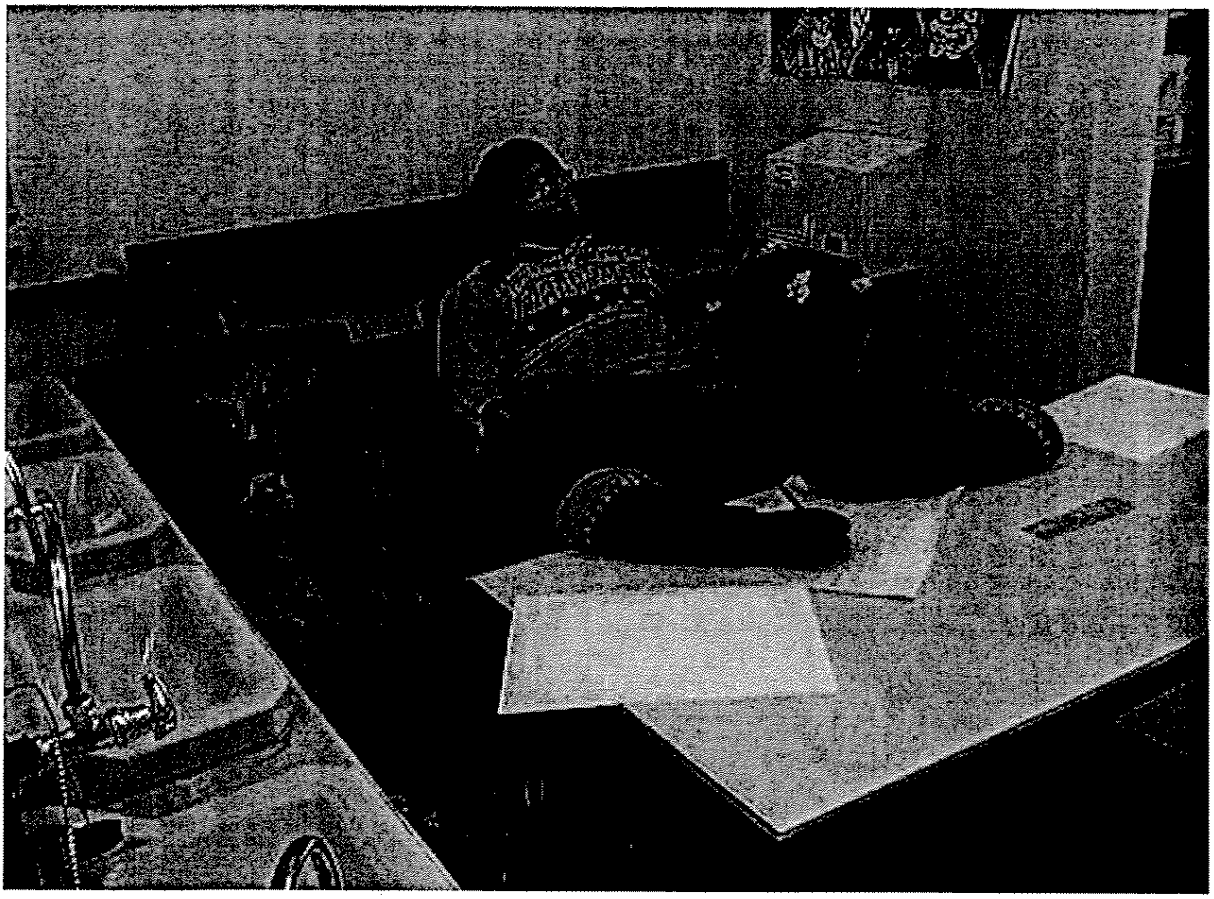
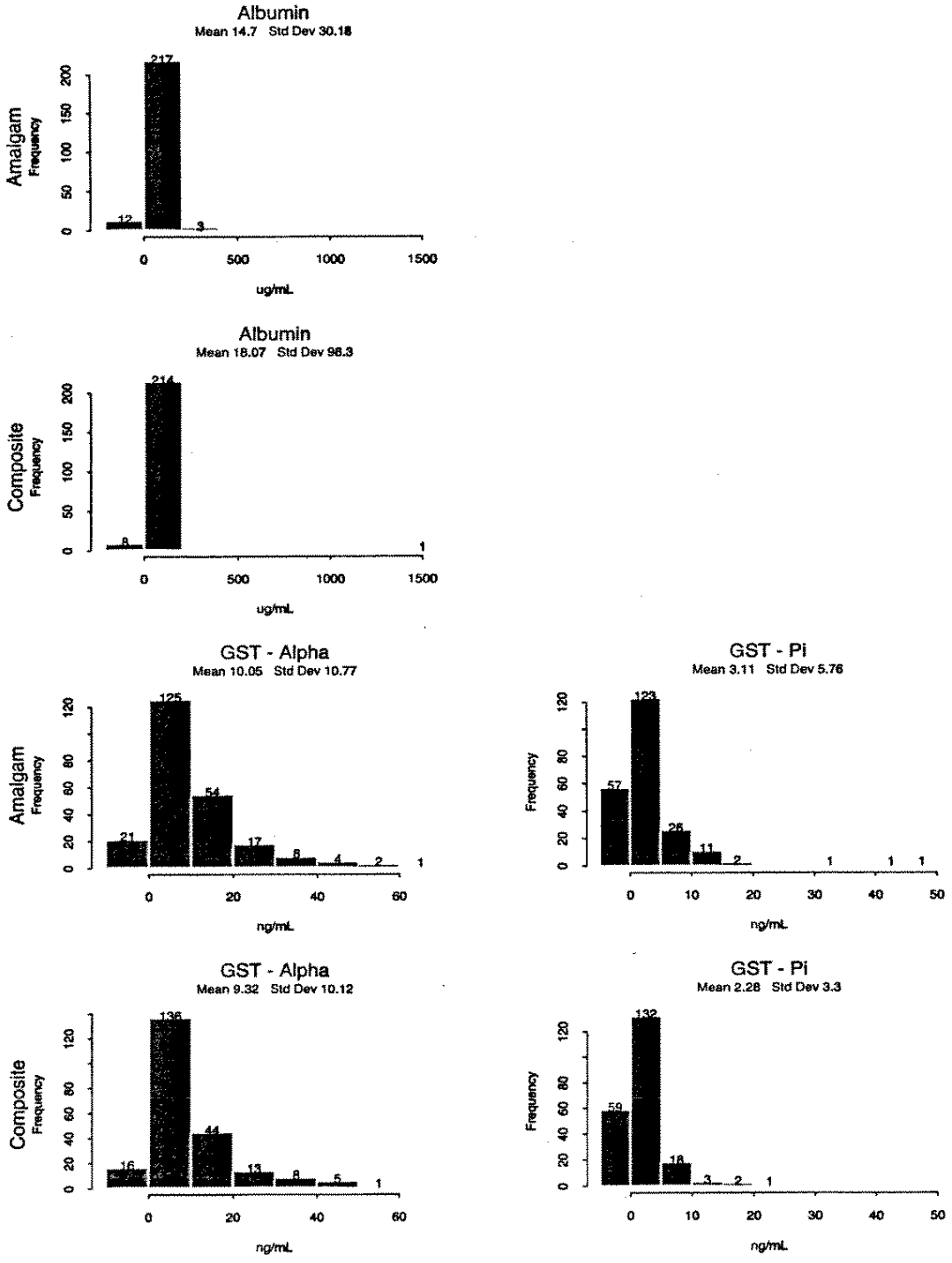
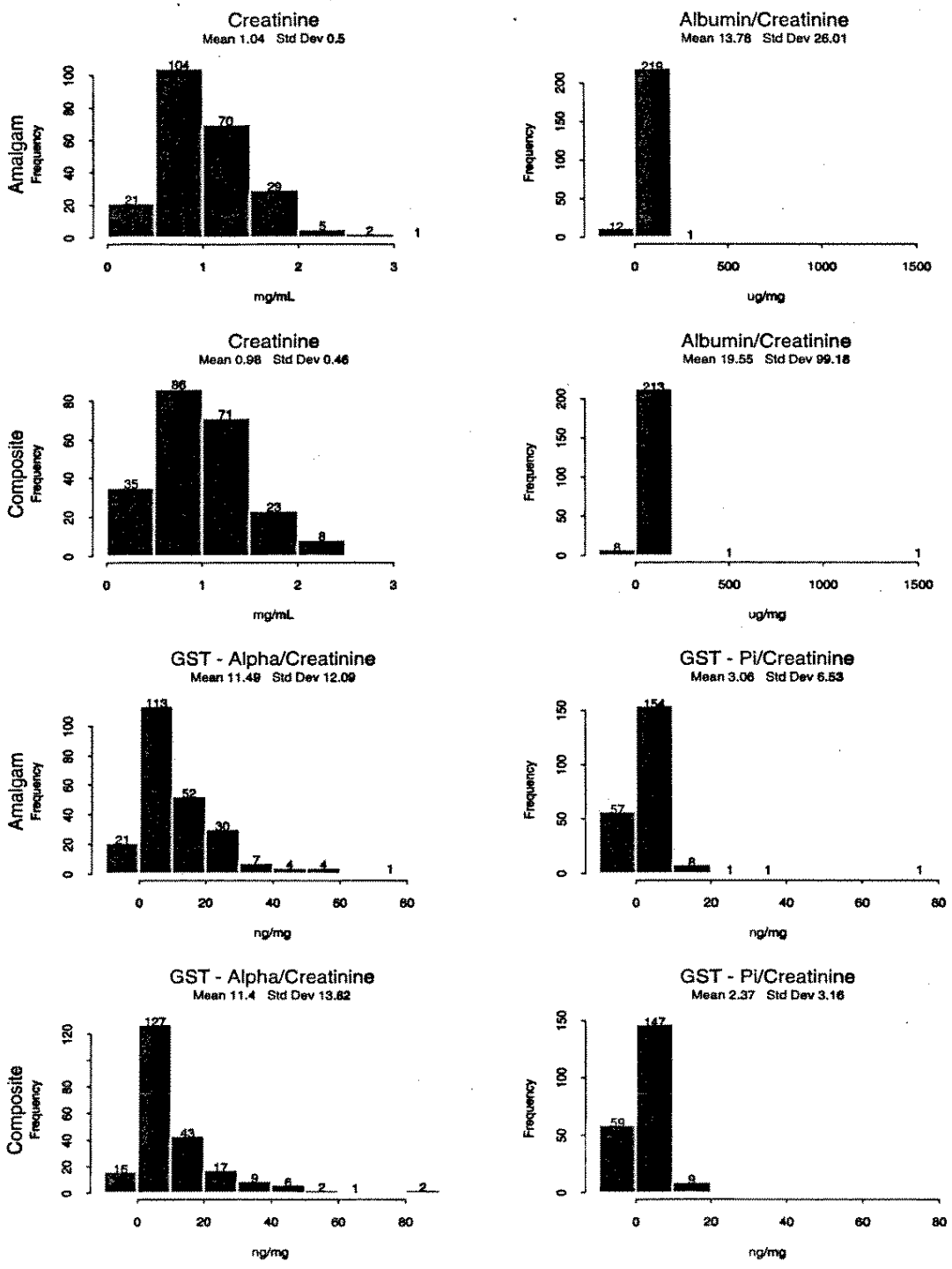


Figure 6.4. The distribution by treatment group of additional follow-up laboratory measures:, albumin, GST-alpha, and GST-pi.



Creatinine levels were also similar in the two groups, and adjustment of laboratory values by creatinine level did not alter the group comparisons for albumin, GST-alpha and GST-pi (Figure 6.5).

Figure 6.5. The distribution by treatment group of additional follow-up laboratory measures: creatinine and creatinine-adjusted values of albumin, GST-alpha, and GST-pi.



7 Quality Assurance Monitoring

7.1 Timeliness of data entry

During the recruitment period and beginning of follow-up, there was a substantial lag time from data collection to entry for neurobehavioral forms (median 33-34 days) due to the need for the two psychometrists to re-score each other's forms before data entry. After recruitment of an additional psychometrist to perform the re-scoring, the median lag time has been reduced to 3 days during the last three months of follow-up. Lag times to data entry have also been high for the oral exam forms (median 5 days during baseline and 114 days during follow-up) because the completed oral exam forms are stored in the patient dental charts until dental treatment is complete; during the last three months of follow-up, this has been reduced to 54 days. Nerve conduction velocity forms and neurological exam forms are typically (i.e., at least 50%) entered on the same day of the visit.

7.2 Completeness of data

Baseline data collection was completed on 11/24/1998. With the exception of a small amount of missing data (itemized below), all baseline data forms and baseline laboratory data have been entered. Missing baseline data:

- nerve conduction velocity: testing was not performed on one subject who dropped out of the study before this testing was done (Figure 2.2)
- neurobehavioral: 25 subjects had some missing data for the three stroop tests, 1 subject had a missing peg non-dominant score, 1 subject had a missing draw score (Figure B.1)
- neurological: one subject did not have a neurological exam; adventitious movements were recorded for only 385 subjects (beginning in March 1997); 24 subjects had missing data for some of the other items (Table 2.2)
- albumin levels are available for only 286 subjects (albumin analyses began in October 1997) (Figure 2.6)
- several non-essential items on the baseline health history form are incomplete (Table A.1)

First-year follow-up data collection was stopped on 7/20/1999 and all data have been entered. No visits were conducted for the rest of the summer due to the difficulty in arranging child visits during this time. Aside from the subjects lost to follow-up as noted in Section 4.1, there are some missing data values as follows:

- 15 subjects who were tested had missing values for one or more of the 3 stroop tests (Figure B.2)
- neurological exams were performed on 468 of the children (39 missing) (Table 6.2)
- urinary mercury, GST-alpha, and GST-pi values are available on 453, 455, and 437 of the subjects, respectively.

All descriptive statistics are computed using raw data, with no imputation of missing values. The statistical analyses reported in section 5.3 were conducted in two ways: using raw data and using an imputed data set.

7.3 Accuracy of data entry

Data entry audits on the baseline forms were performed for a random sample of approximately 15% of the randomized subjects (74/507). The mismatch rate was found to be 0.39% (132 mismatches out of 33,522 data values checked. Some of the mismatched data values identified in these audits were a result of misunderstanding by the Data Manager of notes written on the form in Portuguese or corrections made in red ink that were not visible on the photocopy of the form. Thus, the true error rate is less than 0.39%. For follow-up forms the error rate was reduced to 0.18% (18 errors out of 10,125 data values checked) based on a random sample of 75 subjects.

7.4 Examiner reliability

In our previous report we described a system for accuracy assessment of the neurobehavioral testing. The Senior Psychometrist reviews videotaped testing sessions for a sample of subjects and evaluates the performance of items on a checklist (provided previously). The results are provided in the table below.

	Examiner 1			Examiner 2		
Year of Session	Number of Sessions	Number of Items	Accuracy (%)	Number of Sessions	Number of Items	Accuracy (%)
1997	4	599	95	5	785	97
1998	7	1113	98	10	1576	98
1999	2	332	98	2	333	97

In addition, we have conducted a reliability study to assess inter-examiner agreement and also to determine if time of day may be a factor in the small systematic differences between examiners reported previously. These differences are documented in Table 7.1 for baseline and follow-up results. One possible explanation for the examiner differences shown in Table 7.1 is the fact that Examiner 2 performed testing in the morning and Examiner 1 in the afternoon. Note that these differences are not anticipated to have any substantial impact on the main study results because the treatment groups are balanced according to the number of children tested by each examiner as shown in the table below.

	Group A	Group B
Baseline:		
Examiner 1	105	110
Examiner 2	148	144
Follow-up:		
Examiner 1	102	105
Examiner 2	139	133

Table 7.1. Examiner differences in baseline and follow-up neurobehavioral testing results. Highlighted means for one examiner are significantly better than the other examiner's mean. (indicates items for which a lower score is better; for all others, higher is better)**

	Baseline				Follow-up			
	Examiner 1		Examiner 2		Examiner 1		Examiner 2	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
RAVLT T.8	8.05	3.23	8.37	2.81	7.38	2.65	7.97	2.46
RAVLT TOT	37.56	9.83	39.07	10.09	37.39	8.20	37.91	8.25
WRAML Visual Learning	7.80	2.79	8.10	2.65	7.94	2.84	8.57	2.61
WRAML DE VI	6.28	3.00	6.71	3.15	6.34	3.16	7.34	2.92
Memory Domain	-0.07	0.81	0.05	0.74	-0.12	0.78	0.09	0.77
WRAVMA Drawing	99.06	13.04	102.53	11.87	90.71	12.73	96.46	10.80
WRAVMA Matching	94.95	13.07	96.55	13.05	95.51	13.10	101.10	12.38
WRAVMA Peg –dominant	101.34	17.41	103.16	16.59	109.97	17.96	108.26	14.36
WRAVMA Peg Non-dom	105.96	16.37	106.62	14.04	111.71	16.38	108.76	12.36
Standard Reaction Time**	0.94	0.23	0.85	0.18	0.83	0.16	0.93	0.17
Finger Tapping-Dominant	35.60	6.31	37.13	5.88	38.64	6.30	39.39	5.57
Finger Tapping Non-Dom	31.53	5.51	31.78	5.25	33.83	5.84	33.60	5.22
Visual Motor Domain	-0.10	0.63	0.08	0.60	-0.01	0.62	0.01	0.56
Coding	8.52	3.15	9.09	3.08	8.97	3.11	9.09	3.04
Symbol Search	9.31	2.75	9.49	2.51	9.64	2.78	9.54	2.93
Digit Span	6.94	2.35	7.68	2.78	6.89	2.06	7.54	2.38
Finger Windows	6.76	2.48	7.70	2.27	7.23	2.34	7.13	2.25
Trails A (seconds)**	28.93	12.64	26.97	13.07	22.44	8.72	22.52	8.32
Trails B (seconds)**	60.73	26.50	68.34	39.93	50.88	29.02	52.72	29.36
STROOP WD4	41.62	6.72	42.03	6.29	42.28	7.26	41.20	6.93
STROOP COL4	42.85	6.44	44.12	5.35	40.81	6.70	40.86	6.16
STROOP CWD4	43.43	7.11	43.95	6.81	42.64	7.52	42.72	6.73
Attention Domain Score	-0.08	0.58	0.04	0.59	-0.01	0.63	-0.01	0.65

For the reliability study, a random sample of 24 children was selected from the subjects who had given consent to participate in the main study but had been determined to be ineligible. Each subject was examined on two occasions at the same time of day (morning or afternoon) separated by two weeks. Half of the subjects (chosen randomly) were examined by examiner 1 at the first visit and by examiner 2 at the second visit; the remaining subjects were examined by the two examiners in the opposite order. Time of day was also randomly assigned. Twenty-two subjects completed the testing (11 morning and 11 afternoon). Videotapes of a sample of testing sessions were reviewed for accuracy; the estimated accuracy was 98.4% for Examiner 1 based on 1137 items in 8 sessions and 99.6% for Examiner 2 based on 1108 items in 7 sessions.

The results did not exhibit the same consistent examiner differences as in the main study, possibly due to the small sample size; however the results were not inconsistent with the examiner differences seen in Table 7.1. Similarly, there are no systematic differences between test results in the morning and afternoon. Thus, the study results do not provide any evidence that the inter-examiner differences observed in the main study are explained by time of day of testing. The results of this study can be used to establish the inter-examiner reliability of the measures (Table 7.2). Two methods for assessing

measurement reliability are reported: 1) the measurement error standard deviation (calculated as the square-root of one-half the average squared difference between the two repeat scores on the same child), and 2) the correlation between the scores obtained on the same child by the two examiners. The correlations for the memory domain items ranged from 0.63 to 0.82 and was 0.90 for the memory domain score. For the visual motor domain most of the correlations for individual test items ranged from 0.75 to 0.89 (except for the Drawing item which had a correlation of only 0.41) and the domain score correlation was 0.94. The attention domain items had correlations ranging from 0.49 to 0.79 and the domain score correlation was 0.85. These results indicate that the measures of the primary neurobehavioral outcome variables have a high level of reliability as assessed by measurement error standard deviations (0.23, 0.30, 0.38) and correlations (0.85, 0.90, 0.94).

Table 7.2. Results of the inter-examiner reliability study.

	Measurement Error Standard Deviation[9]	Inter-Examiner Correlation[10]
RAVLT T.8	1.82	0.63
RAVLT TOT	3.76	0.81
WRAML Visual Learning	2.77	0.82
WRAML DE VI	2.58	0.69
Memory Domain Score	0.31	0.90
WRAVMA Drawing	10.40	0.41
WRAVMA Matching	6.22	0.83
WRAVMA Peg –dominant	9.64	0.75
WRAVMA Peg - non-dominant	7.57	0.83
Standard Reaction Time - mean	0.10	0.83
Finger Tapping-Dominant	2.80	0.87
Finger Tapping -Non-Dominant	2.20	0.89
Visual Motor Domain Score	0.23	0.94
Coding	2.01	0.79
Symbol Search	2.61	0.65
Digit Span	1.47	0.55
Finger Windows	1.30	0.74
Trails A (seconds)	5.67	0.50
Trails B (seconds)	16.69	0.49
STROOP WD4	3.68	0.78
STROOP COL4	3.46	0.69
STROOP CWD4	6.14	0.64
Attention Domain Score	0.38	0.85

[9] Calculated as the square-root of one-half of the average squared difference between the two repeat scores on the same subject.
[10] The Pearson correlation coefficient between the two repeat scores on the same subject.

8 Appendix A. Additional Tables.

Table A.1. Baseline health history information by treatment group. (denotes significant difference between groups)**

	Group A		Group B	
	Number	Percent	Number	Percent
First visit to dentist	149	58.9	146	57.5
Pain, infection or injury**	72	28.5	56	22.05
Has had dental injection	83	32.8	88	34.65
Surgery or hospitalization	88	34.8	98	38.6
Being treated by M.D.	31	12.25	26	10.2
Taking medications	29	11.5	29	11.4
Allergies	65	25.7	52	20.5
Has had rheumatic fever	4	1.6	2	0.8
Heart disease or murmur	0	0.0	2	0.8
High blood pressure	2	0.8	0	0.0
Diabetes	0	0.0	0	0.0
Seizure disorder	1	0.4	3	1.2
Bleeding problems	2	0.8	3	1.2
Has had jaundice, hepatitis or STD	9	3.6	11	4.3
Fearful of Dentist	60	23.7	62	24.4
Nursing bottle	2	0.8	4	1.6
Foods child cannot eat	31	12.25	26	10.2
Vitamin-mineral pills**	24	9.5	12	4.7
Poor or unusual appetite	56	22.1	48	18.9
Sucks or bites lips	35	13.8	34	13.4
Sucks thumb or finger	23	9.1	25	9.8
Mouth breaths	41	16.2	44	17.3
Other oral habits	108	42.7	102	40.2

Table A.2. Norms for the key neurobehavioral test items.

	Age	Norm	Std. Dev.
Memory Domain:			
RAVLT T.8	9	7.0	3.0
	10	7.5	3.0
	11	8.0	3.0
RAVLT TOT	9	37.9	9.5
	10	39.8	9.5
	11	41.3	9.5
WRAML Visual Learning	9-11	10.0	3.0
WRAML DE VI		NA	NA
Visual Motor Domain:			
WRAVMA Drawing	9-11	100	15
WRAVMA Matching	9-11	100	15
WRAVMA Peg –dominant	9-11	100	15
WRAVMA Peg Non-dom	9-11	100	15
Standard Reaction Time		NA	NA
Finger Tapping-Dominant	9	33.9	4.12
	10	37.4	5.43
	11	40.9	4.79
Finger Tapping Non-Dom	9	30.2	3.27
	10	33.2	4.86
	11	36.5	4.75
Attention Domain:			
Coding	9-11	10	3
Symbol Search	9-11	10	3
Digit Span	9-11	10	3
Finger Windows		NA	NA
Trails A (seconds)	9	25.1	10.5
	10	22.6	6.6
	11	20.9	6.3
Trails B (seconds)	9	54.9	22.0
	10	52.7	26.3
	11	40.5	16.1
STROOP WD4	9-11	50	10
STROOP COL4	9-11	50	10
STROOP CWD4	9-11	50	10
IQ	9-11	100	15

Table A.3. Results of baseline nerve conduction velocity exam by treatment group.

	Group A				Group B			
	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
Ulnar Distance (mm)	165.03	22.27	100.00	230.00	165.77	23.79	105.00	250.00
Ulnar Latency (msec)	2.78	0.46	1.90	4.80	2.79	0.55	1.50	6.30
Ulnar Velocity (m/sec)	59.57	6.39	42.40	80.60	58.75	6.51	33.70	78.80
Peroneal Distance (mm)	323.51	29.33	255.00	415.00	321.48	32.40	250.00	430.00
Peroneal Latency (msec)	4.50	0.81	3.00	7.40	4.46	0.96	2.50	9.10
Peroneal Velocity (m/sec)	51.12	5.29	33.40	70.00	51.00	5.58	36.80	74.00

Table A.4. Results of follow-up nerve conduction velocity exam by treatment group.

	Group A				Group B			
	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
Ulnar Distance (mm)	192.26	21.57	140.00	320.00	190.61	21.17	145.00	295.00
Ulnar Latency (msec)	2.81	0.46	1.70	5.60	2.84	0.48	1.50	5.30
Ulnar Velocity (m/sec)	60.59	5.47	46.00	74.20	59.95	6.71	41.70	78.90
Peroneal Distance (mm)	348.28	31.84	280.00	430.00	343.96	36.85	109.00	490.00
Peroneal Latency (m*sec)	4.78	1.11	3.10	14.80	4.61	0.88	2.80	7.40
Peroneal Velocity (m/sec)	52.31	4.06	42.00	65.20	51.73	4.36	37.70	66.10

9 Appendix B. Additional Figures

Figure B.1. The distribution by group of baseline scores on neurobehavioral items used to compute domain scores.

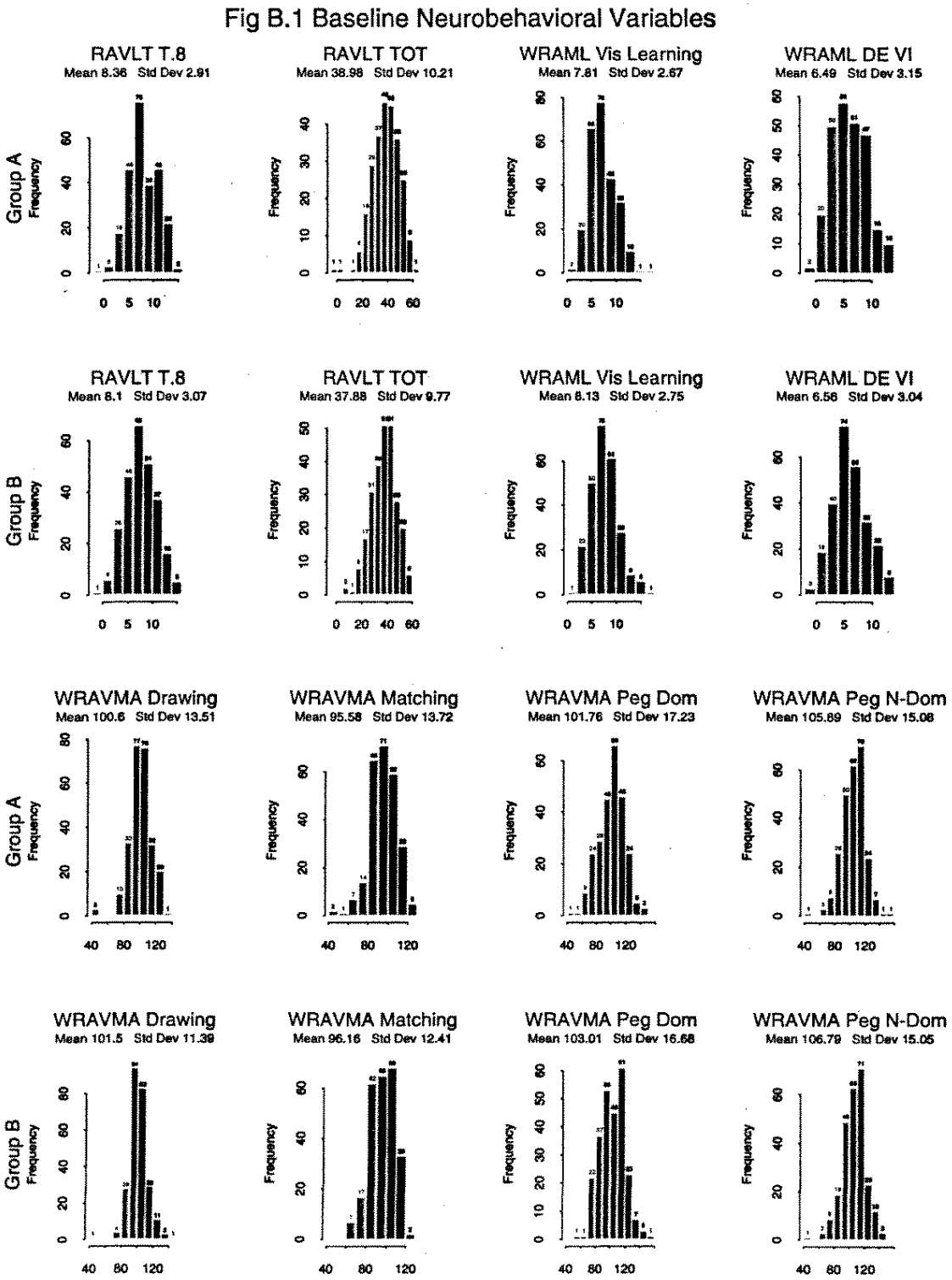


Fig B.1 (continued)

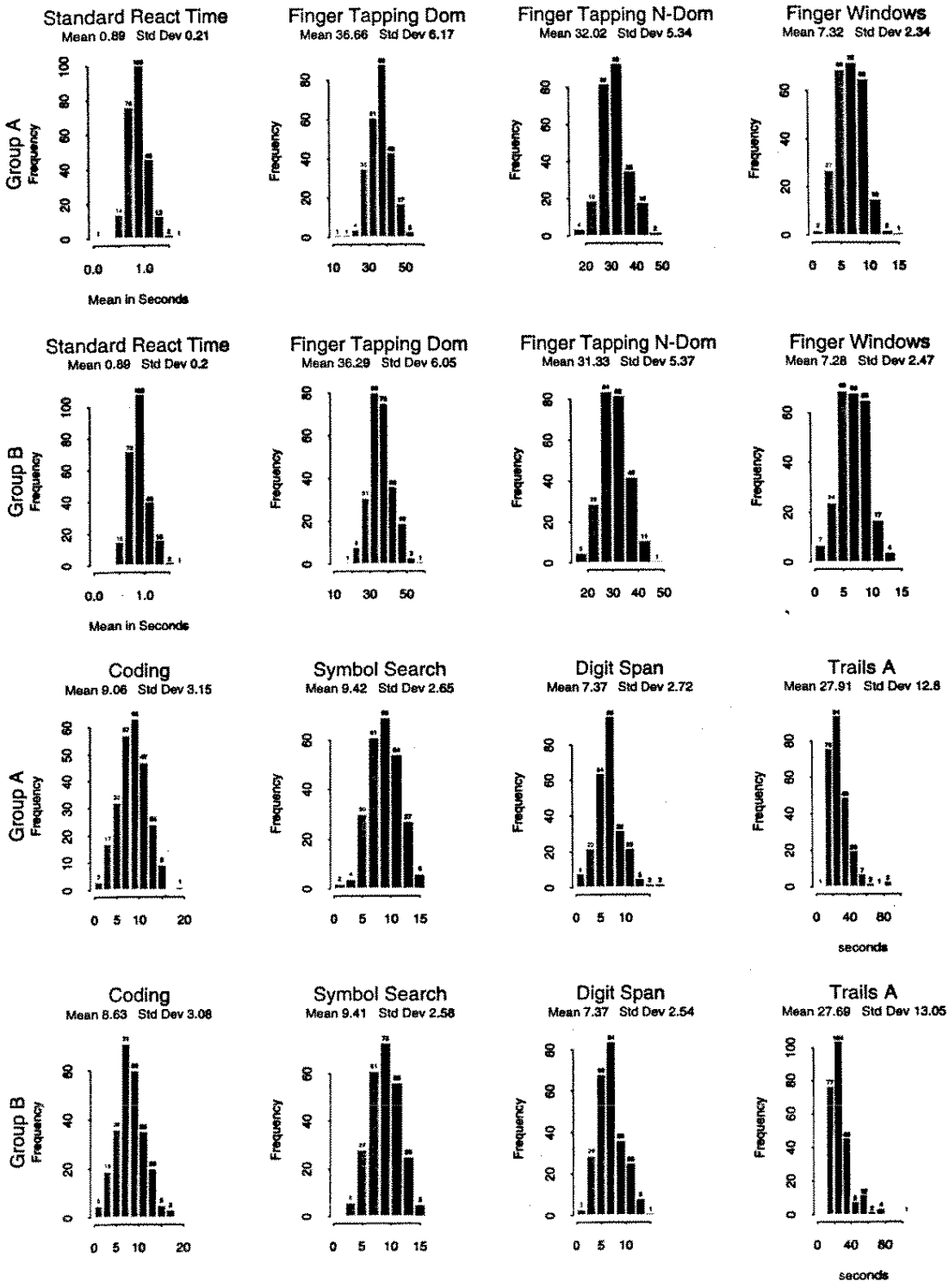


Fig B.1 (continued)

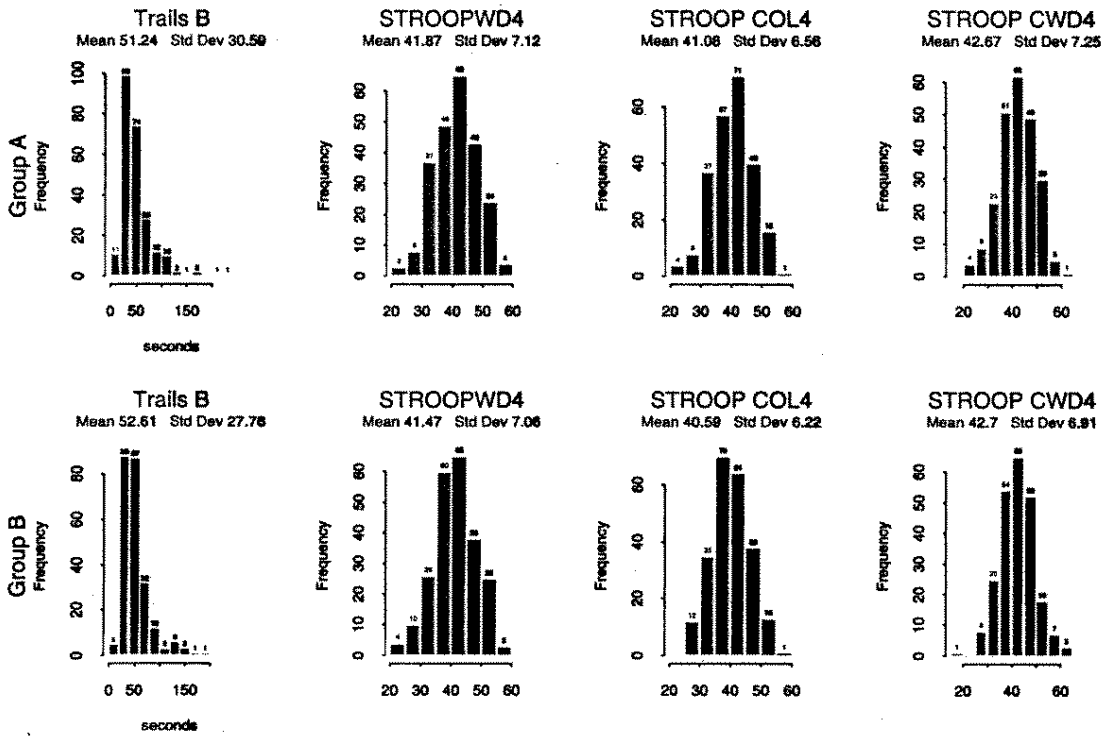


Figure B.2. The distribution by group of first-year follow-up scores on neurobehavioral items used in computation of domain Z-scores.

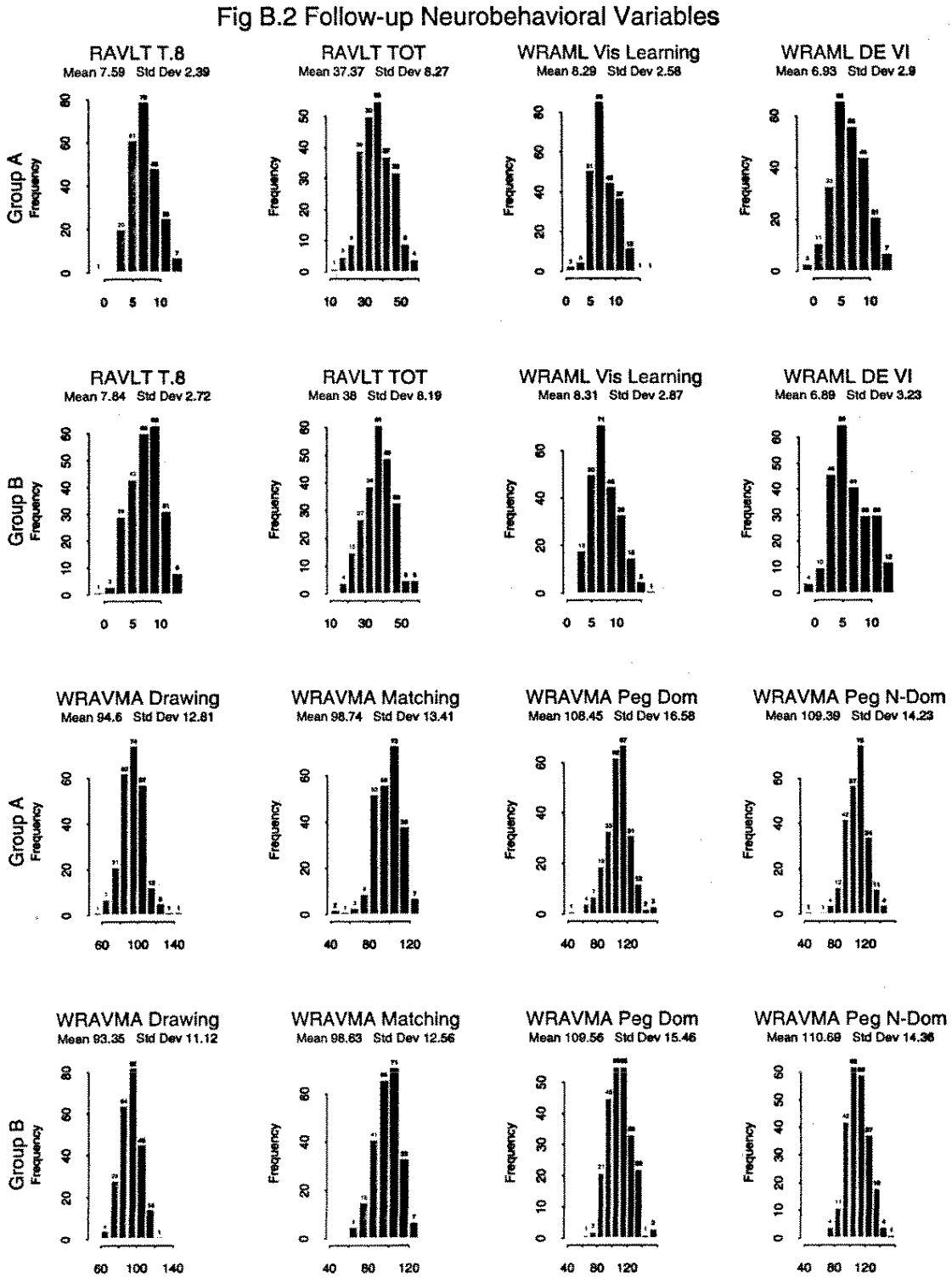


Fig B.2 (continued)

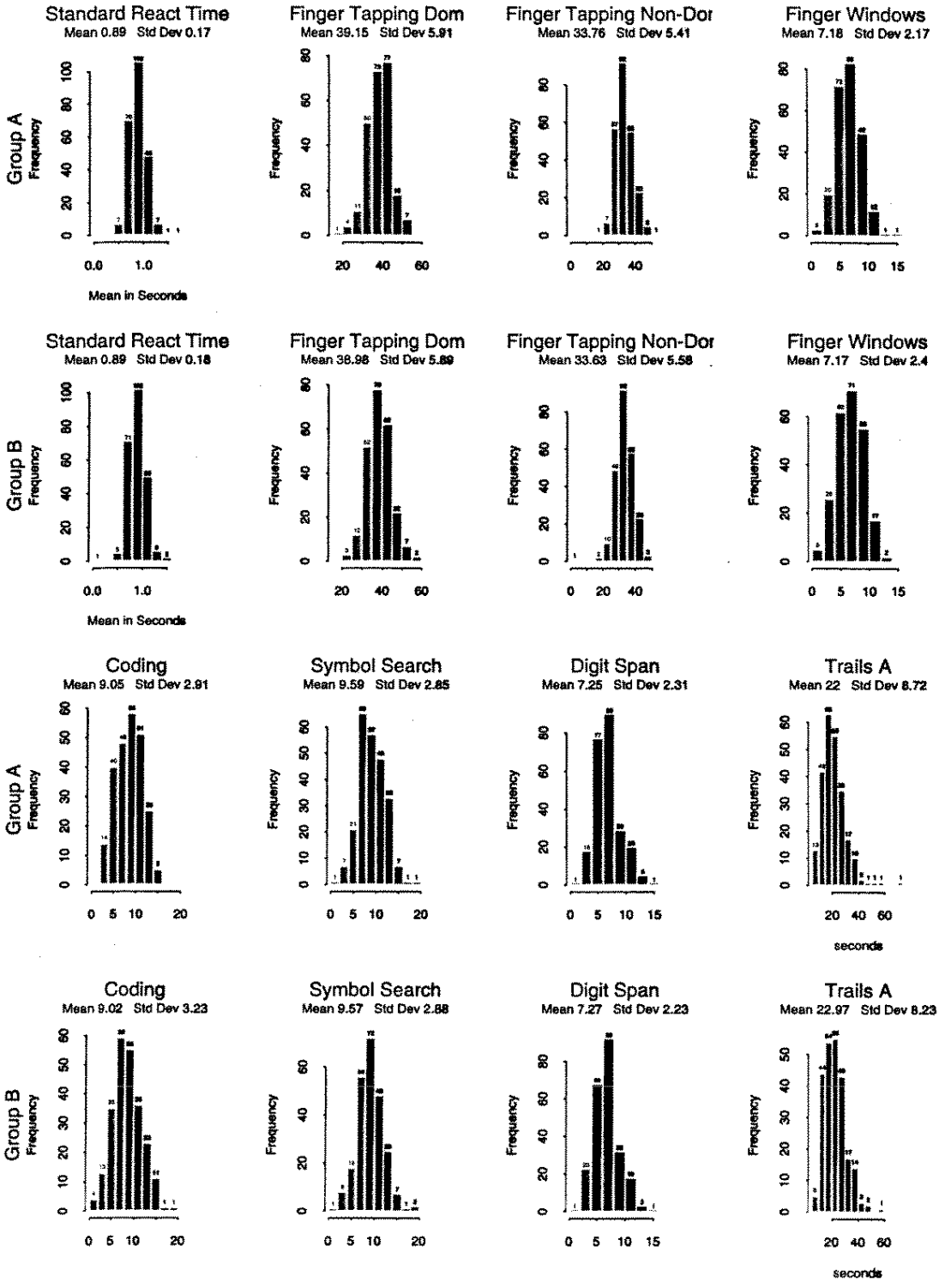


Fig B.2 (continued)

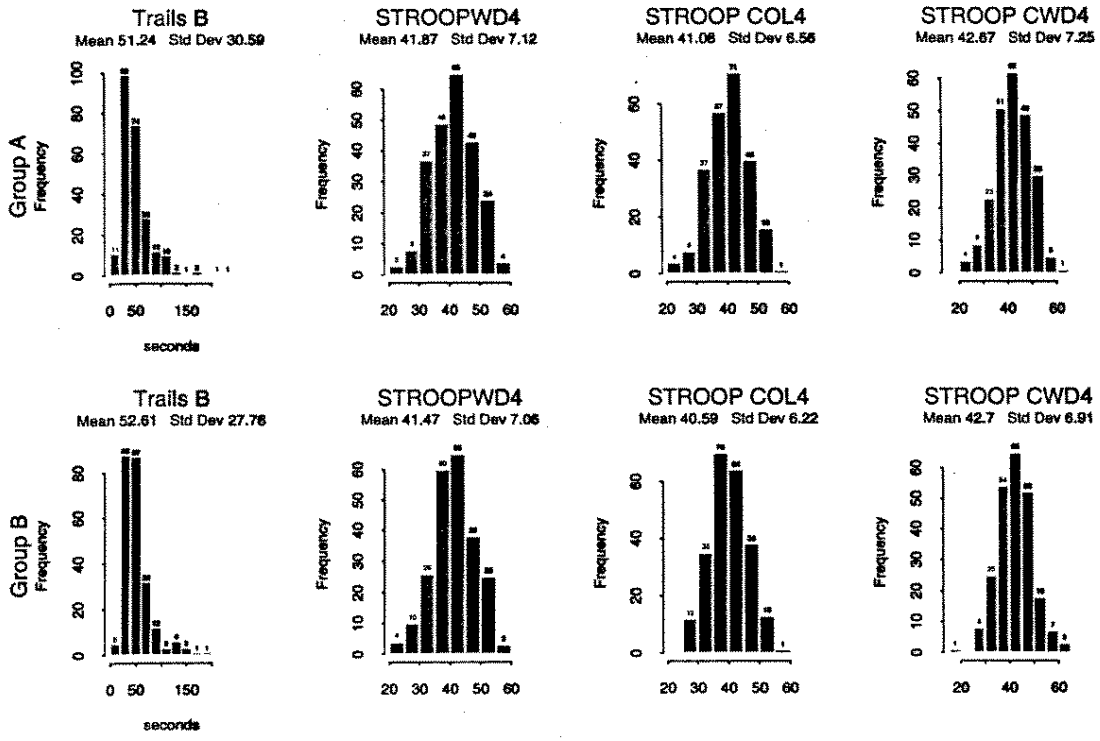


Figure B.3. The distribution by group of changes from baseline (follow-up minus baseline) in neurobehavioral items used in computation of domain Z-scores.

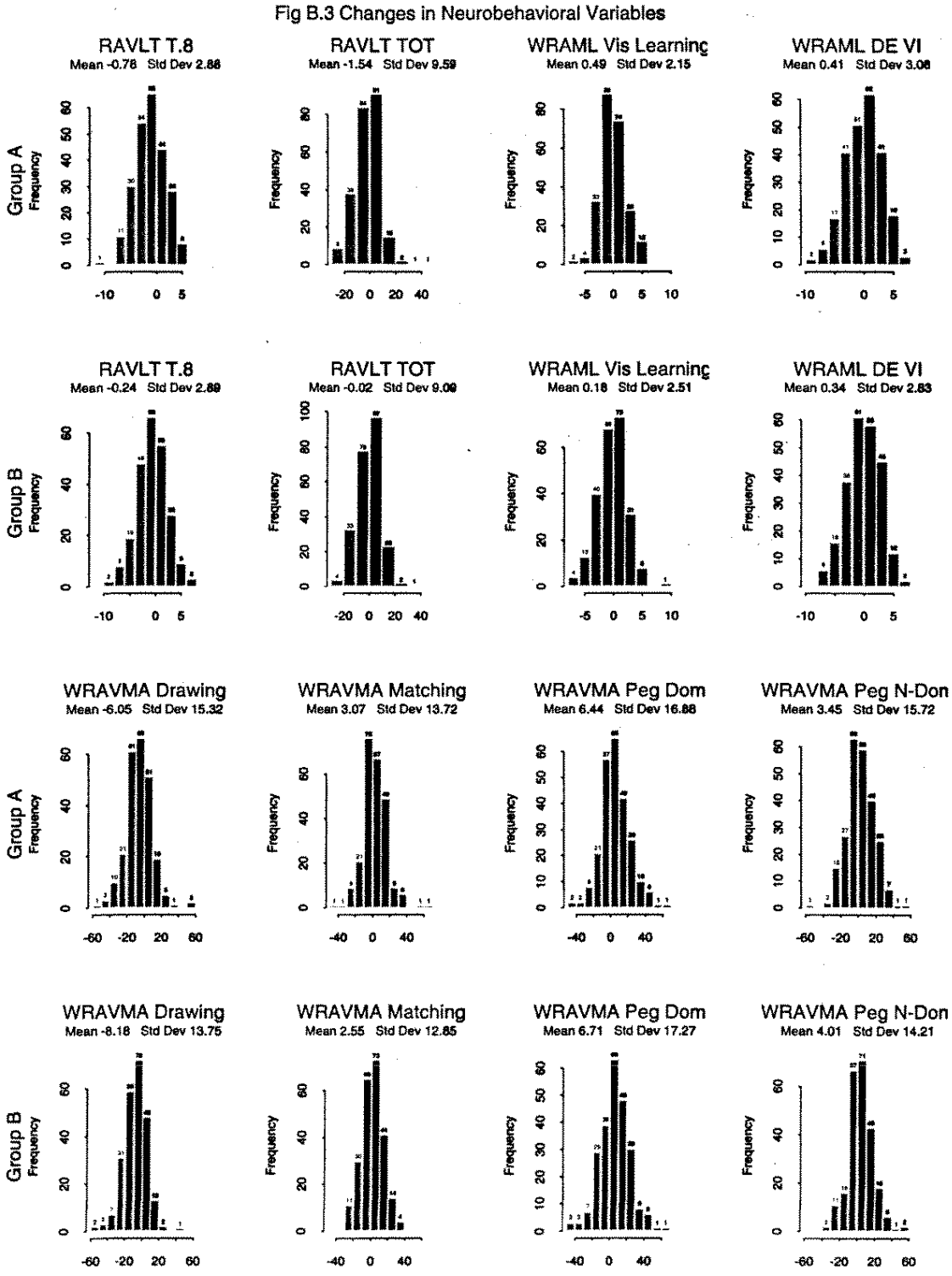


Fig B.3 (continued)

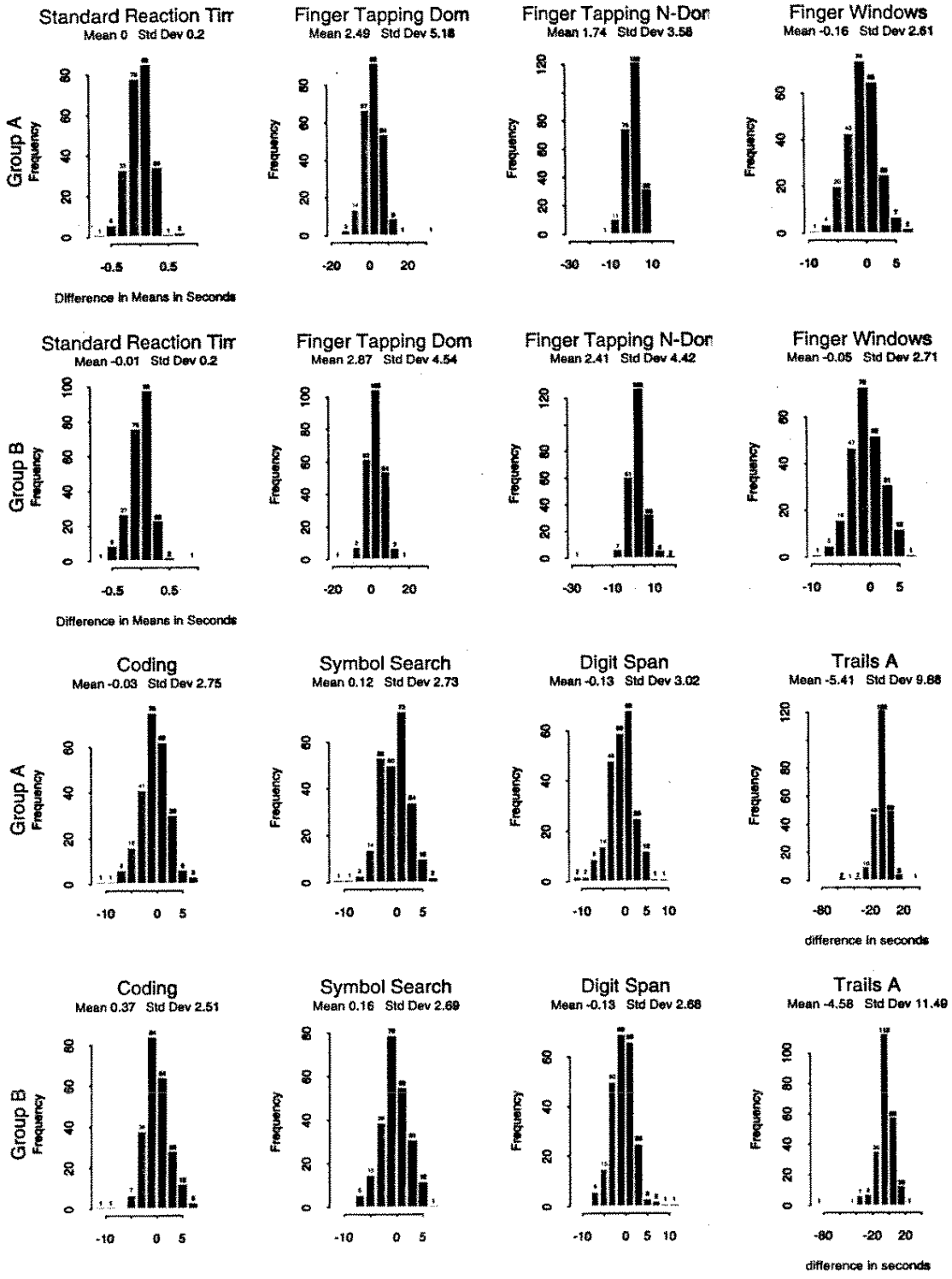
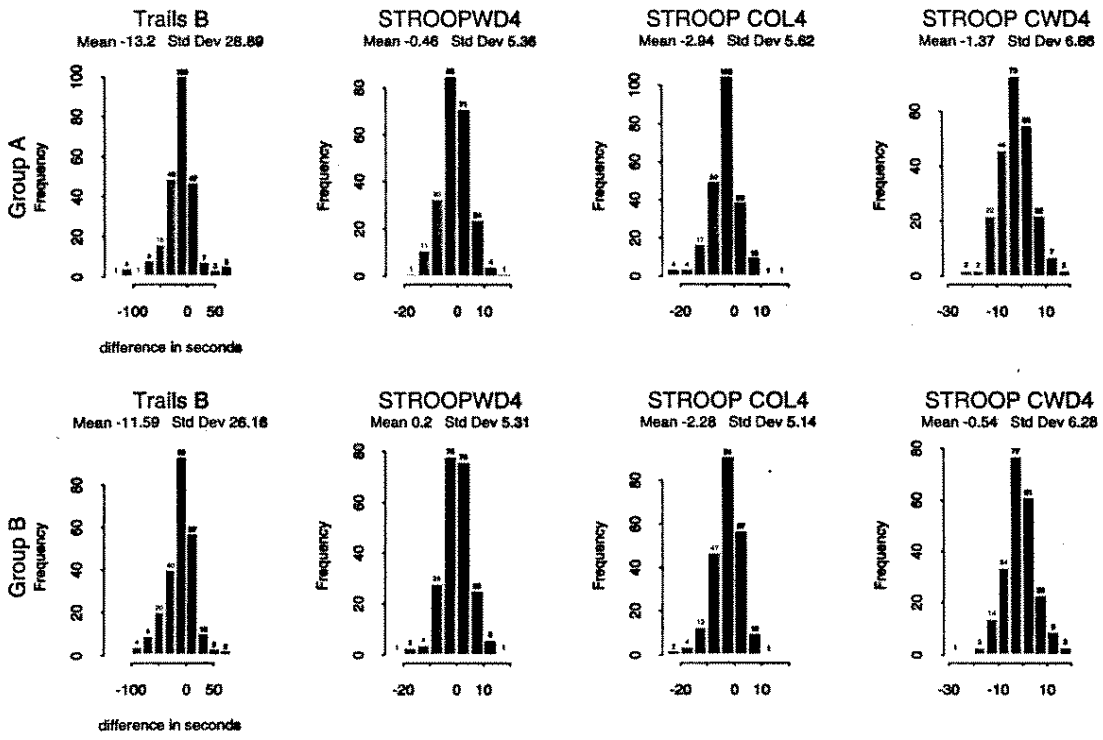


Fig B.3 (continued)

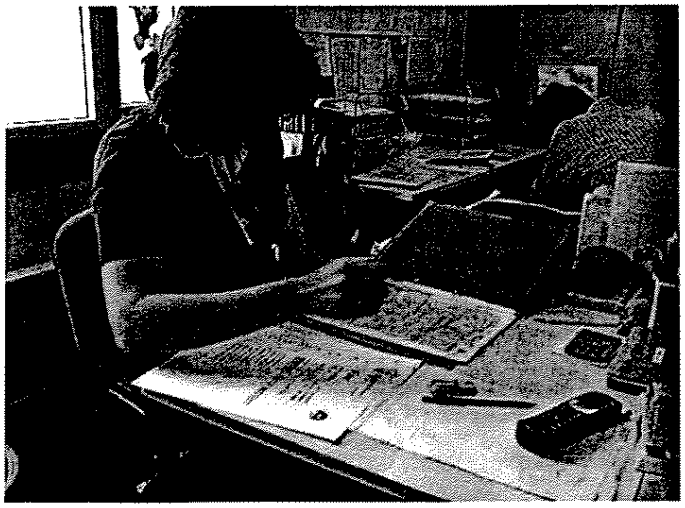


10 Appendix C. The Statistical Stopping Rule.

The total significance level for the trial will be 0.05, which will be divided among the five interim analyses as follows: 1) 0.0125, 2) 0.0025, 3) 0.0025, 4) 0.0025, and 5) 0.03. The significance levels to be used in the interim analyses for each of the components are as follows: test for adverse effects of amalgam (total 0.03): 1) 0.0075, 2) 0.0015, 3) 0.0015, 4) 0.0015, and 5) 0.018; test for adverse effects of composite material (total 0.01): 1) 0.0025, 2) 0.0005, 3) 0.0005, 4) 0.0005, and 5) 0.006; Hotelling's T-squared test (total 0.01): 1) 0.0025, 2) 0.0005, 3) 0.0005, 4) 0.0005, and 5) 0.006. These values are the probabilities of occurrence at each of the interim analyses of each of the possible significant findings described above, given that there are in fact no differences between groups.

The nominal significance levels that are used to conduct the tests differ from these significance levels because of correlations between tests conducted in the successive analyses and because of overlap between the O'Brien and Hotelling tests. The nominal levels were calculated using the estimated correlation matrix from the current data. The results were the same as those obtained in the manual of operations with an assumed correlation matrix: 0.0075, 0.0027, 0.0027, 0.0029, 0.022 for the Amalgam test; 0.0025, 0.0008, 0.00078, 0.0082, 0.007 for the Composite test; and 0.0036, 0.0010, 0.0010, 0.0011, 0.010 for the Hotelling test.

Details on the calculation of the test statistics and power were given in the manual of operations.



11 Appendix D. Results of the Bacterial Resistance Study.

As of the end of June 1999, we have collected 135 one year follow-up samples from the initial 150 children enrolled. The remaining children were either unavailable for the collection or have been lost to the study. The majority of the first year samples were collected within plus or minus 4 weeks from the time the initial samples were collected. We are currently analyzing the 1 year time points and will compare each child's samples with each other as well as look at the 2 treatment groups.

We have now collected and frozen approximately:13,000 aerobic isolates, 1,400 anaerobic isolates, 300 black pigmentors and 1,600 urine isolates. We now have appropriate Hg resistant (Hgr) controls for S. aureus, coagulase negative staphylococci, enterococci (all Gram-postive), and E. coli (Gram-negative).

The analysis of the oral aerobic bacteria suggests that those in group A have slightly higher resistance at the after-treatment (3 months) samples. These differences were mainly not statistically significant and were mostly accounted for by adjusting for slight differences in baseline resistance levels in a multiple linear regression.

In the after-treatment samples 8 children grew black pigmentors in their cheek samples (4 treatment group A and 4 treatment group B) as compared to 6 in the initial samples. None of the culture positive children in the after-treatment samples had black pigmentors cultured from the baseline sample. Eighteen children grew black pigmentors from their gums in the after-treatment samples of which only 1 was positive at baseline. At baseline, 10 children were positive for black pigmentors from their gum samples. In the one year gum samples 12 treatment group A and 6 treatment group B were positive. In contrast, from the cheek sample there were 4 from treatment group A and 4 from treatment group B. The numbers are too small to make anything of this finding, because based on the binomial distribution, there is about a 24% chance of getting a split this uneven with these small numbers. Perhaps by the end of the study, with multiple time points we might be able to see some differences between the 2 treatment groups.

The age of the children with black pigmentors after treatment for both cheek and gum was between 9-11 years of age while those children with black pigmentors in initial sample was between 10-11 years of age. However, since only one child had black pigmentors isolates in both samples it is likely that the carriage of black pigmentors in the cheek and gum is either intermittent or generally low so cultures are positive only occasionally.

We found that screening on nonblood containing media such as BHI or MacConkey agar we could use 200 mM mercury and select for Hg phenotypically resistant isolates. This was done on 84% of the urine samples and 91% of the aerobic oral cultures. There was a reasonably good correlation between ability to grow on 200 mM mercury supplemented media and presence of a known Gram-positive or Gram-negative mercury resistance gene. This will allow us to track the mercury resistant population directly for all bacteria that do not require blood for growth and compare numbers over time.

Only one Gram-negative Hg structural gene has been well characterized, though it is very likely that multiple genes exist in the Gram-negative bacteria. We have now phenotypically characterized 245 Gram-positive isolates and 106 Gram-negative isolates from the urine samples from initial, after-treatment and one year. Only one isolate of any one species from any one sample site (A or B) were counted but if different species from the same sample were found each species was included in the count. The majority have also been tested for known Gram-positive and Gram-negative Hgr genes. Only 16 of

the isolates examined were antibiotic susceptible and Hgr or Hgi while more isolates were antibiotic resistant but Hg susceptible (Hgs).

Of the 245 Gram-positive isolates, 70 (35%) were positive and 128 (65%) were negative for the Hgr Gram-positive oligonucleotide probes in the urine samples. The remaining 47 isolates will be tested for with Hgr Gram-positive oligonucleotide probes. Of the 106 Gram-negative isolates, 48 (53%) were positive and 43 (47%) were negative for the Hgr Gram-negative oligonucleotide probes. The remaining 15 Gram-negative isolates will be examined.

Note that there was a very similar distribution for Hgr Gram-positive and Hgr Gram-negative isolates at initial and 3 months but a slight increase in treatment group A for Hgr Gram-positive bacteria at one year. Whether this is related to the slight difference for baseline resistance in aerobic bacteria in the gum (sample A) or of no significance will be examined later this year. Whether any pattern is evident in the aerobic oral bacteria will be examined after more isolates have been tested for the Hgr oligonucleotide probes. In the preliminary analysis we found no significant difference between two treatment groups for carriage of Hgr Gram-negative bacteria in the urine. We are currently working on the oral bacteria where more differences due to more direct exposure could be expected.

We have screened 8 Hgr antibiotic resistant Gram-negative urine isolates for the ability to transfer both the antibiotic resistance and Hgr characteristics together by conjugation. In all 8 matings both Hgr and antibiotic resistance were transferred together suggesting that both are on the same conjugative unit. We will try moving isolates that just carry the Hgr as well as some of the other of the 44 isolates which were Hgr or HgI. This is following our hypothesis of a single mobile unit carrying both antibiotic resistance genes and genes for Hg resistance.

We have found new classes of oral bacteria including lacotobacilli, anaerobes and opportunistic bacteria hybridizing with the Gram-positive Hg resistance probes as well as Gram-positive cocci. The oligonucleotide probes for the Gram-positive mercury genes have been developed in the laboratory and give positive reactions with all of the different Gram-positive enterococci and staphylococci controls. Finding other groups of bacteria which may carry the Gram-positive Hg genes is very exciting and totally new.

Until we start getting the 2 year samples, characterization of the new classes of oral bacteria that give positive hybridzation reactions to the Gram-positive Hg probes will be the focus in the laboratory.

One paper has been published based on the data collected to date: Luna, V.A., P. Coates, E.A. Eady, J. Cove, T.T.H. Nguyen, and M.C. Roberts. A variety of Gram-positive bacteria carry mobile mef genes. J. Antimicrob. Chemother. 44:19-25, 1999.

12 Appendix E. Results of the Dietary Mercury Study.

Results on the dietary survey portion of this sub-study were reported previously. Here we report results on blood methylmercury levels and mercury levels in fish meals from the Casa Pia schools.

Blood Mercury Results:

One hundred fifty children aged 8-10 were included in this sub-study at baseline, and 132 of them were available at one-year follow-up. Of the 18 children who had missing data at one-year follow-up, 7 dropped out of the main Casa Pia study, 1 moved away, and 10 refused to give blood for a second time.

Table E.1: Comparison of Baseline and One-year Follow-up Methylmercury Levels

	Baseline	**One-Year**	**p-value**
N	132	132	
Mean MeHg (ng/g)	2.7 +/- 1.5	2.4 +/- 1.3	.012

As indicated in Table E.1, there were significant differences between baseline and one-year follow-up methylmercury levels, with methylmercury values lower at one-year follow-up than at baseline.

Table E.2: Comparison of Groups on baseline and first-year follow-up methylmercury.

(ng/g)	**Group 1**	**Group 2**	**p-value**
Mean MeHg (baseline)	2.6 +/- 1.6	2.8 +/- 1.4	.300
Mean MeHg (year 1)	2.2 +/- 1.2	2.6 +/- 1.3	.088

As shown in Table E.2, there were no differences in the methylmercury values between the two groups at baseline. There was a small difference in methylmercury values between the two groups at one-year follow-up, although this was not quite statistically significant.

All of the methylmercury values are well below those of children in dietary studies done in the Faroe Islands (mean 12 ng/g) and the Seychelles (mean 26 ng/g).

Analyses of Mercury Levels in Fish:

Samples of fish meals served during a one week period were analyzed for total, inorganic and methylmercury (organic) from four different Casa Pia schools between 2/23/99 and 2/26/99. Five fish samples from each school were analyzed. Results shown in Table E.3[11] indicate low levels of mercury present, which is consistent with the blood mercury results from the sample of 150 children.

School	Mean Total Hg	Mean Inorganic Hg	Mean Organic Hg
1	25.6 ng/g	2.5 ng/g	23.6 ng/g
2	15.8 ng/g	1.4 ng/g	14.5 ng/g
3	15.2 ng/g	1.4 ng/g	14.6 ng/g
4	14.5 ng/g	1.1 ng/g	13.9 ng/g

[11] Recent studies of fish from non-polluted (Alaska) and relatively polluted (Glomma estuary, Norway) areas found levels of organic Hg ranging from 90-100 ng/g and 80-150 ng/g respectively. The current U.S. FDA regulatory level for seafood is 1,000 ng/g.

13 Appendix F. Progress on the Prepubertal Periodontitis Ancillary Study

From among the 647 children who consented and on whom a baseline oral exam was conducted, 311 having mixed dentition met eligibility criteria for the ancillary study of prepubertal periodontitis which Dr. Paula Marques proposed to conduct as part of her doctoral studies. Bitewing radiographs from the oral exam on the 311 children were digitized and examined for evidence of significant ($\geq$ 2 mm) bone loss. Sites in 22 children were identified which met this case definition. An equal number of children, matched for age, race, and gender, were identified as controls. Paperpoint samples from the case sites, from a non-affected control site within each case, and from the matched control were collected, stored, and shipped to Dr. Roberts' lab for DNA probe analyses to test for the presence/absence of an array of potential periodontal pathogens. Those analyses are now complete.

Data analyses from the 22 case-control comparisons of DNA probe results, as well as the investigation of demographic variables related to bone loss in general in the 311 eligible, are now underway. A poster is being prepared on the bone loss data for a pediatric dentistry conference, and a report on all aspects of the study will be prepared for the doctoral examination.

14 Sentinel Health Event Questionnaire

Annual Health Questionnaire (Sentinel Events)

Name:

Please answer the following questions relative to your child:
Circle the correct answer.

1. Has your child been to the hospital in the last year? yes no unsure

   If so, why?

2. In the last year has your child had any disease or received any treatment other than for the teeth? yes no unsure

   If so, what?

3. In the last year has your child received any medication, for example, antibiotics or injections? yes no unsure

   If so, what?

4. Is your child taking any medication now? yes no unsure

   If so, what?

5. In the last year has your child developed any allergies? yes no unsure

   If so, to what?

Thank you for your cooperation

Casa Pia Children's Amalgam Trial

Response to Action Items from DSMB Meeting September 14–15, 1999

Table of Contents

Introduction

1.0 Lost Teeth

2.0 Replacement of Fillings

3.0 Participants with Outlying Hg Levels or with Significant Hg Changes from Baseline

3.1 Clinical Summary of Data

4.0 Sentinel Event Monitoring (Periodic Health History)

5.0 Policy for Publications

6.0 Policy for Ancillary Studies

7.0 IRB Approval

Summary

Introduction

This document contains the information requested by the Data and Safety Monitoring Board (DSMB) for the Casa Pia Children's Amalgam Trial as a result of its meeting on September 14–15, 1999. The information herein reflects our responses to the "Action Steps" presented in the "Action Items/General Comments" provided to us with a memorandum from Dr. Maryann Redford dated October 15, 1999. There were two Action Steps listed which we understood the DSMB requested be incorporated into all future Annual Reports, but did not require revisions to this year's report:

1. Break behavioral data down by sex and age
2. Put total sample sizes on tables (better labeling)

We confirmed this to be the case in email exchanges with Dr. Redford. Each of the "Action Steps" other than these two have been addressed in this document, although not necessarily in the order presented in the memorandum. We appreciate the overall evaluation of the DSMB, and are pleased to provide this information.

1.0 Lost Teeth

The DSMB requested data on teeth lost by exfoliation or extraction. This information was inadvertently omitted from the 1999 DSMB report, and is here provided.

The following table gives the total number of teeth extracted during initial dental treatment and the number of teeth that were lost between the baseline and follow-up oral exams which were not recorded as extractions. The majority of teeth extracted were primary teeth (650/772). Note that Group A was slightly younger on average. The number of teeth exfoliated is within normal expectations for a population of this age distribution.

	Group A	Group B	Total
Number of teeth extracted	374	398	772
Number of non-extracted teeth lost	658	556	1214

2.0 Replacement of Fillings

The DSMB wanted us to review the replacement of fillings, and determine the reasons for these replacements. There were actually relatively very few teeth which required filling replacement (n=58), and virtually all of those were due to adjacent previously virgin surfaces needing restoration (n=47). This procedure almost always involves removing and replacing the original surface and incorporating it into a new, larger restoration. As mentioned in a prior report, we found an unexpectedly high amount of new decay at the first of the annual dental follow-up screenings. It was for this reason that an aggressive in-class oral hygiene program was implemented. We anticipate that the number of teeth with recurrent decay will decline over time as this prevention program takes effect. Those very

few fillings with a "same-surface-only" filling replacement (n=11) were replaced either for recurrent decay on the same surface, or because a filling failed. The relatively very small number of these occurrences (0.4%) indicates that the quality of clinical restorative dentistry provided is excellent. The numbers of teeth restored were similar in the two treatment groups (1473 versus 1450) as were the numbers of replacements (21 that added a surface and 3 that did not in one group, versus 26 and 8 in the other).

3.0 Participants with Outlying Hg Levels or with Significant Hg Changes from Baseline

The DSMB recommended that data on individuals with outlying mercury levels be examined on an individual basis to "discern if these participants are doing less well in comparison to others or less well in comparison to themselves (over time)". Items of particular interest are 1. Neurobehavioral assessments and 2. The size of restorations.

The DSMB also recommended that we "Examine more closely data on participants who have significant changes in mercury levels from baseline and check their neurobehavioral profiles."

We have combined the information on these two "groups" in this section for several reasons: 1. Because of overlap between some of these individuals, and 2. We wanted to identify individuals with large Hg changes in either creatinine-adjusted or unadjusted Hg levels. Eight children were identified with follow-up urinary mercury values of 10 ug/L or greater or a change from baseline of 5 ug/L or greater. Subsequently, an additional 10 children were identified who had similar large follow-up values or increases from baseline when the creatinine-adjusted mercury values were used in place of raw mercury concentration values. (Note that all children with a follow-up value of 10 or greater also had an increase from baseline of 5 or greater.)

Of the 18 children, 16 are in the amalgam group and 2 are in the composite group. The table below lists all 18 children, the raw and adjusted mercury values at baseline and follow-up, the three domain Z-scores at baseline and follow-up, and the number of restored surfaces broken down by size of restoration (small, medium or large). Highlighted figures indicate whether the raw or adjusted mercury values (or both) meet the criteria for a large follow-up value or large increase from baseline. Both measures are somewhat above average in every case. The only large decreases in Z-scores (greater than 1.25) for these children occurred for subject 19225 in the memory and visual motor domains.

The figures for filled surfaces include both composite and amalgam fillings. The 2 subjects from the composite group included in the table did not receive any amalgam fillings. For subjects in the amalgam group, approximately 93% of restored surfaces were restored with amalgam (2388/2577 for the entire group). The average number of restored surfaces for these 18 children was 12.8 which was slightly higher than the average of 10.0 per child in the entire study sample. These 18 children had a higher proportion of large fillings

compared with the entire sample (20% vs 15%), fewer medium fillings (37% vs 42%), and the same proportion of small fillings (43%).

Table 1

	ID*	Baseline					Follow-up					Filled Surfaces#		
		HG	HG/CRE	MEM	VIS	ATT	HG	HG/CRE	MEM	VIS	ATT	Small	Medium	Large
1	16005	1.6	2.0	0.38	0.24	0.41	13.2	18.4	–0.11	0.07	0.59	10	3	1
2	16545	1.2	1.2	–1.85	–1.42	–1.01	3.6	7.0	–1.88	–1.43	–1.81	10	12	9
3	16562	0.2	0.3	–0.14	–0.83	–0.77	5.3	NA	–0.11	–0.36	–0.78	10	11	3
4	18498	0.7	1.0	0.74	1.22	1.23	11.3	6.4	0.74	0.49	1.21	6	6	5
5	18532	0.5	1.1	0.60	0.72	1.38	10.1	10.2	1.04	0.80	1.51	4	2	1
6	19225	0.7	2.2	0.38	0.81	0.44	6.9	5.0	–0.93	–0.02	0.44	2	2	5
7	19226	1.7	2.9	–0.03	0.44	–0.39	5.4	8.7	–0.41	0.66	0.06	3	2	4
8	19233	1.8	1.8	–0.80	0.18	–0.75	8.5	5.5	0.35	0.25	–0.65	1	1	2
9	19372	0.9	1.0	–0.11	–0.37	0.23	3.0	6.0	0.46	0.15	0.49	6	5	0
10	19448	0.2	0.3	0.19	0.39	0.82	4.3	5.5	–0.57	–0.11	0.62	0	5	0
11	19581	0.7	0.6	–0.23	–0.12	–0.40	10.5	5.4	0.27	–0.23	0.04	4	6	0
12	19606	1.3	1.2	–0.65	–0.85	–1.55	2.8	6.6	–0.36	–1.03	–1.43	4	2	3
13	19618	0.6	0.7	–0.37	–0.50	–0.05	3.3	6.3	–0.18	–0.78	0.16	7	9	2
14	20300	1.1	0.9	–1.09	0.28	0.00	3.5	7.2	–1.37	–0.54	0.01	7	4	0
15	22378	0.8	1.4	0.15	–0.17	–0.67	7.9	4.8	–0.32	0.06	–1.14	3	3	0
16	22385	1.1	0.8	–0.40	0.41	0.92	5.5	6.0	–0.75	0.43	1.02	8	5	1
17	22386	1.1	0.6	0.21	0.72	0.61	5.6	10.4	0.18	0.66	0.78	5	5	8
18	22403	0.8	0.8	0.37	–0.16	0.15	4.7	6.4	1.07	–0.38	0.68	10	2	2

* Scrambled id number (not study id).
# Mercury and composite fillings.

3.1 Clinical Summary of Data

Dr. Brenda Townes, the study Psychologist and director of neurobehavioral studies, has reviewed all data for the individuals in Table 1, including the raw scores. This section presents her summary and conclusions.

Neurobehavioral change (defined as a shift in Z-score ≥ 0.10) revealed that no one domain was negatively impacted by Hg levels at follow up (see Table 2). Overall, there were slightly more improvements (22) than decreases (20).

Table 2

Domain	Improve	No Change	Decrease
Memory	7	4	7
Visual Motor	4	5	9
Attention	11	3	4

In individual patients, domain decreases were intermingled with increases or no change in other domains, providing no consistent pattern and only one patient with decreases in all three domains. Most significantly, review of all raw neurobehavioral test scores revealed no clinically significant changes in performance among these 18 patients.

4.0 Sentinel Event Monitoring (Periodic Health History)

The Action Step for this item refers to making revisions in the Sentinel Event Monitoring form which clarifies questions asked about allergies and medications. In responding to this recommendation, we have attempted to take the comments of the DSMB made at the September meeting into account without adding a long list of items which would be culturally inappropriate, and potentially produce more confusion than clarification. In consultation with Drs. Leitao and Caldas, Portugal site P.I. and Director of Neurobehavior/Neurology respectively, as well as Dr. Bernardo, Portugal site Lead Dentist, we have made slight revisions in several of the questions which we think will increase the probability of gathering the desired information more accurately and completely. The entire questionnaire is presented below. In particular, questions # 1,3 and 6 have been revised as follows:

#1. (Original)
Has your child gone to the hospital at any time in the last year?

#1. (Revised)
Has your child *been diagnosed with any serious illness during the last year?*

#2. (Unchanged)
In the past year has your child had any disease of, or treatment for their teeth?

#3. (Original)
In the past year has your child taken any medications, for example antibiotics or injections?

#3. (Revised)
In the past year has your child taken any medicines given to him/her by a doctor or nurse or clinic, for example antibiotics or injections?

#4. (Unchanged)
Is your child taking any medications now?

#5. (Original)
In the past year has your child developed any allergies?

#5. (Revised)
In the past year has *your child been diagnosed with an allergy to anything?*

We want to reiterate further that although the DSMB did not feel that our originally proposed, extensive Sentinel Event Monitoring system was necessary, and suggested instead the use of an annual questionnaire, we did additionally implement portions of this

proposed system as safeguards for obtaining information on sentinel events. In particular, we hired two part-time dental hygienists to implement an aggressive oral hygiene program on-site in the classrooms. Since these hygienists are regularly in class with study participants, they also check attendance records on every child enrolled in the study on a weekly basis, and follow up on the reasons for absences. In this way we collect good quality, accurate data on any potentially serious health problems in a timely fashion without relying solely on the Sentinel Event Questionnaire.

5.0 Policy for Publications

The DSMB recommended that we develop a policy/protocol for publishing. This follows:

The following policy will apply for anything to be published using any data collected on the children who have been randomized in the Casa Pia Study:

Investigator(s) interested in developing a publication will develop a short (1-2 page) proposal outlining the background, rationale, and overall plan for the paper. That proposal should be circulated to all Casa Pia investigators to identify any other interested collaborators and to make sure that no one has objections to/problems with the proposal. Next, in consultation with Brian Leroux, an outline of an analysis plan will then be added, and the proposal with analysis plan should be sent to Dr. DeRouen for his approval/signature. At that stage, he will send the approved proposal to NIDCR for their information/approval/potential interest in participating. It is only after their acknowledgment (lack of objection) that we should proceed to fully develop the publication.

6.0 Policy for Ancillary Studies

The DSMB recommended that we develop a policy/protocol for developing and implementing ancillary studies. This follows:

An Ancillary Study is defined to be any study that requires the collection of specimens or data, or the introduction of clinical or other methods not already part of the ongoing study protocol.

Investigator(s) interested in developing an ancillary study will develop a brief proposal outlining the proposed research. Sections should include: Background (brief); Purpose (goals); and Methods. The proposal should be submitted to the Primary Investigator (Dr. Timothy DeRouen) for initial review and circulation to the appropriate other study personnel for initial review.

Once this initial review has been completed, and any revisions have been made, Dr. DeRouen will send the brief proposal to NIDCR for their information/approval/potential interest in participating. Following their acknowledgment and response, the initiating investigator(s) will proceed to fully develop the proposal.

The completed proposal will then be reviewed by Dr. DeRouen and other study personnel he deems appropriate, and by NIDCR for final revision and implementation.

7.0 IRB Approval

The DSMB requested evidence of IRB approval for our amended study protocol, and a copy of the modified consent form. We wish to point out that we did appropriately seek IRB approval for the amended study protocol in advance from all involved agencies; the University of Washington, the University of Lisbon, and the Casa Pia School System. The amended protocol and above IRB approvals were also submitted to NIDCR at the time of the submission of Dr. Robert's ancillary study. We did not realize that it was necessary to submit these to the DSMB as well, but these are provided in this section.

Included following this section are:

1. **Modified Consent Form**
2. **University of Washington IRB Approval for Amended Protocol**
3. **University of Lisbon IRB Approval for Dr. Robert's Study**
4. **University of Lisbon IRB Approval for Modified Consent Form**

Summary

We hope the above information will answer the questions/concerns of the DSMB adequately. We appreciate the DSMB's interest in and contributions to the Casa Pia Children's Amalgam Trial.

FACULDADE DE MEDICINA DENTÁRIA UNIVERSIDADE DE LISBOA
UNIVERSITY OF WASHINGTON

CONSENT FORM

FOR PARENTS OR LEGAL GUARDIANS

Casa Pia Study of the Health Effects of Dental Amalgams in Children.

Principal investigator:

Timothy A. DeRouen, Ph.D., Professor and Chair, Dental Public Health Sciences, University of Washington, Telephone: 206 543-2034

Co-Investigators:

Michael D. Martin, DMD, M.P.H., Ph.D., Department of Oral Medicine, University of Washington, Telephone: 206 221-3785

Dr. Jorge Leitão, Coordinator in Portugal, Professor of Biomaterials, Faculdade de Medicina Dentária, Universidade de Lisboa, Telephone: 351 21 793-6949

Dr. Castro Caldas, Director Neurological Services and Professor at the Faculdade de Medicina, Universidade de Lisboa, Telephone: 351 21 794-4480.

INVESTIGATORS' STATEMENT
PURPOSE AND BENEFITS

The Faculdade de Medicina Dentária de Lisboa, in collaboration with the University of Washington (USA) is going to make a dental study in which several tests will be performed. The study will be performed among children who attend the Casa Pia Schools.

These tests will help us on deciding if your child needs dental treatment. If your child needs dental care, the child will be invited to participate in a study of the effects of normal, routine tooth filling materials on humans. There are two main types of tooth filling materials commonly in use today. One is a silver amalgam, and the other is a plastic composite material. Even though these filling materials are the standard ones used in hundreds of millions of fillings each year throughout the world, questions remain about the safety of the materials. We want to see if there is any ill effect of these materials on the health of children who have these materials placed in teeth for fillings.

To do this we will place one type **or** the other of filling material in the teeth that need fillings of each child who participates in the study. Both kinds of materials are normally used in the filling of teeth. We will randomly (by chance, like flipping a coin) assign each child to receive one or the other of the filling materials for all of their teeth needing fillings. We will therefore have two

groups of children who differ in the type of fillings they have in their teeth. Then we will measure various kidney, muscle, nerve and brain functions of the children to see if there is any difference between the two groups.

We also want to see if the placement of the filling materials changes the normal bacteria in the mouth and stomach.

We also want to study what kinds of bacteria in the mouth cause gum disease in children.

We also want to study the diet of the children in the study to see how much mercury comes from the diet.

All tests and treatments will be performed at Faculdade de Medicina Dentária, Universidade de Lisboa, with no costs to you. All examinations and care will be provided by qualified professionals including dentists, dental hygienists, doctors and nurses.

This study has been approved by the Administration of the Casa Pia School system.

PROCEDURES

The study is anticipated to last for as long as five years. Each child will participate in the study according to the following general schedule:

Year One:

First dental and medical examinations (described below).

Oral rehabilitation (cleaning and fixing of all teeth).

One year dental and medical follow-up examinations; necessary additional dental treatment. (Described below)

Years 2-5

Annual dental and medical follow-up examinations; necessary additional dental treatment. (same as One-year procedures)

Year 6

If there is found to be a negative health effect from either of these filling materials, you will be informed.

First Dental Examination:

- First we will examine the child's teeth, gums and other parts of the mouth.
- Then we will take a series of x-rays (8 to 12) to look for decay between the teeth and to see if the adult teeth are forming normally beneath the gums. These x-rays are routine and necessary.
- We will collect small plaque samples to test for bacteria by holding small bits of paper between the tooth and gum for about 10 seconds.
- The first dental examination should take about an hour.

Initial Dental Treatment will consist of the following:

- Cleaning by a dental hygienist, and home care and cleaning of the teeth and mouth instructions.
- Repair (filling) of all decayed teeth which can be repaired. Teeth with decay which is not too large will be repaired with a filling material.
- Removal of teeth which are too decayed to be repaired. If the decay is too large, we cannot fix the tooth, and it would be better to remove the tooth.
- Protective sealant placed on teeth which do not have decay to keep decay out. We will place a clear plastic coating on the back teeth which do not already have decay. This coating should last for at least a year, and will help prevent decay on the chewing surface of the tooth. If the coating falls off, we can replace it.
- The amount of time the dental treatment will take is entirely dependent upon the condition of the child's teeth, but will take several appointments.

First Medical Examination and Testing will consist of the following:

- We will examine your child's medical records to see if there are any problems which may prevent him or her from being in the study or for receiving dental treatment. These may include:

1. mental retardation
2. autism/developmental delay
3. severe motor impairment
4. uncontrolled seizure disorder (controlled with medication defined as no seizure activity within last two years)
5. birth defect with CNS feature (e.g., neurofibromatosis, fragile X syndrome, Down syndrome)
6. inborn error of metabolism (e.g., phenylketonuria, Wilson's disease)
7. tic disorder (e.g., Tourette syndrome)
8. attention deficit disorder (inattentive or hyperactive subtype)
9. neoplastic disease (e.g., leukemia, brain tumor)
10. traumatic brain injury
11. CNS infection resulting in cognitive impairment
12. routine use of psychotropic medication (e.g., antidepressant, antipsychotic, anxiolytic)

- We will ask the child for a small urine sample in a cup.
- After that we will collect one small tube of blood (about 20 milliliters) from an arm vein with a needle and syringe.
- (A nurse will collect the urine and blood samples)
- We will use a soft swab to pass over the gum line and then the surface of some teeth. This will be done by the dental hygienist. This will be repeated at the last treatment visit.
- Bacteria obtained from urine and/or oral swab may be used on other research on the biology of the bacteria.
- Blood obtained will be examined to see how much mercury is in it.
- We will ask you to answer some questions about what kinds of foods your child eats.
- Finally, we will test how well the child's muscles and nerves work with a nerve conduction velocity machine. This is done by placing small foam pads on the skin of the arm or leg, and letting a very small amount of electrical current pass through. We measure how fast the current passes between the pads. It does not hurt, and produces a slight tingling feeling.
- The child will then answer some tests that will evaluate his or her brain function by testing memory, attention span and intelligence. These tests are like games, and there is no pressure for the child to perform well. The questions include vocabulary tests, memorization of word lists, looking at symbols to find ones which are alike, the matching of colors with words, number memorization, and recognition of patterns.
- The child will also take a finger-tapping test to see how fast they can tap a button for a ten second period. The child will also take a test to see how fast they can push a button when a light comes on.
- These tests take about 90 minutes to be completed. The child can refuse to participate in any or all parts of these tests, or may refuse to answer any of the questions. There will be no penalty.

Yearly Follow-up Dental Examinations will consist of the following:

- An examination by a dentist. The dentist will look at all of the teeth and other parts of the mouth to see if there is any new decay, or other problems needing attention.
- A cleaning by a dental hygienist.
- Any necessary treatment such as a filling for new decay or other problems which have occurred.
- Additional oral swab samples will be collected at the first and second follow-up appointment.
- Every two years we will take a minimal number of dental x-rays (two to four) to look for new decay in the teeth.

RISKS, STRESS OR DISCOMFORT

None of these examinations, treatments or tests are experimental. They are all procedures which are commonly in use throughout the world. The only risks or discomfort from these procedures are the ones usually associated with routine dental care or with routine medical examinations and tests.

When the dental examination, cleaning and treatment are performed, there may be slight discomfort from time to time, and your child will feel pressure in his/her mouth. Local anesthesia to numb the teeth and gums (lidocaine, carbocaine) will always be used whenever indicated to keep any discomfort from occurring. This anesthesia (numbing) is given by first applying a cream (topical anesthetic) which numbs a gum, and then injecting a small amount of anesthetic liquid into the gum near the teeth. Although many of these injections are painless, there may be an occasional small stinging feeling. There is a very small chance of an allergic reaction to the anesthetics used for dental care. The reaction to the topical anesthetic would probably show up as a reddened area at the site of application. This should pass within a day.

To get the blood sample, a nurse will use a needle and syringe to draw blood from a vein in your child's arm. The procedure produces a small bit of pain which is gone very quickly. Pressure must be applied to the area of the puncture after the blood draw to avoid letting blood leak out and causing a hematoma or bruise.

The machines which test nerve function cause a small amount of tingling in the leg or arm (depending on which is used for the test). Small foam pads are placed on the limb. A small amount of safe electric current is passed from one pad to the next. The machine measures how fast the current moves through the nerves under the skin.

In order to answer the questions from some of the tests, it will be necessary for the child to sit at a table for up to 90 minutes all together. Each child will also be asked to write a few lines. We will not pressure the child to perform well, and the child will be encouraged to relax and have fun with the game-like tests. It may be tiring. The child can refuse to answer any question.

The amount of x-ray exposure will be very small, and we will adhere strictly to the guidelines for dental x-ray exposure from the Academy of Pediatric Dentistry (U.S.) The amount of radiation the child will receive the first year is approximately equal to that from 3 to 6 days of background radiation from natural sources. The amount of radiation the child will receive during the every-two-year check-ups is approximately equivalent to that from one to three days of background radiation from natural sources.

The child may refuse to participate in any or all of the study procedures, and may choose to leave the study at any time. There will be no penalty.

Other Information

All information collected for each child will be kept strictly confidential. Although we will use the results for scientific reports, no child's name will appear in any written or oral report. Only members of the study team will have access to the data from the study. This includes the study members from the Faculdade De Medicina Dentária Universidade De Lisboa and the University of Washington School of Dentistry and the National Institutes of Health (U.S.) The data will be kept under secure conditions by the investigators. No child will ever be identified by name in any written publication.

If your child is injured as a direct result of the dental treatment given from this study, we will treat the child at no cost to you.

Neither you or the child will be paid for participation in the study, but the child will receive dental care and treatment for the duration of the study at no cost.

You may ask any questions that you wish at any time. You can change your mind about giving permission for the child to be in the study at any time. We will not be upset, and will respect your wishes.

Signature of Investigator _____ Date _____

PARTICIPANT'S STATEMENT

The study described above has been explained to me, and I wish for my child or ward to participate in this activity. I have had an opportunity to ask any questions that I wanted. I understand that I may ask any questions about the study at any time. If I change my mind about having my child or ward participate in the study, I may withdraw my permission at any time. All I have to do is inform any of the persons related to the study.

Signature of Parent or Legal Guardian _____ Date _____

FACULDADE DE MEDICINA DENTÁRIA
UNIVERSIDADE DE LISBOA

MINOR ASSENT FORM

Casa Pia Study of the Health Effects of Dental Amalgams in Children

PURPOSE AND BENEFITS

You may be one of the many children who have decayed or broken teeth. Decay is a hole in teeth that can be filled with some type of materials. There are two types of filling materials. One is named amalgam and the other one is named composite.

We want to make a study to find out if these materials have any other effect on people. We want to know if there is mercury or lead in your body in normal quantities. We also want to learn how well you will do in some examinations, and how well your muscles and nerves work.

If you agree to be in this study you will have your teeth cleaned by a dental professional, all of your decayed or broken teeth will be fixed, and you will have a dental check-up once a year.

PROCEDURES

Here is what we are going to do:

- First we will look at your teeth and take some x-ray pictures of your teeth.
- We will pass a soft swab over places on your gums and teeth to wipe off some bacteria which are normally there.
- Then we will analyze a small sample of urine.
- After that we going to ask you if we can use a needle to collect some blood from your arm.
- Finally we will test how well your muscles and nerves work with a machine. We will place some small sponge-like pads on your arms and legs, and a small bit of electricity will run through these pads to test your nerves. It may make a tingling feeling which is a little uncomfortable.
- You will answer some tests that will evaluate and compare your memory, attention span and intelligence. These tests are like games.

A nurse will collect the urine and blood samples. The test takes about 90 minutes to be completed.

RISKS, STRESS OR DISCOMFORT

When we look at your teeth we are going to use a mirror and a light without causing you pain. If you want to ask us any question or if there is anything bothering you just tell us.

To get the urine sample, we ask you to urinate into a glass and give it to the nurse. To collect blood, the nurse will stick a needle into a vein in your arm. It is going to be real quick and you will feel a slight pain.

To test how well your nerves work, we will use a machine that will make a brief tingling feeling in your leg.

In order to answer the tests, you will be sitting at a table for 90 minutes and answer to several questions; you also have to write a few lines.

Other Information

We are going to make a report of what we have learned in this study; your name will not show on it. This study is anonymous.

You may ask all the questions you want, when and anytime you feel like it. You can change your mind about being in the study. You can give up; just say it to us. Do you have any questions now? Do you understand what we are doing?

_____ _____

Signature of Investigator Date

PARTICIPANT'S STATEMENT

The study described above has been explained to me, and I agree to participate in this activity. I have had an opportunity to ask questions. If I change my mind on participating in this study I understand that I can give up. All I have to do is to inform any of the persons related to this study.

_____ _____

Signature of Participant Date

Human Subjects Division

JUL 29 1997

UW

University of Washington Correspondence

INTERDEPARTMENTAL

School of Dentistry ♦ Department of Dental Public Health Sciences ♦ B509 Health Sciences Center Box 357475 ♦ Telephone: 3-2034

Master Copy ☐
Committee Copy ☐
Reviewer Copy ☐
Investigator Copy ☒

July 29, 1997

TO: **Carol Wilkinson**
Human Subjects Division
Box 354080

FROM: Timothy A. DeRouen, Ph.D.
Professsor and Chair

RE: Approved Human Subjects Application 26-00240B/C

Attached is a revised consent form for the above referenced project. At the request of NIH we are adding a statement notifying the subject that the specimen we are collecting may be used for other studies. The added sentence is in italics.

Date 8-4-97

Modification to 26-0024-B

Approved by Human Subjects Review Committee *

Signature

* Contingent on use of consent form, attached, as modified (page 3).

UNIVERSIDADE DE LISBOA

FACULDADE DE MEDICINA DENTÁRIA

Dr. Timothy DeRouen
Principal Investigator
The Casa Pia Study of the Health Effects
of Dental Amalgam in Children
University of Washington
Box 357475
Seattle, WA 98195-7475

Lisbon, 19 9 / 1997

Dear Dr. DeRouen:

The Ethics Committee of the Faculdade de Medicina Dentária - Universidade de Lisboa has reviewed the proposal entitled "Influence of Dental Amalgams on Mercury and Resistant Bacteria" , and approves this project.

Sincerely,

Vice Chairman of the Ethics Committee

Prof. Dr. Boleo Tomé

UNIVERSIDADE DE LISBOA

FACULDADE DE MEDICINA DENTÁRIA

Dr. Timothy DeRouen
Principal Investigator
The Casa Pia Study of the Health Effects
of Dental Amalgam in Children
University of Washington
Box 357475
Seattle, WA 98195-7475

Dear Dr. DeRouen:

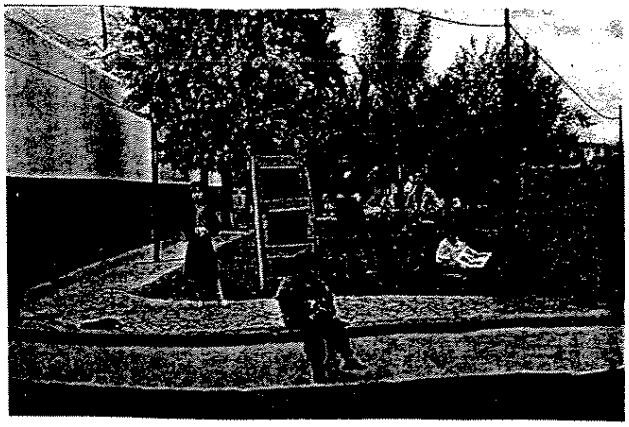
At your request, the Ethics Committee of the Lisbon Dental School examined the revised consent forms submitted for appreciation. The proposed changes received full approval.from this Committee.

Sincerely,

Vice Chairman of the Ethics Committee

Prof. Dr. Boleo Tomé

Lx, 97/9/10



RECEIVED
Human Subjects Division

OCT 3 1 2003

UW

ONLY FILE COPY
Keep with Committee Copy
Send to Reviewer at Renewal
To Records when file is closed
App.# 95-0401-A/C 08

Children's Amalgam Trial
Casa Pia Study of the Health Effects of Amalgam in Children

Report to the Data and Safety Monitoring Board

September 25, 2002

Table of Contents

Section	Page
1. Introduction	3
1.1 Summary of progress reported previously	3
1.2 Site visits	3
1.3 Personnel changes	3
1.4 Protocol changes	3
1.4.1 Informed consent for extension of follow-up	3
1.5 Adjustments to study procedures	4
1.5.1 Methods for identifying outliers	4
1.5.2 Alternative measure of mercury exposure	4
1.5.3 Creatinine adjustment of urinary concentrations	4
1.5.4 Monitoring of conduct disorders	4
1.5.5 Utilization of the third neurobehavioral examiner for conducting tests	5
1.6 Recent publications	5
1.7 Future plans	5
1.7.1 Plans for testing and implementation of computerized tests	5
1.8 Progress on the ancillary study on antibiotic resistance	6
2. Follow-up Visits	7
2.1 Subject retention	7
2.2 Timing of follow-up visits	7
3. Dental Treatment and Mercury Exposure	10
3.1 Dental treatment	10
3.2 Mercury exposure	11
4. Follow-up Results on Primary Outcome Variables	13
4.1 Correlations between child and adult tests	13
4.2 Distribution of primary outcome variables by follow-up period	14
4.3 Monitoring changes in performance of individual children	19
4.4 Statistical stopping rule	21
4.5 Reassessment of power	21
5. Follow-up Results on Secondary Outcome Variables	22
5.1 Sentinel health events and health history	22
5.2 Neurological	23
5.3 Oral exam	24
5.4 Laboratory measures	26
5.4.1 Urinary samples	26
5.4.2 Monitoring children with high urinary mercury values	26
5.4.3 Monitoring children with elevated blood lead values	26
6. Quality Assurance Monitoring	29
6.1 Completeness of data	29
6.2 Accuracy of data	29
6.3 Examiner reliability and accuracy	30
6.3.1 Checklist evaluation	30
6.3.2 Inter-examiner reliability	30
6.3.3 Differences between examiners	31
7. Summary	33
Appendix A. Revised Informed Consent Materials	
Appendix B. Issues in Monitoring Conduct Disorders	
Appendix C. Progress in Pilot Testing of Computerized Tests	
Appendix D.1. Data Tables for Children with High Mercury Exposure	

Appendix D.2. Data Tables for Children with Sentinel Health Events
Appendix D.3. Data Tables for Children with Elevated Urinary Mercury Levels
Appendix D.4. Data Tables for Children with Elevated Baseline Blood Lead Levels
Appendix E. Norms for the Neurobevioral Test Scores
Appendix F. Abstracts and Papers from the Main Study
Appendix G. Report on Children with Clinical Concerns (to be submitted separately)

1. Introduction

The Casa Pia Study of the Health Effects of Amalgam in Children, an arm of the "Children's Amalgam Trial", has completed the fourth year of follow-up evaluation and data collection and is currently in the fifth year of follow-up. The Study is proceeding very well, with all goals being met in a timely fashion, and a subject retention rate that continues to be high.

In this report, we summarize the progress made on the trial since the last DSMB meeting in December, 2001. We begin by summarizing some of the key results presented in our previous reports. Following that, we describe our visits to the study site in the past year, personnel changes, protocol changes, changes in study procedures, recent publications, future plans, and a brief report of progress on the ancillary study on antibiotic resistance. In the subsequent sections we report results of the trial. Results are reported in a cumulative fashion by including data from baseline and all follow-ups to avoid as much as possible the need to refer to our previous reports. Data included in this report are current as of September 6, 2002.

1.1. Summary of Progress Reported Previously

A total of 507 children were randomized to one of the two treatment groups (group sample sizes 253 and 254). The treatment groups were similar at baseline in demographics, health history, and the distributions of the baseline values of primary and secondary outcome variables, oral exam results, blood lead levels, urinary mercury levels, and other urinary markers. (Group comparisons on baseline variables are presented along with follow-up data throughout this report.) Three protocol errors occurred in which subjects that should have been declared ineligible were randomized: 1) one subject with abnormal neurological findings (cerebral palsy) was randomized on 4/25/97; 2) one subject was randomized on 12/23/97 with a blood lead value of 14 ug/dl that was later adjusted by the laboratory to 16ug/dl (above our cut-off of 15 ug/dl); and 3) one subject with existing amalgam was randomized on 1/7/98.

1.2. Site Visits

We continue to have key personnel from the University of Washington make regular, periodic site visits to Lisbon for oversight and other specific purposes. During the past year, site visits were made by T. DeRouen, Principal Investigator; M. Martin, Project Director; B. Townes, Neuropsychologist; J. Woods, Toxicologist; and G. Rosenbaum, Senior Psychometrist (2 visits).

1.3. Personnel Changes

A physician has been hired at the Lisbon site to assist with monitoring and facilitating recommendations based on neurobehavioral, neurological or laboratory findings. Dra. Rita Silva is a pediatric neurologist.

1.4. Protocol Changes

When a member of the study team determines that a change in protocol may be beneficial or necessary, a proposal is prepared for the change, which includes rationale and details of the proposed change. This proposal is reviewed at both the Lisbon and Seattle sites, and following a final determination of the details and start date, the P.I. (Dr. DeRouen) in Seattle, and the Co-P.I. and director of the Lisbon site, Dr. Leitao "sign off" on the newly established protocol. Protocol changes implemented in the past year are described in the following sub-sections.

1.4.1. Informed Consent for Extension of Follow-Up

Based on considerations detailed in our April 2002 letter to the Board, we plan to obtain active informed consent from each parent/guardian for continuance of their child in the study through seven years of follow-up. Each parent/guardian will receive a letter explaining the continuance of the project. With the letter they will receive an opportunity to either have their child continue in the study beyond five years or to withdraw their child after five

years of follow-up by returning a signed form to the study coordinator. The re-consent packet sent to each parent/guardian will include a photocopy of their original, signed consent form (to refresh their memory regarding the details of the study), a cover letter, a brief report of dental care delivered to date, and a response form along with a postpaid envelope. Copies of these are included in Appendix A. We have received approval from both the Ethics Committee at the University of Lisbon and the IRB at the University of Washington for this procedure. The letter of approval from the Ethics Committee in Lisbon and the approved human subjects application for the IRB at the University of Washington were appended to the April 2002 letter to the Board and are not included here.

1.5. Adjustments to Study Procedures

In addition to changes to study protocol, minor adjustments to study procedures or methods of analyzing or reporting results are made. These changes do not require a formal approval process but instead are implemented as decisions made by the investigators, often in response to suggestions or requests from the DSMB. Adjustments made during the past year are described below.

1.5.1. Methods for identifying outliers

In our April 2002 letter to the Board we outlined revised procedures for identifying outliers in neurobehavioral test scores, which consist of the Reliable Change Index (RCI) method that we had already been using and a method used by the NERI group based on changes in composite scores. For the Casa Pia trial, we will identify children with a decrease of one unit or more in any one of the domain Z-scores for memory, visual-motor, and attention domains, which are three of the four primary outcome variables for the trial. Results are given in Section 4.3 of this report.

1.5.2. Alternative measure of mercury exposure

As described in our April 2002 letter to the Board, we have developed a measure of mercury exposure based on the amount of time children are exposed to amalgam fillings and also on the size of the fillings and the type of tooth surface in which they are placed. Exposure is measured in units of surface-years, according to which 1 surface-year would represent exposure for one year to one surface restored with a medium-size amalgam filling. Data on this measure will be reported in two ways: 1) each year we will identify children with outlying values of this exposure measure and report data on the neurobehavioral performance and other outcomes for these children in comparison with other children; and 2) we will list values of the mercury exposure measure along with other information that we report on children identified as having poor performance on neurobehavioral tests. An explanation of the method of calculation of mercury exposure and the results are given in Section 3.2 of this report.

1.5.3. Creatinine adjustment of urinary concentrations

In response to a request made by the DSMB during the December 2001 meeting, results of laboratory analyses of urine samples are displayed with adjustment for creatinine concentration. Similar plots based on unadjusted concentrations have also been produced but are not shown here.

1.5.4. Monitoring of conduct disorders

One of the recommendations made by the Board after the December 2001 meeting concerned development of a plan to uncover and monitor conduct disorders. This recommendation was a surprise to us, since it was not conveyed to us in the verbal debriefing at the end of the December meeting. There was discussion during our presentation about some of the children we identified as having clinical declines purportedly exhibiting some behaviors that may be indicative of conduct disorder. In the discussion, Dr. Townes made it clear that these were second-hand reports by the patient care coordinators or teachers, and that reliable first-hand information required to make a true clinical diagnosis of conduct disorder was not available. While we understand the DSMB interest in this issue, we do not think it feasible to try to implement a surveillance system to detect conduct disorders among all 507 children in the study; such a system would require extensive interviews with parents and teachers

for each child regarding specific behaviors and sensitive issues for which we have neither the resources nor the consent/authority (issues involved in assessment of conduct disorders are described in Appendix B). What we propose to do is continue to acquire whatever information may be available about children who fall into the pool of "outliers" for whom we review cumulative records, perform neurological exams, and query patient care coordinators and teachers. This information may contain second-hand reports of behaviors that may be suggestive of conduct disorders, and we will continue to record that information when provided.

1.5.5. Utilization of the Third Neurobehavioral Examiner for Conducting Tests

During the past year, our third examiner, Susana Margalhos Rodrigues, has begun performing annual neurobehavioral tests of some children. Ms. Rodrigues had been added to the study personnel prior to last year to help with scoring of neurobehavioral tests and also to be available to perform testing if the other two examiners were unavailable or to allow scheduling of three children to be tested at the same time. Ms. Rodrigues was fully trained by Ms. Rosenbaum and has provided tapes of testing sessions for evaluation, and has also participated in reliability studies.

1.6. Recent Publications

Copies of all abstracts and papers published or in press or accepted for publication in the past year are given in Appendix E.

Publications:

Evens CC, Martin MD, Woods JS, Soares HL, Bernardo M, Leitao J, Simmonds PL, Liang L, DeRouen T. Examination of dietary methylmercury exposure in the Casa Pia study of the Health Effects of Dental Amalgams in Children. J Tox Environ Health 64 (7): 521-530, 2001.

DeRouen TA, Leroux BG, Martin MD, Townes BD, Woods JS, Leitao J, Castro-Caldas A, Braveman N. Issues in Design and Analysis of a Randomized Clinical Trial to Assess the Safety of Dental Amalgam Restorations in Children. Controlled Clinical Trials 23:301-320, 2002.

Abstracts:

Martin MD, Leroux BG, Benton T, Leitao J, Bernardo M, Soares Luis H, Braveman N, Woods JS, Simmonds L, DeRouen TA. Longitudinal assessments of urinary mercury levels in the Casa Pia study. J Dent Res 81:3445, 2002.

Martin MD, Leroux BG, Woods JS, Benton T, DeRouen TA, Bernardo MF, Leitao J, Braveman N. The effect of creatinine adjustment on urinary mercury concentration in children during a longitudinal study. Submitted AADR 2002.

1.7. Future Plans

1.7.1. Plans for testing and implementation of computerized tests

In the coming year we plan to conduct pilot testing in Lisbon of the computerized tests in order to clarify test instructions and procedures and to determine whether the procedures are appropriate for Casa Pia children. To date, we have pilot tested the battery with children 13-17 years of age in Seattle. The clarity of the instructions and procedures was determined and adjustments to the protocol were made as needed. Preliminary work on the first phase is described more fully in Appendix C. In the second phase, participants will be children in the Casa Pia schools who were ineligible to participate in the trial. Analyses of performance among participants in Lisbon will determine the appropriateness of the instructions in Portuguese for Casa Pia children. It should be emphasized that these measures are not intended to replace the primary outcome variables of the trial. The computerized test scores will be administered after completion of the regular test battery, and in the analysis they

will be treated as secondary outcome variables. Further details on the computerized tests and plans for pilot testing were provided in our April 2002 letter to the Board.

1.8. Progress on the Ancillary Study on Antibiotic Resistance

The ancillary study on antibiotic resistance (Marilyn C. Roberts, PI) has continued to make good progress. Results through the third-year follow-up showed no differences between treatment groups in antibiotic-resistant or mercury-resistant bacteria. The proportion of children successfully followed continues to be high through the fourth year of follow-up (Table 1.8.1). Statistical analyses will be conducted after fourth-year data are complete. Work has continued on characterization of individual isolates collected over the first 5 visits and recording of the information in computer databases. PCR evidence has been obtained to suggest that the gram-positive mef(A) gene coding efflux of macrolides may be linked to a gram-positive mercury reductase gene which confers mercury resistance in gram-positive bacteria. This is a novel finding and resulted in two manuscripts accepted for publication. Additional data will be gathered including DNA sequencing of PCR products to verify this finding. Additional work has begun on macrolide resistance in gram-negative oral and urine isolates and genotyping using gene probes for rRNA methylases, efflux, esterases, and phosphorylases genes. This work has been accepted as a poster presentation at ICAAC 2002. One additional publication examined the persistence of candida albicans in the study participants.

Table 1.8.1. The number of children successfully followed for the antibiotic resistance study.

	Baseline	Post-treatment	Year 1	Year 2	Year 3	Year 4 (preliminary as of 8/31/02)
Number of Children (%)	150 (100%)	145 (96.7%)	136 (90.7%)	140 (93.3%)	132 (88.0%)	123 (82.0%)

Recent Publications on the Study of Antibiotic Resistance

V.A. Luna, K. Judge, H. Luis, M. Bernardo, J. Leitao, and M.C. Roberts. Mercury reductase merA gene in oral and urogenital Gram-positive bacteria: Development of oligonucleotide probes. Submitted.

V.A. Luna, M. Heiken, K. Judge, C. Uelp, N. Van Kirk, H. Luis, M. Bernardo, J. Leitao, and M.C. Roberts. Distribution of mef(A) gene in gram-positive bacteria from healthy Portuguese children. Submitted.

M.C. Roberts, K. Judge, c. Ulep, H. Luis, M. Bernardo and J. Leitao. Acquired macrolide resistance, mef(A), and erm(B) genes, in gram-negative bacteria isolated from healthy Portuguese children. To be presented at 42nd Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC), Sept 2002.

J.R. Starr, T. C. White, B.G. Leroux, H.S. Luis, M. Bernardo, J. Leitao, and M.C. Roberts. Persistence of oral *Candida albicans* carriage among healthy Portuguese schoolchildren followed for three years. *Oral Microbiol Immunol* in press, 2002.

2. Follow-Up Visits

2.1. Subject retention

We continue to achieve a high rate of subject retention. A total of 32 children have been lost to follow-up (18 in the amalgam group and 14 in the composite group). The reasons for loss to follow-up include the following: child removed from the study by written request from parents/guardians (10), child refused to cooperate (1), child either moved away from Lisbon or left the Casa Pia school system and cannot be contacted (20), and child died (1). The number of children who completed neurobehavioral (NB) testing continues to be high (440, 87% of all enrolled children) and approximately equal in the two treatment groups (percentages by treatment group are 85.8% and 87.8%). A large number of these children (420, 83% of 507) have actually completed all 4 follow-up tests. There are 487 children (96%) who have completed at least one follow-up NB test. Scheduling of NB tests is given highest priority because these comprise 3 of the 4 primary outcomes. Data are slightly less complete for nerve conduction velocity (NCV) tests (80%) and neurological exams (77%) because of difficulties with arranging these tests for children outside the Casa Pia school system when they are brought to the study site for NB testing. During the past year, alternate neurologists and NCV technicians were identified to cover for the study neurologist and the regular NCV technicians when they are unavailable. These alternate personnel have been available only for the latter period of year 4 follow-up and hence have had little impact on follow-up to date; the number of children seen for NCV or neurological testing at year 4 is either the same or only slightly higher than the number at year 3 (Table 2.1.1). Larger increases are expected for year 5. Oral exams were conducted on 84% and urinary mercury analyses completed for 86% of the children (Table 2.1.1). Health history questionnaires were obtained for 88% of the children in year 3; the fourth-year survey is on-going (an update on these results will be available at the meeting).

Table 2.1.1. Number of subjects with successful data collection.

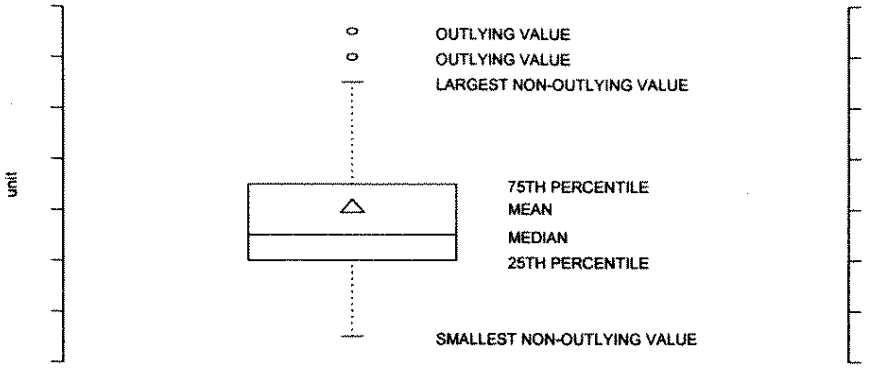
Data Type	Number with Data At Follow-up Years:				
	1	2	3	4	5
Neurobehavioral	478	468	455	440	197
Nerve Conduction Velocity	457	446	407	407	179
Neurological	467	453	383	389	158
Oral Exam	452	468	448	428	181
Urinary Mercury	473	453	438	438	183
Health History	479	454	447	NA	NA

2.2. Timing of follow-up visits

Annual follow-up tests are scheduled at regular one-year intervals (plus or minus two weeks) starting from the anniversary date of the baseline exam. The intervals between tests have adhered fairly closely to this target in most cases; the variation is small relative to the overall time scale of the study (Figure 2.2.1). (See key for interpretation of boxplots below.) Most importantly, the distribution of time between visits is very similar in the two treatment groups (see Figure 2.2.1; compare boxplots for Amalgam and Composite groups for each follow-up year). When scheduling visits, priority is given to the anniversary date for neurobehavioral (NB) testing because this provides data for three of the primary outcome variables and could potentially be influenced by the timing of the visit. As a result, the variation in timing is lowest for the NB visits. Other tests may also be performed at the NB visit even if the date falls outside of the two-week target window for the other tests. Late visits occur for a variety of reasons such as absences from school or children being temporarily away from Lisbon. Arranging follow-up visits is unusually difficult for some children, including some of those who have moved outside of the

Casa Pia school system; testing for a few of these children has on occasion been performed whenever the opportunity existed even if it was far outside of the target time interval.

Key for Interpretation of Boxplots



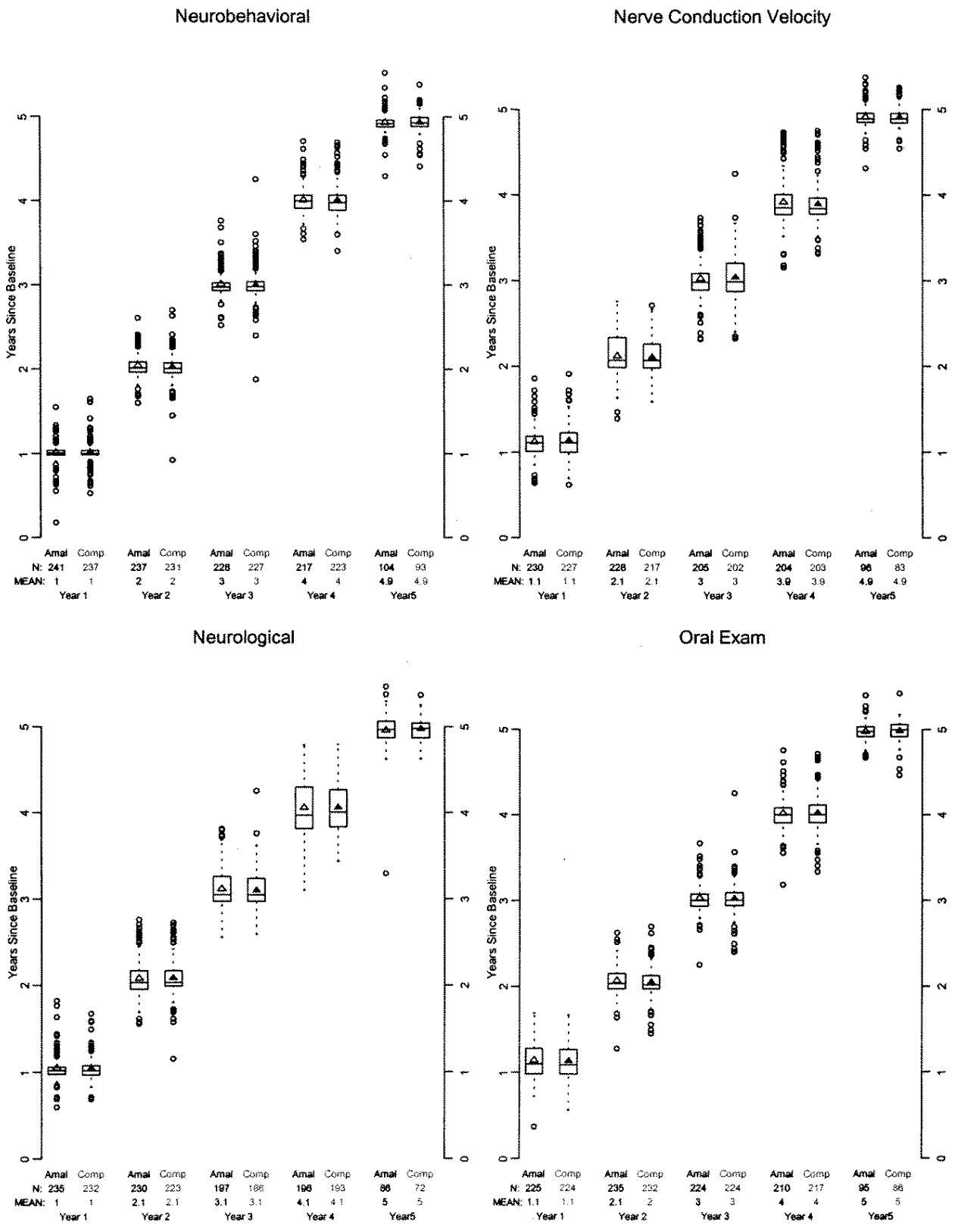


Figure 2.2.1. Time from baseline to follow-up visits in years by type of visit, study year, and treatment group.

3. Dental Treatment and Mercury Exposure

3.1. Dental treatment

During the initial treatment period, approximately 10 surfaces were restored per child on average in both treatment groups. In each annual follow-up period, approximately one-quarter of the children have required additional treatment, with an overall average of approximately 1 restored surface per enrolled child per year (Table 3.1). Children in the amalgam group have required slightly less treatment than children in the composite group during the 2nd and 3rd follow-up periods in accordance with results on caries incidence (Section 5.3); preliminary data for 4th year follow-up do not show the same trend. One child in the Composite group received amalgam restorations on 10 surfaces initially and 3 additional surfaces in year 2 (the 10 initial surfaces were previously reported); this problem has been corrected so that the child will not receive any additional amalgam fillings if treatment is needed in future. A total of 766 teeth were extracted during the initial treatment and an additional 97 teeth have been extracted in subsequent treatment periods; the numbers are similar in the two treatment groups (Table 3.1). A small number (298) of restorations have had to be repaired or replaced to date (132 in the amalgam group, 166 in the composite group).

Table 3.1. Information on dental treatment by treatment group and treatment period. Data for 4th-year follow-up are preliminary because 4th-year treatment is incomplete.

	Amalgam			Composite		
	No.	Mean	SD	No.	Mean	SD
Initial Treatment Period:						
Children treated	249			248		
Treatment visits	1154	4.6	2.0	1237	4.9	2.4
Restored surfaces	2553	10.1	5.5	2484	9.8	6.2
Surfaces restored with amalgam	2359	9.3	5.0	10	0.0	0.6
Surfaces restored with other materials	194	0.8	1.8	2474	9.7	6.2
Teeth extracted	373	1.5	1.9	393	1.5	2.0
1st-Year Follow-up Treatment Period:						
Children treated	62			69		
Treatment visits	112	0.5	1.2	117	0.5	1.0
Restored surfaces	207	0.9	2.9	211	0.9	2.2
Surfaces restored with amalgam	172	0.8	2.5	0	0.0	0.0
Surfaces restored with other materials	35	0.2	0.9	211	0.9	2.2
Teeth extracted	24	0.1	0.5	29	0.1	0.5
2nd -Year Follow-up Treatment Period:						
Children treated	66			69		
Treatment visits	96	0.4	0.8	125	0.5	1.0
Restored surfaces	225	1.0	2.3	272	1.2	2.4
Surfaces restored with amalgam	164	0.7	1.9	3	0.0	0.2
Surfaces restored with other materials	61	0.3	1.2	269	1.2	2.4
Teeth extracted	12	0.1	0.3	6	0.0	0.2
3rd - Year Follow-up Treatment Period:						
Children treated	59			77		
Treatment visits	100	0.4	1.0	144	0.6	1.2
Restored surfaces	222	1.0	2.7	308	1.4	2.9
Surfaces restored with amalgam	148	0.7	1.8	0	0.0	0.0
Surfaces restored with other materials	74	0.3	1.5	308	1.4	2.9
Teeth extracted	13	0.1	0.4	10	0.0	0.2
4th - Year Follow-up Treatment Period:						
Children treated	12			27		
Treatment visits	20	0.1	0.5	29	0.1	0.4
Restored surfaces	47	0.2	1.2	56	0.3	1.0
Surfaces restored with amalgam	39	0.2	1.1	0	0.0	0.0
Surfaces restored with other materials	8	0.0	0.4	56	0.3	1.0
Teeth extracted	0	0.0	0.0	3	0.0	0.2

3.2. Mercury exposure

In our April 2002 letter to the DSMB, we proposed a measure of mercury exposure based on the amount of time children are exposed to amalgam fillings and also on the size of the fillings and the type of tooth surface in which they are placed. Exposure is measured in units of surface-years, with one surface-year representing exposure for one year to one surface restored with a medium-sized amalgam filling. Each surface that is treated with an amalgam restoration is assigned 1) a mercury exposure score based on the surface type and the size of the restoration (see table below), and 2) an exposure time. For each patient, the total surface-years is calculated as the sum of the products of exposure scores multiplied by exposure times over all restorations.

The mercury exposure score is based on the surface type and size of the restoration as given in the following table:

	Mesial	Distal	Buccal	Lingual	Occlusal	Incisal
Small	0.5	0.5	0.5	0.5	0.75	0.25
Medium	1	1	1	1	1.5	0.5
Large	1.5	1.5	1.5	1.5	2.25	0.75

The scores for occlusal fillings are 1.5 times as large as the scores for Mesial, Buccal, Distal, and Lingual surfaces because the occlusal surface is larger and the size categorization is based on the fraction of the surface that is filled. This scoring system also allows for the increased mercury release from chewing surfaces. For incisal surfaces, the scores are one-half of the scores for M, D, B, and L surfaces. The precise values of the scores used may not be critical because of the wide variation in mercury exposure values between children within the amalgam group.

The mercury exposure time is calculated by the following formula: Exposure Time = (End – Start)/365.25, where Start is the date of placement of the restoration, End is the date of loss of the restoration (see below), and division by 365.25 converts from days to years. The definition of End is the latest date at which exposure is known to be continuing, which could be determined in one of several ways, including the date of the last treatment on the tooth, the date of extraction of the tooth, the date of the last oral exam at which the tooth was recorded as being present, or the mid-point of the interval between two visits during which the tooth was lost.

The distribution of the mercury exposure values to date is shown in Figure 3.2.1 using histograms for each of the two treatment groups. In the amalgam group there is an average 33.3 surface-years of mercury exposure and a wide variation between children (standard deviation of 21.4). In the composite group there is one child (a boy) who received amalgam fillings by mistake (previously reported) and has accumulated 33.6 surface-years of mercury exposure as a result. We also have examined a mercury exposure measure that is identical to the one described above except that it includes restorations on occlusal surfaces only. The two exposure measures are highly correlated ($r = 0.945$), because many of the restorations are on occlusal surfaces and these surfaces are given high weights in the exposure measure; hence, we report results only for the original (inclusive) measure.

We are monitoring 13 children (8 girls, 5 boys) in the amalgam group with the highest 2.5% of mercury exposure values (> 64 surface-years). Based on a review of the data histories for these 13 children (Appendix D.1, IDs 1005, 1029, 1033, 1074, 1089, 1096, 3166, 3512, 4227, 5338, 5645, 7396, 7411), Dr. Townes determined that five of these children (IDs 1005, 1029, 1089, 1096, 5645) have significant decrements in neurobehavioral performance and should be evaluated, one child (ID 3166) should be monitored next year, one child (3512) had endorsed multiple negative feelings on the VAMS and should be evaluated, and the other six children (IDs 1033, 1074, 4227, 5338, 7396, 7411) present no concerns.

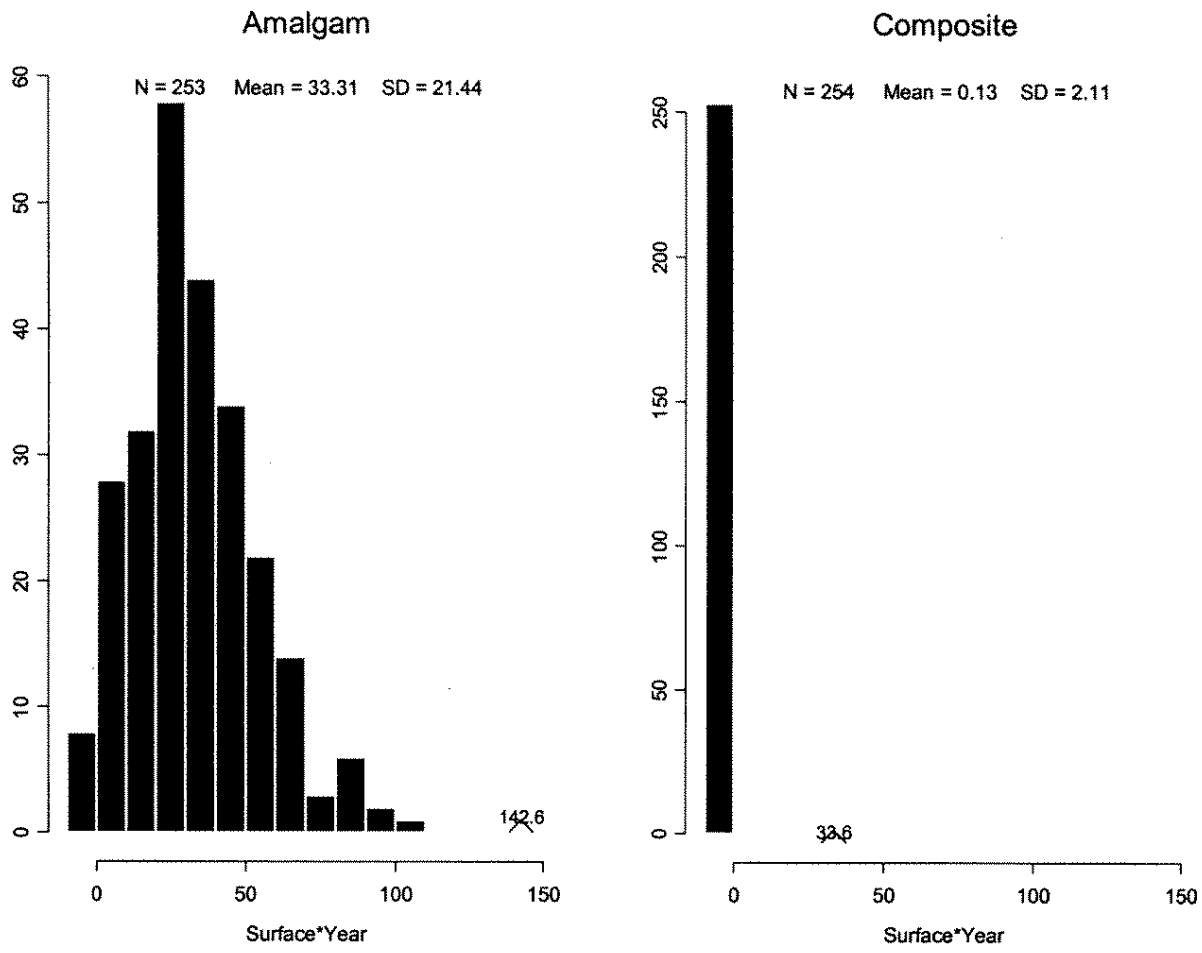


Figure 3.2.1. Distribution of the mercury exposure measure by treatment group.

4. Follow-up Results on Primary Outcome Variables

4.1. Correlations between child and adult tests

In the 4th year of follow-up, we started administering adult versions of some of the neurobehavioral tests that are more age-appropriate than the child versions that we started with (see Table 4.1.1). The child tests listed in Table 4.1.1 have been discontinued starting with follow-up year 5. Norms for all of these tests are provided in Tables A.1 and A.2.

Correlations between child and adult tests at year 4 and year 5 are given in Table 4.1.1. Some of the adult tests are identical to the corresponding child tests except that the words or symbols used are different (WISC Digit Span and Symbol Search vs. WAIS-III Digit Span and Symbol Search, Intermediate Trails A and B vs. Adult Trails A and B); for these the correlations between adult and child versions are moderately high (0.53-0.73). In other cases, the adult test is similar to the child test (WRAML/Finger Windows vs. WMS-III/Spatial Span, WISC/Coding vs. WAIS-III/Digit Symbol); the correlation is only moderate (0.48) in the first case, but quite high (0.78) in the second case. A few of the adult tests have a distinctly different structure than the corresponding child test (WRAML/Visual Learning vs. WMS-III/Visual Learning, WRAML/Visual Delay vs. WMS-III/Visual Delay, WRAVMA/Matching vs. WASI/Matrices); for these, the correlations are low to moderate (0.24-0.52). The most reassuring aspect of our findings is that there are high correlations (0.76-0.96) between neurobehavioral Z-scores computed using the new test battery that includes the adult test scores and the original Z-scores computed using child test scores (Table 4.1.1). Within-group correlations (not shown) are similar (small to moderately high correlations for individual tests and high correlations for Z-scores within each treatment group). The correlations observed here are quite similar to the correlations we have observed between repeat examination of the *same tests* by different examiners. Additional analyses showed that year 4 and year 5 adult test scores correlate reasonably well with the (child) test scores at baseline; in fact, the correlations are comparable to the correlations for the child scores themselves at year 4 except for WRAML/Visual Learning and WRAML/Visual Delay (Table 4.1.1). Because of the high correlations between child and adult Z-scores, results on primary outcome variables are essentially unchanged if the child tests are replaced with adult tests (see Section 4.2).

Table 4.1.1. Correlations between child test scores and corresponding adult test scores and correlations between Z-scores computed using child tests only and Z-scores computed using adult tests in place of child tests.

Variables		Correlations			
Child Test	**Adult Test**	**Child vs Adult at Year 4**	**Child at Baseline vs Child at Year 4**	**Child at Baseline vs Adult at Year 4**	**Child at Baseline vs Adult at Year 5**
Z-Memory	Z- Memory	0.76	0.65	0.53	0.58
Z-Visual Motor	Z- Visual Motor	0.96	0.67	0.67	0.62
Z-Attention	Z- Attention	0.93	0.78	0.75	0.74
WRAML/Visual Learning	WMS-III/Visual Learning	0.24	0.54	0.22	0.33
WRAML/Visual Delay	WMS-III/Visual Delay	0.30	0.48	0.20	0.28
WRAVMA/Matching	WASI/Matrices	0.52	0.48	0.46	0.45
WISC/Coding	WAIS-III/Digit Symbol	0.78	0.61	0.55	0.60
WISC/Symbol Search	WAIS-III/Symbol Search	0.65	0.46	0.41	0.38
WISC/Digit Span	WAIS-III/Digit Span	0.73	0.37	0.30	0.35
WRAML/Finger Windows	WMS-III/Spatial Span	0.48	0.32	0.27	0.16
Intermediate Trails A	Adult Trails A	0.53	0.48	0.40	0.59
Intermediate Trails B	Adult Trails B	0.57	0.49	0.57	0.59

4.2. Distribution of primary outcome variables by follow-up period

There continues to be no evidence of group differences in the primary outcome variables; the group means are approximately 0 for both groups (Figure 4.2.1). (Because the Z-scores are standardized to have overall mean 0 in the entire sample for each year, a treatment effect would result in a difference between the group means, with one being positive and the other negative.) The sample means of the Z-scores for memory and visual-motor domains tend to be lower for Group A than Group B at follow-ups, whereas the means for attention and nerve conduction velocity domains tend to be larger for Group A than for Group B at most follow-ups. However, these group differences are all small and within the range of sampling variation. The neurobehavioral Z-scores are calculated using child tests for year 4; adult tests will be used for calculation of Z-scores starting in year 5. The results for year 4 change very little if adult tests are used in place of child tests (Table 4.2.1) because of the high correlations between Z-scores using child and adult tests (see Section 4.1).

Table 4.2.1. Group comparisons (mean with SD in parentheses) based on neurobehavioral Z-scores calculated using child and adult test scores with 4th-year follow-up data.

	Memory		Visual-Motor		Attention	
	Group A	Group B	Group A	Group B	Group A	Group B
Mean Z-scores calculated using child test scores	-0.029 (0.789)	0.028 (0.810)	-0.001 (0.672)	-0.005 (0.601)	0.016 (0.723)	-0.016 (0.688)
Mean Z-scores calculated using adult test scores	0.012 (0.776)	-0.012 (0.738)	0.001 (0.678)	-0.010 (0.611)	0.037 (0.736)	-0.036 (0.678)

Results for boys (Figure 4.2.2) and girls (Figure 4.2.3) exhibit differences in the anticipated direction: boys tend to score higher than girls on visual motor tests whereas girls score higher than boys on attention tests. Note that higher scores for one gender are reflected in mean Z-scores that are positive for both treatment groups. Scores for memory and nerve conduction velocity show no consistent patterns by gender. Memory domain Z-scores were lower for boys in the amalgam group than boys in the composite group at all follow-ups, but the differences were not statistically significant. There were no other systematic treatment group differences within genders. All treatment group differences within genders were less than 0.21 (NCV result for boys in year 1, Figure 4.2.2) and none were statistically significant.

Although the primary endpoints for the study are the four Z-scores, we have also performed analyses to compare the treatment groups on the individual neurobehavioral test items at each time period. This decision was motivated by the desire to monitor changes in absolute performance (whereas the Z-scores are standardized at each time point and do not reflect changes in absolute scores). As in previous years, the results for 4th-year follow-up (Table 4.2.2) do not provide evidence for treatment group differences; the t-tests were significant at a nominal 0.05 level of significance for 1 of the 28 variables for girls, and for 2 of the 28 variables for boys.

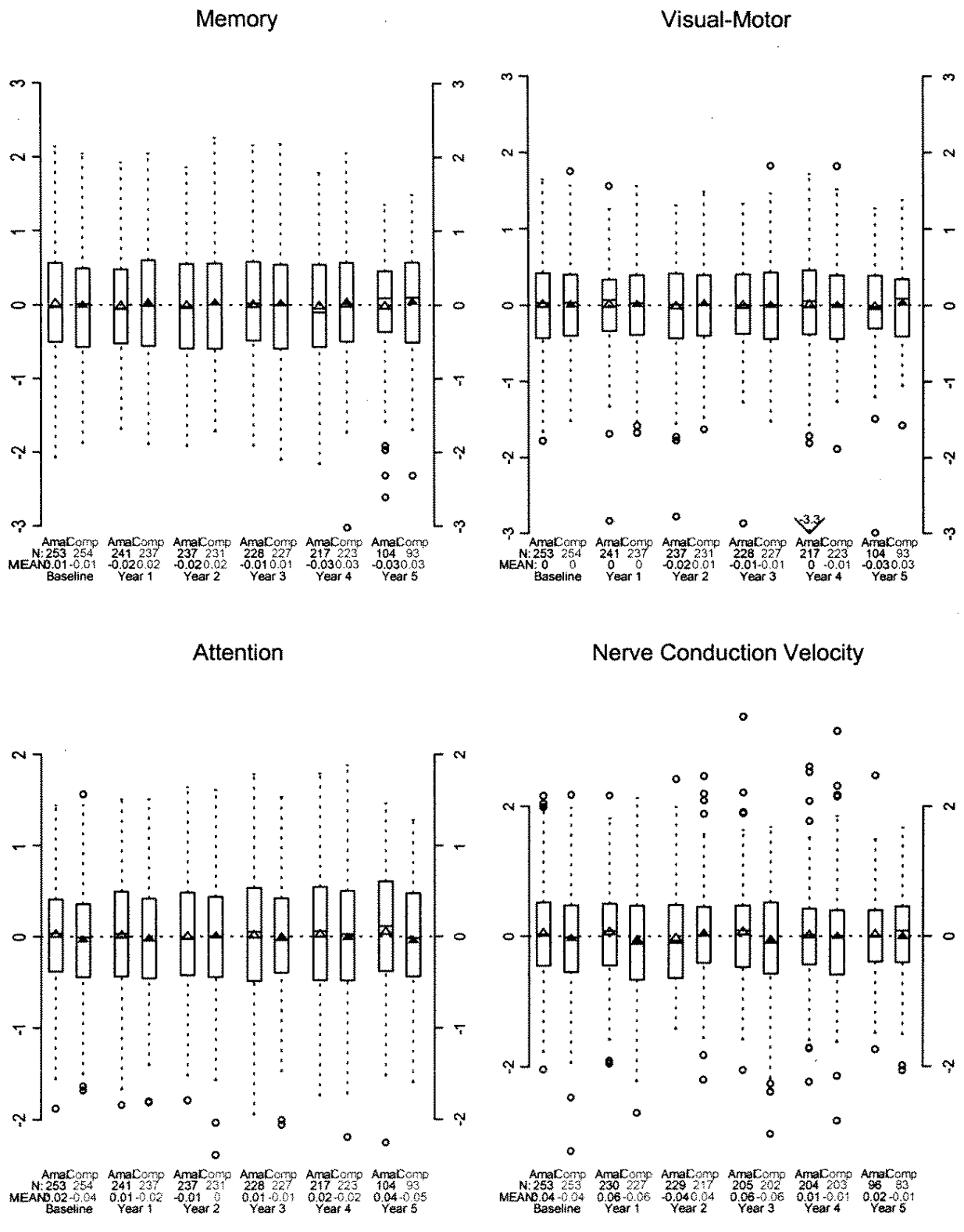


Figure 4.2.1 Distribution of Z-scores for neurobehavioral domains (memory, visual motor, and attention) and nerve conduction velocity by time period and treatment group.

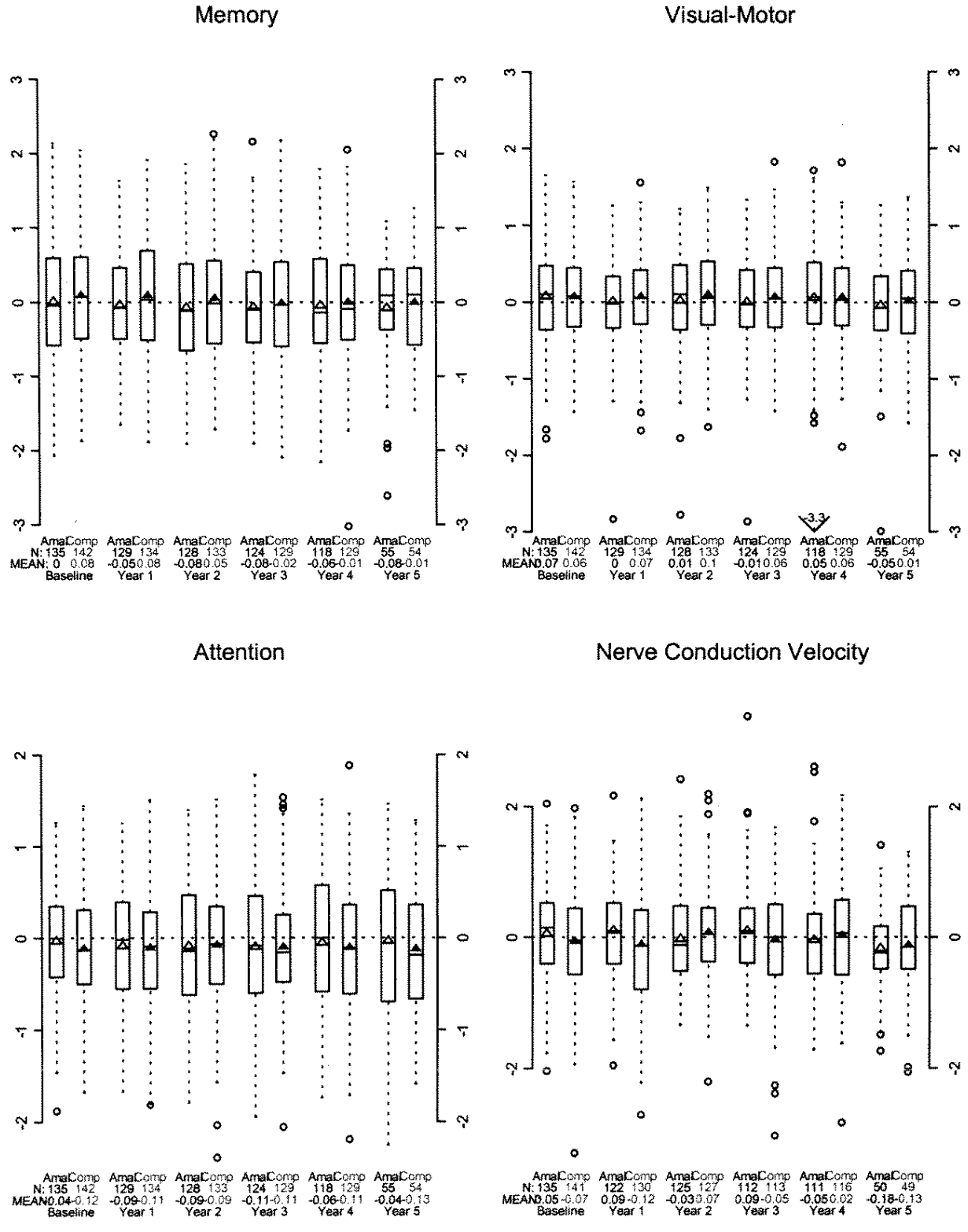


Figure 4.2.2. Distribution of Z-scores for neurobehavioral domains (memory, visual motor, and attention) and nerve conduction velocity by time period and treatment group for male participants.

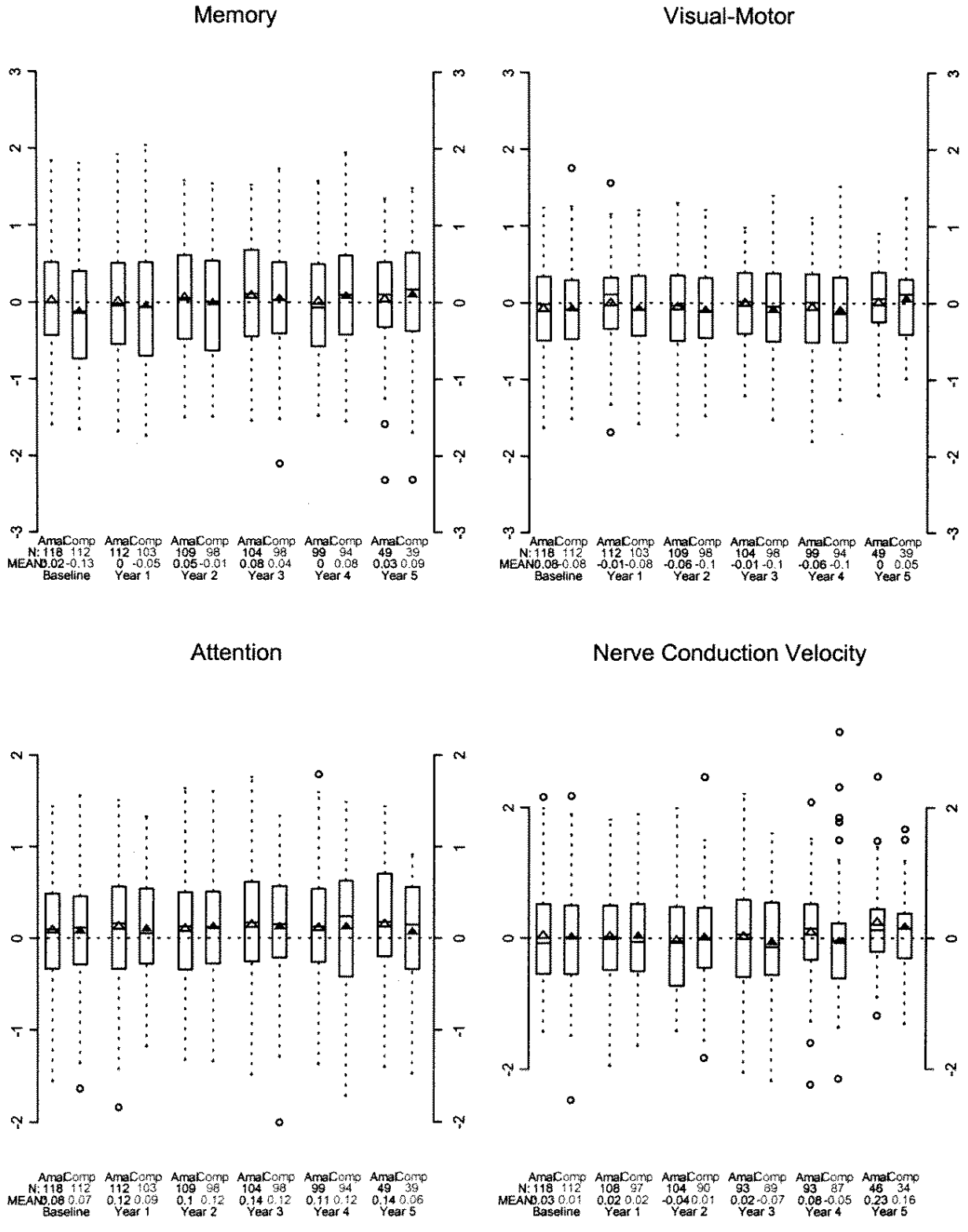


Figure 4.2.3. Distribution of Z-scores for neurobehavioral domains (memory, visual motor, and attention) and nerve conduction velocity by time period and treatment group for female participants.

Table 4.2.2. Comparison of treatment groups on individual neurobehavioral test scores at 4th year follow-up for girls and boys. Significantly better mean scores (p < 0.05 by t-test) are indicated by underlining and highlighting. (Higher scores are better except for items with units of seconds, as indicated below.)

Variable	Girls					Boys				
	Group A		Group B			Group A		Group B		
	Mean	Std Dev	Mean	Std Dev	P-value	Mean	Std Dev	Mean	Std Dev	P-value
RAVLT T.8	9.41	2.83	9.80	2.65	0.326	9.12	2.77	8.98	2.75	0.685
RAVLT TOT	45.24	8.35	46.95	8.54	0.161	43.77	8.41	43.74	8.67	0.976
WRAML visual Learning	8.65	2.78	8.43	2.96	0.586	8.68	3.09	9.10	3.02	0.276
WRAML Visual Delay	7.24	3.14	7.39	3.28	0.739	7.47	3.33	7.78	3.34	0.463
WRAVMA Matching	102.09	12.30	100.38	12.12	0.332	101.94	13.91	100.98	11.39	0.555
WRAMA Peg - dominant	120.99	15.69	121.96	15.32	0.665	115.91	17.97	117.12	16.52	0.583
WRAMA Peg - non-dominant	118.75	14.24	116.73	14.73	0.334	116.72	15.65	118.30	15.39	0.428
Standard Reaction Time (sec)	0.81	0.14	0.83	0.15	0.467	0.81	0.18	0.81	0.16	0.836
Finger Tapping - dominant	43.63	5.31	44.34	5.50	0.360	47.67	6.54	47.16	6.13	0.525
Finger Tapping - non-dominant	38.41	4.91	38.47	5.54	0.943	41.68	6.53	41.84	5.72	0.833
WISC - Coding	9.60	3.36	10.18	3.42	0.234	8.33	3.40	8.16	3.63	0.692
WISC - Symbol Search	10.80	3.17	10.84	3.11	0.929	10.01	3.47	10.09	3.30	0.845
WISC - Digit Span	7.60	2.80	7.27	2.56	0.388	7.03	2.59	7.24	2.44	0.521
WRAML - Finger Windows	7.68	2.47	7.19	2.25	0.152	7.03	2.78	7.57	2.42	0.105
Intermediate Trails A (sec)	16.84	6.55	15.38	6.00	0.109	14.96	6.01	16.73	6.83	0.033
Intermediate Trails B (sec)	32.08	12.76	32.16	14.58	0.968	30.56	13.84	32.74	14.12	0.225
STROOP WD4	40.44	7.86	40.81	6.57	0.724	38.89	7.76	38.08	7.76	0.413
STROOP COL4	40.67	7.64	40.49	6.86	0.863	37.68	7.56	37.11	7.20	0.548
STROOP CWD4	44.19	8.04	44.36	8.19	0.883	42.74	9.60	41.58	7.94	0.304
WMS III - Visual Learning	33.02	4.52	32.32	4.42	0.278	32.18	4.33	31.95	4.68	0.694
WMS III - Visual Delay	28.53	7.24	27.29	6.44	0.212	26.15	7.26	26.01	7.47	0.883
WASI Matrices	19.70	6.07	19.64	6.15	0.947	20.38	6.63	18.90	6.54	0.080
WAIS III - Digit Symbol	58.45	12.17	60.55	12.70	0.240	54.74	14.12	52.32	13.22	0.167
WAIS III - Symbol Search	26.55	8.74	27.36	6.57	0.466	26.67	8.61	25.69	7.68	0.348
WAIS III - Digit Span	12.80	3.61	12.07	2.86	0.125	12.54	3.35	12.38	3.37	0.712
WMS III - Spatial Span	14.17	3.08	13.24	3.27	0.044	13.56	3.62	13.62	2.86	0.886
Adult - Trails A (sec)	30.37	11.05	29.66	10.96	0.653	27.98	9.48	31.37	13.98	0.029
Adult - Trails B (sec)	73.56	32.56	70.17	31.52	0.464	75.74	33.47	78.47	31.81	0.514

4.3. Monitoring changes in performance of individual children

Step 1: Detection of Outliers

In our April 2002 letter to the Board, we outlined methods for monitoring decreases in performance in study participants to be used in the Casa Pia trial and the trial performed by NERI. The methods include a new criterion based on changes in Z-scores as well as the continued application of our Reliable Change Index (RCI) criterion. For follow-up year 4, we examined changes over 4 different time intervals for each variable: 1) baseline to year 4, 2) year 1 to year 4, 3) year 2 to year 4, 4) year 3 to year 4. Using the Z-score criterion, we identified children with a decrease of more than one in any one of the three neurobehavioral Z-scores over any of the four time periods. For the RCI method, we identified children having decreases in one of the time periods in any four or more individual test scores. By these analyses, a total of 82 children were identified by one of these two methods, including 25 identified by both methods, 3 identified by the RCI alone, and 54 identified by changes in Z-scores alone (Table 4.3.1). Of the children identified by the Z-score criterion, 72 had decreases in exactly 1 of the Z-scores, 7 children had decreases in exactly 2 Z-scores, and no children had decreases in 3 Z-scores. In our 2001 report we identified 41 children who had experienced decreases in neurobehavioral test scores in the first three years of follow-up, including 13 children who showed a decline that was of clinical concern; the number of these who were also identified in Year 4 was 12 and 8, respectively (Table 4.3.1).

Table 4.3.1. Number of children identified as having decreases in neurobehavioral test performance in Year 4 and comparison with children identified in 2001.

Method	Overall	Among 41 children identified in 2001	Among 13 children of clinical concern in 2001
Children with decreases in Z-scores but without decreases in RCI	54	5	4
Children with decreases in RCI but not decreases in Z-scores	3	2	2
Children with decreases in *both* Z-scores and RCI	25	5	2
Children with decreases in *either* Z-scores or RCI (total of above three rows)	82	12	8
Children without decreases	358	24	4
Children not tested in Year 4	67	5	1
Total	507	41	13

There are only minor differences between the children identified with decreases and the rest of the participants (Table 4.3.2). Among those identified by the RCI method, there are more in the Amalgam group (16) than the Composite group (12), whereas there is a higher proportion in the Composite group (32/54) among those with decreases in Z-scores only. Boys are overrepresented among those identified with decreases as in the previous year. There are no appreciable differences in age, baseline blood lead, or urinary mercury (Table 4.3.2). The average mercury exposure (surface-years) is lower in the group identified by decreases in Z-scores only (due to the low proportion in the Amalgam group) and is even lower among those not tested (due to less known exposure time among drop-outs).

Table 4.3.2. Comparison of children identified as having decreases in test performance with other children.

Variable	Decreases in Z-scores (n = 54)	Decreases in RCI (n = 3)	Decreases in Both Z-scores and RCI (n = 25)	Other Children (n = 358)	Not Examined in Year 4 (n = 67)
% Group A	22 (40.7%)	2 (66.7%)	14 (56.0%)	179 (50.0%)	36 (53.7%)
% Male	70.4	100	56.0	53.6	44.8
Mean Age (yrs)	14.3	14.6	14.0	14.1	14.1
Mean Baseline Blood Lead (mg/dL)	4.8	3.8	4.2	4.6	4.8
Mean Surface-Years Hg Exposure:	13.5	20.8	18.4	18.4	9.2
Mean Creatinine-Adjusted Urinary Hg (ug/g):					
Baseline	2.1	2.4	2.1	1.8	1.8
Year 1	1.9	1.1	1.9	2.0	1.8
Year 2	2.2	1.0	2.7	2.4	2.8
Year 3	2.2	0.9	2.6	2.3	2.3
Year 4	1.4	0.9	1.6	1.8	2.7*

* Results for 3 children who have not yet been tested but for whom laboratory data are available.

Step 2: Clinical Assessment

After identification of children with decreases in test performance based on statistical criteria, Dr. Townes performed a masked clinical assessment of the entire data histories for the 82 children identified. This assessment identified a subset of 22 children who exhibited decrements in test performance that were of clinical concern and required evaluation, an additional 21 children of less concern who should be evaluated/monitored, 26 children with minimal concerns, and 13 with no concerns. Of the 22 children of most concern, 9 were identified by both the Z-score and RCI criteria, 12 were identified by Z-scores only, and 1 was identified by RCI only. Most (20 of 22) of these children are boys. The distribution between the two treatment groups is 13 Amalgam and 9 Composite. Dr. Townes also assessed the data for the 41 children identified as outliers in previous years and reported on in our 2001 report as well as children identified by neurologic exam findings or unusual values on any neurobehavioral test. These assessments identified one additional child of clinical concern and 13 children of less concern who should be evaluated/monitored.

In late September to early October 2002, Dr. Townes will be in Lisbon to evaluate children of concern for the purpose of determining the underlying causes of the children's conditions and to make recommendations for any necessary care and services that they need. A total of 57 children will be targeted for examination, including 23 of highest clinical concern (the 22 identified as outliers this year and one of the 41 identified in previous years) and 34 other children of less concern (the 21 identified as outliers this year and 13 children identified in previous years or identified based on other findings this year). Priority will be given to examination of the 23 children of most concern. The findings will be added as Appendix G to this report and sent to the Board in October prior to the meeting.

4.4. Statistical stopping rule

The primary analysis for the study is based on the four primary outcomes (memory, visual motor, and attention domain scores from the neurobehavioral testing, and a combined nerve conduction velocity score). Two separate tests are performed: an O'Brien type test that will detect small group differences that are similar across the four endpoints and a Hotelling T^2 test that will detect inconsistent group differences, e.g., a difference in just one of the variables. The overall significance level of the test accounts for the overlap between these two tests (i.e., it is possible that both tests reject the null hypothesis). The O'Brien test statistic is a two-sample t-statistic computed using a derived variable that weights the four endpoints according to the inverse of the pooled correlation matrix. A 1-sided p-value is reported for this test. The required significance levels for this test component depend on which treatment group has the worse outcome (a higher significance level is allowed for amalgam so that we will have higher power to detect adverse effects of amalgam). Note that if the O'Brien statistic p-value was intermediate between the two required significance levels, the treatment groups would have to be unmasked in order to determine if the required significance level was achieved; in this event, we would provide the treatment identities to the DSMB only in order to maintain masking of study personnel. Differences in either direction are treated the same by Hotelling's test; therefore, a 2-sided p-value is reported for this test. Neither of the two test components is significant (Table 4.4.1).

Table 4.4.1. Results of statistical analysis.

	Test Statistic	P value	Required Significance Levels	
			Hypothesis: Amalgam Harmful	Hypothesis: Composite Harmful
O'Brien	T = -0.16	0.44 (1-sided)	0.0027	0.0008
Hotelling's T^2	F = 0.30	1.00 (2-sided)	0.0020	

4.5. Reassessment of power

There have been no changes in our assessment of the power of the study since our previous report. Attrition rates have been maintained at values that are slightly less than the anticipated values of 5% in the first year and 3% in each subsequent year. The power continues to be sufficient to detect small effects that are roughly uniform across outcomes with five years of follow-up and to detect small effects that occur in only one outcome with seven years of follow-up.

5. Follow-up Results on Secondary Outcome Variables

5.1. Sentinel health events and health history

In our health events surveillance system, health related information from several sources is collected by the Study Records Manager in Lisbon for the purpose of identifying serious health problems. Three sentinel health events, all reported previously, have been identified to date: one child suffered a stroke, one was diagnosed with epilepsy, and one child died of causes unrelated to participation in the study. None of these children were identified as an outlier based on neurobehavioral performance (Section 4.3), mercury exposure (Section 3.2) or urinary mercury (Section 5.4.1). There are no new data since our last report on two of the children (the child who died and the child with epilepsy). Based on examination of the recent data for the child who suffered a stroke, Dr. Townes determined that this child does not show significant decrements in performance. Data histories for all three of these children are given in Appendix D.2 (IDs 3178, 4251, 5318). One of the components of the surveillance system is the annual health history questionnaire. In previous years, there were no significant differences between treatment groups in responses to the questionnaire (Table 5.1.2).

Table 5.1.2. Follow-up health history information by treatment group.

Survey Question	First-Year Follow-up			
	Group A		Group B	
	Positive Responses	Percent	Positive Responses	Percent
Hospitalized in the last year?	24	10.0	20	8.4
Have any disease in the last year?	54	22.4	44	18.5
Taking any medications in the last year?	90	37.3	82	34.5
Taking any medications today?	34	14.1	26	10.9
Had any allergies in the last year?	30	12.4	25	10.5
Survey Question	**Second-Year Follow-up**			
	Group A		Group B	
	Positive Responses	Percent	Positive Responses	Percent
Hospitalized in the last year?	19	8.5	19	8.3
Have any disease in the last year?	43	19.2	35	15.2
Taking any medications in the last year?	67	29.9	64	27.8
Taking any medications today?	14	6.3	18	7.8
Had any allergies in the last year?	23	10.3	20	8.7
Survey Question	**Third-Year Follow-up**			
	Group A		Group B	
	Positive Responses	Percent	Positive Responses	Percent
Hospitalized in the last year?	33	14.7	25	11.2
Have any disease in the last year?	43	19.2	41	18.4
Taking any medications in the last year?	66	29.5	64	28.7
Taking any medications today?	12	5.4	21	9.4
Had any allergies in the last year?	22	9.8	14	6.3

5.2. Neurological

Prevalences of neurological abnormalities are similar in the two treatment groups. There are no consistent trends over time (Table 5.2.1).

Table 5.2.1. Prevalence of neurological abnormalities by study year and treatment group (N is the number of children with positive response).

Abnormality	Baseline				First-Year Follow-up				Second-Year Follow-up			
	Group A		Group B		Group A		Group B		Group A		Group B	
	N	%	N	%	N	%	N	%	N	%	N	%
Mental Status	0	0.0	0	0.0	1	0.4	3	1.3	2	0.9	2	0.9
Standard Cranial Nerve Exam	6	2.4	3	1.2	4	1.7	6	2.6	6	2.6	6	2.7
Motor Exam	1	0.4	1	0.4	0	0.0	1	0.4	1	0.4	0	0.0
Sensory	0	0.0	1	0.4	0	0.0	0	0.0	1	0.4	0	0.0
Coordination	2	0.8	1	0.4	0	0.0	1	0.4	2	0.9	0	0.0
Reflexes	3	1.2	0	0.0	3	1.3	2	0.9	1	0.4	2	0.9
Gait/Station	1	0.4	0	0.0	0	0.0	0	0.0	3	1.3	0	0.0
Adventitious Movement	0	0.0	0	0.0	3	1.3	1	0.4	5	2.2	3	1.3
Any Abnormality	9	3.6	6	2.4	10	4.3	11	4.7	12	5.2	12	5.4

Abnormality	Third-Year Follow-up				Fourth-Year Follow-up				Fifth-Year Follow-up			
	Group A		Group B		Group A		Group B		Group A		Group B	
	N	%	N	%	N	%	N	%	N	%	N	%
Mental Status	1	0.5	1	0.5	2	1.0	1	0.5	0	0.0	0	0.0
Standard Cranial Nerve Exam	1	0.5	3	1.6	1	0.5	0	0.0	2	2.3	3	4.2
Motor Exam	3	1.5	2	1.1	3	1.5	0	0.0	2	2.3	0	0.0
Sensory	0	0.0	1	0.5	0	0.0	0	0.0	1	1.2	0	0.0
Coordination	2	1.0	1	0.5	1	0.5	0	0.0	1	1.2	0	0.0
Reflexes	1	0.5	3	1.6	3	1.5	3	1.6	2	2.3	1	1.4
Gait/Station	3	1.5	2	1.1	4	2.0	0	0.0	3	3.5	0	0.0
Adventitious Movement	1	0.5	2	1.1	1	0.5	0	0.0	3	3.5	1	1.4
Any Abnormality	4	2.0	7	3.8	6	3.1	4	2.1	5	5.8	5	6.9

The prevalence of soft signs was approximately the same in the two groups (Table 5.2.2). Most (72.9%) of the positive signs were coded as 1 ("Mild, Inconsistent"), 26.0% were coded as 2 ("Present"), and 1.1% were coded as 3 ("Exuberant or Very Evident").

Table 5.2.2. Prevalence of neurological soft signs by study year and treatment group (N is the number of children with positive response).

Soft Sign	Third-Year Follow-up				Fourth-Year Follow-up				Fifth-Year Follow-up			
	Group A		Group B		Group A		Group B		Group A		Group B	
	N	%	N	%	N	%	N	%	N	%	N	%
Mirror Movements	54	27	63	34	47	24	57	30	24	28	19	26
Syncinesias	76	39	78	42	74	38	65	34	24	28	22	31
Fine Finger	44	22	48	26	25	13	23	12	11	13	8	11
Lack of Balance	14	7.1	9	4.8	5	2.6	7	3.6	4	4.7	1	1.4
Motor Impersistence	34	17	38	20	28	14	20	10	5	5.8	7	9.7
Hyperactivity	0	0.0	4	2.2	2	1.0	0	0.0	2	2.3	0	0.0
Any soft sign	122	62	118	63	118	60	112	58	42	49	37	51
Total	197	100	186	100	196	100	193	100	86	100	72	100

5.3. Oral exam

The Amalgam group has experienced less tooth decay during follow-up than the Composite group (Table 5.3.1). The group difference in average number of carious surfaces is statistically significant ($p < 0.05$ by t-test) at Years 2, 3, and 4; results for year 5 are preliminary (n=181). As requested by the DSMB, starting in October 2000, we have distinguished between recurrent decay and new decay. Results in Year 4 show that 91.4% of the decay found is new decay and 8.6% is recurrent decay. Most of the recurrent decay has been in the Composite group (41 decayed surfaces in the Composite group versus 11 surfaces in the Amalgam group).

Table 5.3.1. Results of caries exams by study year and treatment group.

	Baseline		1st Year Follow-up		2nd Year Follow-up		3rd Year Follow-up	
	Amalg.	Comp.	Amalg.	Comp.	Amalg.	Comp.	Amalg.	Comp.
Children examined	253	254	226	224	236	232	224	224
Children with caries	252	254	120	134	87	97	89	109
Carious surfaces present	3947	4031	368	462	229	392	256	457
Mean number of carious surfaces	15.60	15.87	1.63	2.06	0.97	1.69	1.14	2.04

	4th Year Follow-up		5th Year Follow-up	
	Amalg.	Comp.	Amalg.	Comp.
Children examined	210	217	95	86
Children with caries	72	98	37	38
Carious surfaces present	207	344	101	119
Mean number of carious surfaces	0.99	1.59	1.06	1.38

The numbers of primary and permanent teeth present are similar in the two treatment groups (Table 5.3.2). By the fourth year of follow-up, very few primary teeth are remaining.

Table 5.3.2. Number of primary and permanent teeth present by study year and treatment group.

Tooth Type	Study year											
	Baseline				First-Year Follow-up				Second-Year Follow-up			
	Group A		Group B		Group A		Group B		Group A		Group B	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Primary	8.4	3.7	8.2	3.9	5.5	3.5	5.8	3.8	4.5	3.1	4.3	3.3
Permanent	16.5	5.1	16.9	5.3	21.1	5.4	20.8	5.6	24.0	4.3	23.6	4.6

Tooth Type	Study year											
	Third-Year Follow-up				Fourth-Year Follow-up				Fifth-Year Follow-up			
	Group A		Group B		Group A		Group B		Group A		Group B	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Primary	2.6	2.3	3.3	2.9	1.9	1.8	2.2	1.8	2.2	1.1	1.2	0.4
Permanent	26.2	2.7	25.4	3.5	27.0	1.8	26.8	2.0	27.5	1.3	27.2	1.1

5.4. Laboratory Measures

5.4.1. Urinary Samples

During follow-up, the mean creatinine-adjusted urinary mercury concentration has increased in the amalgam group but there has been little or no change in the composite group (Figure 5.4.1). After a steady rise in average values in the Amalgam group from baseline to year 3, there has been a slight downward trend from year 3 to year 5. Similar trends were found in the unadjusted mercury levels (not shown). There is a moderate association between urinary mercury levels in year 4 and the mercury exposure values (correlation coefficient = 0.345, $p < 0.001$).

In both groups, the levels of GST-alpha have increased slightly and the levels of GST-pi show a clear increasing trend (Fig 5.4.1); changes for GST-alpha are not statistically significant whereas changes for GST-pi are significant at the 0.05 level for all follow-ups in the amalgam group and for years 2-4 in the Composite group. The average levels of the GST parameters tend to be slightly larger in the amalgam group than the composite group at most follow-ups, but these differences are not statistically significant. Albumin levels (Fig 5.4.1) do not show any clear trends over time or systematic differences between groups. The levels of three porphyrins thought to be related to mercury exposure are displayed in Figure 5.4.2. Average levels of these porphyrins increased from baseline to first year ($p < 0.05$) and then declined slightly in later follow-ups (Fig. 5.4.2). Average levels of the porphyrins are slightly higher for the amalgam group than the composite group but these differences are not statistically significant. Levels of uroporphyrin, heptaporphyrin, and hexaporphyrin exhibit the same time trends as illustrated in Fig 5.4.2 but no differences between treatment groups (data not shown); these latter three porphyrins are not considered sensitive to mercury exposure. Creatinine levels have increased over time in both treatment groups, with no apparent systematic differences between the two groups (Fig. 5.4.2).

5.4.2. Monitoring of Children with High Urinary Mercury Values

A total of 28 children (23 in the Amalgam group, 5 in the Composite group) have had a follow-up urinary mercury level exceeding 10 ug/L. In the past year, 7 children had high values, including one child who had also had a high value in a previous year. This group has higher proportions of females (20/28, 71%), and black children (7/28, 25%) than the study population as a whole. Samples with mercury levels above 10 ug/L are re-analyzed to attempt to confirm the high value. To date, there have been 7 cases (5 in the Amalgam group and 2 in the Composite group) of repeat mercury analyses exceeding 10 ug/L, including 3 reported previously and 4 new ones. All 7 of these children are girls and 3 are black. One of these children (ID 1074) has a high mercury exposure derived score (Section 3.2) but there are no other overlaps with other monitoring lists (sentinel health events, high baseline blood lead, and neuropsychological outliers). No cause for the elevated urinary mercury levels is apparent; a source of mercury exposure other than from amalgams is possible. Later urine samples have been obtained and analyzed for four of the children (ID 1068, 1074, 1075, 4219); mercury concentrations were below 10 ug/L for all four samples. Additional samples will be obtained for the other children and results will be made available at the meeting. Dr. Townes found no clinical concerns for five of the children (IDs 1068, 1074, 1583, 3187, 4219) but determined that two of the children (IDs 1075, 3506) required careful monitoring next year. We will continue to monitor all seven of these children. Detailed case reports for these children are provided in Appendix D.3 (IDs 1068, 1074, 1075, 1583, 3187, 3506, 4219).

5.4.3. Monitoring Children with Elevated Blood Lead Values

We have continued to monitor the outcome data for the twelve children who had blood lead values larger than 10ug/dl at baseline (one of our exclusion criteria was blood lead level exceeding 15ug/dl but levels between 10 and 15 may also be of some concern). The results of Dr. Townes' assessment of the data histories for these children (Appendix D.4; IDs 1014, 1021, 1056, 1124, 1135, 2138, 3190, 4237, 4270, 5295, 5340, 6354) were as follows: two of the children (IDs 1135, 2138) had significant decrements and should be evaluated in October, one child (ID 4270) showed some decrements and should be carefully monitored next year, five of the children showed no cause for concern, and four of the children (IDs 1021, 1056, 5340, 6354) were not tested this year. We will continue to monitor all 12 of these children in future follow-ups.

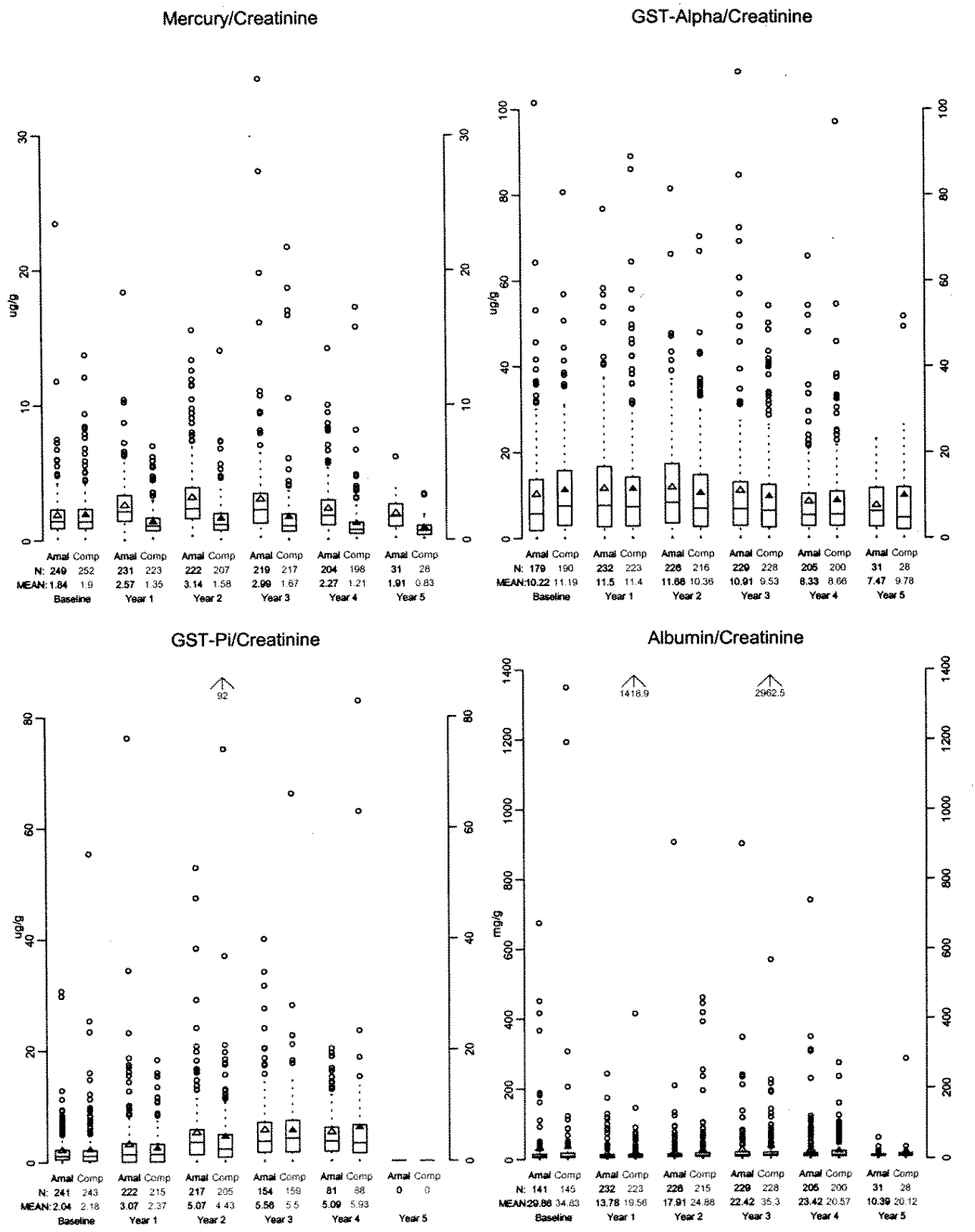


Figure 5.4.1. Distribution of creatinine-adjusted values of urinary mercury, GST-alpha, GST-pi, and albumin by time period and treatment group.

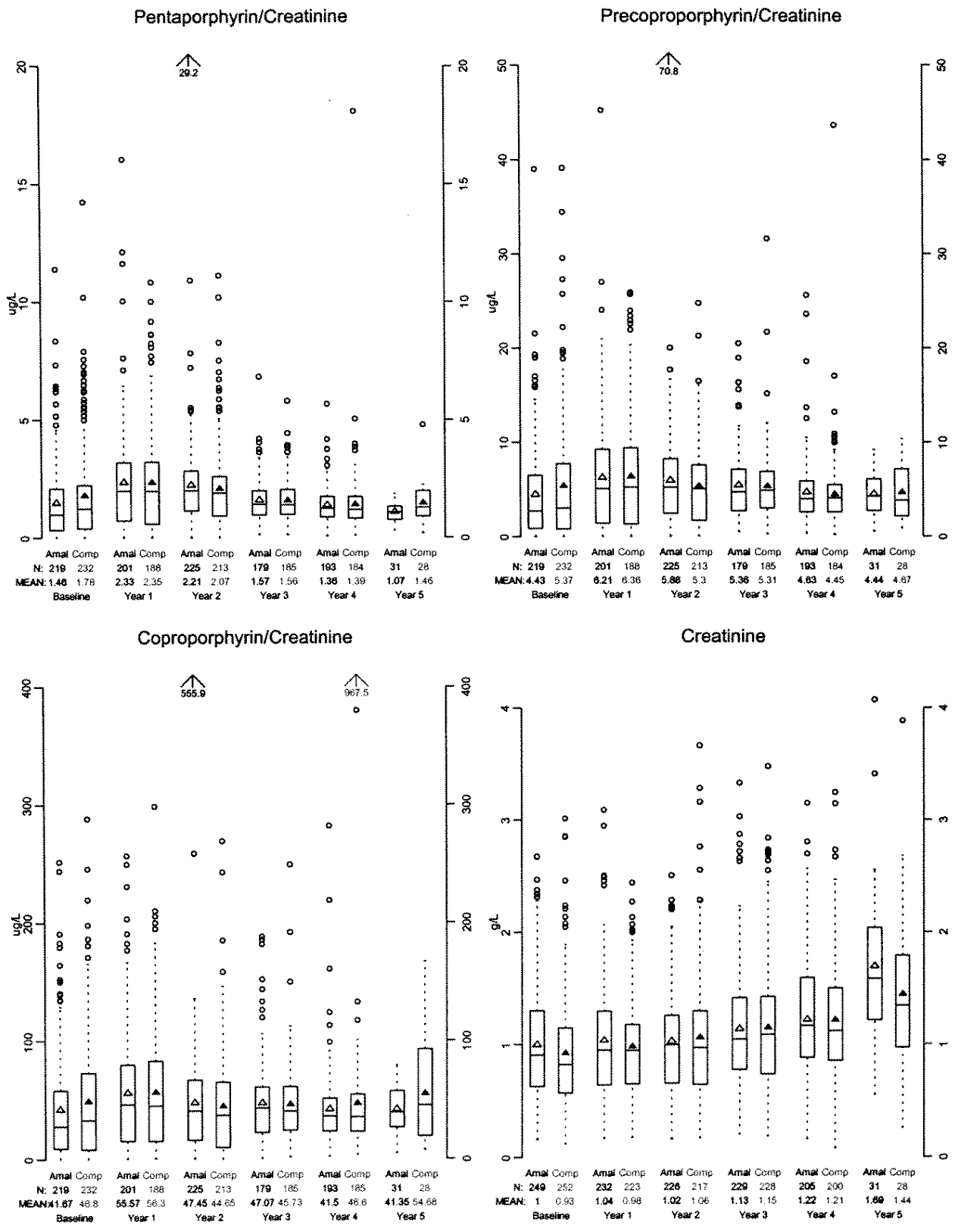


Figure 5.4.2. Distribution of creatinine-adjusted values of pentaporphyrin, precoproporphyrin, and coproporphyrin, and levels of creatinine, by time period and treatment group.

6. Quality Assurance Monitoring

6.1. Completeness of data

Information on loss to follow-up and missed visits was given in Section 2.2. This section provides information on item-specific missing data for children who were examined, item-specific non-response on health history questionnaires received, and missing laboratory measures. Information is provided only where the percentage of missing data is 5% or more, that is, data are at least 95% complete except where noted below.

Table 6.1.1. Data items with 5% or more missing values at a single time period. Values given are number of children with missing data on the item indicated. Footnotes are given below the table.

	Baseline	Year 1	Year 2	Year 3	Year 4
Standard Reaction Time[1]			56	84	
Adventitious movements[2]	122				
Albumin[3]	221				
GST – Alpha[4]	138				
GST – Pi[5]				145	237
Porphyrins[6]	50	84		85	26

Notes:

1. Standard reaction time data are missing because of equipment failure (equipment has now been replaced).
2. Recording of adventitious movements was started after initiation of the study during the enrollment period.
3. Albumin analyses began during the enrollment period in October, 1997.
4. Missing GST-alpha data at baseline are due to problems with assays that have been corrected.
5. Missing GST-pi data in year 3 and 4 is due to discontinuation of GST-pi analyses by agreement with NIDCR to accommodate a reduction in the budget for the trial.
6. Missing porphyrin data at baseline and in year 1 are due to instrument failure that has been solved. Missing values in year 3 are due to freezer failure. Missing values in year 4 are for samples that have not been analyzed yet.

6.2. Accuracy of data

There are three different types of checks on the accuracy of the data as described below.

1) Range Checks: The main study data are entered into Access© databases in Lisbon with built-in range checks to ensure that only valid data values are entered. Ranges of all data are checked again in Seattle before analysis as an extra precautionary measure.
2) Random Audits: In each follow-up year, approximately 15% of the subjects are selected at random for comparison of all data forms against the computerized database. Estimated error rates have been less than 0.2% throughout follow-up (Table 6.2.1).
3) Data-Generated Audits: The data are checked for unusual values and for consistency of data values collected on the same subject over time. Unusual values and large changes are checked for accuracy.

When an error in the computerized database is identified based on any of the checks described above, the information is transmitted to the Study Records Manager in Lisbon who identifies the source of the error and makes any necessary correction on the original data forms and in the computer database. The corrected version of the database is transmitted to Seattle for subsequent analyses.

Table 6.2.1. Results of random data audits.

	Baseline	1st-Year Follow-up	2nd-Year Follow-up	3rd-Year Follow-up	4th-Year Follow-up
Number of Subjects	74	78	70	75	79
Number of Data Items Checked	33522	20555	25791	28725	28732
Number of Mismatches	132	32	19	45	55
Mismatch Rate	0.39%	0.16%	0.07%	0.16%	0.19%

6.3. Examiner reliability and accuracy

In January 2002, Ms. Rosenbaum visited Lisbon to train and re-evaluate the examiners for the beginning of the 5th year follow-up. At that time, each of the primary examiners was observed testing two subjects. The third examiner administered the tests to Ms. Rosenbaum one time in Portuguese and was also observed administering the tests to two subjects. Comments and critiques were done after each testing. There was also a discussion of scoring issues. Test administration questions and scoring issues are addressed regularly by e-mail and FAX. The examiners continue to send videos of their testing and videos of all three examiners are made on the reliability study. The examiners' ability to administer the tests in a consistent and standardized manner remains high. The data is verified by another examiner and additionally, unusual values are double checked by Dr. Townes. In September, Ms. Rosenbaum returned to Lisbon and worked with the examiners to refresh their enthusiasm and investment in the study. Each examiner was observed one time and performance was critiqued. They were also instructed in the use of the computerized tests.

6.3.1. Checklist evaluation

The neurobehavioral examiners continue to send videotapes of testing sessions to Ms. Rosenbaum who uses them to evaluate the examiners' performance with the Test Procedure Checklist. When errors are noted, they are discussed with the examiner via e-mail or in person. The percentage of items correctly performed continues to be high (Table 6.3.1).

Table 6.3.1. Results of checklist-scoring of neurobehavioral examiners' performance.

	Examiner 1			Examiner 2			Examiner 3		
Year of Session	# Sessions	# Items	Accuracy (%)	# Sessions	# Items	Accuracy (%)	# Sessions	# Items	Accuracy (%)
1997	4	599	95	5	785	97			
1998	7	1113	98	10	1576	98			
1999	3	488	98	4	624	98			
2000	5	773	99.6	5	785	99.5			
2001 (Main Study)	3	663	97.4	4	882	99.0			
2001 (Reliability Study)	3	408	97.3	3	408	97.8	3	272	94.5
2002	4	544	96.0	4	544	98.3	3	4-8	99.3

6.3.2. Inter-examiner reliability

The first inter-examiner reliability study was performed in 1999. Starting in 2001 we initiated annual reliability studies to evaluate the reliability of the examiners in conducting the new tests that were started in follow-up year 4 and additional new tests to be implemented in later years, and to assess the reliability of the third examiner, so that we might be able to employ her in conducting follow-up testing for the trial if the need arises to replace one of the original two examiners. Each year, children are selected from those who gave consent to participate in the

trial but were found to be ineligible. The 2002 reliability study is on-going and expected to be completed by the end of the year; results will be reported in next year's report.

6.3.3. Differences between examiners

As described in Section 1.5.5, we now have 3 different examiners performing tests. There continue to be similar distributions of examiners in the two treatment groups (Table 6.3.3), which provides protection against the possibility of bias in the treatment comparison due to differences between examiners in test scores.

Table 6.3.3. Distribution of neurobehavioral examiner by treatment group and follow-up year.

	Baseline		Year1		Year2	
	Group A	Group B	Group A	Group B	Group A	Group B
Examiner 1	105	110	102	104	110	111
Examiner 2	148	144	139	133	127	120
Examiner 3	.	.	.	.	.	.

	Year3		Year4		Year5	
	Group A	Group B	Group A	Group B	Group A	Group B
Examiner 1	109	108	94	89	47	34
Examiner 2	119	118	102	113	42	47
Examiner 3	.	.	21	21	15	12

As in previous follow-ups, there were some small differences between the original two examiners in average scores on some fourth-year neurobehavioral tests (Table 6.3.4). In year 4, there were statistically significant differences (0.05 level of significance) in 4 tests, with one examiner having a better average score on 3 tests (including 2 of the 4 tests in the Memory domain) and the other examiner having the better average score on 1 test. The continual re-assessment and re-training that Gail Rosenbaum performs and verbal feedback provided to the psychometrists in Lisbon has eliminated the consistent differences between the original two examiners that was noted at baseline and has ensured that differences between these two examiners throughout follow-up are small. There is a concern with differences between the third examiner and the original two examiners in year 4 testing (Table 6.3.4). Review of the taped sessions by Ms. Rosenbaum has suggested that, despite high scores of technical correctness among all three examiners, the newest examiner shows more energy and attention to the testing process than the two original examiners. Observation of the children reviewed on the tapes, and review of the raw testing data, suggest that the children as well as the original examiners may be putting forth less effort than previously. This was particularly evident on the simpler tests from the attention domain. During the September 2002 visit, Ms. Rosenbaum did a formal retraining of all the examiners, focusing on the need for an energetic and enthusiastic approach.

Table 6.3.4. Examiner differences in neurobehavioral test results in year 4. (Higher scores are better except for items with units of seconds, as indicated.)

Variable	Examiner 1		Examiner 2		Examiner 3		P-value (ANOVA)
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	
Memory Domain Score	-0.13	0.82	0.07	0.76	0.21	0.82	0.011
Visual Motor Domain Score	-0.05	0.64	-0.01	0.65	0.21	0.55	0.055
Attention Domain Score	-0.01	0.69	-0.03	0.71	0.20	0.72	0.164
RAVLT T.8	9.37	3.04	9.12	2.52	9.81	2.63	0.297
RAVLT TOT	44.87	8.95	44.35	8.02	46.50	9.50	0.326
WRAML visual Learning	8.02	2.76	9.27	3.04	9.19	2.90	0.000
WRAML Visual Delay	6.51	3.19	8.14	3.14	8.45	3.27	0.000
WRAVMA Drawing	N/A	N/A	N/A	N/A	N/A	N/A	N/A
WRAVMA Matching	99.98	12.54	102.09	12.23	103.62	12.68	0.112
WRAMA Peg - dominant	118.66	15.66	118.66	17.50	119.21	16.52	0.979
WRAMA Peg - non-dominant	117.86	14.58	116.74	15.71	121.36	13.12	0.186
Standard Reaction Time (sec)	0.81	0.18	0.83	0.15	0.76	0.13	0.066
Finger Tapping - dominant	45.09	6.77	46.36	5.71	46.95	5.29	0.062
Finger Tapping - non-dominant	40.11	5.94	40.07	5.95	42.24	5.83	0.084
WISC - Coding	8.80	3.37	8.98	3.64	9.62	3.89	0.402
WISC - Symbol Search	10.26	3.16	10.53	3.36	10.26	3.50	0.686
WISC - Digit Span	7.20	2.80	7.23	2.46	7.81	2.27	0.367
WRAML - Finger Windows	7.31	2.43	7.19	2.40	8.62	3.00	0.003
Intermediate Trails A (sec)	15.52	6.27	16.42	6.63	15.90	5.93	0.378
Intermediate Trails B (sec)	31.69	13.82	32.20	13.82	31.17	14.18	0.879
STROOP WD4	39.36	7.50	39.01	7.56	41.71	8.02	0.107
STROOP COL4	38.32	7.42	38.76	7.53	40.98	7.29	0.116
STROOP CWD4	43.00	8.40	42.72	8.59	45.24	8.70	0.213
WMS III - Visual Learning	31.98	4.21	32.55	4.63	32.71	5.00	0.386
WMS III - Visual Delay	26.31	7.36	27.37	7.05	26.90	7.26	0.350
WMS III - Spatial Span	13.65	3.17	13.19	3.03	16.02	3.38	0.000
WASI Matrices	18.15	6.37	20.14	6.17	23.45	5.61	0.000
WAIS III - Digit Span	12.29	3.38	12.47	3.22	13.05	3.51	0.403
WAIS III - Digit Symbol	55.28	13.37	55.91	13.42	60.79	13.57	0.055
WAIS III - Symbol Search	26.48	7.87	26.37	7.81	27.29	9.25	0.792
Adult - Trails A (sec)	29.35	9.11	30.20	13.27	30.50	12.60	0.721
Adult - Trails B (sec)	71.93	26.47	79.19	37.06	65.29	26.64	0.011

7. Summary

In summary, the Casa Pia Study of the Health Effects of Dental Amalgams in Children continues to progress satisfactorily in all respects. Subject retention continues to be high, maintenance of the timetable continues to be acceptable, there have been no adverse events related to the study, and there were no additional protocol deviations during the current reporting year.

ONLY FILE COPY
Keep with Committee Copy
Send to Reviewer at Renewal
To Records when file is closed
App.# 95-0401-A/C 09
DeRouen, Timothy

CHILDREN'S AMALGAM TRIAL: UNIVERSITY OF WASHINGTON / UNIVERSITY OF LISBON STUDY OF THE HEALTH EFFECTS OF DENTAL AMALGAM IN CHILDREN

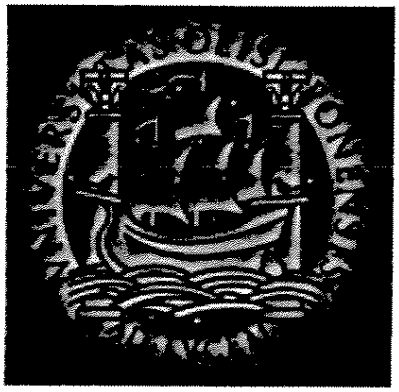
RECEIVED
Human Subjects Division
OCT 0 6 2004
UW

REPORT TO THE DATA AND SAFETY MONITORING BOARD

RECEIVED
Human Subjects Division
OCT 0 6 2004
UW

OCTOBER 22, 2004

UW School of Dentistry
University of Washington



Date 11/3/04
Modification to 95-0401-A/C 09
Approved by Human Subjects Review Committee
Signature

RECEIVED
Human Subjects Division
OCT 0 6 2004
UW

Table of Contents

Section ..... **Page**
1. Introduction 1
   1.1 Summary of Progress Reported Previously 2
   1.2 Site Visits 2
   1.3 Recent Publications 2
   1.4 Adjustments to Study Procedures 3
   1.5 Protocol Violations 3
2. Follow-Up Visits 3
   2.1 Subject Retention 3
   2.2 Timing of follow-up visits 4
3. Dental Treatment and Mercury Exposure 6
   3.1 Dental treatment 6
   3.2 Mercury exposure 8
4. Follow-up Results on Primary Outcome Variables 9
   4.1 Correlations between child and adult tests 9
   4.2 Monitoring changes in performance of individual children 14
   4.3 Statistical Stopping Rule 15
   4.4 Reassessment of Power 16
5. Follow-up Results on Secondary Outcome Variables 17
   5.1. Sentinel health events and health history 17
   5.2 Neurological 17
   5.3 Oral exam 20
   5.4 Laboratory Measures 23
      5.4.1 Urinary Samples 23
      5.4.2 Monitoring of Children with High Urinary Mercury Values 23
      5.4.3 Monitoring Children with Elevated Blood Lead Values 24
6. Quality Assurance Monitoring 27
   6.1 Completeness of data 27
   6.2 Accuracy of data 27
   6.3 Examiner reliability and accuracy 28
      6.3.1 Checklist evaluation 28
      6.3.2 Differences between examiners 29
7. Summary 31

Appendix A. Abstracts and Papers from the Main Study
Appendix B.1. Data Tables for Children with High Mercury Exposure Values
Appendix B.2. Data Tables for Children with Sentinel Health Events
Appendix B.3. Data Tables for Children with Elevated Urinary Mercury Levels
Appendix B.4. Data Tables for Children with Elevated Baseline Blood Lead Levels

1. Introduction

The Casa Pia Study of the Health Effects of Dental Amalgam in Children, an arm of the Children's Amalgam Trial, has entered its last follow-up year of data collection. Although the study has weathered several unusual events during the last two years, the trial remains scientifically on track, and subject retention continues to be high.

In this report we summarize the progress made on the trial since the last DSMB meeting held in December of 2003. Results are reported in a cumulative fashion by including data from baseline and all follow-ups to avoid as much as possible the need to refer to our previous reports. Data included in this report are current as of mid-September, 2004.

In the DSMB report for December 3, 2003, we reported on the unfortunate pedophilia activities which were uncovered in the Casa Pia schools in late 2002, as well as on a brief series of somewhat unfavorable articles which came out in a Portuguese newspaper, Publico, in November 2003 covering the participation of the Casa Pia children in the Children's Amalgam Trial.

Regarding the pedophilia scandal, the legal process continues but the level of publicity surrounding it has subsided. Some prominent citizens of Lisbon implicated in earlier reports have now been released from custody and presumably cleared of charges. One former high-level administrator in the Casa Pia system (from a period several years prior to initiation of our study) remains in custody, as does the driver who was initially the primary person implicated. As we reported last year, the effects of this scandal on the study seem to be minimal. To our knowledge, only two of the study participants have been directly affected: the one student who initially reported the driver and caused the investigation to be conducted, and one other student who is likely to be a witness because he is accompanied by security guards when visiting the dental school for testing or treatment. Both students continue as participants in the study.

Regarding the somewhat negative publicity in newspaper articles describing the amalgam controversy and criticizing the participation of Casa Pia students in the study, the University of Lisbon study team and administration responded immediately to these articles by holding a press conference and by contacting the officials of the Casa Pia school system. In addition to the press conference and meetings, a comprehensive compendium of materials covering these issues was compiled by the University of Lisbon study team, and provided to Casa Pia officials as well as to the IRB of the University of Lisbon, and to NIDCR. The articles and controversy regarding the study disappeared after the first week. Because of this immediate and thorough response by the University of Lisbon, there has been no discernable effect on the status of the study. There are some challenges to maintaining the high retention levels we have experienced up to now, but there is little indication that these challenges are due to or influenced by the publicity surrounding the pedophilia activities or the amalgam articles.

During the past year it was necessary to go through a re-consent process due to the extension of the study to 7 years of follow-up, a situation with which the DSMB is familiar. We reported last year that we had data on 429 children for follow-up year 5. As a result of the re-consent process, we have obtained re-consent for 400 subjects to date, so that is the number who are currently officially eligible to continue in the study for the sixth and seventh year of follow-up (down from the original 507). This number does not include four participants who, in a protocol violation, had some test data obtained after receiving a verbal promise from the family to mail in the re-consent, but prior to receipt of the signed form (this is addressed in the section on Protocol Violations, 1.5). Therefore, as the report is read, it is important to remember that our denominator for the last 2 years of data collection is now 400.

At this time, the biggest hurdles we face are involved with subject retention, in that most of the subjects are now late adolescent/young adults, with all of the personality "challenges" that accompany this stage of life. This makes it more difficult to assure continued participation in the study, or at the very least to

Additionally, an abstract has been submitted for the 2005 IADR: Martin MD, Benton T, Leroux B, Bernardo MF, Soares H, Leitao J, DeRouen TA. Childhood Lead Exposure and Dental Caries. A copy is appended.

1.4 Adjustments to Study Procedures

For the seventh (and last) follow-up year of testing, the test battery changed to include the short computerized version of the Wisconsin Card Sorting Test and the Wechsler Abbreviated Scale of Intelligence (WASI). The purpose of adding these tests is to provide data for purposes of comparison with the NERI study. Although we have been administering one component of the WASI all along, we need to add 3 additional components, and this adds between 30-40 minutes to the overall testing time.

1.5 Procotol Violations

There was a protocol violation involving four subjects near the end of the re-consent process. The violation involved the administration of some of the study procedures prior to receipt of the signed re-consent form for four participants whose families had given a verbal promise to mail in the signed re-consent form. Those cases were put on hold once the error was realized. No further testing of them will occur until there is resolution of the re-consent process. Phone follow-up with the parents of the four participants again confirmed their re-consent, along with their promise to mail in the signed re-consent form. If the signed re-consent forms are not received, the partial data collected for year 6 for those subjects will be deleted

The Lisbon team has confirmed that in the future no testing for years 6 and 7 will proceed for a participant without a signed re-consent form in hand. This will not be a continuing issue, since we are at the end of the re-consent process. While not an excuse, the logistical difficulties in scheduling participants for testing, along with the slow response by some parents to mail in the re-consent form even though they gave verbal assurances over the phone that they would do so, led to the instances cited. Those participants were scheduled for testing with the understanding that the re-consent forms were "in the mail", and they underwent some testing before it was realized that the re-consent forms had not been received.

2. Follow-Up Visits

2.1. Subject retention

As described above, 400 participants have provided written re-consent to participate in the study beyond the 5th-year follow-up; this number represents 93.2% of the 429 participants who were seen at 5th year, and 78.9% of the original sample of 507 randomized participants. The numbers of children re-consented were similar in the two treatment groups (195 Amalgam, 205 Composite). Not included are the four participants described above in protocol violations of the re-consent process. We hope to confirm their continued participation in the study through receipt of signed re-consent forms very soon.

We have neurobehavioral data in year 6 for 375 participants (Table 2.1.1), including 373 who provided re-consent and 2 of the four children for whom verbal consent only was given. Of the remaining 27 participants who provided re-consent, 1 completed year 6 testing but the data have not yet been entered, 5 are still expected to complete year 6 testing, 18 missed the time window for year 6 but are expected to return and complete year 7 testing (including 3 who have already done year 7), and 3 refused to do the tests. Based on these figures, we project that the final numbers of participants for neurobehavioral testing will be approximately 380 in years 6 and 7. In power re-calculations in Section 4.4, we use n=375 for years 6 and 7 as conservative estimates.

Smaller numbers (Table 2.1.1) have been seen for nerve conduction velocity (276, 68.3% of the 400 re-consented) and neurological exams (285, 70.5%), for the reasons described above in the introduction. Sixth year NCV and neurological exam data on an additional 30 participants have been collected or will be obtained in the next two weeks, but are not included in these tables. Participation rates continue to be nearly identical in the two treatment groups.

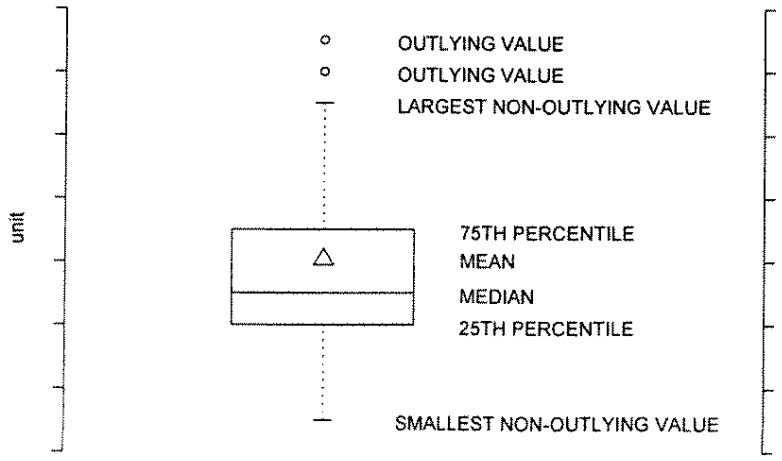
Table 2.1.1. Number of subjects participating in data collection visits by follow-up year.

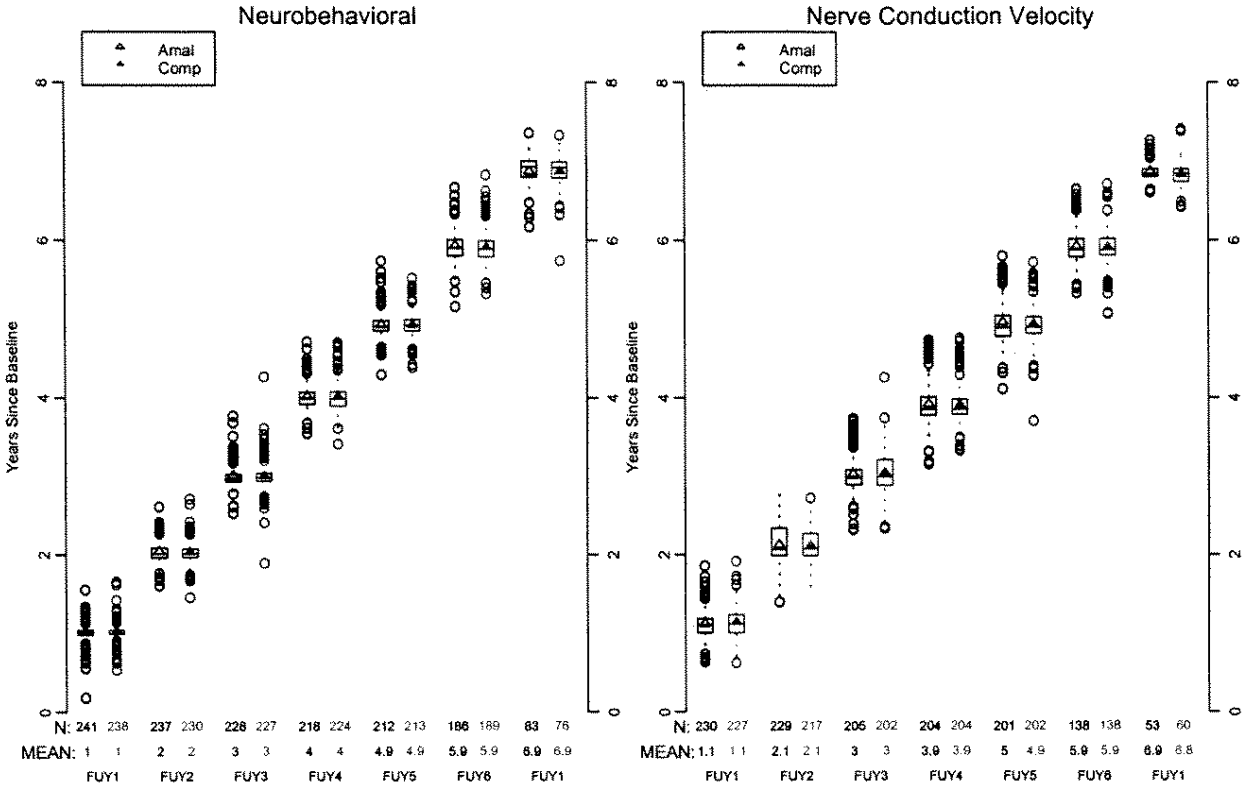
Data Type	Number with Data At Follow-up Years:						
	1	2	3	4	5	6	7
Neurobehavioral	479	467	455	442	425	375	159
Nerve Conduction Velocity	457	446	407	408	403	276	113
Neurological	467	453	383	391	395	285	121
Oral Exam	449	468	448	431	416	357	171
Urinary Mercury	473	456	438	441	412	357	94

2.2. Timing of follow-up visits

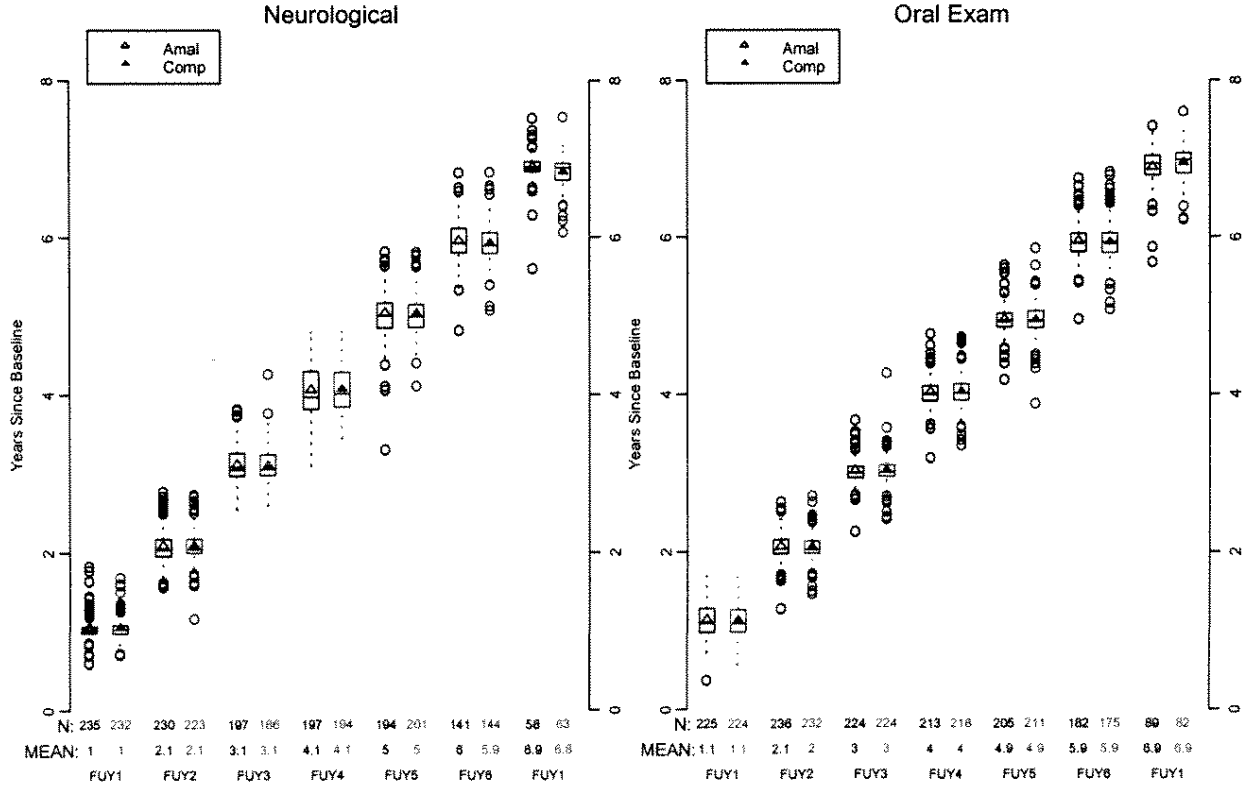
Annual follow-up tests occur at roughly one-year intervals (Fig 2.2.1). Irregular intervals occur in some cases due to rescheduling of missed visits, and for children who are difficult to track because they have left the school system.

Key for Interpretation of Boxplots




Neurobehavioral
Amal
Comp
Years Since Baseline
N: 241 238 237 230 228 227 218 224 212 213 186 189 83 76
MEAN: 1 1 2 2 3 3 4 4 4.9 4.9 5.9 5.9 6.9 6.9
FUY1 FUY2 FUY3 FUY4 FUY5 FUY6 FUY1

Nerve Conduction Velocity
Amal
Comp
Years Since Baseline
N: 230 227 229 217 205 202 204 204 201 202 138 138 53 60
MEAN: 1.1 1.1 2.1 2.1 3 3 3.9 3.9 5 4.9 5.9 5.9 6.9 6.8
FUY1 FUY2 FUY3 FUY4 FUY5 FUY6 FUY1


Neurological
Amal
Comp
Years Since Baseline
N: 235 232 230 223 197 186 197 194 194 201 141 144 58 63
MEAN: 1 1 2.1 2.1 3.1 3.1 4.1 4.1 5 5 6 5.9 6.9 6.8
FUY1 FUY2 FUY3 FUY4 FUY5 FUY6 FUY1

Oral Exam
Amal
Comp
Years Since Baseline
N: 225 224 236 232 224 224 213 216 205 211 182 175 89 82
MEAN: 1.1 1.1 2.1 2 3 3 4 4 4.9 4.9 5.9 5.9 6.9 6.9
FUY1 FUY2 FUY3 FUY4 FUY5 FUY6 FUY1

3. Dental Treatment and Mercury Exposure

3.1. Dental treatment

After dropping from an average of approximately 10 surfaces restored per child during the initial treatment period, there has been a trend of increasing restorative treatment in follow-up years 1 through 5 (Figure 3.1). There have been more children treated and more treatment in total for the composite group than the amalgam group. The treatment figures for years 6 and 7 are incomplete because treatment is on-going.

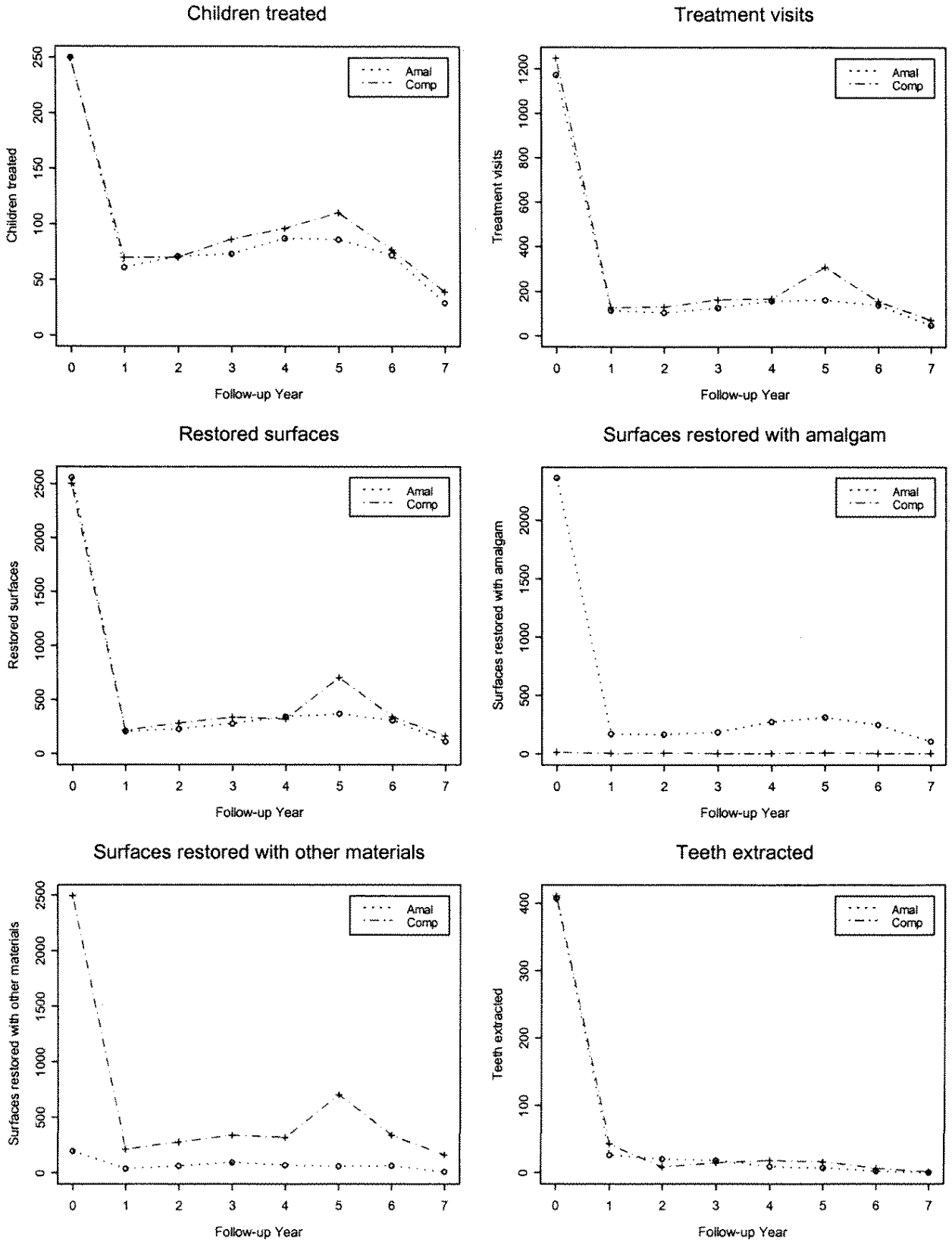


Figure 3.1. Dental treatment for randomized participants, by treatment group and treatment period.

3.2. Mercury exposure

For each participant we compute a measure of mercury exposure (surface-years) that accounts for the number of tooth surfaces restored with amalgam as well as the size and type of surface. For the reasons detailed in last year's report, the exposure measure does not assign different weights to time intervals that are closer to placement than later times.

To date, children in the amalgam group average approximately 50 surface-years of exposure (Fig 3.2.1) with a maximum of 252.7 surface years in one child. As reported previously, two children in the composite group received mercury fillings by mistake. We are monitoring the 13 children whose mercury exposure values are in the top 5% in the amalgam group (at least 117.5 surface-years). Data tables for these children are provided in Appendix B.1.

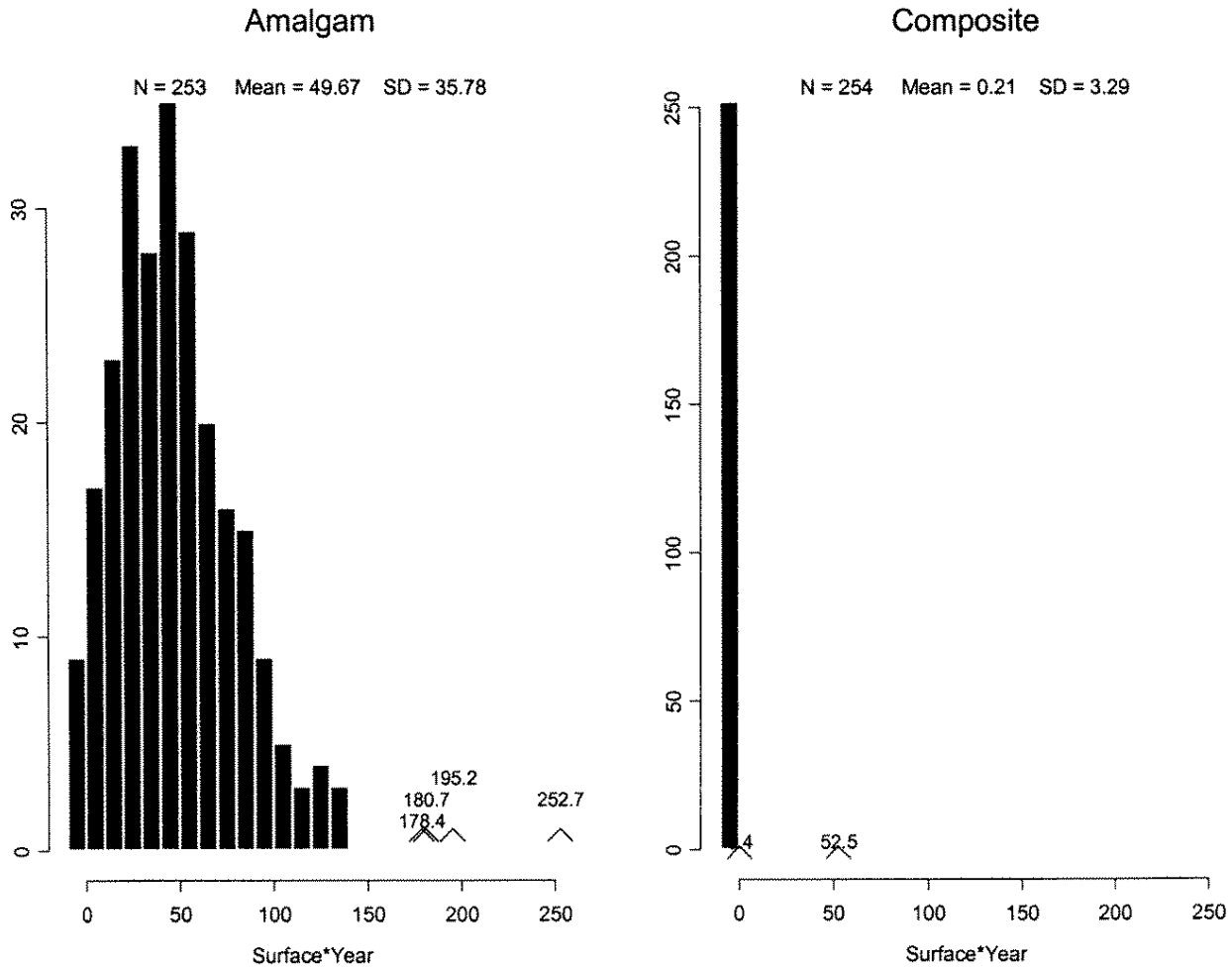


Figure 3.2.1. Distribution of the mercury exposure measure by treatment group.

4. Follow-up Results on Primary Outcome Variables

4.1. Distribution of primary outcome variables by follow-up period

Three of the primary outcomes for the trial are Z-scores for memory, visual-motor, and attention domains. The distributions of the primary outcomes continue to be very similar in the two treatment groups as in previous years (Figure 4.1.1). Results for boys (Figure 4.1.2) and girls (Figure 4.1.3) separately also show no evidence for treatment group differences.

The Z-scores at baseline and years 1, 2, and 3 are based on the original child test battery, whereas Z-scores for years 4 through 7 are based on the current test battery that includes child and adult tests. The justification for the adoption of the new test battery and the discontinuation of the original battery was provided in our previous reports; in particular, results from year 4 showed that corresponding child and adult tests were moderately positively correlated (correlations ranged from 0.25 to 0.80), and Z-scores based on the child and adult test batteries were highly correlated (correlations 0.76, 0.97, and 0.93 for memory, visual-motor, and attention domains, respectively). In addition, group comparisons based on Z-scores using the adult test battery were very similar to comparisons based on the original child test battery.

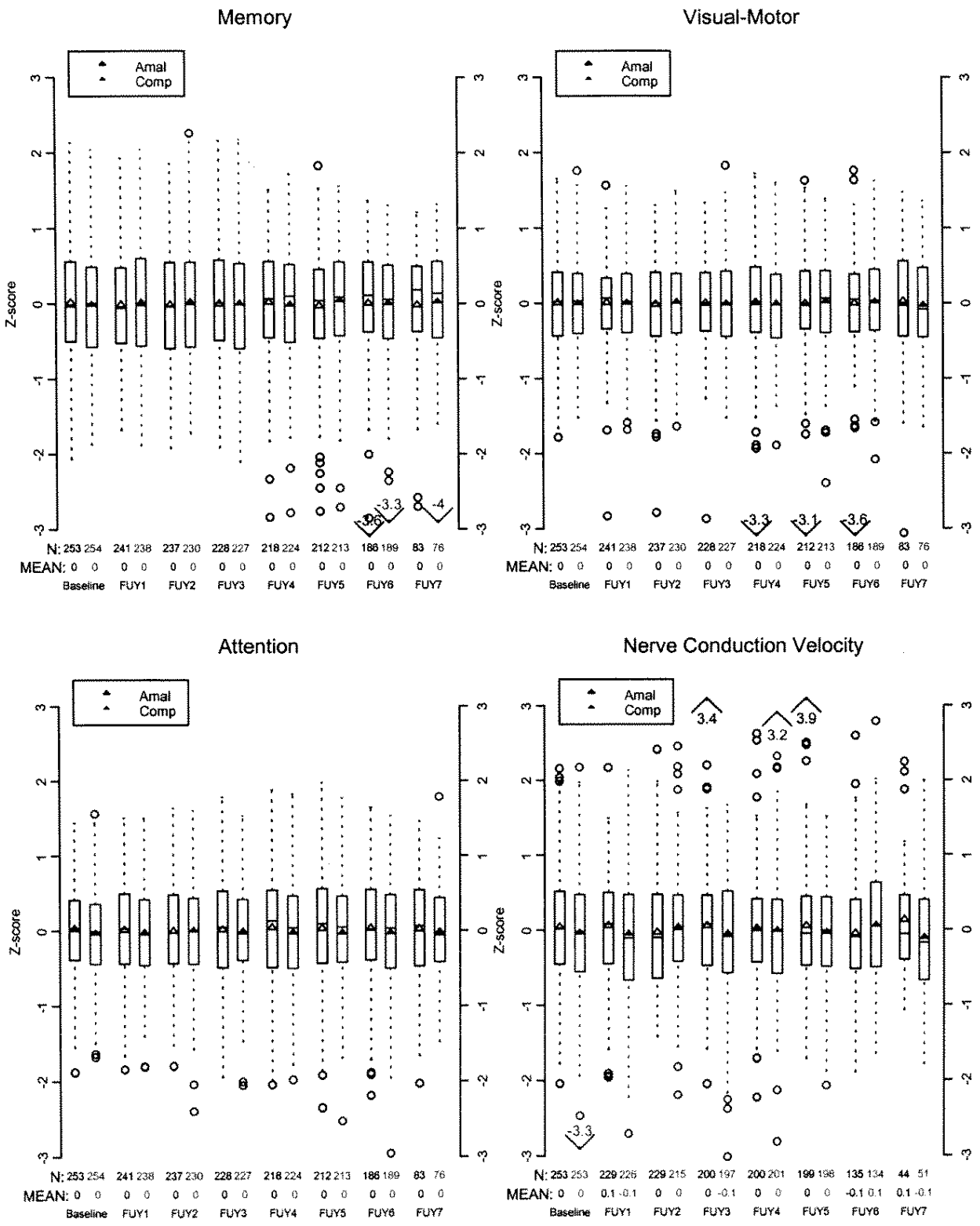


Figure 4.1.1 Distribution of Z-scores for neurobehavioral domains (memory, visual motor, and attention) and nerve conduction velocity by time period and treatment group.

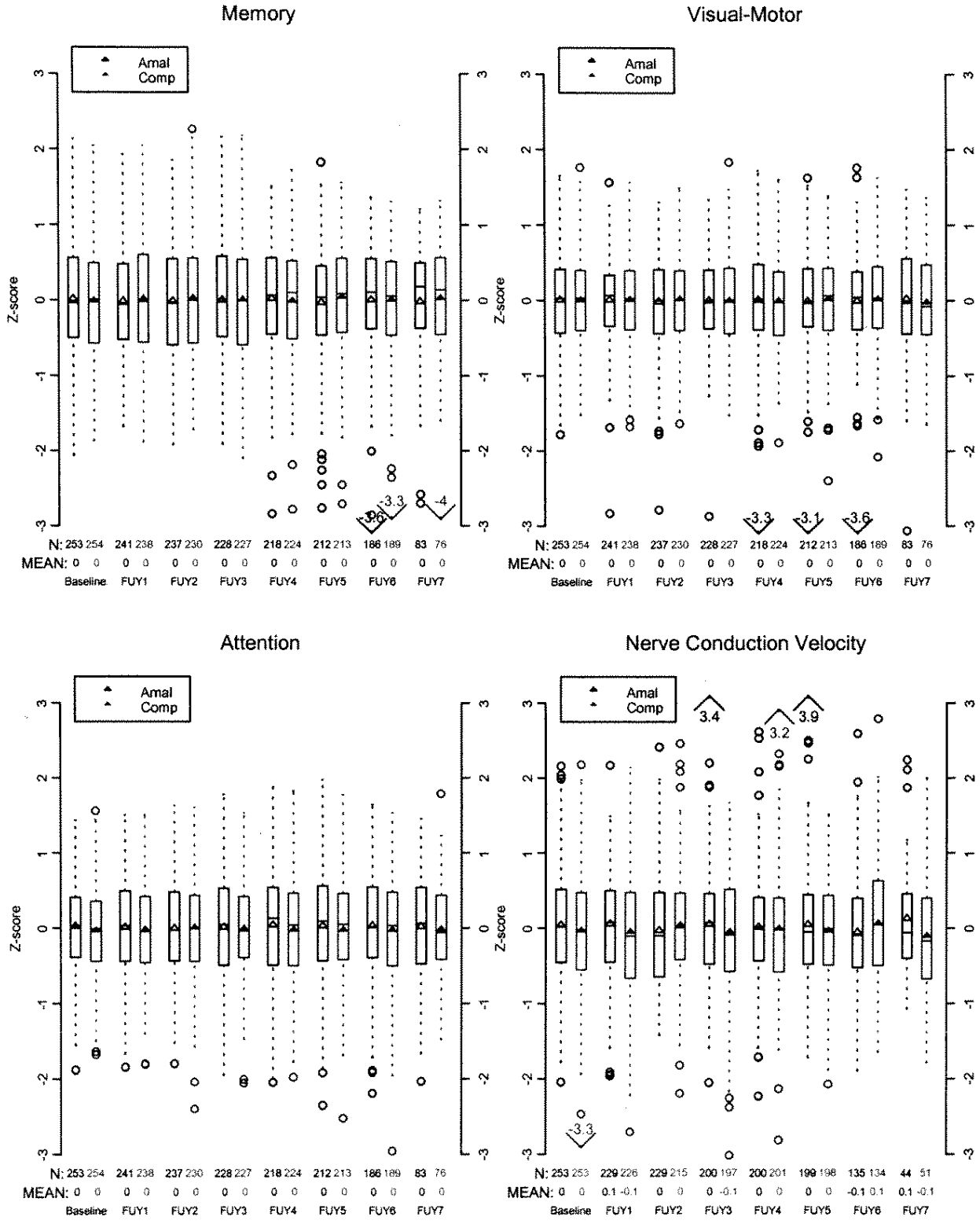


Figure 4.1.1 Distribution of Z-scores for neurobehavioral domains (memory, visual motor, and attention) and nerve conduction velocity by time period and treatment group.

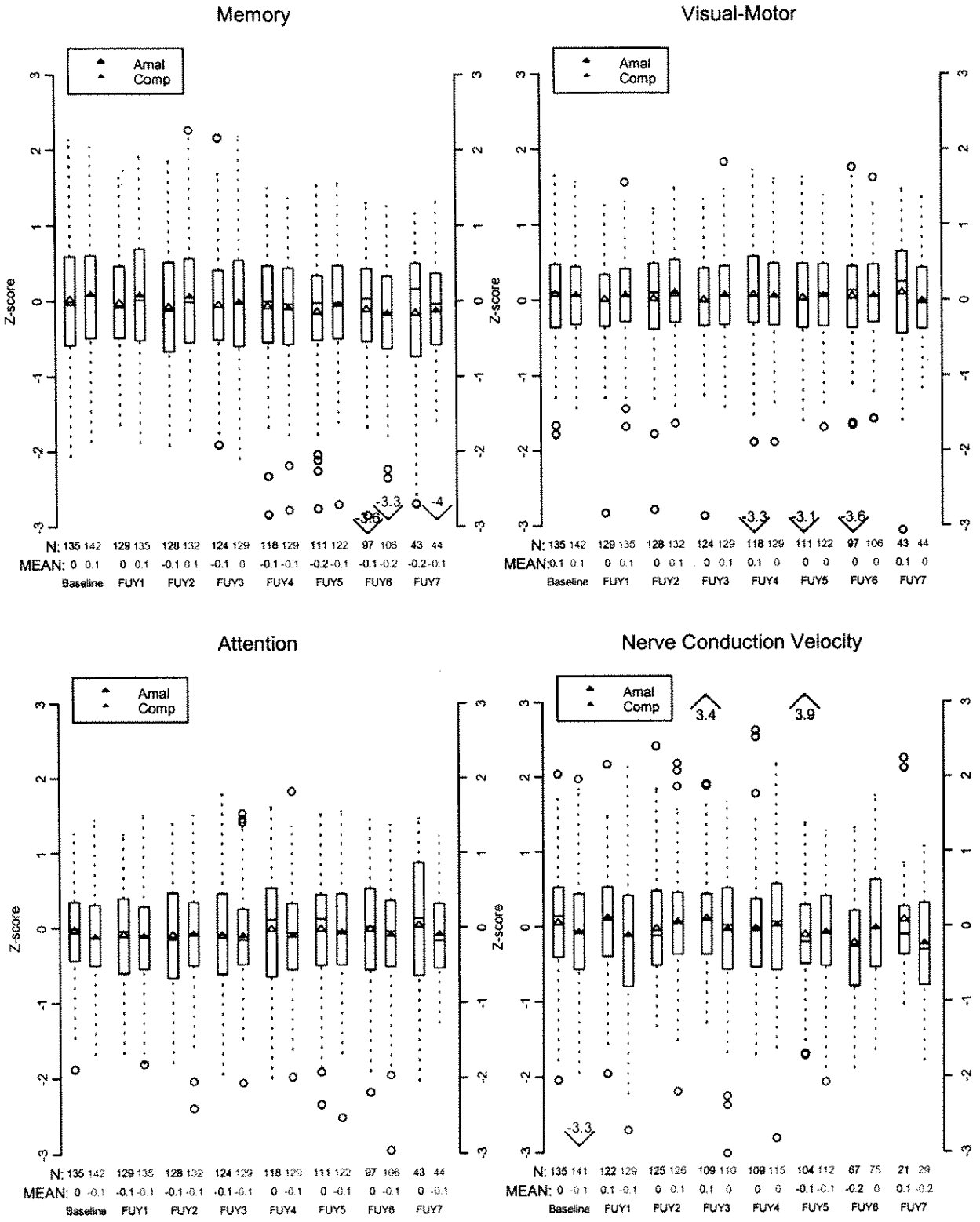


Figure 4.1.2. Distribution of Z-scores for neurobehavioral domains (memory, visual motor, and attention) and nerve conduction velocity by time period and treatment group for male participants.

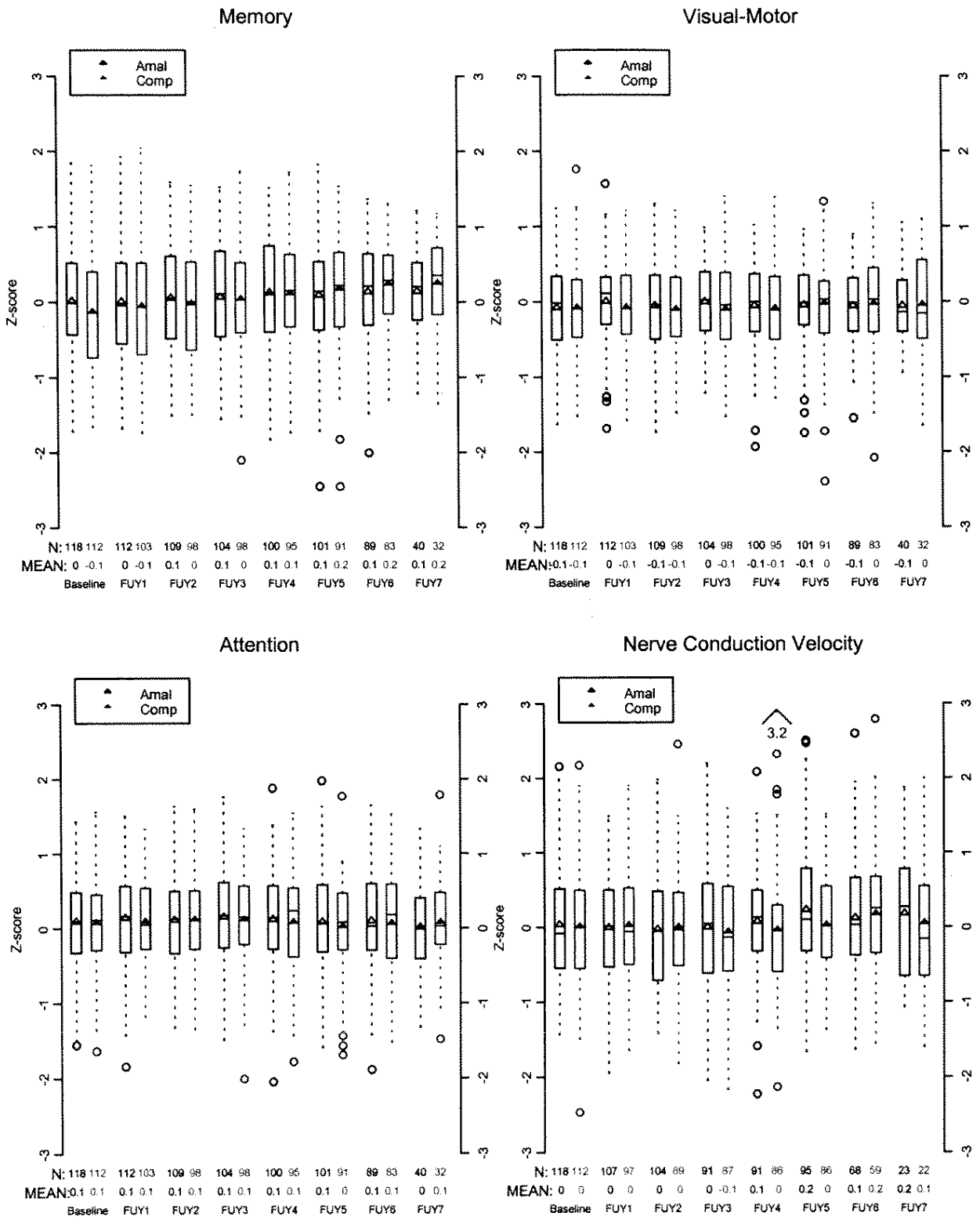


Figure 4.1.3. Distribution of Z-scores for neurobehavioral domains (memory, visual motor, and attention) and nerve conduction velocity by time period and treatment group for female participants.

Although the primary endpoints for the study are the four Z-scores, we have also performed analyses to compare the treatment groups on the individual neurobehavioral test items at each time period. There was one treatment group comparison significant at the 0.05 level (Table 4.1.2, Row 4), but this is interpreted as most likely a type I error, given the large number of tests performed (38).

Table 4.1.2. Comparison of treatment groups on individual neurobehavioral test scores at 5th year follow-up for girls and boys. The domain to which each score belongs is denoted as <M> (memory), <V> (visual-motor), or <A> (attention). Higher scores are better except for items with units of seconds, as indicated.

Variable<Domain>	Girls					Boys				
	Group A		Group B		P-value	Group A		Group B		P-value
	Mean	Std Dev	Mean	Std Dev		Mean	Std Dev	Mean	Std Dev	
RAVLT T.8 <M>	10.22	2.53	10.37	2.42	0.694	9.64	2.78	9.23	2.71	0.291
RAVLT TOT <M>	47.92	7.52	49.51	7.35	0.164	45.45	8.99	44.94	8.40	0.676
WRAMA Peg - dominant <V>	124.45	14.61	123.14	17.18	0.591	113.53	19.81	116.26	15.76	0.275
WRAMA Peg - non-dominant <V>	120.80	12.27	119.43	15.88	0.528	114.72	16.95	119.33	14.90	0.041
Standard Reaction Time (sec) <V>	0.80	0.16	0.79	0.17	0.638	0.77	0.18	0.77	0.15	0.962
Finger Tapping - dominant <V>	47.04	4.94	48.40	5.82	0.102	51.80	7.40	51.09	6.11	0.456
Finger Tapping - non-dominant <V>	41.36	4.75	41.39	6.10	0.975	45.39	6.81	44.78	5.96	0.498
STROOP WD4<A>	40.63	8.01	41.61	7.59	0.410	40.51	8.73	40.55	8.24	0.972
STROOP COL4 <A>	42.39	7.49	41.54	7.53	0.459	40.35	8.53	39.82	8.20	0.652
STROOP CWD4 <A>	48.04	8.74	46.25	8.06	0.165	45.63	10.56	44.10	9.48	0.280
WMS III - Visual Learning <M>	35.88	4.44	36.19	3.59	0.610	35.11	4.66	35.12	4.23	0.988
WMS III - Visual Delay <M>	33.92	5.34	34.58	4.04	0.367	32.12	6.71	32.11	6.36	0.991
WASI Matrices <V>	23.82	4.27	24.28	4.54	0.498	24.22	5.67	23.31	4.91	0.224
WAIS III - Digit Symbol <A>	72.70	14.43	74.66	14.26	0.370	69.59	17.24	67.81	16.27	0.451
WAIS III - Symbol Search <A>	33.20	8.56	31.80	7.57	0.256	31.55	8.54	29.94	8.26	0.176
WAIS III - Digit Span <A>	13.83	3.32	13.95	3.26	0.811	13.23	3.08	13.99	3.95	0.128
WMS III - Spatial Span <A>	15.22	2.75	14.96	3.12	0.561	15.35	2.91	15.14	2.71	0.597
Adult - Trails A (sec) <A>	30.29	10.93	30.25	9.63	0.980	30.18	12.78	30.21	12.12	0.985
Adult - Trails B (sec) <A>	67.60	27.53	66.58	24.61	0.798	69.26	28.20	76.11	33.35	0.117

4.2. Monitoring changes in performance of individual children

Step 1: Detection of Outliers

A total of 118 children were found to have decreases according to either the RCI method or the Z-score method (Table 4.2.1). The large number of children with outliers can be explained partly by the large number of comparisons that are done as part of our outlier detection method (i.e., comparison of baseline with year 7, year 1 with year 7, year 2 with year 7, etc.). Most of these children (91 of 118) were identified by the Z-score method only (and not by the RCI method), whereas only 27 were identified by the RCI method, including 18 who were identified by both methods. Approximately half (62 of 118) of the children identified this year had also been identified last year. Six of the children who were of clinical concern last year were identified as outliers in this year.

Table 4.2.1. Number of children identified as having decreases in neurobehavioral test performance in Year 6 and comparison with children identified in 2003.

Method	Overall	Identified in 2003	Clinical Concern in 2003
Z score decrease only	91	45	2
RCI decrease only	9	4	0
Z score & RCI decrease	18	13	4
Z score or RCI decrease	118	62	6
No decrease	253	59	9
Not tested	136	21	2
Total	**507**	**142**	**17**

The children identified as outliers do not differ appreciably from the children not identified or from the children who were not tested this year on treatment group membership, age, baseline blood lead, mercury exposure, or urinary mercury concentrations (Table 4.2.2). Note that approximately half of the children identified as outliers by any of the methods are in the amalgam group. A possible exception is that, of children identified by the RCI method only, 66.7% (6 of 9) are in the amalgam group (Table 4.2.2, column 3), but these numbers are too small to represent a meaningful association. There is an association with gender as we have noted in the past: boys are slightly more likely to be identified as outliers than girls.

Table 4.2.2. Comparison of children identified as having decreases in test performance with other children.

Variable	Z only	RCI only	Z & RCI	No Decrease	Not tested (NB)
N	91	9	18	253	136
Group A	44	6	8	126	69
%Group A	48.4	66.7	44.4	49.8	50.7
%Male	62.6	55.6	61.1	51	55.1
Mean Age	16	15.8	16.1	16.1	16.1
Baseline Blood Pb	4.8	4.2	5.8	4.6	4.4
Mean Hg Exposure (surface-years)	29.7	40	24.2	27.9	15.3
Baseline Mean Urinary Hg	2	2.5	1.7	2	1.5
Year 1	2	2.1	1.6	2.1	1.7
Year 2	2.5	2.3	1.8	2.5	2.3
Year 3	2.6	5.9	3.9	2	2.3
Year 4	1.7	0.9	1.6	1.8	1.8
Year 5	1.3	1.1	2.3	1.4	1.5
Year 6	1.4	3	0.6	1.2	3.9

Step 2: Clinical Assessment

The records of the 118 children identified as having significant RCI and/or Z-score decreases were examined by Dr. Townes. Of these, Dr. Townes identified 6 children of clinical concern because of neurobehavioral decline, 3 with neurologic decline, 3 with neurologic and emotional decline and 8 with emotional concerns. Emotional concern was suggested if the child endorsed 4 out of 8 of the scales on the VAMS in the pathological direction (T equal to or greater than 68).

Throughout the year all of the children's neurobehavioral profiles were examined by Dr. Townes. Of those children not on the RCI/Z-score list, one child was judged to have possible neurologic and emotional decline, 7 possible neurologic decline and 11 possible emotional decline. An additional three children were considered at risk by Casa Pia staff.

On 28/29 September 2004, Dr. Pavao-Martins, Helena Brunheta, and Dr. Townes met to review the 42 identified children (20 with RCI and/or Z-score decreases and 22 without). Only five children could be scheduled for examination; of those, one did not appear and two cancelled. Neurologic examinations were performed on two children. The study records and previous examinations were discussed for the remaining children. Recommendations for further clinical assessments/follow-up were made for a small number of the children discussed.

The potential neurologic declines could be accounted for in most cases by differences in scoring neurologic soft signs between the neurologists evaluating the children. The emotional concerns of the children probably are typical of most teenagers across cultures. Finally, there was no consistent pattern of neurobehavioral decline observed among this group of 42 children. Data tables for these children will be available at the DSMB meeting.

4.3 Statistical Stopping Rule

The primary analysis for the study is based on the four primary outcomes (memory, visual motor, and attention domain scores from the neurobehavioral testing, and a combined nerve conduction velocity score). Two separate tests are performed: an O'Brien type test that will detect small group differences that are similar across the four endpoints and a Hotelling T^2 test that will detect inconsistent group differences, e.g., a difference in just one of the variables. The overall significance level of the test accounts for the overlap between these two tests (i.e., it is possible that both tests reject the null hypothesis). The O'Brien test statistic is a two-sample t-statistic computed using a derived variable that weights the four endpoints according to the inverse of the pooled correlation matrix. This test provides no evidence against the null hypothesis of no group difference: the 1-sided p-value for this test is larger than the required significance levels for evidence of adverse effect of amalgam or composite (Table 4.3.1). The Hotelling test also provides no evidence against the null hypothesis; the 2-sided p-value is larger than the required significance level (Table 4.3.1). The overall conclusion is that there is no evidence for a difference between treatment groups based on data through year 6.

Table 4.3.1. Results of statistical analysis including data through year 6.

Test Component	Test Statistic	P value	Required Significance Levels	
			Hypothesis: Amalgam Harmful	Hypothesis: Composite Harmful
O'Brien	T = 0.11	0.45 (1-sided)	0.0027	0.0008
Hotelling's T^2	F = 0.09	0.99 (2-sided)	0.0020	

4.4 Reassessment of Power

Table 4.4.1 shows the number of children with follow-up at each year including results to date for the current follow-up year (year 6) and the projected number of children for follow-up year 7.

Table 4.4.1. The number of children with follow-up in each year (projected number for year 7).

Follow-up Year	**1**	**2**	**3**	**4**	**5**	**Re-consent obtained**	**6**	**7**
Number	489	476	471	448	429	400*	375*	375^
%	96.4	93.9	92.9	88.4	85.8	78.9	74.0	71.8

* Numbers current as of mid-September.
^ Projected number as explained in Section 2.1.

The power calculations done for the original grant proposal were based on a projection that we would have data on 400 trial participants through 5 years of follow-up. Our actual follow-up experience through year 5 exceeded that projection (429 participants), and follow-up was extended to 7 years. Thus, the power is higher than estimated in our original power calculations. However, the power is not as high as previously, when we had projected approximately 400 participants with 7 years of follow-up, whereas we now expect approximately 375. The reduction in sample size from 400 to 375 leads to a reduction in power of only about 2 percentage points; for instance, the power to detect an adverse effect of amalgam (ratio of mean difference to standard deviation) of 0.3 in the three neurobehavioral domains and equal to 0.15 in nerve conduction velocity is reduced from approximately 90% to 88%. This calculation assumes that all children with neurobehavioral testing are used in the analysis even if their NCV data are not available. If the analysis is restricted to those with NCV data at year 6 (276 participants), the power would be reduced further to 79%. We conclude that it may be necessary to impute values for year 6 NCV using neurobehavioral test scores and previous values of NCV in order to maintain the sample size at 375 or higher. Because of the large number of children with missing NCV at year 6, additional analyses will be conducted using only the three neurobehavioral endpoints to be sure that the results are not adversely affected by the missing NCV data. We conclude that the power is still adequate to meet the original aims of the study.

5. Follow-up Results on Secondary Outcome Variables

5.1. Sentinel health events

A total of seven sentinel health events have occurred, including six reported previously (a stroke, two cases of epilepsy, a case of asthma, a death due to hepatitis, a death due to an accidental gunshot), and one new event this year, a subject in the composite group was reported to have a thyroid nodule with hyperthyroidism. This subject is being treated by physicians in Lisbon for this ailment. Although there is no evidence in the scientific or medical literature linking thyroid disease with low-level mercury exposure, an Adverse Event form was completed and submitted to the IRB at the University of Washington. This form was also submitted to NIDCR and to the DSMB, along with an accompanying review of the relevant literature.

Three of the first six events occurred in the Amalgam group and three in the Composite group. The available data for all seven children are given in Appendix B.2.

5.2. Neurological Exams

Prevalences of neurological abnormalities are highly variable from year to year due to the small numbers of events, but no consistent treatment group differences are apparent (Figure 5.1.2). In year 7, the prevalence of abnormalities in the cranial nerve exam is much higher in the composite group (8%) than the amalgam group (0%), but these results are preliminary and should not be overinterpreted because of the small sample sizes for year 7. The downward trend in the prevalences of soft signs has continued (Figure 5.1.3). The only consistent pattern in the data on soft signs is the higher prevalence of mirror movement soft signs exhibited for the composite group compared with the amalgam group.

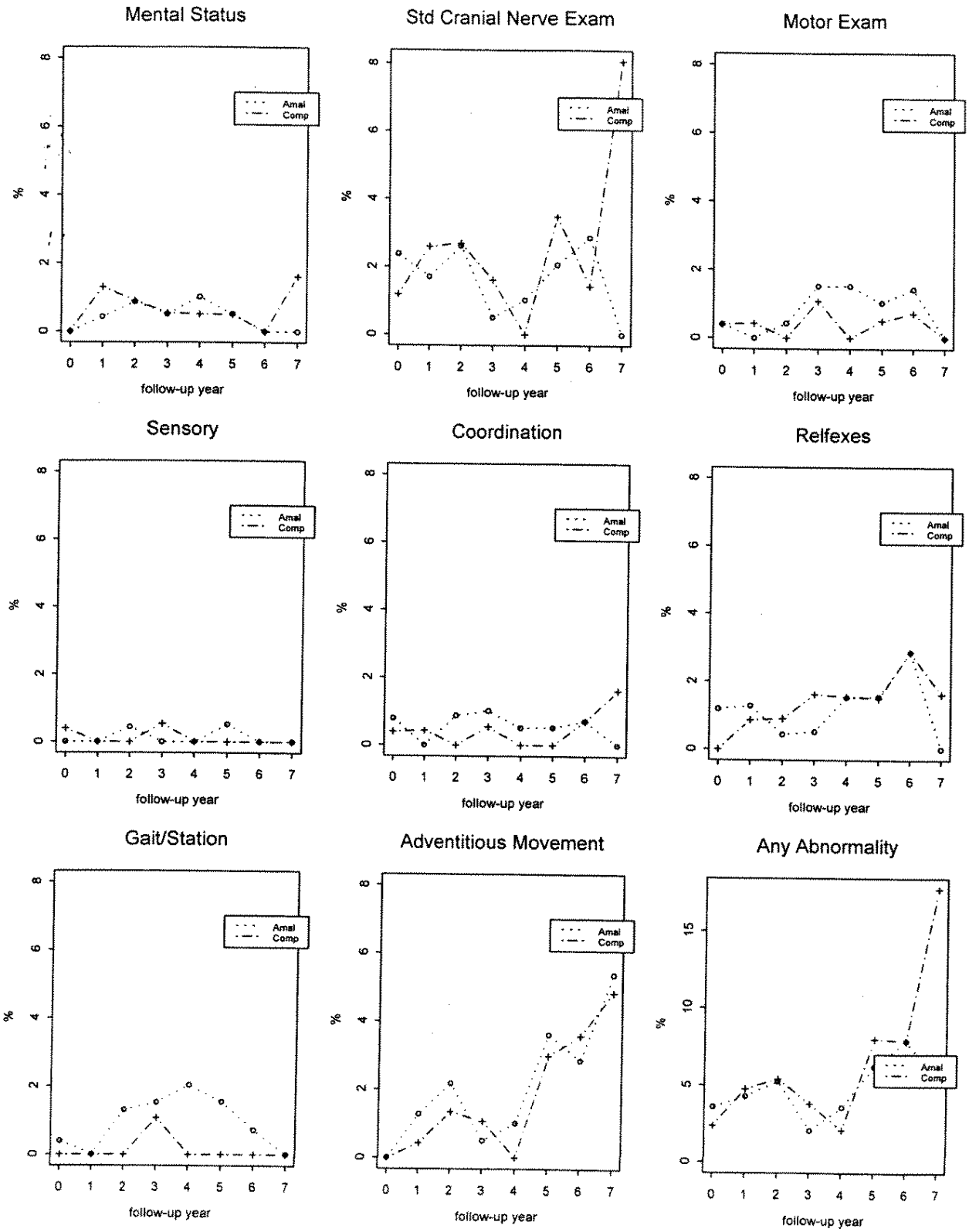


Figure 5.2.1. Prevalence of neurological abnormalities by study year and treatment group (N is the number of children with positive response).

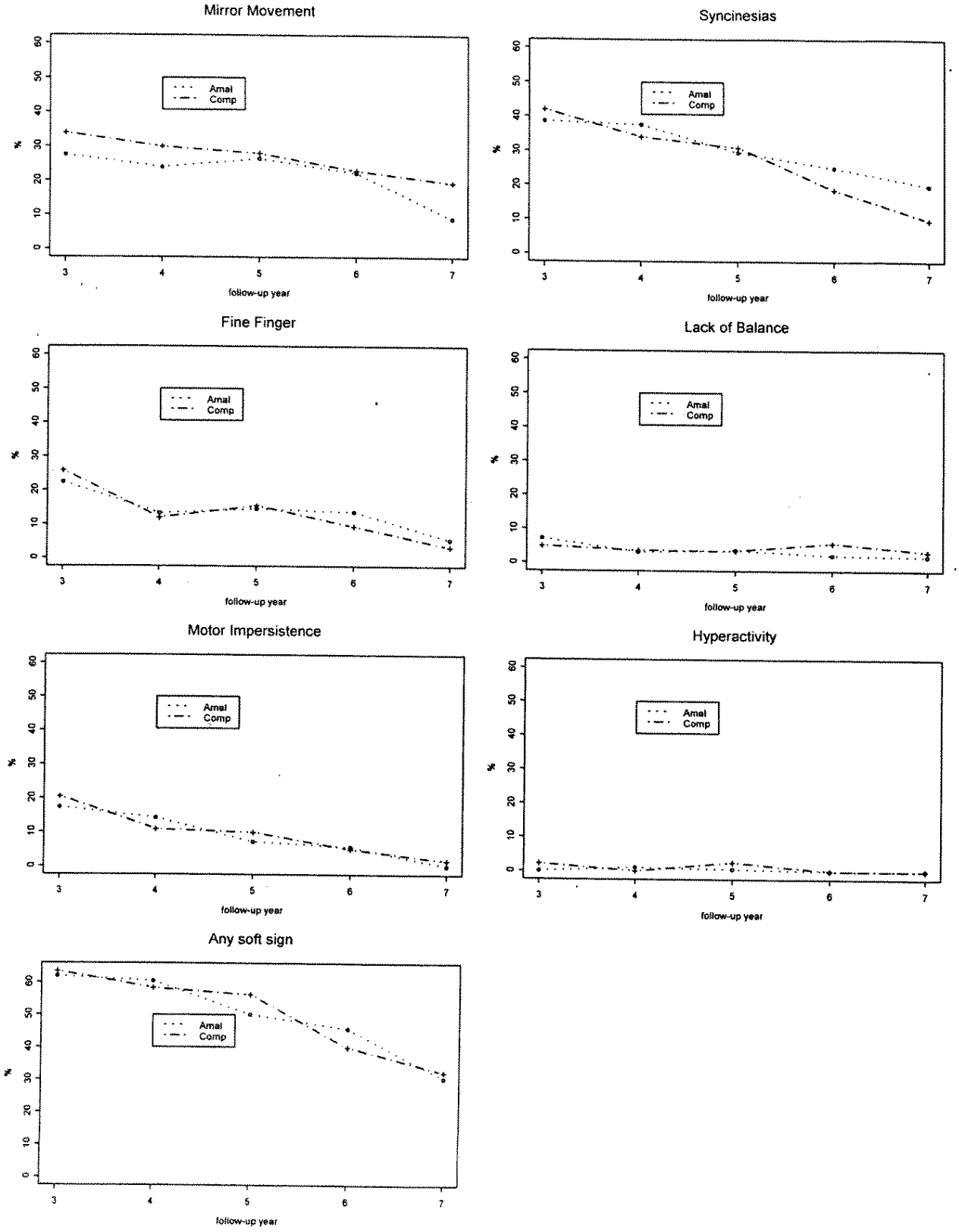


Figure 5.2.2. Prevalence of neurological soft signs by study year and treatment group.

5.3.Oral exam

The number of carious surfaces observed in oral exams has remained fairly constant throughout follow-up, after declining sharply from baseline (Figure 5.3.1). There has been slightly more caries observed during follow-up in the composite group than in the amalgam group. The group difference is particularly noteworthy for recurrent caries, for which the caries rate is approximately 3 times as high for the composite group than the amalgam group; however, in absolute terms the difference is quite small (about 0.1 vs 0.3 surfaces per child). The participants in both groups have achieved adult dentition status (Figure 5.3.2).

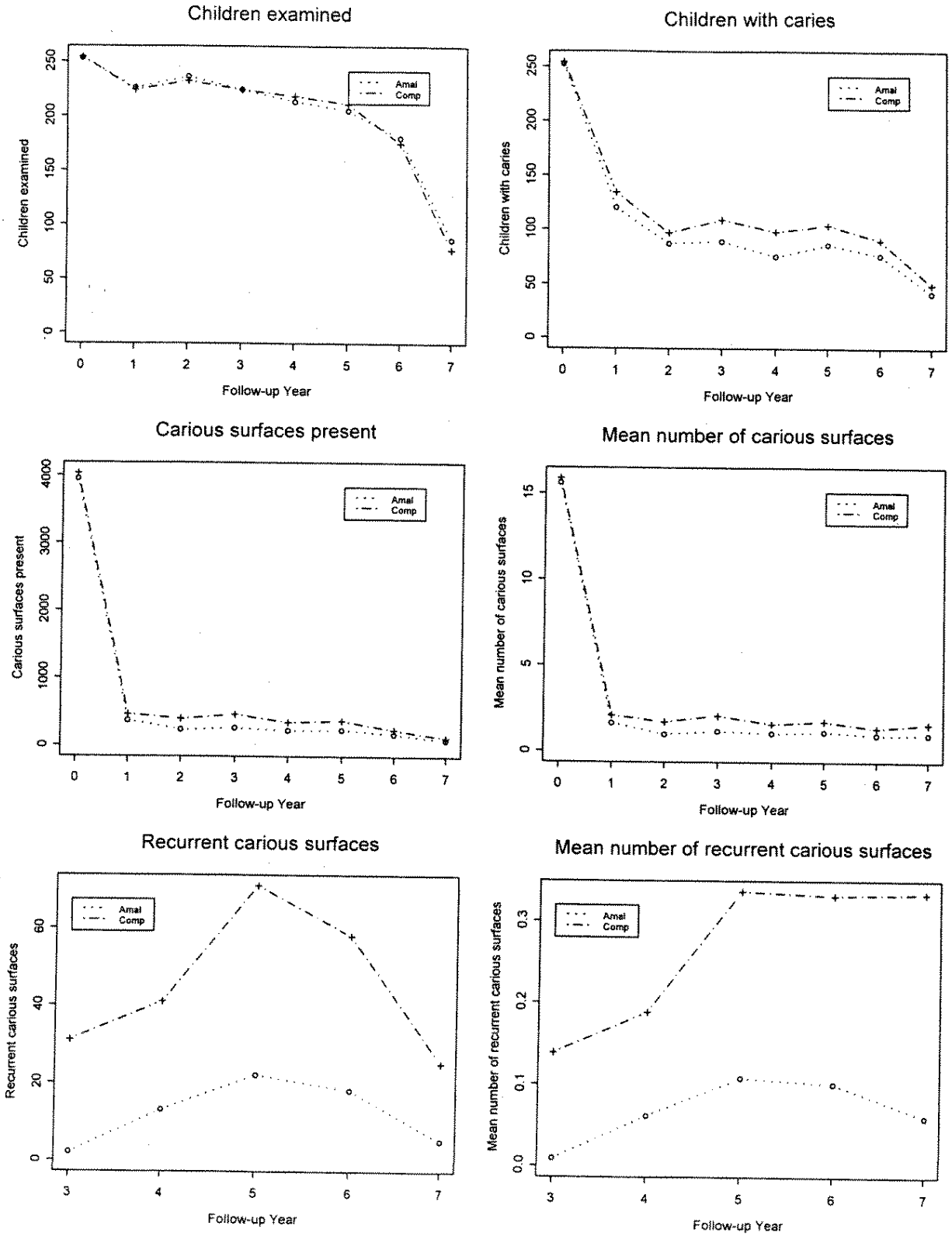


Figure 5.3.1. Results of caries exams by study year and treatment group. Note that recurrent caries was recorded starting in year 3 (bottom two panels).

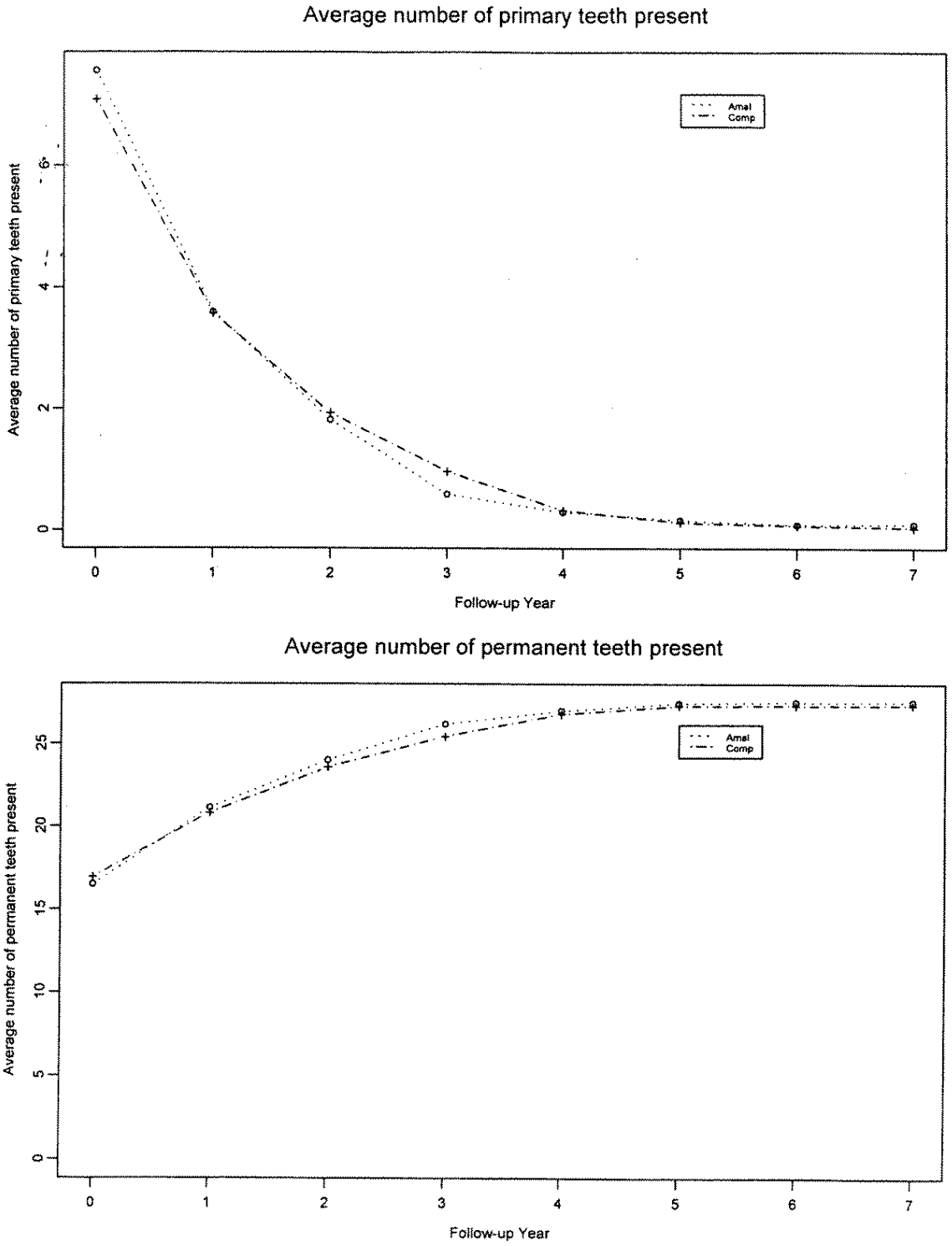


Figure 5.3.2. Number of primary and permanent teeth present by study year and treatment group.

5.4. Laboratory Measures

5.4.1. Urinary Samples

Average creatinine-adjusted urinary mercury levels in the amalgam group achieved a peak of approximately 3 ug/L in year 2 and then declined almost to the baseline level (Figure 5.4.1, top-left panel). In the composite group, average levels have dropped substantially from baseline. The declines in adjusted mercury values can be explained in part by an increase in average creatinine values throughout follow-up (Figure 5.4.2, bottom-right panel). (Creatinine values from year 7 have not been analyzed yet, thus only lab results up to year 6 are included here.)

In both groups, the levels of GST-alpha have increased slightly and the levels of GST-pi show a clear decreasing trend (Fig 5.4.1). The average levels of the GST parameters are similar in the two treatment groups and the differences are not statistically significant. Albumin levels (Fig 5.4.1) do not show any clear trends over time or systematic differences between groups. The levels of three porphyrins thought to be related to mercury exposure are displayed in Figure 5.4.2. Average levels of these porphyrins increased from baseline to first year ($p < 0.05$) and then declined slightly in later follow-ups (Fig. 5.4.2). Average levels of the porphyrins are slightly higher for the amalgam group than the composite group but these differences are not statistically significant. Levels of uroporphyrin, heptaporphyrin, and hexaporphyrin exhibit the same time trends as illustrated in Fig 5.4.2 but no differences between treatment groups (data not shown); these latter three porphyrins are not considered sensitive to mercury exposure. Creatinine levels have increased over time in both treatment groups, with no apparent systematic differences between the two groups (Fig. 5.4.2).

5.4.2. Monitoring of Children with High Urinary Mercury Values

There were 15 children with a high urinary mercury level during the past year (13 values above 10 ug/L and two values slightly under 10). In one case, a second mercury analysis gave a low value (2.27) and no further follow-up was done for this case. In 10 cases, the second analysis gave a value above 10 ug/L, and in 4 cases, the second analysis result is not yet available. Two of the mercury values were extremely high (288 and 277) and remained very high after re-analysis (253 and 266, respectively). Third analyses were performed on these samples and gave results 110 and 101, respectively. For one child a second sample has been obtained and this sample gave a mercury value of 11.08; for the other child the second sample has not yet been obtained.

Data tables for the 14 children with a confirmed high value or whose 2nd value is not available are given in Appendix B.3. Two of these children have high mercury exposure scores (Section 3.2) and one child (#1121) is of concern based on psychological test scores (Section 4.2). Prior to this year there had been 12 children (8 in the amalgam group and 4 in the composite group) with high urinary mercury values, including one child with a high mercury exposure score.

5.4.3. Monitoring Children with Elevated Blood Lead Values

We have continued to monitor the outcome data for the twelve children who had blood lead values larger than 10ug/dl at baseline (Appendix B.4). None of these children has had a high mercury value (5.4.2), or is of concern based on neuropsychological test scores (4.2). We will continue to monitor all 12 of these children in future follow-ups.

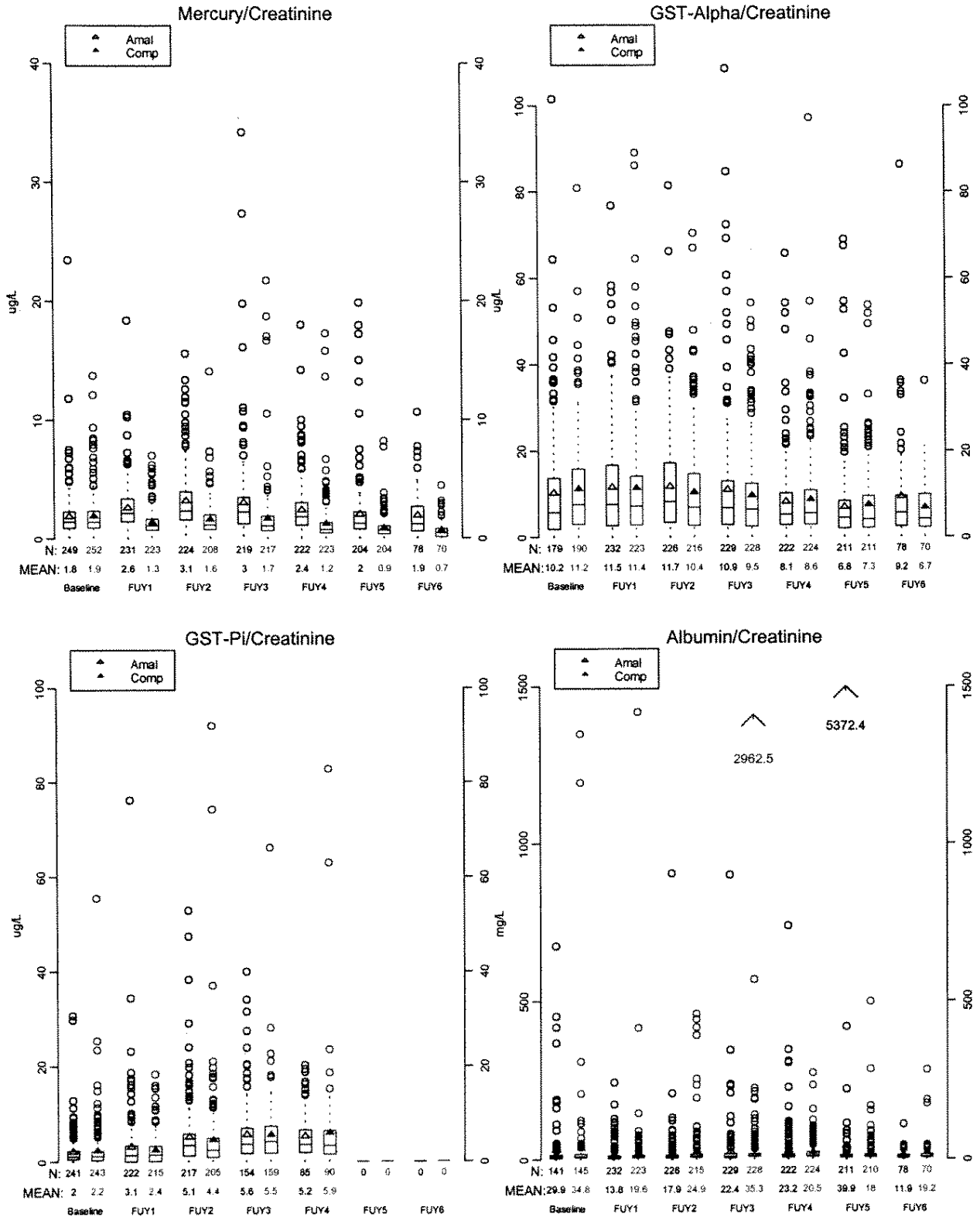


Figure 5.4.1. Distribution of creatinine-adjusted values of urinary mercury, GST-alpha, GST-pi, and albumin by time period and treatment group. (Plots of unadjusted values will be available at the meeting.)

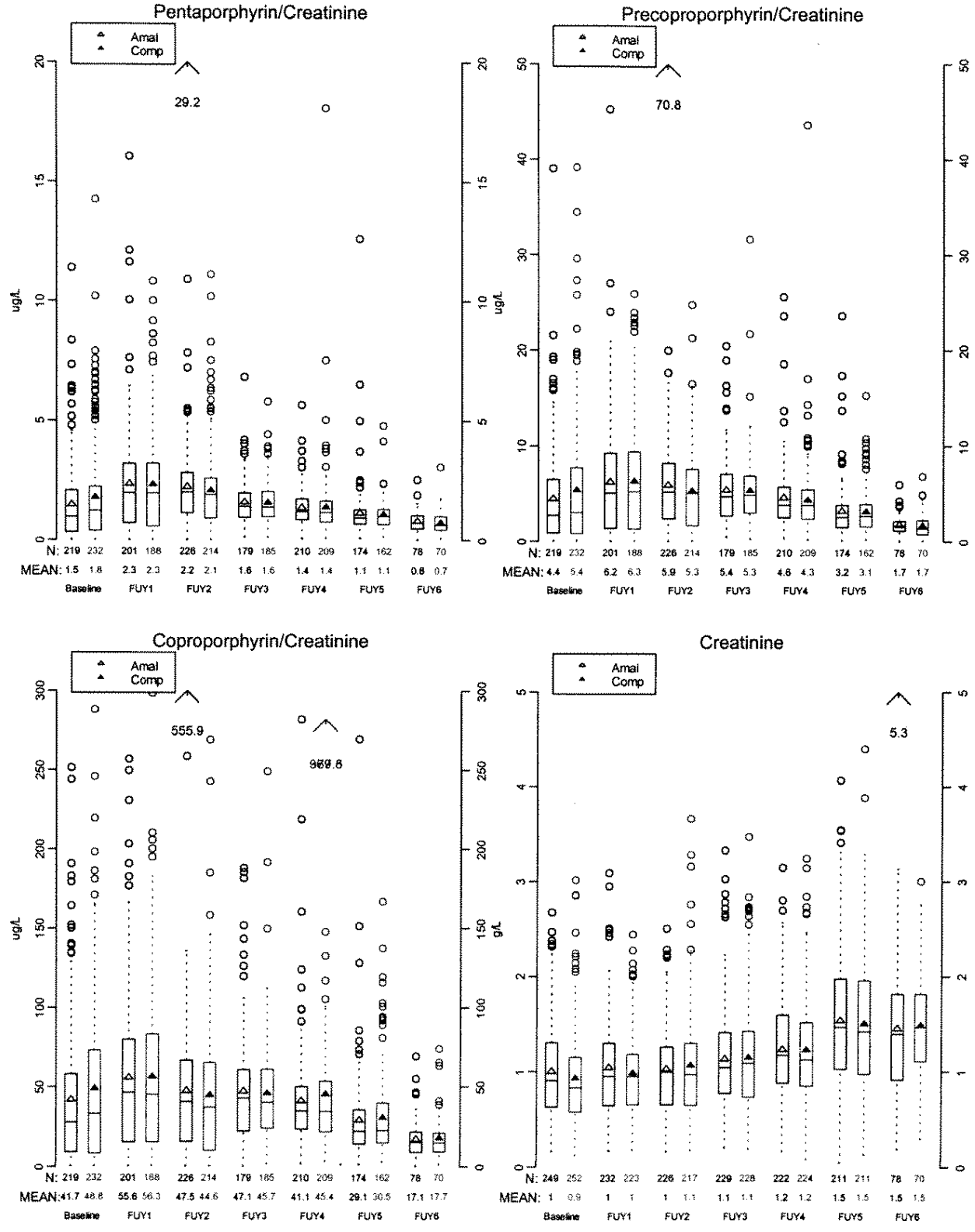


Figure 5.4.2. Distribution by time period and treatment group of creatinine and creatinine-adjusted values of pentaporphyrin, precoproporphyrin, coproporphyrin. (Plots of unadjusted values will be available at the meeting.)

6. Quality Assurance Monitoring

6.1. Completeness of data

Information on loss to follow-up and missed visits was given in Section 2.2. This section provides information on item-specific missing data for children who were examined, item-specific non-response on health history questionnaires received, and missing laboratory measures. Information is provided only where the percentage of missing data is 5% or more, that is, data are at least 95% complete except where noted below.

Table 6.1.1. Data items with 5% or more missing values at a single time period. Values given are number of children with missing data on the item indicated. Footnotes are given below the table.

	Baseline	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Standard Reaction Time[1]			56	84			
Adventitious movements[2]	121						
Albumin[3]	241						
GST – Alpha[4]	144						
GST – Pi[5]				145	270	401	148
Porphyrins[6]	50	84		85	26	78	

Notes:

1. Standard reaction time data are missing because of equipment failure (equipment has now been replaced).
2. Recording of adventitious movements was started after initiation of the study during the enrollment period.
3. Albumin analyses began during the enrollment period in October, 1997.
4. Missing GST-alpha data at baseline are due to problems with assays that have been corrected.
5. Missing GST-pi data is due to discontinuation of GST-pi analyses by agreement with NIDCR to accommodate a reduction in the budget for the trial.
6. Missing porphyrin data at baseline and in year 1 are due to instrument failure that has been solved. Missing values in year 3 are due to freezer failure. Missing values in year 5 are for samples that have not been analyzed yet.

6.2.Accuracy of data

There are three different types of checks on the accuracy of the data as described below.

1) Range Checks: The main study data are entered into Access© databases in Lisbon with built-in range checks to ensure that only valid data values are entered. Ranges of all data are checked again in Seattle before analysis as an extra precautionary measure.
2) Random Audits: In each follow-up year, approximately 15% of the subjects are selected at random for comparison of all data forms against the computerized database. Estimated error rates have been less than 0.2% throughout follow-up (Table 6.2.1).

3) Data-Generated Audits: The data are checked for unusual values and for consistency of data values collected on the same subject over time. Unusual values and large changes are checked for accuracy.

When an error in the computerized database is identified based on any of the checks described above, the information is transmitted to the Study Records Manager in Lisbon who identifies the source of the error and makes any necessary correction on the original data forms and in the computer database. The corrected version of the database is transmitted to Seattle for subsequent analyses.

Table 6.2.1. Results of random data audits.

	Baseline	1st-Year Follow-up	2nd-Year Follow-up	3rd-Year Follow-up	4th-Year Follow-up	5th-Year Follow-up
Number of Subjects	74	78	70	75	79	60
Number of Data Items Checked	33522	20555	25791	28725	28732	20642
Number of Mismatches	132	32	19	45	55	23
Mismatch Rate	0.39%	0.16%	0.07%	0.16%	0.19%	0.11%

6.3. Examiner reliability and accuracy

6.3.1. Checklist evaluation

The neurobehavioral examiners continue to send videotapes of testing sessions to Ms. Rosenbaum who uses them to evaluate the examiners' performance with the Test Procedure Checklist. When errors are noted, they are discussed with the examiner via e-mail or in person. The percentage of items correctly performed continues to be high (Table 6.3.1).

Table 6.3.1. Results of checklist-scoring of neurobehavioral examiners' performance.

	Examiner 1			Examiner 2			Examiner 3		
Year of Session	# Sessions	# Items	Accuracy (%)	# Sessions	# Items	Accuracy (%)	# Sessions	# Items	Accuracy (%)
1997	4	599	95	5	785	97			
1998	7	1113	98	10	1576	98			
1999	3	488	98	4	624	98			
2000	5	773	99.6	5	785	99.5			
2001	6	1071	97.4	7	1290	98.6	3	272	94.5
2002	4	544	96.0	4	544	98.3	3	408	99.3
2003	4	544	98.3	4	544	97.0	4	544	98.0
2004	3	408	96.5	2	296	96.5	2	296	97.0

6.3.2. Differences between examiners

As in previous follow-ups, there were some differences between examiners in average scores on some year 6 neurobehavioral tests (Table 6.3.4). Differences between examiners were statistically significant (at level 0.05) for 3 tests and the memory domain score. For WMS-Visual Learning, WMS-Visual Delay, and the memory domain score, the second examiner had the highest mean scores. For WAIS III-Digit Span, the first examiner had the highest mean score. These examiner differences do not follow the same pattern as those observed in year 5, when the third examiner had higher mean scores than the original two examiners on 3 tests and the memory domain score. The effect of examiner differences on the study is mitigated by the fact that the examiners are fairly well balanced across treatment groups (Table 6.3.3).

Table 6.3.3. Distribution of neurobehavioral examiner by treatment group and follow-up year.

Examiner	Baseline		Year1		Year2		Year3	
	Group A	Group B	Group A	Group B	Group A	Group B	Group A	Group B
1	105	110	102	105	110	110	109	108
2	148	144	139	133	127	120	119	118
3	.	.	.	.	.	.	.	.

Examiner	Year4		Year5		Year6		Year7	
	Group A	Group B	Group A	Group B	Group A	Group B	Group A	Group B
1	94	89	94	74	67	78	29	27
2	103	114	87	113	98	89	43	38
3	21	21	31	26	19	20	11	10

Table 6.3.4. Examiner differences in neurobehavioral test results in year 6. When statistical significance at level 0.05 is reached, the p-value and the highest mean score is highlighted by bolding and underlining. (Higher scores are better except for items with units of seconds, as indicated.)

Variable	Examiner 1		Examiner 2		Examiner 3		
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	P-value
Attention Domain Score	0.04	0.74	-0.01	0.67	-0.10	0.94	0.528
Memory Domain Score	-0.11	0.76	**0.09**	0.71	-0.06	0.87	**0.046**
Visual Motor Domain Score	0.05	0.59	-0.02	0.67	-0.08	0.66	0.426
RAVLT T.8	9.81	2.68	9.86	2.61	9.69	2.85	0.936
RAVLT TOT	46.92	8.26	46.38	8.32	48.38	8.82	0.386
WRAML visual Learning	N/A	N/A	N/A	N/A	N/A	N/A	N/A
WRAML Visual Delay	N/A	N/A	N/A	N/A	N/A	N/A	N/A
WRAVMA Drawing	N/A	N/A	N/A	N/A	N/A	N/A	N/A
WRAVMA Matching	N/A	N/A	N/A	N/A	N/A	N/A	N/A
WRAMA Peg - dominant	121.72	16.07	117.11	18.27	117.87	18.74	0.055
WRAMA Peg - non-dominant	120.52	15.16	117.57	14.83	116.10	17.32	0.124
Standard Reaction Time (sec)	0.77	0.16	0.78	0.17	0.82	0.15	0.240
Finger Tapping - dominant	49.94	6.24	49.63	6.44	49.36	7.47	0.852
Finger Tapping - non-dominant	43.63	6.13	43.18	6.17	43.21	6.97	0.793
WISC - Coding	N/A	N/A	N/A	N/A	N/A	N/A	N/A
WISC - Symbol Search	N/A	N/A	N/A	N/A	N/A	N/A	N/A
WISC - Digit Span	N/A	N/A	N/A	N/A	N/A	N/A	N/A
WRAML - Finger Windows	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Intermediate Trails A (sec)	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Intermediate Trails B (sec)	N/A	N/A	N/A	N/A	N/A	N/A	N/A
STROOP COL4	41.20	7.68	40.79	7.76	40.44	10.32	0.833
STROOP CWD4	45.56	8.95	46.22	9.22	45.44	11.94	0.780
WMS III - Visual Learning	34.48	4.60	**36.45**	3.91	35.08	4.01	**0.000**
WMS III - Visual Delay	31.99	6.02	**34.32**	5.24	31.62	6.27	**0.000**
WMS III - Spatial Span	14.78	3.05	15.53	2.62	15.18	2.91	0.057
WASI Matrices	23.25	5.62	24.30	4.27	24.31	4.89	0.132
WAIS III - Digit Span	**14.34**	3.95	13.43	2.96	12.92	3.37	**0.016**
WAIS III - Symbol Search	31.57	8.49	31.67	8.01	30.41	9.15	0.686
WAIS III - Digit Symbol	71.87	15.92	70.22	15.13	70.26	19.03	0.624
Adult - Trails A (sec)	29.44	11.20	30.24	10.44	33.92	16.17	0.096
Adult - Trails B (sec)	67.95	28.63	71.88	27.86	72.56	35.88	0.423

7. Summary

In summary, as the Casa Pia Study of the Health Effects of Dental Amalgams in Children moves into it's 7th and last year of follow-up testing, it continues to progress satisfactorily in all respects. Subject retention continues to be high, maintenance of the timetable continues to be acceptable, there have been no adverse events related to the study, and there were no additional protocol deviations during the current reporting year. The problems encountered due to the pedophilia activity and the publicity from the newspaper articles in the Portuguese newspaper, Publico, have proven to be quite limited, and to have had very little impact on the study. Our current greatest challenge is to determine ways in which to make participation by the increasingly independent-minded adolescent subjects acceptable and convenient so that we can complete the trial with sufficient data to meet the original goals.

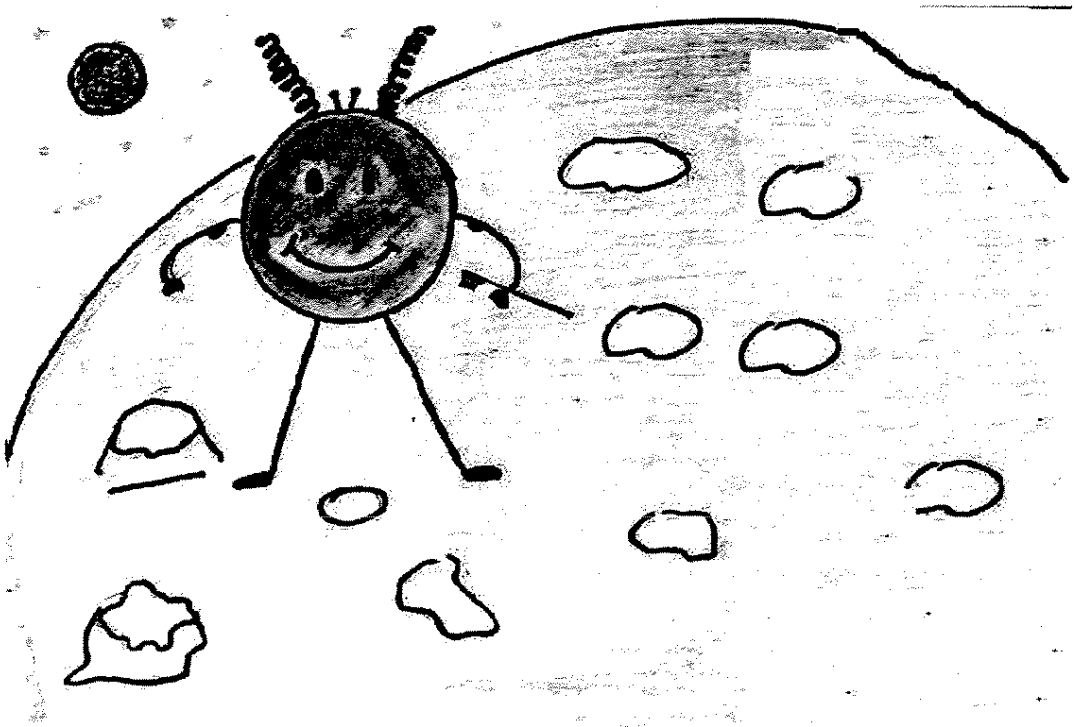
RECEIVED
Human Subjects Division
OCT 31 2003
UW

Children's Amalgam Trial
Casa Pia Study of the Health Effects of Amalgam in Children

Report to the Data and Safety Monitoring Board

October 17, 2000

ONLY FILE COPY
Keep with Committee Copy
Send to Reviewer at Renewal
To Records when file is closed
App.# 95-0401-A/C08



"Brushing Teeth on the Moon"
Drawing by Casa Pia Study participant.

Table of Contents

1. Introduction and Updates	1
1.1 Summary of recruitment and baseline comparisons	2
2. Follow-up Visits	3
2.1 Subject retention	3
2.2 Timing of follow-up visits	3
3. Dental Treatment and Mercury Exposure	4
4. Follow-up Results on Primary Outcome Variables	6
4.1 Distribution of primary outcome variables by follow-up period	6
4.2 Monitoring changes in performance of individual children	9
4.3 Additional follow-up of children with decreases in performance.	11
4.4 Group comparisons based on individual neurobehavioral test items.	12
4.5 Statistical stopping rule	12
4.6 Reassessment of power	13
5. Follow-up Results on Secondary Outcome Variables	13
5.1. Health history and sentinel events	13
5.2. Neurological	14
5.3. Oral exam	15
5.4. Laboratory Measures	16
5.4.1 Urinary Mercury	16
5.4.2 Other Urinary Measures	19
5.4.3 Monitoring Children with Elevated Blood Lead Values	23
6 Quality Assurance Monitoring	23
6.1 Completeness of data	23
6.2 Timeliness of data entry	24
6.3 Accuracy of data	24
6.4 Examiner reliability	25
7 Changes to Study Protocol	26
7.1 Additions to the Sentinel Event Monitoring System	26
7.2 Extension of follow-up to 7 years.	28
7.3 Modification of the neurobehavioral test items	28
8 Appendix A. Additional Figures	31
9 Appendix B. Additional Tables.	32
10 Appendix C. Description of the Reliable Change Index Method	36
11. Appendix D. Sentinel Health Events Checklist: English Version	37
12 Appendix E. Description of the proposed new neurobehavioral tests.	38
13 Appendix F. Abstracts and Papers from the Main Study	39
14 Appendix G. Progress on the Ancillary Study on Antibiotic Resistance	47

1. Introduction and Updates

The Casa Pia Study of the Health Effects of Amalgam in Children, an arm of the "Children's Amalgam Trial", is currently in the third year of follow-up evaluation and data collection. The Study is proceeding very well, with all goals being met in a timely fashion, and a subject retention rate which is better than that anticipated in the original proposal.

A competing renewal funding application was submitted in February, 2000, and reviewed in July, 2000. It was recommended for approval, with a priority score of 168.

Reports on non-endpoint aspects of the Study are given at scientific meetings each year. In March, 2000, Dr. Michael Martin presented a poster on urinary mercury changes observed as a function of treatment in the Study at the annual meeting of the International Association for Dental Research. A report on dietary methylmercury exposures for Study participants is proposed for presentation at the American Association for Dental Research and Society of Toxicology meetings in March, 2001. Abstracts for those presentations are given in Appendix F. Also, in May, 2000, Dr. Timothy DeRouen presented an overview of the Study at the meeting of the Task Force on Design and Analysis.

Manuscripts for the Study exist in two stages of development, "concepts" and "approved for analysis". Concepts are cleared for further development and analysis following review by Study principals and authors, and by NIDCR. The following manuscripts are currently in the analysis and writing stages:

- Neurobehavioral findings in Portuguese children, 8-11 years of age (baseline data).
- Dietary exposure to mercury: relationship to blood mercury levels in Portuguese children.
- Neurobehavioral measures: examiner effects.

The following are "concepts" which are being developed for review:

- Determinants of dental caries in a population of Portuguese schoolchildren.
- Epidemiology of oral diseases in a Portuguese children 8-11 years of age.
- Neurologic findings in a population of Portuguese children, 8-11 years of age.
- Urinary mercury and porphyrin levels in Portuguese children, 8-11 years of age.
- Pre-pubertal periodontitis in Portuguese children.
- The relationship between blood lead and dental caries in a population of Portuguese schoolchildren.

Unfortunately, one paper was published this year without going through the system. For the ancillary study on dietary exposure to methylmercury, blood mercury analyses were done through a subcontract with a private laboratory, Cebam Analytical, Inc. of Seattle. The owner of the lab, Dr. Lian Liang, had earlier presented a poster on the analytical method used for the blood mercury analyses at the 5th International Conference on Mercury as a Global Pollutant. In that poster, she listed some of the Casa Pia investigators as coauthors, and mentioned that the method was being used in the Casa Pia Study. She subsequently wrote a manuscript on the methods and submitted it to the Journal of Analytical Toxicology. A pure analytical methods paper would not have been subject to review within the Study system, since development of the method was independent of the Study. The problem is that she decided to include an illustration of the application of the method by publishing at the very end a table giving the baseline data on blood mercury levels of the subset of 150 Casa Pia children who were in the dietary study. This was done without getting permission from any of the Casa Pia investigators, and the manuscript was published without two of the Casa Pia coauthors (Martin, DeRouen) even being sent a draft or

notified of its submission. Fortunately, the publication of the baseline blood mercury levels for both groups combined does not compromise any significant part of the study, and there remains considerable additional information to publish on changes in blood mercury levels after treatment, and the relationship to dietary information. Dr. Liang has been seriously chastised for this breach of ethics, and she tried to withdraw the paper when the problem was pointed out to her, but it had already gone to press. It is included in Appendix F.

We have continued to have key personnel from the University of Washington make regular, periodic site visits to the Lisbon site for oversight purposes. Dr. DeRouen, Principal Investigator, made two site visits in the past year; Dr. Martin, Project Director, made a site visit in September, 2000; Gail Rosenbaum, Senior Psychometrist, also makes regular site visits for the purpose of examiner recalibration, testing evaluation and data checks, with the most recent visit in September, 2000.

During the past year, Aimee Loar, MS, formerly working on the study as a Graduate Research Assistant, has taken over as Data Manager at the UW at 25% effort, replacing Natalie Hawkins, who left the University. Tonya Benton, MA, was appointed to the study in February 2000, as a Data Analyst at 25% effort, to assist Ms. Loar and Dr. Leroux, Biostatistician, with data analyses and report preparation. There have been no other changes in study personnel.

There have been no substantive changes in protocol for the study since the last report other than those made to satisfy the desire of the Board to have a more comprehensive "Sentinel Event" monitoring system in place (Section 7.1). Two major changes in protocol were successfully proposed in the competing renewal application and are described in this report: 1) extension of the length of follow-up to 7 years (Section 7.2) and 2) changes in the neurobehavioral tests beginning in the 4th year of follow-up (Section 7.3).

1.1. Summary of recruitment, randomization and baseline comparisons

A total of 507 children were randomized to one of the two treatment groups (group sample sizes 253 and 254). Three protocol errors occurred in which subjects that should have been declared ineligible were randomized: 1) 1 subject with abnormal neurological findings (cerebral palsy) was randomized on 4/25/97; 2) 1 subject with existing amalgam was randomized on 1/7/98; and 3) 1 subject was randomized on 12/23/97 with a blood lead value of 14 ug/dl that was later adjusted by the laboratory to 16ug/dl. The treatment groups were similar at baseline in demographics, health history, and the distributions of the baseline values of primary and secondary outcome variables, oral exam results, blood lead levels, urinary mercury levels, and other urinary markers. A summary of the baseline comparisons is provided in Tables B.1, B.2, and B.4 (Appendix B); other baseline data are included with follow-up data in graphs and tables in the report. Norms for the neurobehavioral and nerve conduction velocity endpoints are given in Table B.3 (Appendix B). Baseline urinary mercury levels were low: mean 1.48 ug/ml, 7 out of 507 randomized subjects had levels > 5 ug/L. The average urinary mercury level for the US population is estimated to be 3.1 ug/ml. Blood lead levels were low; only 12 exceeded 10 ug/dl (5 in Group 1 and 7 in Group 2).

2. Follow-Up Visits

2.1. Subject retention

Data included in this report are current as of September 14, 2000. As of that date, second-year follow-up was complete; third-year follow-up is on-going. As reported in the 1999 report to the DSMB, a 94.9% retention rate was achieved at the first-year follow-up: 481 of the 507 randomized had at least one of the primary end-points measured (neurobehavioral, nerve conduction velocity). In the second year of follow-up, endpoint data was obtained on 92.5% (469/507). For 485 subjects (95.7%), at least one primary end-point has been obtained during follow-up; the percentage is similar in the two treatment groups (96.8% vs. 94.5%). The number of subjects lost to further follow-up is 11; reasons for loss to follow-up are a) written request to remove the child from the study (10), b) death of the child (1). No further attempts at follow-up will be made for any of these 11 children. Attempts are made to maintain contact with all other children and to obtain follow-up data whenever possible, even if the child has moved away from Lisbon. Because NCV testing is done by having the team visit the schools to test groups of participants in one visit, missed NCV visits are more difficult to reschedule than missed neurobehavioral visits; as a consequence, follow-up rates are slightly lower for NCV than for neurobehavioral testing. In response to a suggestion of the DSMB, an annual follow-up health history questionnaire was implemented in February 1999. A good response rate of 94.3% (478/507) was achieved in the first follow-up, due to intensive efforts to remind parents/guardians to complete the questionnaire. Mailing of questionnaires and follow-up of non-responders occurs during February-July of each year, during the last 6 months of each study follow-up period. Collection of 2nd-year follow-up health history questionnaires is still on-going (373 were received by September 14); 3rd-year follow-ups are scheduled for 2001. Detailed information on retention as of September 14, 2000 for each data type and year of follow-up is provided in Table 2.1.1.

Table 2.1.1. Number of subjects with successful follow-up.

Data Type	Number with Data At Follow-up Years:				
	1	2	3	1 AND 2	1, 2, OR 3
Neurobehavioral	478	469	189	464	484
Nerve Conduction Velocity	464	445	108	430	480
Health History	478	373	NA	369	482
Neurological	468	453	103	439	483
Oral Exam	454	469	173	441	483
Urinary Mercury	474	425	112	417	483

2.2. Timing of follow-up visits

Visits are scheduled within a target of two weeks on either side of the anniversary date. Particular attention is paid to the anniversary date for neurobehavioral testing because this provides data for three of the primary outcome variables and could potentially be influenced by the timing of the visit. However, late visits sometimes occur because of absences from school. Visits are sometimes scheduled early if it is known or suspected that a child will not be available during the target time interval. The majority of neurobehavioral visits have occurred within a 6-week time window around the anniversary date. Importantly, the timing of follow-up visits is similar in the two treatment groups (Table 2.2.1).

Table 2.2.1. Time from baseline to follow-up visits in weeks by study year, treatment group and type of visit (mean and standard deviation).

	First-Year Follow-up				Second-Year Follow-up				Third-Year Follow-up*			
	Group A		Group B		Group A		Group B		Group A		Group B	
Type of Visit	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Neurobehavioral	52	6	53	6	107	7	106	8	155	4	154	7
Nerve Conduction Velocity	59	12	59	11	111	13	110	12	155	4	155	9
Neurological	55	8	54	8	108	12	109	11	157	5	157	4
Oral Exam	59	9	59	11	108	9	107	9	158	5	158	5
Urinary Sample	51	9	52	8	105	11	105	9	154	3	154	2

* Incomplete (40%).

3. Dental Treatment and Mercury Exposure

Table 3.1 presents information on dental treatment during the baseline treatment period (up to the first follow-up oral exam) and during the first follow-up treatment period (between the first and second follow-up oral exams). The information provided for the first follow-up period is preliminary and no information is presented on treatment provided after the second-year oral exam (still in progress). There were approximately 10 surfaces restored per child on average during baseline dental treatment. The number of restorations was similar in the two treatment groups. Children in the amalgam group received on average less than one restoration using non-amalgam materials (190 total surfaces). One child in the composite group received amalgam fillings on 10 surfaces by mistake (this protocol error was previously reported). There was a small amount of treatment provided during follow-up, slightly more in the composite group than the amalgam group. The number of replacement restorations was similar in the two groups; these accounted for 12 surfaces replaced during baseline treatment and 74 during follow-up. All replaced restorations used the same material as the original restoration with one exception: one surface restored with a non-amalgam, non-composite material was replaced with composite.

A total of 781 teeth were extracted during the baseline treatment period (379 in the amalgam group and 402 in the composite group), most of which were primary teeth (Table 3.2). The number of extractions during follow-up treatment was much smaller (32 and 39 respectively). An additional large number of teeth were lost for reasons other than extraction; most of these were primary teeth presumed to have exfoliated. The total number of lost or extracted teeth during baseline treatment was 956 primary and 70 permanent in the amalgam group and 862 primary and 102 permanent in the composite group. The total numbers of teeth lost or extracted during the follow-up treatment period is 496 and 423, respectively, almost all of which have been primary teeth. Very few permanent teeth have been lost (21) or extracted (19) during follow-up. Table 3.1 gives the number of restored surfaces that have been lost due to extraction or loss of previously restored teeth.

Table 3.1. The number of restored surfaces placed, then subsequently replaced, extracted, or lost, by time period, treatment group, tooth type and material.

# of Surfaces			Baseline		Follow-up	
			Amalgam Group	Composite Group	Amalgam Group	Composite Group
Placed	Primary	Amalgam	950	8	6	0
		Other	33	788	1	21
	Permanent	Amalgam	1448	2	126	0
		Other	154	1698	34	190
	Overall		2585	2496	167	211
Replaced	Primary	Amalgam	0	0	3	0
		Other	0	2	0	6
	Permanent	Amalgam	2	0	29	0
		Other	0	5	4	32
	Overall		2	7	36	38
Extracted	Primary	Amalgam	13	0	2	0
		Other	0	14	0	4
	Permanent	Amalgam	0	0	0	0
		Other	0	4	0	4
	Overall		13	18	2	8
Lost	Primary	Amalgam	176	4	327	4
		Other	10	130	9	245
	Permanent	Amalgam	2	0	7	0
		Other	0	8	0	4
	Overall		188	142	343	253

Table 3.2. Teeth extracted or lost by time period and treatment group.

# of teeth		Baseline		Follow-up	
		Amalgam Group	Composite Group	Amalgam Group	Composite Group
Extracted	Primary	334	325	22	30
	Permanent	45	77	10	9
	Overall	379	402	32	39
Lost	Primary	622	537	452	375
	Permanent	25	25	12	9
	Overall	647	562	464	384

4. Follow-up Results on Primary Outcome Variables

4.1. Distribution of primary outcome variables by follow-up period

The distribution of the primary outcome variables at baseline and during follow-up is shown using boxplots in Figure 4.1.1 and Figure 4.1.2. The differences between the groups are small at all follow-up periods. Note that the Z-scores are standardized to have mean 0 in the entire sample at each time point. Therefore, the group means are approximately[1] equal in magnitude with opposite signs. A treatment effect would be displayed as a difference between the group means; if there are no group differences, the mean values will be approximately 0 for both groups.

The data are also being examined descriptively to assess longitudinal changes over time. In particular, a regression model is fit to the outcome data for each child versus time since randomization. The regression slopes (annual rates of change) are plotted in Figure A.1 (Appendix A). These histograms support the conclusions from the boxplots that the groups are very similar in changes over time in the primary end-points. The histograms also illustrate the fact that a small number of children have large rates of change that are potentially of concern. Such changes in individual children are examined more closely in the following section.

[1] The group means would be exactly equal in magnitude if the group sample sizes were equal and there was no missing data.

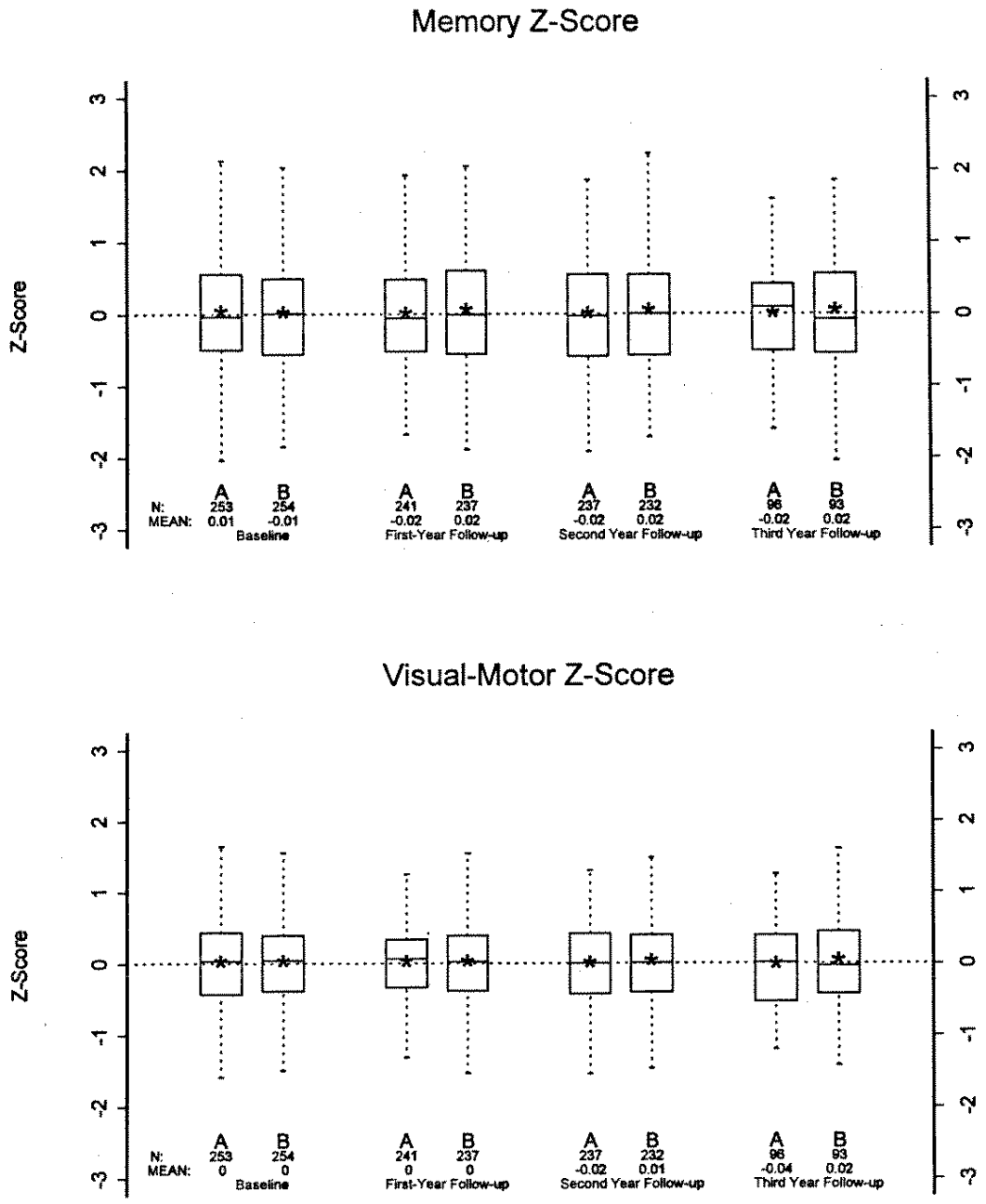


Figure 4.1.1 Distribution of neurobehavioral (memory and visual motor) Z-scores by time period and treatment group. The mean value is denoted by X, median by a line inside the box, 25th and 75th percentiles by lower and upper edges of the box. Tails extend to minimum and maximum values.

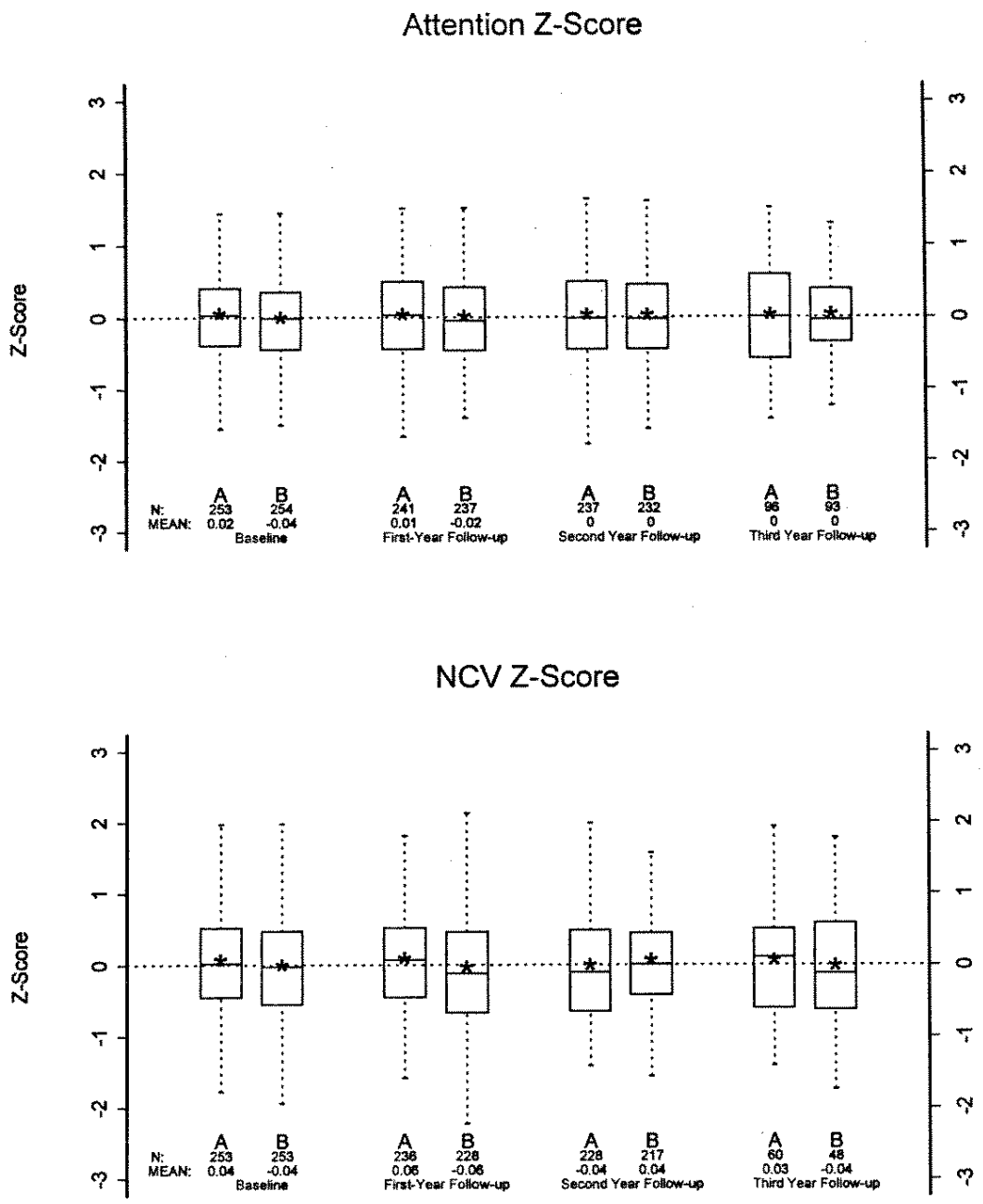


Figure 4.1.2 Distribution of neurobehavioral (attention) and nerve conduction velocity Z-scores by time period and treatment group. The mean value is denoted by X, median by a line inside the box, 25th and 75th percentiles by lower and upper edges of the box. Tails extend to minimum and maximum values.

4.2. Monitoring changes in performance of individual children

To monitor changes in performance at 2nd-year follow-up, we use a Reliable Change Index (RCI), which provides a systematic method for identifying children whose neurobehavioral test scores show a decline across time. The RCI takes into account the standard error of measurement as well as the correlation between performance on a test at two time points, to identify children whose performance falls outside a one-sided 95% confidence interval for their predicted value (see Appendix C for a detailed description of this method). Table 4.2.1 shows the number of children with decreases on 1, 2, 3, and 4 or more test scores using the RCI.

Table 4.2.1. The number of children with decreases based on the reliable change index in 1, 2, 3, and 4 or more test scores excluding Drawing. There were 20 children with decreases on 4 or more scores either from baseline to 2nd-year follow-up, or from 1st to 2nd-year follow-up, and an additional 2 children with decreases on 4 or more scores including Drawing at both 1st and 2nd year follow-up.

	Total with each # of reliable changes								Total for the year
	One		Two		Three		Four or more		
	Group A	Group B	Group A	Group B	Group A	Group B	Group A	Group B	
Baseline to First-Year Follow-up	67	69	52	38	11	9	8	3	257
Baseline to Second-Year Follow-up	72	68	26	23	7	14	12	2	224
First to Second-Year Follow-up	51	67	26	27	13	11	6	3	204

We have focused on 22 children who had a pattern of decreases that may be indicative of a true decline, namely, either a) decreases in 4 or more test scores exclusive of the drawing test either from baseline to 2nd year follow-up or from 1st to 2nd-year follow-up (20 subjects total including 3 overlaps), or b) decreases in 4 or more test scores inclusive of drawing, *both* from baseline to 1st year follow-up *and* from 1st to 2nd year follow-up (2 additional subjects). Of these 22 children, 17 are in one treatment group and 5 are in the other group. The difference between groups is statistically significant (p=0.02 by Fisher's exact test), but should be interpreted with caution because this does not represent a pre-defined definition of "true decline", it was not suggested as a primary endpoint for the study, and this analysis was not pre-planned. Nevertheless, the imbalance between groups is of concern, although some perspective is provided by the fact that there is an association with gender that is almost as strong (there are more boys, 16, than girls, 6, in this sub-group). These children were from 6 of the 7 participating schools, and approximately equal numbers of these children were tested by each of the two neurobehavioral examiners. Urinary mercury levels for these children were in the normal range (< 10 ug/L) and mean values were similar to the mean values for their respective treatment groups (data not shown because it reveals the treatment group identities).

Table 4.2.2 summarizes the data on other key variables for these children. IQ scores were fairly close to the average of 85 for the entire sample. It is notable that the decreases in performance in these children can be detected by examining the Z-scores, particularly for the attention and memory domains. The nerve conduction velocities for these children have increased from baseline.

Table 4.2.2. Descriptive information for the 22 children identified using the reliable change index.

		17 Ss in Group A		5 Ss in Group B	
		Mean	Std Dev	Mean	Std Dev
Age at Baseline in Years		10.16	0.84	9.68	0.90
IQ at Baseline		84.82	8.71	78.00	6.48
Blood Lead at Baseline (ug/dL)		5.06	3.04	4.20	1.60
Z-Attention	Baseline	-0.37	0.75	-0.49	1.01
	First-year Follow-up	-0.56	0.67	-0.28	0.75
	Second-year Follow-up	-0.94	0.55	-0.85	0.79
	Third-year Follow-up	-0.71	0.69	-0.65	0.41
Z-Memory	Baseline	0.01	0.93	-0.14	0.83
	First-year Follow-up	-0.27	0.73	0.01	0.88
	Second-year Follow-up	-0.65	0.79	-0.29	0.54
	Third-year Follow-up	-0.32	0.74	0.19	0.48
Z-Visual Motor	Baseline	-0.51	0.85	-0.52	0.37
	First-year Follow-up	-0.40	0.75	-0.37	0.77
	Second-year Follow-up	-0.67	0.80	-0.62	0.60
	Third-year Follow-up	-0.70	0.75	-0.26	1.03
Tibial Velocity (m/sec)	Baseline	50.76	4.67	49.98	6.96
	First-year Follow-up	51.89	2.71	54.04	7.70
	Second-year Follow-up	52.08	3.27	54.33	7.42
	Third-year Follow-up	50.40	3.66	49.17	1.59
Ulnar Velocity (m/sec)	Baseline	57.14	7.13	60.02	5.78
	First-year Follow-up	60.78	6.56	60.24	3.31
	Second-year Follow-up	60.05	5.00	61.80	9.29
	Third-year Follow-up	58.30	6.06	56.53	4.80
Z-Nerve	Baseline	-0.18	0.83	-0.03	0.77
	First-year Follow-up	0.03	0.67	0.24	0.85
	Second-year Follow-up	-0.01	0.65	0.43	1.25
	Third-year Follow-up	-0.12	0.64	-0.46	0.28

Decreases in performance have occurred most frequently for the RAVLT, Finger Windows, Drawing and Stroop tests (Table 4.2.3).

Table 4.2.3. The number of children, among the 22 identified using the reliable change index, with decreases on each neurobehavioral test item.

	Number with Change from Baseline to 2nd-Year Follow-up	Number with Change from 1st to 2nd-Year Follow-up	Total
RAVLT T.8	7	4	11
RAVLT TOT	9	2	11
WRAML Visual Learning	1	2	3
WRAML DE VI	4	2	6
WRAVMA Drawing	11	7	18
WRAVMA Matching	1	3	4
WRAVMA Peg - dominant	1	3	4
WRAVMA Peg - non-dominant	0	4	4
Standard Reaction Time (mean)	3	2	5
Finger Tapping-Dominant	2	2	4
Finger Tapping - Non-Dominant	1	1	2
Coding	1	1	2
Symbol Search	5	4	9
Digit Span	4	5	9
Finger Windows	6	7	13
Trails A (seconds)	0	2	2
Trails B (seconds)	1	2	3
STROOP WD4	8	6	14
STROOP COL4	15	8	23
STROOP CWD4	8	5	13

As an aid to interpretation of the reliable change results, the RCI was also used to identify 67 children with increases in performance on five or more individual neurobehavioral test items. These children are divided relatively evenly between Group A (27) and Group B (40). The 67 children with increases are very similar to the 22 children with decreases in performance as far as age, gender, baseline blood lead, baseline IQ, neurobehavioral examiners and distribution by school.

4.3. Additional follow-up of children with decreases in performance.

A detailed examination of the data was made by Dr. Townes for the 22 children identified by the RCI and an additional 4 children who had been judged to be of clinical concern one year ago based on previous data. On neurologic examination, 10 of these children had two or more "soft neurologic signs" (mirror movements and syncinesias) at one or more of the examinations. Eight of these 10 children were rated as having "normal neurologic" examinations; one child had abnormal neurologic findings (cerebral palsy) at the baseline examination (this protocol error was reported previously); and one child was thought to have poor hearing, speech disturbance and distractibility. One of the children was rated as neurologically normal but hyperactive. In summary, we have identified a small group of participants who have suffered declines in test performance and/or soft neurologic signs; these individuals will be monitored closely. Individual data for the 26 children discussed here will be available at the meeting.

4.4. Group comparisons based on individual neurobehavioral test items.

Although the primary endpoints for the study are the four Z-scores, we have also performed analyses to compare the groups on the individual neurobehavioral test items at each time period (Tables B.4 -B.5). This decision was motivated by the desire to monitor changes in absolute performance (whereas the Z-scores are standardized and do not reflect changes in absolute scores) and also by the results of the reliable change analysis presented in the previous section. No significant group differences were found for either 1st or 2nd year follow-up (the smallest p-value was 0.099 for 2nd-year follow-up finger windows). Results for 3rd year follow-up are tentative due to small sample size and show no convincing evidence of group differences; there was one significant group difference for Symbol Search: mean values 10.61 for Group A vs 9.65 for Group B (p=0.033).

4.5. Statistical stopping rule

The primary analysis for the study is based on the four primary endpoints (memory, visual motor, and attention domain scores from the neurobehavioral testing, and a combined nerve conduction velocity score). Two separate tests are performed: an O'Brien type test that will detect small group differences that are similar across the four endpoints and a Hotelling T^2 test that will detect inconsistent group differences, e.g., a difference in just one of the variables. The overall significance level of the test accounts for the overlap between these two tests (i.e., it is possible that both tests reject the null hypothesis). The O'Brien test statistic is a two-sample t-statistic computed using a derived variable that weights the four endpoints according to the inverse of the pooled correlation matrix. A 1-sided p-value is reported for this test. The required significance levels for this test component depend on which treatment group has the worse outcome (a higher significance level is allowed for amalgam so that we will have higher power to detect adverse effects of amalgam). Note that if the O'Brien statistic p-value was intermediate between the two required significance levels, the treatment groups would have to be unmasked in order to determine if the required significance level was achieved; in this event, we would provide the treatment identities to the DSMB only in order to maintain masking of study personnel would remain masked. Differences in either direction are treated the same by Hotelling's test; therefore, a 2-sided p-value is reported for this test. The significance levels for the entire follow-up period are given in Section 7.2.

Neither of the two test components is significant (Table 4.5.1). The primary statistical analysis is based on the intent to treat principle, using imputed values for missing observations, and with no adjustment for covariates. Imputation of a small number of missing values in the baseline data was performed using linear regression with age, gender, dominant hand, and examiner (1 or 2) as predictors; imputation of follow-up data was done using linear regression with the previous visit's values as predictors. A complete-case analysis (i.e., using only the subjects with complete data on the four primary endpoints) gave similar results for both the O'Brien test (T=-1.31, P=0.10) and Hotelling test (F=0.23, P=0.99). No adjustment for covariates was done because the treatment groups were balanced on all important covariates.

Table 4.5.1. Results of intent-to-treat analysis using imputed data at 1st and 2nd-year follow-up.

	Test Statistic	P value	Required Significance Levels	
			Hypothesis: Amalgam Harmful	Hypothesis: Composite Harmful
O'Brien	T=-1.14	0.13 (1-sided)	0.0027	0.0008
Hotelling's T^2	F=0.12	1.00 (2-sided)	0.0020	

4.6. Reassessment of power

Revised power calculations were reported in the 1999 report to the DSMB. The revised calculations used the retention rates and correlation values that had been observed up to that time and showed that the study power values were slightly lower but close to the values given in the original grant application. The calculations used in the 1999 report assumed the number of subjects with data at years 2, 3, 4, and 5 would be 450, 438, 426, and 410, respectively. We have 2nd year follow-up data on 469 subjects, more than assumed. Therefore, the power of the study assuming a 5-year follow-up continues to be adequate for most effect sizes of interest. However, a 5-year study has inadequate power to detect a *small* effect on only one of the primary end-points. This was documented in our original grant application and again in our grant renewal application, where we proposed to extend follow-up to seven years (Section 7.2).

5. Follow-up Results on Secondary Outcome Variables

5.1. Health history and sentinel events

An annual health history questionnaire is completed for each child by the parents or guardians for the purpose of monitoring hospitalizations, illnesses, medications, and allergies. The results from the first two follow-ups are provided in Table 5.1. There were slightly more events in Group A, but there were no significant differences between treatment groups.

Table 5.1.1. Follow-up health history information by treatment group.

	First-Year Follow-up				Second-Year Follow-up			
	Group A		Group B		Group A		Group B	
Survey Question	Positive Responses	%	Positive Responses	%	Positive Responses	%	Positive Responses	%
Has your child been diagnosed with any serious illness in the last year?	24	10.00	20	8.40	18	9.68	16	8.56
In the past year, has your child had any disease or treatment other than for the teeth?	54	22.50	44	18.49	37	19.89	29	15.51
In the past year has your child taken any medicines given by a doctor or nurse or clinic, for example, antibiotics or injections?	90	37.50	82	34.45	57	30.65	51	27.27
Is your child taking any medications now?	34	14.17	26	10.92	12	6.45	16	8.56
In the past year has your child been diagnosed with an allergy to anything?	30	12.50	25	10.50	16	8.60	14	7.49

The health history questionnaire is used as one of the screening devices for identifying potential sentinel health events (Section 7.1). All randomized subjects in the study except for the 10 who have requested to be removed have been followed using these methods. When serious health problems are suspected, the health of the child is investigated with the school officials and parents. Three sentinel health events have been identified to date: one child suffered a stroke, one has been diagnosed with epilepsy, and one child died (date of death: December 8, 1999; cause of death: hepatitis). The first two cases were reported previously. A listing of all available data on primary outcome variables, urinary mercury levels, and dental treatment for these children is presented in Table 5.1.2. No follow-up data have been collected on the child with epilepsy because The child with epilepsy had left Lisbon and could not be followed-up, but has recently returned to Lisbon and will resume participation in study examinations.

Table 5.1.2. Data listing for children with sentinel health events.

Encrypted Identification		Treatment Group		
		Amalgam		Composite
		18168-stroke	20308-death	19241-epilepsy
Amalgam surfaces		3.00	13.00	0.00
Other surfaces		0.00	0.00	3.00
Memory Z-Score	Baseline	0.32	0.20	0.47
	First-Year	-0.30	0.42	.
	Second-Year	-0.31	0.31	.
	Third-Year	0.43	.	.
Visual-Motor Z-Score	Baseline	1.07	0.48	0.17
	First-Year	1.19	0.15	.
	Second-Year	-0.51	0.28	.
	Third-Year	-0.63	.	.
Attention Z-Score	Baseline	-0.38	-0.83	-0.19
	First-Year	-0.64	-0.68	.
	Second-Year	-0.77	-0.41	.
	Third-Year	-0.49	.	.
NCV Z-Score	Baseline	-0.79	-0.83	0.14
	First-Year	.	-0.98	.
	Second-Year	0.07	.	.
	Third-Year	.	.	.
Mercury Level	Baseline	4.74	0.89	2.25
	First-Year	4.79	1.97	.
	Second-Year	4.70	0.40	.
	Third-Year	.	.	.

5.2. Neurological

The prevalence of neurological abnormalities by treatment group is given in Table 5.3. In addition the prevalence of any abnormality is given in the last line of the table. Prevalences of abnormalities remain low and similar in the two treatment groups.

Table 5.2.1. Prevalence of neurological abnormalities by study year and treatment group.

Abnormality	Baseline				First-Year Follow-up				Second-Year Follow-up				Third-Year Follow-up			
	Group A		Group B		Group A		Group B		Group A		Group B		Group A		Group B	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Mental Status	0	0.0	0	0.0	2	0.8	3	1.3	1	0.4	2	0.9	1	1.8	0	0.0
Cranial Nerves	6	2.4	3	1.2	4	1.7	6	2.6	6	2.6	6	2.7	0	0.0	2	4.3
Motor Exam	1	0.4	1	0.4	0	0.0	1	0.4	1	0.4	0	0.0	2	3.6	1	2.1
Sensory	0	0.0	1	0.4	0	0.0	0	0.0	1	0.4	0	0.0	0	0.0	1	2.1
Coordination	2	0.8	1	0.4	0	0.0	1	0.4	2	0.9	0	0.0	1	1.8	1	2.1
Reflexes	3	1.2	0	0.0	3	1.3	2	0.9	1	0.4	2	0.9	1	1.8	1	2.1
Gait/Station	1	0.4	0	0.0	0	0.0	0	0.0	3	1.3	0	0.0	2	3.6	2	4.3
Adventitious Movement	0	0.0	0	0.0	3	1.3	1	0.4	5	2.2	3	1.3	1	1.8	2	4.3
Any Abnormality	9	3.6	6	2.4	11	4.7	11	4.7	11	4.8	12	5.4	2	3.6	4	8.5

5.3. Oral exam

Information on the number of decayed surfaces found at the oral exams is provided in Table 5.3.1. New decay has occurred at a fairly constant rate during the follow-up period. The amount of new decay is less in Group A compared with Group B.

Table 5.3.1. Number of decayed surfaces identified at oral exams by study year, treatment group and type of tooth surface.

# of carious surfaces		Baseline				First-Year Follow-up			
		Group A		Group B		Group A		Group B	
		Total	Mean	Total	Mean	Total	Mean	Total	Mean
Primary	Occlusal	704	2.78	688	2.71	29	0.13	25	0.11
	Non-Occlusal	1675	6.62	1624	6.39	130	0.57	164	0.73
Permanent	Occlusal	839	3.32	926	3.65	73	0.32	94	0.42
	Non-Occlusal	729	2.88	793	3.12	142	0.62	177	0.79
Overall		3947	15.60	4031	15.87	374	1.63	460	2.04

# of carious surfaces		Second-Year Follow-up				Third-Year Follow-up			
		Group A		Group B		Group A		Group B	
		Total	Mean	Total	Mean	Total	Mean	Total	Mean
Primary	Occlusal	13	0.06	15	0.06	1	0.01	6	0.07
	Non-Occlusal	54	0.23	88	0.38	4	0.04	25	0.30
Permanent	Occlusal	35	0.15	93	0.40	20	0.22	26	0.31
	Non-Occlusal	126	0.54	222	0.95	122	1.36	141	1.70
Overall		228	0.97	418	1.79*	147	1.63	198	2.39

* This mean value is reduced after exclusion of an outlying value equal to 42 surfaces.

5.4. Laboratory Measures

5.4.1. Urinary Mercury

The distributions of urinary mercury levels and creatinine-adjusted mercury levels are shown in Figure 5.4.1. Urinary mercury levels were similar at baseline in the two treatment groups. During follow-up, the mean mercury concentration increased in the amalgam group. In the composite group, mercury levels have been unchanged through the first two years of follow-up; the sample size is too small to allow conclusions about third-year follow-up. The creatinine-adjusted mercury levels show similar trends (Fig 5.4.1). Nineteen children have had a mercury level exceeding 10 ug/L. A listing of the primary outcome variables, urinary mercury levels and baseline dental treatment for these children is presented in Table 5.4.1. None of these 19 subjects belongs to the group of children who were identified as being of clinical concern or with decreases in test performance (Section 4.3). Examination of the raw data for these children (data will be available at the meeting) confirmed that there are no clinical concerns. Three of the 19 subjects have had urinary mercury levels larger than 15 ug/L; additional urine samples have been collected for these children and all mercury values have been below 10 ug/L. Note that two of these children had unusually high urinary mercury levels (31.90 and 38.65 ug/L) despite having not received amalgam fillings; their re-test values were normal (6.90 and 0.44, respectively).

Additional information on increases in mercury is provided in Figure 5.4.2 that shows the distribution of the annual rate of change of mercury (regression coefficient of mercury against time). The results indicate that urinary mercury levels increased by 0.765 ug/L/year on average in the amalgam group (95% CI: 0.59 to 0.94). Note that the increase was larger in the first year and lower in subsequent years as can be seen in Fig 5.4.1. There is no evidence for an increase in the mean mercury level in the composite group (mean annual rate of change 0.077, 95% CI: -0.11 to 0.26).

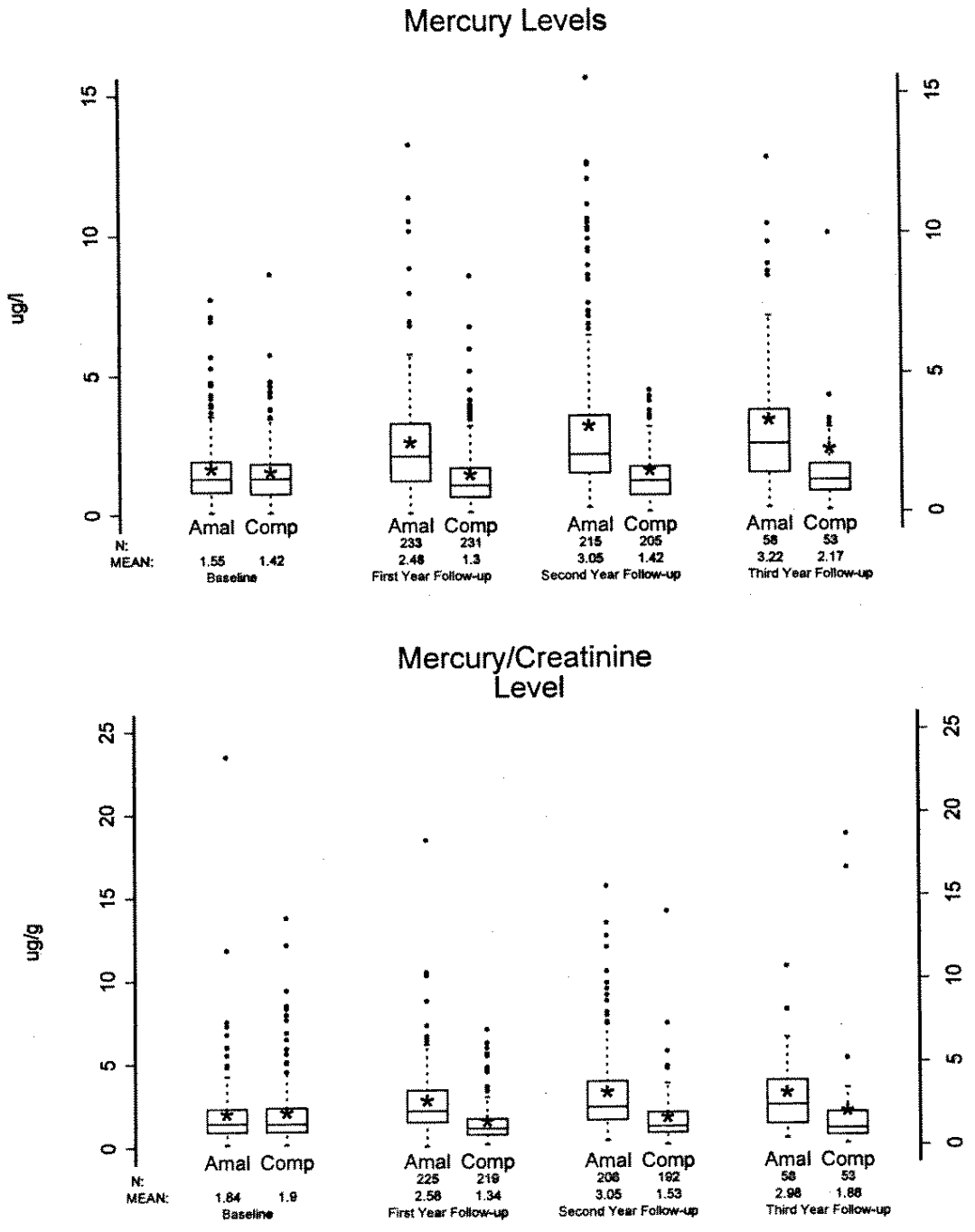


Figure 5.4.1. Distribution of urinary mercury values by time period and treatment group. The mean is marked by X, 25th and 75th percentiles by lower and upper edges of the box, median by a line inside the box, and tails extend to minimum and maximum values. Outliers are plotted by individual points except for the following values not plotted: mercury (ug/L): 31.9 (composite group, 2nd year); 38.65 (composite group, 3rd year).

Table 5.4.1. Data listing for children with high urinary mercury values.

Encrypted Identification		Treatment Group											
		Amalgam											
		16005	16019	16101	16112	18181	18183	18496	18498	18503	18532	19260	19581
Amalgam surfaces		14.00	14.00	6.00	14.00	9.00	13.00	11.00	11.00	11.00	7.00	9.00	10.00
Other surfaces		0.00	0.00	3.00	0.00	0.00	0.00	0.00	6.00	0.00	0.00	0.00	0.00
Memory Z-Score	Baseline	0.38	-0.28	1.03	1.15	0.46	-0.29	-0.79	0.74	-0.47	0.60	0.85	-0.23
	First-Year	-0.12	-0.13	0.47	1.29	1.03	0.25	-0.25	0.59	-0.77	0.92	0.70	0.22
	Second-Year	-0.50	-0.16	0.58	0.80	0.74	-0.48	-0.54	0.13	-1.15	1.64	0.09	0.64
	Third-Year	0.40	0.38	0.43	1.69	1.02	.	.	.	.	.	.	.
Visual-Motor Z-Score	Baseline	0.24	0.07	-0.23	0.08	0.90	-0.00	-0.34	1.22	-1.10	0.72	0.86	-0.12
	First-Year	0.25	-0.25	-0.25	0.31	1.68	0.45	-0.07	0.64	-0.82	0.92	1.22	-0.06
	Second-Year	0.20	-0.05	0.49	-0.10	1.43	0.83	-0.16	1.06	-0.22	0.78	0.94	-0.04
	Third-Year	0.57	-0.10	-0.28	0.82	1.42	.	.	.	.	.	.	.
Attention Z-Score	Baseline	0.41	-0.34	0.16	-0.00	0.61	0.49	0.10	1.23	-0.46	1.38	-0.09	-0.40
	First-Year	0.58	-0.33	0.54	0.00	0.69	-0.02	-0.15	1.17	-0.05	1.54	0.05	0.12
	Second-Year	0.57	-0.15	0.52	-0.14	1.02	0.31	-0.19	1.19	0.03	1.68	0.17	-0.05
	Third-Year	0.79	-0.56	0.52	-0.01	1.29	.	.	.	.	.	.	.
NCV Z-Score	Baseline	-0.12	0.08	-0.40	-0.64	-0.01	-0.79	0.49	0.62	-0.51	0.23	0.80	0.41
	First-Year	-0.45	-0.93	0.47	-0.53	-0.11	-1.12	-0.01	0.38	-0.40	-0.43	0.18	-0.35
	Second-Year	0.14	-0.26	0.92	0.24	-0.13	-1.21	-0.91	-1.33	-0.95	-0.74	0.47	-0.11
	Third-Year	-0.06	-1.08	-0.40	0.29	-0.83	.	.	.	.	.	.	.
Mercury Level	Baseline	1.56	1.09	1.17	1.65	2.34	2.35	0.32	0.72	1.13	0.52	0.95	0.66
	First-Year	13.20	4.02	3.32	3.24	4.95	5.14	5.07	11.30	4.45	10.11	3.42	10.46
	Second-Year	4.80	.	1.67	15.56	12.54	12.46	11.29	6.59	10.39	1.19	10.12	5.82
	Third-Year	5.77	12.70	10.33	5.28	.	.	.	.	.	.	.	.

Encrypted Identification		Treatment Group						
		Amalgam					Composite	
		19618	20313	22386	22387	22421	16017	20526
Amalgam surfaces		18.00	9.00	14.00	8.00	4.00	0.00	0.00
Other surfaces		0.00	4.00	4.00	1.00	0.00	4.00	11.00
Memory Z-Score	Baseline	-0.37	0.57	0.21	-0.37	-0.57	1.93	0.64
	First-Year	-0.13	0.29	0.16	-0.01	-0.01	1.77	1.85
	Second-Year	0.52	-0.14	0.42	-0.42	0.37	1.95	0.57
	Third-Year	.	.	.	.	.	1.91	.
Visual-Motor Z-Score	Baseline	-0.50	0.92	0.72	0.24	0.96	-0.59	0.15
	First-Year	-0.59	0.97	0.78	0.26	1.24	-0.08	0.54
	Second-Year	-0.26	1.15	1.23	0.28	1.40	-0.10	0.86
	Third-Year	.	.	.	.	.	-0.34	.
Attention Z-Score	Baseline	-0.05	0.61	0.61	0.72	0.98	-0.64	0.86
	First-Year	0.28	0.73	0.75	0.14	1.24	-1.02	0.95
	Second-Year	0.27	0.53	0.77	0.00	1.14	-1.18	0.46
	Third-Year	.	.	.	.	.	-0.39	.

NCV Z-Score	Baseline	-0.45	-0.28	1.96	-0.76	-0.91	0.17	-0.66
	First-Year	0.17	-0.69	-0.76	-0.25	-0.55	0.09	0.08
	Second-Year	0.78	-0.19	0.35	-1.08	-1.23	1.06	0.21
	Third-Year	.	.	.	.	.	.	.
Mercury Level	Baseline	0.64	0.95	1.10	0.17	1.37	1.16	0.19
	First-Year	3.27	3.24	5.64	4.54	.	0.52	1.32
	Second-Year	10.53	11.04	11.95	12.52	10.22	1.60	31.90
	Third-Year	.	.	.	.	.	38.65	.

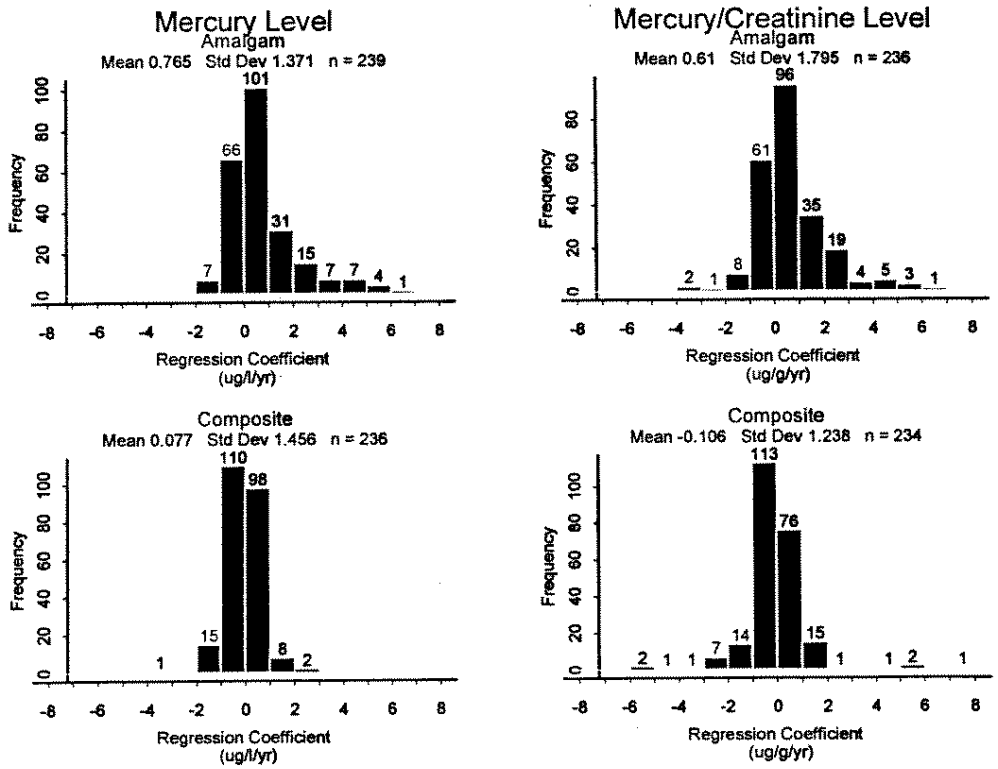


Figure 5.4.2. Histograms of annual rates of change of urinary Hg and creatinine-adjusted urinary mercury (regression coefficients from regression against years since randomization).

5.4.2. Other Urinary Measures

Boxplots of other urinary markers over time are shown in Figures 5.4.3 – 5.4.5. There are increases from baseline in GST-alpha and GST-pi levels in both groups and the levels tend to be slightly larger in the amalgam group than the composite group. The number of elevated albumin values (larger than 500) has been small: 5 at baseline, 1 at year 1, 4 at year 2, 0 at year 3 so far. There is a very slight upward trend in creatinine levels but no differences between the two treatment groups.

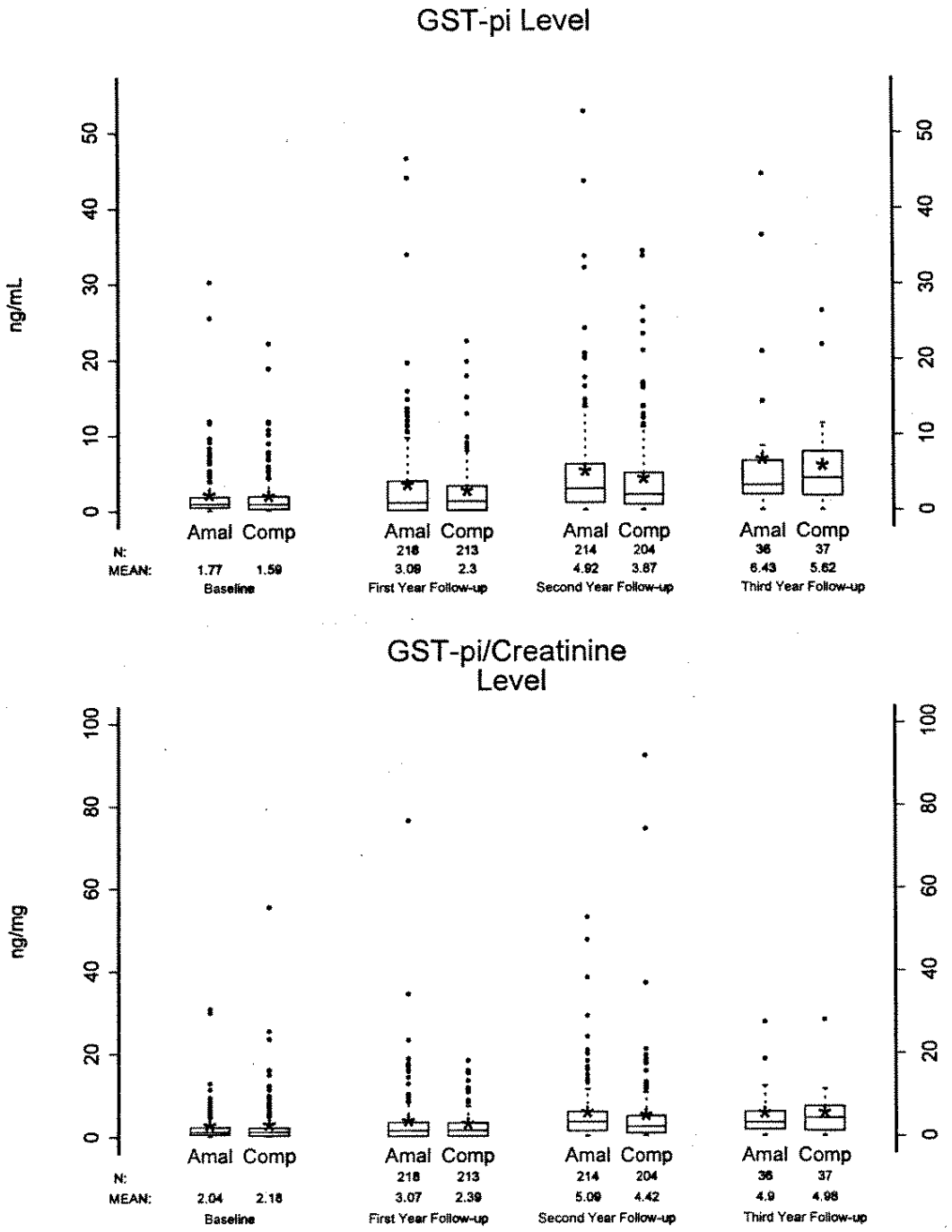


Figure 5.4.3. Distribution of GST-pi with and without creatinine adjustement by time period and treatment group. The mean is marked by X, 25th and 75th percentiles by lower and upper edges of the box, median by a line inside the box, and tails extend to minimum and maximum values. Outliers are plotted by individual points.

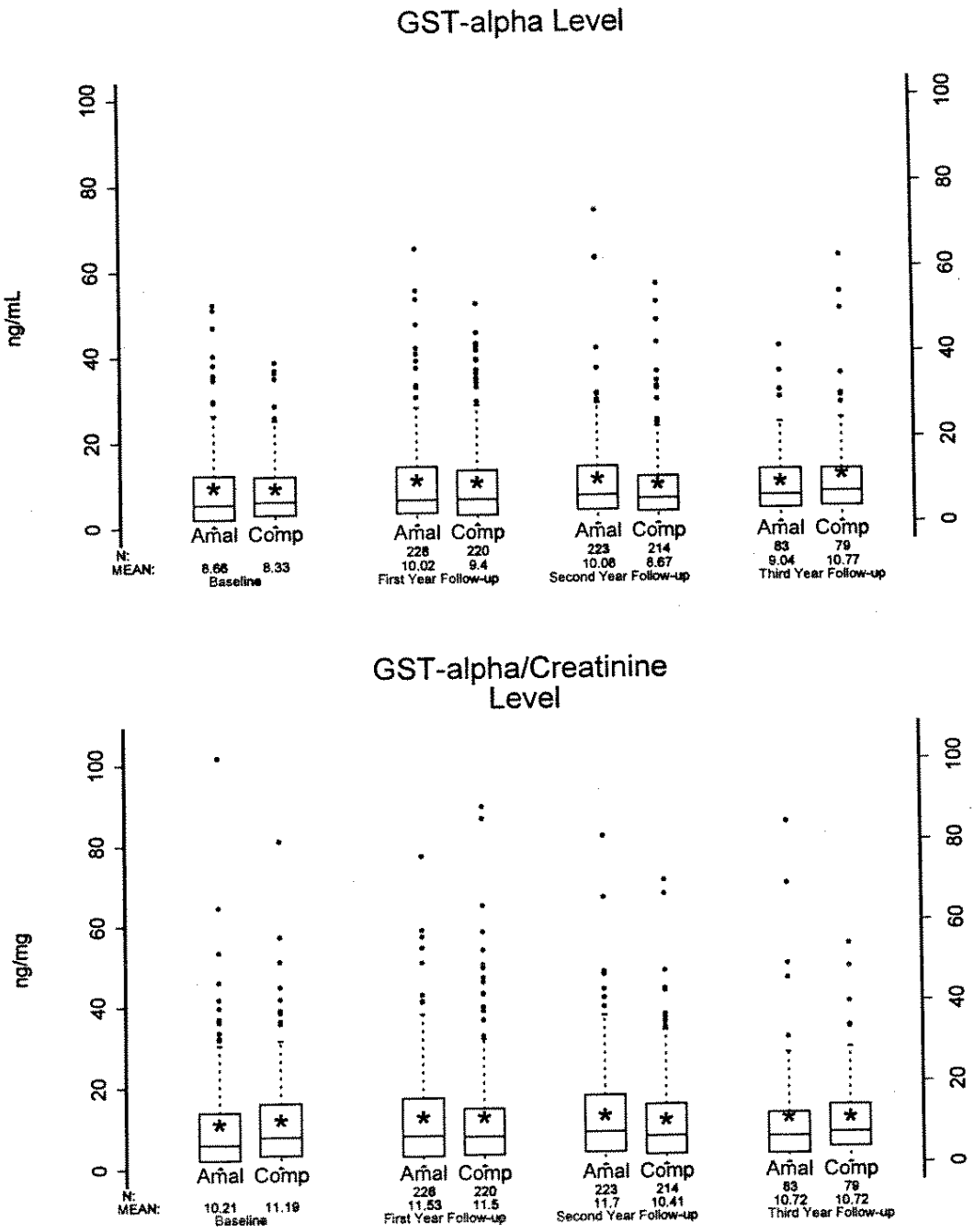


Figure 5.4.4. Distribution of GST-alpha with and without creatinine adjustement by time period and treatment group. The mean is marked by X, 25th and 75th percentiles by lower and upper edges of the box, median by a line inside the box and tails extend to minimum and maximum values. Outliers are plotted by individual points.

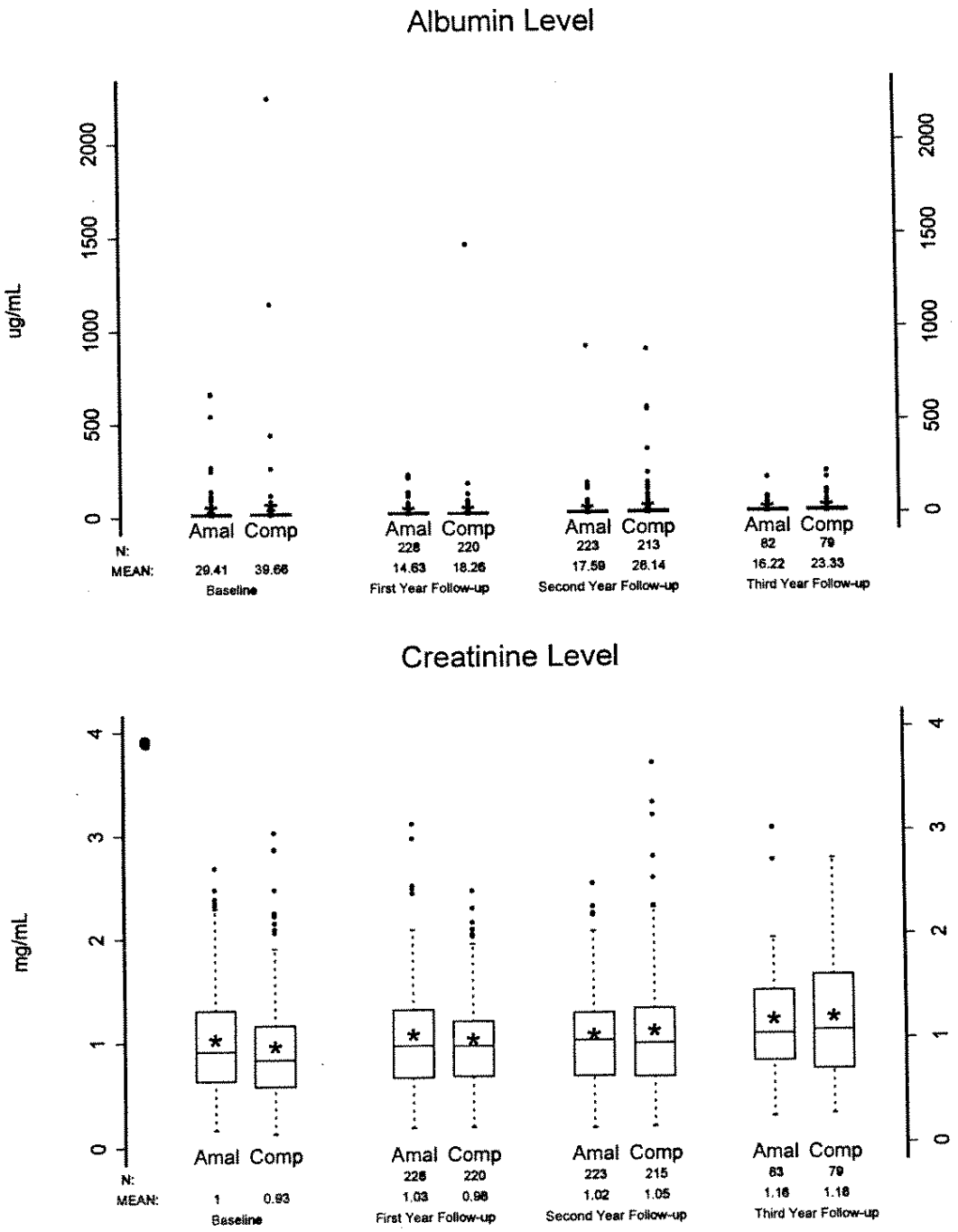


Figure 5.4.5. Distribution of albumin and creatinine levels by time period and treatment group. The mean is marked by X, 25th and 75th percentiles by lower and upper edges of the box, median by a line inside the box. Tails extend to minimum and maximum values; outliers are plotted by individual points.

5.4.3. Monitoring Children with Elevated Blood Lead Values

Baseline blood lead values of some children had to be re-calibrated by the laboratory that performed the analyses. One of the adjusted values was 16 ug/dl, which exceeds our eligibility criterion of 15 ug/dL. A follow-up blood lead evaluation yielded a value of 28 ug/dL, and a third blood sample gave a value of 14 ug/dL. This child was referred to the Casa Pia physician and the parents were informed of the child's status, and additional medical follow-up was recommended. Although this child has not experienced large decreases in test performance, an additional follow-up neurological exam was requested. Additional blood samples were requested of 11 other children with blood lead values larger than 10ug/dl (but less than 15ug/dl), including one child who was included in the group of 22 children with decreases in test performance based on the reliable change index (Section 4.2); the re-test values for 7 of these subjects have been completed and all were less than 10 ug/dL. Examination of the raw test data for children with elevated blood lead values (data will be available at the meeting) indicate that there are no clinical concerns for these children. They will continue to be monitored in future follow-ups.

6. Quality Assurance Monitoring

6.1. Completeness of data

Second-year follow-up has been completed and third-year follow-up is on-going. Information on loss to follow-up and missed visits was given in Section 2.2. This section provides information on item-specific missing data for children who were examined, item-specific non-response on health history questionnaires received, and information on completeness of laboratory data.

Neurobehavioral: 1) data are missing for the stroop tests for 25 subjects examined at baseline, 15 at first year, and 4 at 2nd year follow-up; 2) standard reaction time results are not available for 2 subjects at 1st year and 43 subjects at 2nd year due to equipment failure; 3) small numbers (1-4) of isolated missing values for WRAVMA Draw, WRAVMA Peg and Finger Tapping scores.

Neurologic: baseline adventitious movements recorded for 385 subjects (started March 1997), item-specific missing values: 0-10 subjects (0-2%) at baseline, 0-1 subjects at 1st, 2nd-year follow-up.

Health History Questionnaire: item-specific missing values for 0-12 subjects (0-3%) on individual items.

Laboratory: Completeness of urinary mercury analyses was given in Table 2.1.1. Completeness of other laboratory analyses of samples is described in Table 6.1.1. Incompleteness of laboratory data is due to lack of follow-up of subjects (as described in Table 2.2.1), change in protocol (albumin analyses began in October 1997), and temporary instrument failures (porphyrins) or problems with assays (GST).

Table 6.1.1. Number of subjects with laboratory data.

	Baseline	First-Year Follow-up	Second-Year Follow-up	Third-Year Follow-up
GST - Alpha	369	455	442	163
GST – Pi	484	437	422	73
Albumin	286	455	441	162
Creatinine	501	455	443	163
Porphyrins	408	285	318	67

6.2. Timeliness of data entry

During the second year of follow-up, data from neurologic and nerve conduction velocity exams has consistently been entered on the date of the exam. Neurobehavioral test forms are re-scored prior to entry; the typical lag time is 3-7 days between exam and entry date. Health history questionnaires are entered on the date they are received. Oral exam and dental treatment forms are entered on the date they are given to the Study Records Manager by the Lead Dentist.

6.3. Accuracy of data

There are four different checks on the accuracy of the data: 1) range checks built into the data entry system, 2) random audits of the computerized database, 3) consistency checks, and 4) checks of unusual data values. The main study data are entered into Access databases in Lisbon with built-in range checks to ensure that only valid data values are entered. Ranges of all data are checked before analysis as an extra precautionary measure. Approximately 15% of the subjects are selected at random for comparison of all data forms against the computerized database. Results for data audits up to and including 2nd-year follow-up are presented in Table 6.3.1. Audits of third-year follow-up have not been completed yet.

Table 6.3.1. Results of random data audits.

	Baseline	First-Year Follow-up	Second-Year Follow-up
Number of Subjects	74	78	70
Number of Data Items Checked	33522	20555	25791
Number of Mismatches	132	32	19
Mismatch Rate	0.39%	0.16%	0.074%

Consistency checks are performed to ensure that data values collected on the same subject over time are not inconsistent with each other. For example, teeth are followed through the annual oral exam and dental treatment databases to make sure that teeth are not misrecorded as being missing or extracted. Large changes in primary end-points or urinary Hg values are double checked for accuracy. When an error in the computerized database is identified, the information is transmitted to the Study Records Manager in Lisbon who identifies the source of the error and makes any necessary correction on the original data forms and in the computer database. The corrected version of the database is transmitted to Seattle for subsequent analyses.

6.4. Examiner reliability

The Senior Psychometrist continues to make annual visits to Lisbon to re-calibrate the examiners and to evaluate their performance using videotaped neurobehavioral testing sessions for a sample of subjects. The results are provided in the table below as the number of items checked and the percentage of items that are correctly performed.

Table 6.4.1. Accuracy of neurobehavioral testing.

	Examiner 1			Examiner 2		
Year of Session	Number of Sessions	Number of Items	Accuracy (%)	Number of Sessions	Number of Items	Accuracy (%)
1997	4	599	95	5	785	97
1998	7	1113	98	10	1576	98
1999	3	488	98	4	624	98
2000	5	773	99.6	5	785	99.5

We have also continued to monitor the balance of the treatment groups by examiner (Table 6.4.2) and differences between examiners during follow-up testing (Table 6.4.3).

Table 6.4.2. Distribution of neurobehavioral examiner by treatment group and follow-up year.

	Baseline		First-Year Follow-up		Second-Year Follow-up		Third-Year Follow-up	
	Group A	Group B	Group A	Group B	Group A	Group B	Group A	Group B
Examiner 1	105	110	102	104	110	112	47	40
Examiner 2	148	144	139	133	127	120	49	53

The differences between examiners for 2nd and 3rd year follow-up testing are presented in Table 6.4.3. At baseline testing, consistent examiner differences existed, with one examiner scoring higher on most test items. These consistent differences do not appear in the 2nd and 3rd year follow-up testing, although differences do exist on some test items.

Table 6.4.3. Examiner differences in neurobehavioral test results at 2nd and 3rd year follow-up. Highlighted means for one examiner are significantly better than the other examiner's mean. (indicates items for which a lower score is better; for all others, higher is better)**

	Second-Year Follow-up					Third-Year Follow-up				
	Examiner 1		Examiner 2		P-value	Examiner 1		Examiner 2		P-value
Variable	Mean	Std Dev	Mean	Std Dev		Mean	Std Dev	Mean	Std Dev	
Coding	9.40	3.11	9.32	3.21	0.794	9.34	3.60	8.89	3.12	0.365
WRAML DE VI	6.43	3.18	7.68	3.09	0.000	6.50	3.16	7.60	2.92	0.014
Digit Span	7.47	2.51	7.42	1.99	0.820	6.95	2.56	7.25	2.52	0.418
WRAVMA Drawing	84.40	10.41	90.68	8.62	0.000	84.33	9.28	88.44	7.16	0.001
Finger Windows	8.05	2.26	7.33	1.96	0.000	7.87	2.33	7.26	2.30	0.075
Finger Tapping-Dominant	41.57	5.84	42.15	5.69	0.278	44.27	6.09	43.09	5.68	0.173
Finger Tapping - Non-Dominant	36.45	6.08	36.02	5.59	0.434	38.21	5.48	37.90	5.69	0.707
WRAVMA Matching	97.61	11.95	99.96	11.97	0.034	96.40	13.73	102.14	11.62	0.002
WRAVMA Peg - dominant	116.84	17.20	110.98	14.81	0.000	118.26	14.27	112.24	17.82	0.012
WRAVMA Peg - non-dominant	114.43	16.24	112.44	13.52	0.150	117.43	13.63	113.72	14.61	0.076
RAVLT T.8	8.16	2.92	8.42	2.67	0.316	9.22	2.83	7.87	2.57	0.001
RAVLT TOT	40.24	8.56	40.09	8.95	0.853	42.52	8.17	40.51	8.45	0.100
Standard Reaction Time (mean)**	0.83	0.17	0.90	0.17	0.000	0.84	0.15	0.91	0.15	0.010
STROOP COL4	39.91	7.15	38.99	6.17	0.136	39.72	6.57	37.53	6.20	0.020
STROOP CWD4	42.01	7.30	41.93	6.88	0.905	41.74	8.33	41.06	7.04	0.542
STROOP WD4	40.58	7.22	39.72	6.79	0.184	39.05	7.35	38.24	6.16	0.411
Symbol Search	9.77	2.71	10.14	3.10	0.175	10.09	2.95	10.18	3.31	0.854
Trails A (seconds)**	19.49	8.90	19.87	7.74	0.618	19.00	7.30	18.56	6.60	0.664
Trails B (seconds)**	38.41	16.06	42.61	19.41	0.011	36.50	17.05	38.95	15.75	0.307
WRAML Visual Learning	7.93	2.89	8.78	2.79	0.001	7.72	2.77	8.86	2.74	0.005
Attention Domain Score	0.04	0.68	-0.04	0.66	0.166	0.05	0.71	-0.04	0.67	0.335
Memory Domain Score	-0.10	0.79	0.09	0.78	0.009	-0.01	0.79	0.00	0.79	0.922
Visual Motor Domain Score	-0.01	0.62	0.00	0.59	0.870	0.01	0.63	-0.03	0.64	0.717

7. Changes to Study Protocol

7.1. Additions to the Sentinel Event Monitoring System

The occurrences of serious adverse health events, called "Sentinel Events", are identified through a short health questionnaire completed annually by the parents or guardians of the subjects. This was added to the protocol at the suggestion of the DSMB, and the occurrence of these events is reported in our annual DSMB report. We have made efforts to improve the information collected on sentinel events in two ways, utilizing the community hygienists who work closely with the schools. First, the hygienists will monitor the attendance records of children in the study, identifying any extended absences and inquiring through teachers as to any health reasons associated with such absences, and whether any hospitalizations occurred or any serious illnesses were diagnosed. Also, at the time of the annual neurological exam, the community hygienist will interview each child about their health history during the previous year, prompted by a structured

checklist of potentially relevant signs and symptoms (Appendix D). For any signs and symptoms experienced, it will be determined if they were serious enough to qualify as a sentinel health event. The use of these three mechanisms will most likely identify the occurrence of any serious health events.

The Casa Pia school physician will serve as a medical consultant to the study, and will assist the Lead Dentist and the community hygienists in interpreting the health implications of any medications reported and in characterizing the nature of any reported illnesses. For any reported illness or diagnosis, the name of the treating physician will be obtained and recorded, along with the places and dates of any hospitalizations, all for potential future reference. In case of a significant number of reported episodes of a similar nature, permission will have to be sought to obtain medical and hospital records in order to verify the diagnosis and nature of the illness, since the consent given for this study does not include authorization to obtain such records.

On a monthly basis, a summary of all information collected during that period that was determined to potentially constitute the presence of a sentinel event, or that may require further investigation is forwarded to Dr. Martin. He reviews this information, seeks consultation from appropriate Seattle collaborators (i.e., Dr. Townes or Dr. Sarnat for a neurological problem) and provides feedback to the Lisbon site with recommendations for followup.

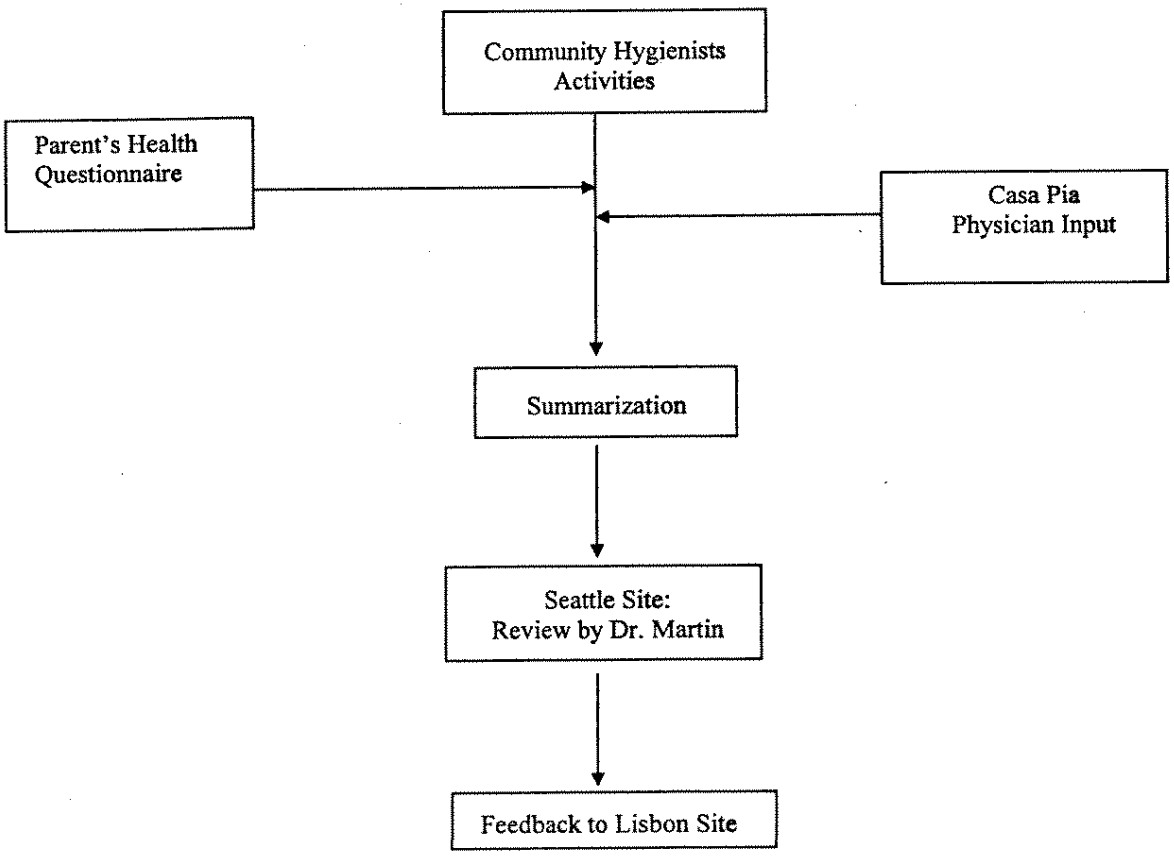


Figure 7.1.1. Sentinel health event monitoring system.

7.2. Extension of follow-up to 7 years.

In the competing renewal proposal recently reviewed and recommended for funding, it was proposed that the follow-up period be extended to 7 years after intitial treatment. Power calculations were provided assuming either 5 or 7 years of follow-up. The annual attrition rate was 5% in the first year, and assumed to be 3% subsequently. The results showed that the study is on track to have sufficient power with a five-year follow-up period to detect small effects that are roughly uniform across end-points. For instance, the power is 94% to detect cumulative adverse amalgam effects over five years equal to 0.3 for memory, attention, and visual-motor, and 0.15 for neurological. The power to detect adverse effects of composite material of these magnitudes is smaller but adequate (87%). The original grant application suggested that seven years of follow-up may be necessary to detect small effects that occur in only one end-point. The power to detect such effects was increased by dropping two end-points and adopting the Hotelling test component; these changes produced adequate power (94-95%) to detect a linear effect in just one end-point equal to 0.1 per year over five years. However, five-year follow-up is still inadequate to detect a linear effect in just one end-point equal to 0.06 per year (cumulative effect size 0.3 over 5 years or 0.42 over 7 years). The power is just 49% with five years of follow-up but increases to 85% with seven years of follow-up. Proposed signficance levels for the statistical stopping rule with 7 years of follow-up are given in Table 7.2.1.

Table 7.2.1 Significance levels for annual analyses and total significance levels.

Test Component	1999	2000	2001	2002	2003	2004	2005	Total
Test for adverse effects of amalgam	0.0075	0.0009	0.0009	0.0009	0.0009	0.0009	0.018	0.03
Test for adverse effects of composite	0.0025	0.0003	0.0003	0.0003	0.0003	0.0003	0.006	0.01
Hotelling's T-squared test	0.0025	0.0003	0.0003	0.0003	0.0003	0.0003	0.006	0.01
Overall Test	0.0125	0.0015	0.0015	0.0015	0.0015	0.0015	0.030	0.05

7.3. Modification of the neurobehavioral test items

Based on our experience with the current battery of neurobehavioral tests, and in order to accommodate the maturation of our study population, some modifications were proposed for the neurobehavioral test battery in the competing renewal. These changes are designed to eliminate unnecessary tests, and to provide continuity over time by introducing equivalent tests for older subjects as study participants reach the age limit of existing tests. The following modifications were approved in the review of our renewal application: 1) replace specific neurobehavioral tests with tests for the same purpose for older subjects; 2) drop two tests currently being administered because they have not been found to add sufficient additional information to warrant continuation in the test battery; and 3) add new tests as secondary endpoints, including computer-based neurobehavioral tests to take advantage of recent advances in technology and recent discoveries. The schedule for these changes is given in Table 7.3.1. The changes are introduced briefly below, and more detail on the new tests is given in Appendix E.

Replacement of tests for older subjects: At baseline, the study subjects were between 8 and 12 years of age. They will be 15 to 19 years old at the end of the proposed 7-year follow-up period. To assure that tests are age appropriate, some tests will be replaced with late adolescent/adult versions of the currently used tests as follows: 1) adult Trail Making test will be substituted for the child version; 2) adult Digit Span test will be substituted for the child version; 3) the Visual Reproduction subtest of the Wechsler Memory Scale-Revised will be substituted for the

WRAML: Visual Learning subtest; 4) the Matrices subtest from the Wechsler Abbreviated Scale of Intelligence will be substituted for the WRAVMA Matching subtest; 5) the Wechsler Memory Scale-III Spatial Span subtest will be substituted for the WRAML Finger Windows subtest; and 6) the written Symbol Digit test will be substituted for the Coding test. The late adolescent/adult versions of these tests will be co-administered with the child versions at the 4th-year follow-up and then substituted for the child versions at the 5th through 7th year of follow-up (Table 7.3.1). Where age appropriate norms are available through age 19, the tests currently in use will not be replaced: Rey Auditory Verbal Learning Test, Finger Tapping Test, Pegs, Stroop Test and Reaction Time. All tests are standard published tests used in clinical practice (Lezak, 1995[2]; Spreen & Strauss, 1998[3]). The correlation between the corresponding child and adult tests would be assessed in the 4th –year follow-up data. Provided the correlations were sufficiently high (above 0.90); the adult tests would be used in the calculation of the primary end-points (domain Z-scores) for follow-up years 5 to 7.

Tests to be dropped: We plan to drop the WRAVMA drawing test in the Visual/Motor domain, and the Wisconsin Card Sorting Test. The drawing test is the most subjective of the tests used and peaks in adolescence. The Wisconsin test is neither a primary or secondary endpoint, and was included originally because it is being used at the NERI site. It has proven to provide a high level of frustration for the participants in this study, and had the potential to discourage participation. Based on recommendations from the renewal application review, and in consultation with NIDCR, this test will be discontinued until the final follow up year (07), at which time it will be given again. The dropping of these two tests will reduce administration time modestly to permit time for overlapping the child/adult tests as described above (Table 7.3.1).

New tests to be added as secondary endpoints: Two existing tests are proposed to be added: the Visual Analog Mood Scales (Stern, 1997[4]) beginning in the 4th-year of follow-up, and the Motor Steadiness Battery in the 5th-year of follow-up (with pre-testing in the 4th year). The rationale for adding the motor steadiness test is that decrements in performance on this test have been related to very low mercury exposure in adults (Bittner, Echeverria, Woods, et al, 1998[5]). The rationale for the Mood Scales is that mood changes have been associated with low mercury exposures (Haut, Morrow, Pool et al, 1999[6]). The Mood Scales test has been found useful for assessing mood states in non-English speaking clinical populations. Because we want to maintain consistency over the entire study period in the primary endpoints, the motor steadiness and mood tests would be secondary endpoints. In addition, the following four computerized tests are proposed for adaptation for adolescents and administration at the final two follow up evaluations (follow-up years 6 and 7): Visual Retention Test, Symbol Digit Test, Spatial Span and Trails B. These are computerized tests comparable to the individually administered tests in the neurobehavioral battery (see Table 7.3.1). The computerized tests have been found to be sensitive to mercury exposure in adults (Echeverria, Aposhian, Woods et al, 1998[7]). They would serve *as secondary endpoints* to the study as well as provide a potential bridge to link the effects, if any, of dental mercury from childhood into adulthood. Also, in response to recommendations from the renewal review, we will add the Wechsler Adult Intelligence Scale in the last follow-up year of the study because the NERI site is using IQ as its primary endpoint.

---

[2] Lezak, M.D. (1995). *Neuropsychological Assessment (3rd Ed.)* New York: Oxford University Press

[3] Spreen, O & Strauss, E. (1998). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary (2nd Ed.)* New York: Oxford University Press.

[4] Stern, R. (1997). Visual Analogue Mood Scale. Odessa, FL: Assessment Resources Inc.

[5] *Neurotoxicology and Teratology*, 20, 429-439, 1998.

[6] *Applied Neuropsychology*, 6, 193-200, 1999.

[7] *The FASEB Journal*, 12, 971-980, 1998.

Table 7.3.1. Proposed Schedule of Neurobehavioral Tests by Follow-up Year.

Domain/ Test	Follow-up Year			
	4	5	6	7
Memory Domain				
Rey Auditory Verbal Learning	x	x	x	x
WRAML: Visual Learning	x			
WMS-R: Visual Reproductions	x	x	x	x
Visual/Motor Domain				
WRAVMA: Draw				
WRAVMA: Match	x			
WAISI: Matrices	x	x	x	x
Finger Tapping (Dominant, Non-dom.)	x	x	x	x
Pegs (Dominant, Non-dom.)	x	x	x	x
Attention Domain				
Digit Span	x			
Digit Span (Adult)	x	x	x	x
Reaction Time	x	x	x	x
Stroop	x	x	x	x
Coding	x			
Symbol Digit (Written)	x	x	x	x
WRAML: Finger Windows	x			
WMS-III: Spatial Span	x	x	x	x
Trails A, B (Intermediate)	x			
Trails A, B (Adult)	x	x	x	x
Non-Domain Tests (2^0 endpoints)				
Wisconsin Card Sorting				x
WAIS				x
Mood (Not part of a domain)	x	x	x	x
Motor Steadiness	pre-test	x	x	x
Computer-Based Tests (2^0 endpoints)				
Visual Retention (Memory Domain)	develop	pre-test	x	x
Symbol Digit (Attention Domain)	develop	pre-test	x	x
Spatial Span (Attention Domain)	develop	pre-test	x	x
Trails B (Attention Domain)	develop	pre-test	x	x

8. Appendix A. Additional Figures

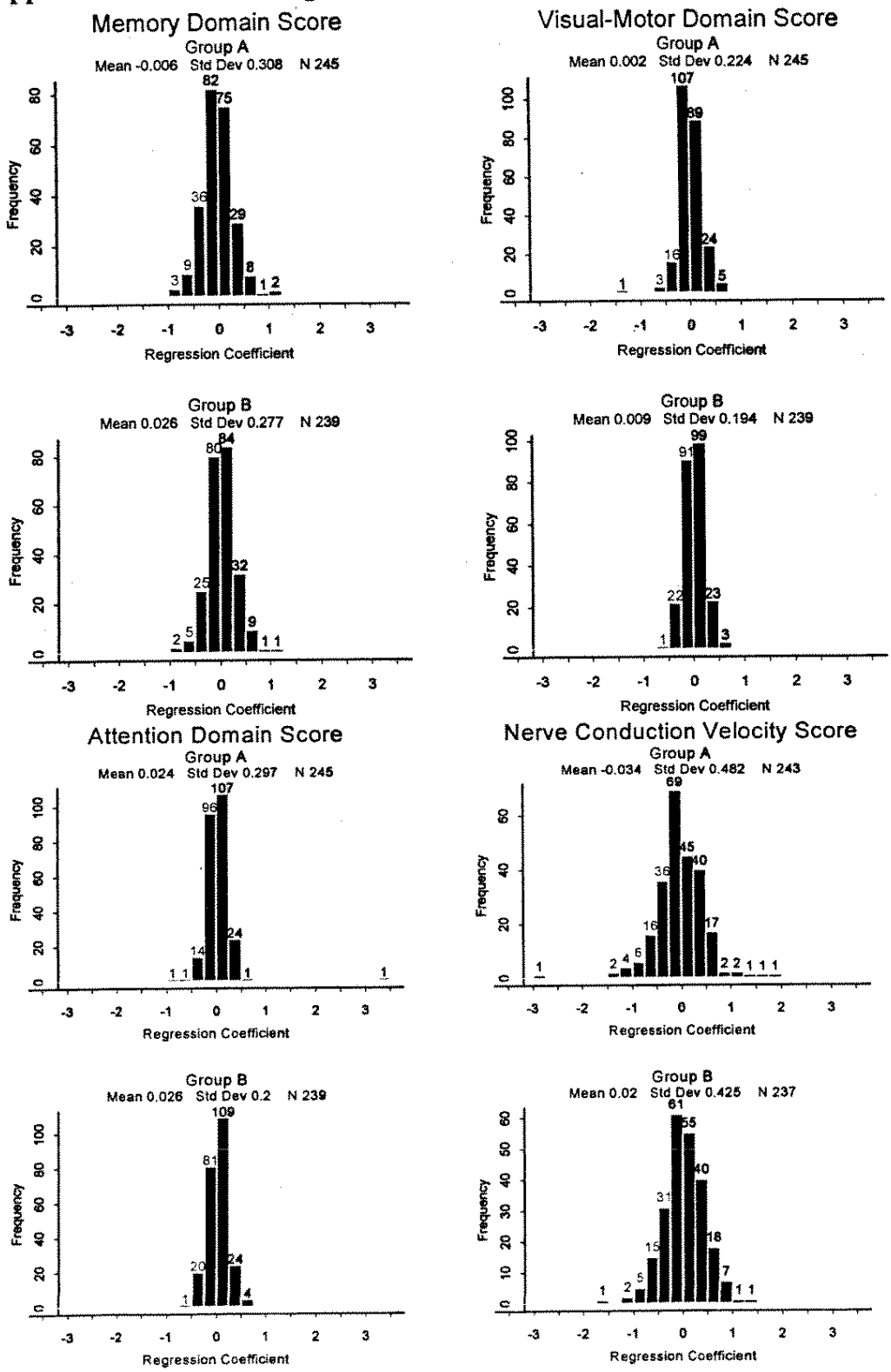


Figure A.1. Distribution of annual rates of change of neurobehavioral and nerve conduction velocity Z-scores (regression coefficients of Z-scores versus time since randomization).

9. Appendix B. Additional Tables.

Table B.1. Comparison of treatment groups on demographics.

	Group A		Group B	
	Number	Percent	Number	Percent
Year of Birth:				
1986	90	35.6	74	29.1
1987	85	33.6	93	36.6
1988	54	21.3	55	21.7
1989	24	9.5	32	12.6
Sex:				
Male	135	53.4	142	55.9
Female	118	46.6	112	44.1
Race:				
White	176	69.6	179	70.5
Black	75	29.6	67	26.4
Asian or Other	2	0.8	8	3.2

Table B.2. Comparison of treatment groups on primary endpoints at baseline.

	Group A				Group B			
	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
Memory Z-Score	0.01	0.77	-2.04	2.12	-0.01	0.78	-1.85	2.03
Visual-Motor Z-Score	0.00	0.65	-2.60	1.64	0.00	0.59	-1.50	1.75
Attention Z-Score	0.02	0.60	-1.88	1.44	-0.04	0.58	-1.68	1.56
Nerve Conduction Velocity Z-Score	0.04	0.74	-2.04	2.16	-0.04	0.80	-3.28	2.18
Ulnar Velocity (m/sec)	59.57	6.39	42.40	80.60	58.75	6.51	33.70	78.80
Tibial Velocity (m/sec)	51.12	5.29	33.40	70.00	51.00	5.58	36.80	74.00

Table B.3. Norms for neurobehavioral test scores.

	Age	Norm	Std. Dev.
IQ	9-11	100	15
Memory Domain:			
RAVLT T.8	9	7.0	3.0
	10	7.5	3.0
	11	8.0	3.0
RAVLT TOT	9	37.9	9.5
	10	39.8	9.5
	11	41.3	9.5
WRAML Visual Learning	9-11	10.0	3.0
WRAML DE VI		NA	NA
Visual Motor Domain:			
WRAVMA Drawing	9-11	100	15
WRAVMA Matching	9-11	100	15
WRAVMA Peg –dominant	9-11	100	15
WRAVMA Peg Non-dom	9-11	100	15
Standard Reaction Time		NA	NA
Finger Tapping-Dominant	9	33.9	4.12
	10	37.4	5.43
	11	40.9	4.79
Finger Tapping Non-Dom	9	30.2	3.27
	10	33.2	4.86
	11	36.5	4.75
Attention Domain:			
Coding	9-11	10	3
Symbol Search	9-11	10	3
Digit Span	9-11	10	3
Finger Windows	9-11	10	3
Trails A (seconds)	9	25.1	10.5
	10	22.6	6.6
	11	20.9	6.3
Trails B (seconds)	9	54.9	22.0
	10	52.7	26.3
	11	40.5	16.1
STROOP WD4	9-11	50	10
STROOP COL4	9-11	50	10
STROOP CWD4	9-11	50	10
Nerve Conduction Velocity (m/sec):			
Ulnar Nerve		67.6	6.0
Tibial Nerve		57.6	7.3

Table B.4. Comparison of treatment groups on neurobehavioral test scores at baseline and 1st year follow-up.

Variable	Baseline					First-Year Follow-up				
	Group A		Group B		P-value	Group A		Group B		P-value
	Mean	Std Dev	Mean	Std Dev		Mean	Std Dev	Mean	Std Dev	
Coding	9.06	3.15	8.63	3.08	0.125	9.05	2.91	9.03	3.23	0.943
WRAML DE VI	6.49	3.15	6.56	3.04	0.802	6.93	2.90	6.91	3.23	0.949
Digit Span	7.37	2.72	7.37	2.54	0.995	7.25	2.31	7.28	2.23	0.903
WRAVMA Drawing	100.60	13.51	101.50	11.39	0.420	94.60	12.81	93.42	11.08	0.284
Finger Windows	7.32	2.34	7.28	2.47	0.821	7.18	2.17	7.17	2.40	0.963
Finger Tapping-Dominant	36.66	6.17	36.29	6.05	0.493	39.15	5.91	39.01	5.89	0.788
Finger Tapping - Non-Dominant	32.02	5.34	31.33	5.37	0.150	33.76	5.41	33.66	5.58	0.847
WRAVMA Matching	95.58	13.72	96.16	12.41	0.618	98.74	13.41	98.68	12.56	0.955
WRAVMA Peg – dominant	101.76	17.23	103.01	16.68	0.406	108.45	16.58	109.60	15.48	0.435
WRAVMA Peg - non-dominant	105.89	15.08	106.79	15.05	0.499	109.39	14.23	110.72	14.38	0.311
RAVLT T.8	8.36	2.91	8.10	3.07	0.334	7.59	2.39	7.87	2.70	0.224
RAVLT TOT	38.98	10.21	37.88	9.77	0.215	37.37	8.27	38.04	8.19	0.378
Standard Reaction Time (mean)	0.89	0.21	0.89	0.20	0.741	0.89	0.17	0.89	0.18	0.949
STROOP COL4	44.12	6.02	43.02	5.67	0.040	41.08	6.56	40.57	6.23	0.392
STROOP CWD4	44.15	6.99	43.29	6.87	0.173	42.67	7.25	42.70	6.92	0.956
STROOP WD4	42.16	6.55	41.53	6.40	0.286	41.87	7.12	41.46	7.07	0.539
Symbol Search	9.42	2.65	9.41	2.58	0.981	9.59	2.85	9.57	2.89	0.941
Trails A (seconds)	27.91	12.80	27.69	13.05	0.848	22.00	8.72	22.92	8.20	0.238
Trails B (seconds)	65.25	34.41	64.98	35.74	0.932	51.24	30.59	52.52	27.78	0.633
WRAML Visual Learning	7.81	2.67	8.13	2.75	0.180	8.29	2.58	8.31	2.88	0.944

Table B.5. Comparison of treatment groups on neurobehavioral test scores at 2nd and 3rd year follow-up.

Variable	Second-Year Follow-up					Third-Year Follow-up				
	Group A		Group B		P-value	Group A		Group B		P-value
	Mean	Std Dev	Mean	Std Dev		Mean	Std Dev	Mean	Std Dev	
Coding	9.45	3.21	9.26	3.11	0.509	9.39	3.31	8.79	3.38	0.227
WRAML DE VI	6.98	3.10	7.20	3.28	0.448	7.22	3.09	6.97	3.07	0.576
Digit Span	7.38	2.17	7.51	2.33	0.536	6.95	2.67	7.29	2.39	0.355
WRAVMA Drawing	86.96	10.52	88.46	9.42	0.105	86.66	9.03	86.44	7.82	0.861
Finger Windows	7.51	2.22	7.84	2.04	0.099	7.26	2.43	7.84	2.19	0.090
Finger Tapping-Dominant	42.03	5.84	41.71	5.69	0.550	43.39	5.63	43.88	6.16	0.570
Finger Tapping - Non-Dominant	36.20	5.49	36.26	6.15	0.912	37.37	5.51	38.74	5.59	0.097
WRAVMA Matching	98.92	12.13	98.78	11.90	0.906	98.92	13.66	100.10	12.15	0.532
WRAVMA Peg - dominant	114.17	16.61	113.36	15.88	0.593	115.96	16.67	114.08	16.38	0.436
WRAVMA Peg - non-dominant	113.19	14.40	113.59	15.40	0.772	115.36	14.45	115.53	14.11	0.933
RAVLT T.8	8.24	2.86	8.35	2.73	0.674	8.21	2.81	8.78	2.71	0.153
RAVLT TOT	39.91	8.74	40.43	8.79	0.521	41.27	8.09	41.60	8.67	0.786
Standard Reaction Time (mean)	0.87	0.18	0.86	0.16	0.375	0.89	0.17	0.87	0.14	0.541
STROOP COL4	39.45	6.89	39.40	6.43	0.934	38.43	6.93	38.64	5.94	0.821
STROOP CWD4	42.35	7.55	41.57	6.54	0.235	41.27	7.96	41.48	7.34	0.853
STROOP WD4	40.08	7.25	40.17	6.74	0.885	38.30	7.09	38.92	6.34	0.528
Symbol Search	9.93	2.94	10.00	2.91	0.791	10.61	3.30	9.65	2.90	0.033
Trails A (seconds)	19.94	8.45	19.44	8.16	0.514	18.95	7.33	18.57	6.48	0.705
Trails B (seconds)	39.60	15.22	41.66	20.45	0.218	37.50	15.72	38.17	17.08	0.778
WRAML Visual Learning	8.35	2.76	8.40	2.98	0.861	8.29	2.62	8.39	3.00	0.808

10. Appendix C. Description of the Reliable Change Index Method for Identifying Children with Decreases in Neurobehavioral Test Performance.

The calculations for the reliable change index are given below:

R = reliability coefficient (from the inter-rater reliability study reported previously)

r = correlation between scores at first and second assessments

SD = standard deviation of scores at second assessment

$SEP = SD\ (\ 1 - r^2\ )^{1/2}$

M = sample mean of the scores in the first assessment

Y_1 = observed test score in the first assessment

Y_2 = observed test score in the second assessment

$Y_{TRUE} = M + R\ (\ Y_1 - M\ )$

A reliable change is declared if Y_2 does not lie in the prediction interval $Y_{TRUE} \pm 1.645$ SEP. Reliable changes are interpreted as decreases or increases accordingly, depending on whether Y_2 falls above or below the interval, and on whether higher or lower scores represent improvements in performance for the particular test.

11. Appendix D. Sentinel Health Events Checklist: English Version

Follow-up Year: _____ **Date:** _____

Name: _____ **School:** _____ **Class:** _____

Instructions: Complete the checklist with the subject. Make sure the subject understands all terms used in the checklist. For any items which produce a "yes" response, determine if the problem was serious enough to require a visit to a doctor or hospital. If so, contact parents or guardians to find out reported diagnosis, doctor and/or hospital names and dates, and record at the bottom of this form.

Symptom	Yes	No	Don't know
Diarrhea			
Hand tremors			
Loss of balance while walking			
Speaking difficulties			
Vision problems			
Numb hands or feet			
Feeling sad			
Irritated			
Nervous			
Headaches			
Dizziness			
Hear humming			
Difficulties in hearing			
Difficulty remembering things			
Frequent cough			
Allergies			
Asthma			
Skin diseases			
More colds than usual			
Feeling weak			
Pain in the joints			
Restlessness			
Convulsions			

12. Appendix E. Description of the proposed new neurobehavioral tests.

Matrix Reasoning, WASI: The Matrix Reasoning Test is a series of 35 incomplete gridded patterns that the examinee completes by pointing to or stating the number of the correct response from 5 possible choices. Matrix Reasoning is a measure of the nonverbal fluid reasoning and general intellectual ability.

Visual Reproductions, Wechsler Memory Scale-Revised: The subject must reproduce from memory geometric designs which he has been exposed to for ten seconds. After 30 minutes, he is asked to reproduce the designs again from memory.

Spatial Span, Wechsler Memory Scale-III: The subject is asked to replicate an increasingly long series of visually presented spatial locations. The examiner points to a series of blocks at a rate of one block per second and asks the examinee to point to the same blocks in the same order (Spatial Span Forward). In the second part of the subtest, the examiner points to a series of blocks and asks the examinee to point to the same blocks in the reverse order (Spatial Span Backward). The Spatial Span subtest taps an examinee's ability to hold a visual-spatial sequence of locations in working memory and then reproduce the sequence.

Digit Span, Wechsler Memory Scale-III: The examinee is asked to repeat a series of digits forward (Digit Span Forward) and then repeat a series of digits backward (Digit Span Backward). The series begins with easy, two number, recalls and progresses to much longer sequences.

Symbol Digit Modalities: The SDMT is a test of visual attention and concentration. The examinee is asked to look at a row of boxes with symbols in the upper part of the box and a number underneath each symbol. He is then shown the rest of the boxes where the symbols appear in the upper box, but the space beneath is empty. He is asked to copy the correct number below each symbol for 90 seconds.

Motor Steadiness Battery: The subject is asked to use a stylus and move it through a maze as quickly as he can without hitting the sides of the maze. The number of times the stylus hits the side of the maze is tabulated, as is the length of time the stylus stays on the side of the maze. The subject is asked to place the stylus in a series of holes that decrease in size. He holds the stylus in the hole for several seconds, trying not to hit the sides of the hole.

Trail Making Test: This test measures planning ability, visual-motor speed and concentration. Trail Making B also requires cognitive flexibility and sufficient short-term memory capacity in order to monitor two sequences simultaneously. On Trails A, the subject is asked to draw a line connecting numbers in order from 1-25. In Trails B, the subject is asked to draw a line sequencing numbers and letters in order, i.e., 1 to A, A to 2, 2 to B, B to 3, 3 to C, and so on, until he reaches the end at L to 13.

Visual Analog Mood Scales (VAMS): Subjects are given a set of 8 scales, each comprised of a drawing of two faces; the first face is neutral, the second is afraid (confused, sad, angry, energetic, tired, happy, tense). Each pair of faces (e.g., neutral-afraid; neutral-happy) is joined by a line 4 inches in length. For each scale, subjects make a mark across the line at the point which best describes how he feels. Scores range from 0 (neutral) to 100 (very afraid; confused; energetic, etc.).

13. Appendix F. Abstracts and Papers from the Main Study

Amalgam exposure and urinary mercury concentration in children

M.D. Martin[1]*, J.S. Woods[1], M. Bernardo[2], N. Hawkins[1], T. DeRouen[1], J. Leitão[2], N. Braveman[3]

(University of Washington, Seattle[1]; University of Lisbon, Lisbon[2]; NIDCR[3])

ABSTRACT

The Casa Pia Study is a large clinical trial (a component of the Children's Amalgam Trial) sponsored by the NIDCR to examine the health effects of mercury amalgam exposure in children. Children were enrolled at age 8 to **12** (n = 507), randomly assigned to either amalgam or alternate materials treatment group, and will be periodically evaluated over time for neurobehavioral, neurological or renal effects. This trial offers a unique opportunity to evaluate the effect of mercury amalgam exposure on urinary mercury (HgU) levels. First annual follow-up evaluations have been completed on **239** subjects in the amalgam treatment group, with repeat HgU sample taken at an average of **6.97** months post treatment completion. Initial treatment involved placing a mean of **9.4** surfaces of amalgam (SD **4.9**), yet the mean HgU value for the amalgam group increased by only a small (mean **0.94** ug/L) although statistically significant (**$p < 0.001$**) amount during this period. Mean first-year follow-up HgU levels remain relatively low at **2.50** ug/L. Total number of surfaces of amalgam provided during treatment was positively correlated (**$r = 0.18$, $p = 0.008$**) with (uncorrected) change in HgU from baseline to first-year follow-up value, as well as with creatinine-corrected changes (**$p = 0.01$**). **The number of occlusal surfaces alone was also marginally correlated with unadjusted HgU ($r=0.13$, $p = 0.048$).** Multiple linear regression showed that, after adjusting for gender, age, and race, the contribution of each surface of amalgam to the change in urinary Hg concentration was **0.084ug/L (95% C.I.: 0.0245, 0.143).** Despite these positive associations, the number of amalgam surfaces explained only **3%** of the variability in HgU changes. These clinical observations indicate that placement of a large number of amalgam restorations results in only a small increase in HgU, and while the amount of HgU increase is related to the number of amalgam restorations placed, the relationship is

2001 AADR ABSTRACT FORM—CHICAGO, IL

Mail form to: AADR, 1619 Duke Street, Alexandria, VA 22314-3406, USA **(FAX SUBMISSIONS WILL BE REFUSED.)**

Type perfect original of abstract here: **Deadline for submission: September 15, 2000**

(Also submitted to Society of Toxicology, March, 2001)

R# _____

Dietary methylmercury exposure in the Casa Pia Trial. C. Evens*, T.A. DeRouen, J.S. Woods, L. Liang, H. Luis, P.L. Simmonds, J. Leitao, M. Bernardo, M. Redford, M.D. Martin (Univ. of Washington, Seattle; Univ. of Lisbon, Lisbon, Portugal; NIDCR)

The Casa Pia Study of the Health Effects of Dental Amalgams in Children is a randomized clinical trial (n=507) designed to evaluate whether low-level elemental Hg exposure from amalgam has health consequences. A potential confounding variable is exposure to methylmercury (MeHg) from dietary consumption of fish and other seafood. This study examined MeHg concentrations in blood of a subset of 150 children over a one year period and related them to their diets. Parents or caregivers completed a food frequency questionnaire designed specifically for this study at baseline. Children provided urinary and blood samples for Hg determination both at baseline and at (post treatment) one-year follow-up. Hg levels in dietary fish samples were also obtained at baseline. The mean value of baseline MeHg in blood was 2.7 ng/g, and increased as the report of seafood consumption increased, although not significant statistically. Blood MeHg and total Hg concentrations were lower at one year follow-up than at baseline (p=.01 and <.01, respectively). These children consumed moderate amounts of fish and other seafood, with 63% of parents/caregivers reporting that their children consumed fish or other seafood once a week or less. The fish from meals at a sample of the schools contained low levels of MeHg (range 13.9-23.6 ng/g). Thus, these children were exposed to low levels of MeHg, resulting in the low baseline and one year followup mean blood MeHg levels in both amalgam and non-amalgam treatment groups, (baseline, 2.6 vs 2.8 ng/g, and one year followup, 2.2 vs 2.6 ng/g), and makes dietary MeHg exposure not of concern in the study. (N01 DE72623)

P# _____

For Office Use Only: (1) ☐ Special Scheduling (2) ☐ Poster Session (3) ☐ Oral Presentation

Presenter Information:

Type or print legibly in black or blue ink.)

1. Complete name and mailing address of the PRESENTER. All correspondence will be mailed to this address.

Name: Carina Evens

In care of: _____
Students, give cooperating instructor's name.)

Univ. / Co.: University of Washington

Dept.: Dental Public Health Sciences

Street Address: Box 357475

City: Seattle

State/Country: Washington, USA

ZIP/Postal Code: 98177

E-mail: cevens@u.washington.edu

2. Daytime Phone Number (include area code):
(206) 221-3785

3. FAX Number (include area code):
(206) 685-4258

4. Membership ID Number (if presenter is a member):
none

5. Area of Review (check only one):
(1) ☐ Behavioral Sciences/Health Services Research
(2) ☐ Behavioral Sciences/Epidemiological Methods
(3) ☐ Cariology Research
(4) ☐ Craniofacial Biology: I—Molecular
(5) ☐ Craniofacial Biology: II—Other Studies
(6) ☐ Dental Anesthesiology Research
(7) ☐ Dental Materials: I—Adhesion–Composite Bond Strength
(8) ☐ Dental Materials: II—Adhesion–Other
(9) ☐ Dental Materials: III—Ceramics and Cements
(10) ☐ Dental Materials: IV—Clinical Trials
(11) ☐ Dental Materials: V—Polymer Materials–Chemistry
(12) ☐ Dental Materials: VI—Polymer Materials–Mechanical Properties and Degradation
(13) ☒ Dental Materials: VII—Others–Metallic
(14) ☐ Dental Materials: VIII—Others–Non-metallic
(15) ☐ Diagnostic Systems
(16) ☐ Educational Research
(17) ☐ Experimental Pathology
(18) ☐ Geriatric Oral Research
(19) ☐ Implantology Research
(20) ☐ Microbiology/Immunology & Infection Control
(21) ☐ Mineralized Tissue
(22) ☐ Neuroscience/TMJ
(23) ☐ Nutrition
(24) ☒ Oral Health Research
(25) ☐ Oral & Maxillofacial Surgery
(26) ☐ Periodontal Research—Diagnosis/Epidemiology
(27) ☐ Periodontal Research—Pathogenesis
(28) ☐ Periodontal Research—Therapy
(29) ☐ Pharmacology, Therapeutics, & Toxicology
(30) ☐ Prosthodontics Research
(31) ☐ Pulp Biology
(32) ☐ Salivary Research

6. Mode of Presentation Preferred (check only one):
(1) ☐ Oral Presentation (2) ☒ Poster Presentation
(3) ☐ Poster Discussion (4) ☐ No Preference

7. Do you wish to withdraw your abstract if you do not get the mode of your choosing? (1) ☐ yes (2) ☒ no

8. Are you interested in being an oral session chair? (Must be an IADR member.)
(1) ☐ yes (2) ☒ no

9. Are you a student?
☐ Post-graduate ☐ Graduate ☐ Undergraduate
(Undergraduates: Have your faculty advisor sign here.)

_____

10. List five descriptors by number (see reverse side). If existing descriptors do not fit your research, then write one new word under "Other".
(1) 110 (2) 804
(3) 185 (4) 820
(5) 186 (write numbers)
Other _____

For American Students Only

11. Are you applying for a Competition? (Check one only if you meet the eligibility requirements listed in the instruction booklet.)

(1) ☐ AADR Hatton
☐ Junior
☐ Senior
☐ Post-doctoral

(2) ☐ Caulk/Dentsply
☐ Basic Science
☐ Clinical Science

12. Was this research supported by NIH? (If yes, print grant number as last line in abstract box.)
—NIH/NIDCR Employees check yes.
(1) ☒ yes (2) ☐ no

All Presenters Must Complete

13. The present scientific work may involve a real or perceived financial conflict of interest. (See page 5.)
☐ yes ☒ no If yes:

_____

_____

14. **I have proofread this abstract and understand the obligation of submission. (Presenter's Signature)**

_____

OVER ☞

Provided by Battelle Seattle Library
Information Services
(206) 528-3370
FAX (206) 528-3553

Journal of Analytical Toxicology, Vol. 24, July/August 2000

Determination of Methyl Mercury in Whole Blood by Ethylation–GC–CVAFS after Alkaline Digestion–Solvent Extraction

L. Liang[1,*], C. Evens[2], S. Lazoff[1], J.S. Woods[3], E. Cernichiari[4], M. Horvat[5], M.D. Martin[6], and T. DeRouen[2]

[1]Cebam Analytical, Inc., Seattle, Washington; [2]Dental Public Sciences, University of Washington, School of Dentistry, Seattle, Washington; [3]Department of Environmental Health, University of Washington, Seattle, Washington; [4]Department of Environmental Medicine, University of Rochester, New York; [5]Department of Environmental Sciences, J. Stefan Institute, Slovenia; and [6]Department of Oral Medicine, University of Washington, Seattle, Washington

Abstract

A method for the determination of methyl mercury in whole blood samples based on ethylation–gas chromatography–cold vapor atomic fluorescence spectrometry after alkaline digestion–solvent extraction is described. The extraction procedure and conditions were optimized, and the matrix interference after extraction was critically investigated. The storage stability of MeHg in blood samples and a series of extracts was determined. The method detection limit was found to be approximately 0.02 ng/g for a 0.5-g blood sample with relative standard deviations of less than 10%. The accuracy and precision were evaluated by summarizing the quality-control (QC) data generated over a one and one half year period. Appropriate procedures for sample collection, transportation, and storage were adapted to the method. Using this method accompanied by explicit QC protocols and procedures, background levels of MeHg and total mercury in blood for 150 8–10-year-old Portuguese children with nonoccupational and nonamalgamal exposure were determined and reported with summarized QC data.

Introduction

Blood mercury levels have been used extensively for the assessment of the health effects of human exposure to mercury. In order to identify the sources of exposure and assess the health effects caused by different mercury species, the quantitation of concentrations for different mercury species in whole blood samples is necessary. It has been well documented that methyl mercury (MeHg) is a neurological poison affecting primarily brain tissue (1). Ingestion of fish or grain contaminated with MeHg resulted in epidemics of severe neurotoxicity and death in Japan in the 1950s and 1960s (2) and in Iraq in 1972 (3). A large population in the world is exposed to MeHg through fish consumption (4). Monitoring MeHg concentration in whole blood samples has drawn increasing attention in recent years and has become an important tool in mercury health-effects research.

Because of severe matrix interference, the most sensitive technique, ethylation–gas chromatography (GC)–cold vapor atomic fluorescence spectrometry (CVAFS) (5–7), is unsuitable for the determination of MeHg in whole blood samples without specific sample preparation. Sample preparation by alkaline digestion/solvent extraction (8) reduced the interference significantly, making the use of this technique for the analysis of blood samples possible. In this work, the extraction procedure and conditions were optimized and the matrix interference after extraction was critically investigated. The storage stability of MeHg in blood samples and in a series of extracts of the procedure was observed. The accuracy and precision were evaluated over a long period. The method detection limit was found to be approximately 0.02 ng/g (20 parts per trillion) for a 0.5-g blood sample with relative standard deviations of less than 10%. The procedures for sample collection, transportation, and storage were also adapted to the analysis.

This technique has been used for the determination of MeHg in red blood protein of dolphins for the Mediterranean Sea Study and in whole blood for comparison in the Seychelles Study. Results from these studies indicated the method to be promising. The method has recently been used to monitor mercury in whole blood in an ongoing, randomized clinical trial of dental amalgam safety among children (T. DeRouen, PI, "The Casa Pia Study of Dental Amalgams in Children", University of Washington; clinical trial supported by the National Institute of Dental and Craniofacial Research, DE11894). The background levels of total mercury (THg) and MeHg concen-

* Author to whom correspondence should be addressed. Email Lliang@aquaticresearchinc.com.

Reproduction (photocopying) of editorial content of this journal is prohibited without publisher's permission.

trations in whole blood samples collected from 150 8–10-year-old children in Portugal with nonoccupational and nonamalgamal exposure were reported.

Experimental

Instrumentation, materials, and reagents

The instrumentation, materials, and reagents have been detailed elsewhere (5,6,8). In addition, 5-mL heparinized "Vacutainers" with freezable tubes were used for sample collection and storage. If freezable tubes were not available, 5-mL polypropylene tubes were used to store frozen samples.

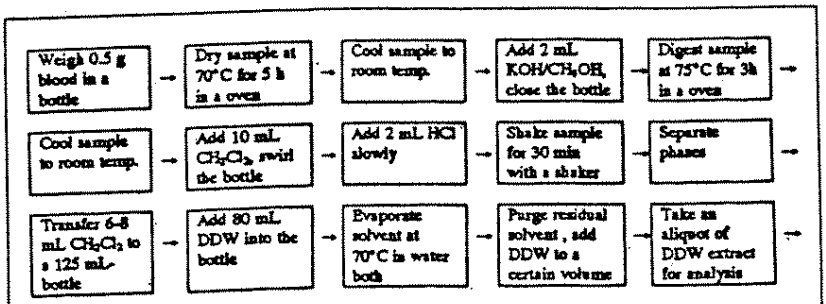


Figure 1. Sample preparation flow chart.

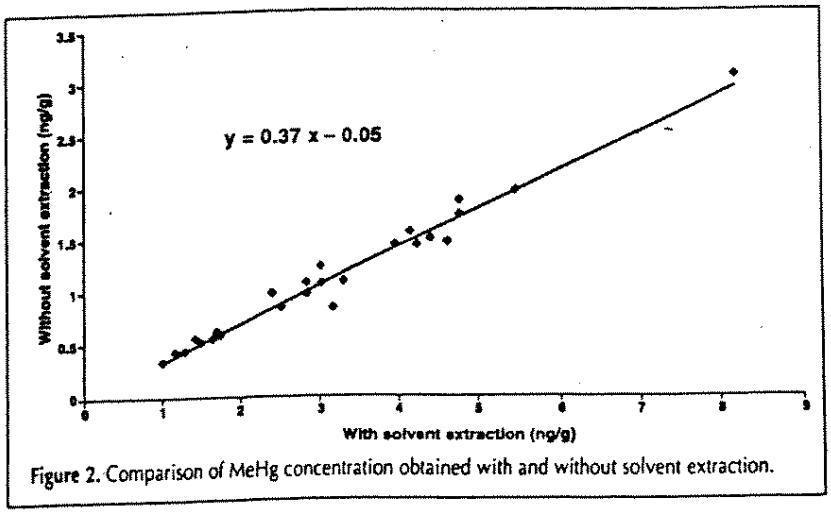


Figure 2. Comparison of MeHg concentration obtained with and without solvent extraction.

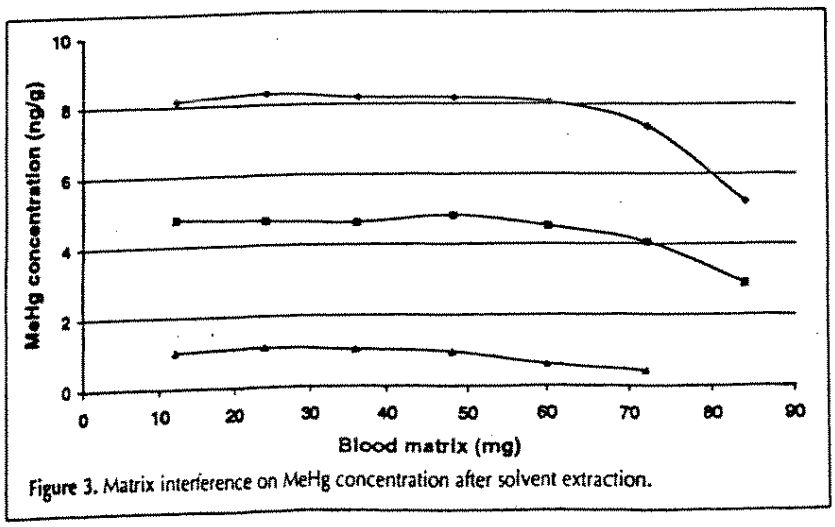


Figure 3. Matrix interference on MeHg concentration after solvent extraction.

Sample collection, transportation, and storage

Blood samples were collected by venipuncture using commercial heparinized Vacutainers. If MeHg and THg are requested to be determined in whole blood samples, it is preferable to freeze the samples for storage and transportation. If it is difficult to keep the samples frozen during transportation, the samples should be shipped by express to ensure that the samples arrive in the analytical laboratory within a few days. After a blood sample is drawn, the tube is shaken immediately and thoroughly to ensure sample homogenization. The tube is then double bagged with new plastic bags and stored frozen until shipment to the laboratory. When freezable tubes were not available, blood samples were transferred into 5-mL polypropylene tubes after homogenization.

If MeHg and THg are requested to be determined in plasma and red cells separately, it is preferable to refrigerate rather than freeze the samples because freezing can cause the cells to rupture. The samples should be analyzed within a few days after sample collection. If hemolysis does occur, the mercury can be determined in whole blood only. In this case, samples should be homogenized before aliquots are taken for analysis. If both MeHg and THg are requested to be determined in whole blood, a sample size of 1 mL is the minimum, and 5 mL is preferable.

Solvent-extraction procedure

The solvent extraction procedure described in a previous paper (8) has been improved as follows: about 0.5 g of thawed and well-homogenized blood was weighed into a 30-mL Teflon FEP bottle (8). The bottle was placed in an oven while covered with a piece of tissue, and the sample was dried at 70°C for about 5 h. When the sample was cooled to room temperature, 2 mL of saturated KOH/CH_3OH solution was added, and the bottle was closed tightly. The sample was digested in the oven at 75°C for 3 h. After the sample was cooled to room temperature, 10 mL of CH_2Cl_2 was added, and the bottle was swirled to mix the solutions. Two milliliters of concentrated HCl was added slowly and the bottle was swirled once again. This procedure must be performed in an acid fume hood. When the strong neutralization reaction was completed, the bottle was tightly capped and shaken in a shaker for 30 min. The bottle was allowed to stand at room temperature for several hours to separate the phases. Then 6 to 8 mL of solvent from the solvent (lower) layer was trans-

ferred to a 125-mL Teflon bottle using a pipette. About 80 mL of double deionized water (DDW) was added into the bottle, and the bottle was placed in a waterbath at 60–75°C for solvent evaporation. In order to avoid bumping, a small piece of Teflon tubing with closed ends was inserted for back-extraction. Residual solvent remaining in DDW after the back-extraction was purged with nitrogen for 4 min. The final volume of the extract was brought to 125 mL with DDW. The back-extraction must be performed at the same day as analysis. Depending on the concentration of MeHg in the DDW extract, an appropriate aliquot such as 5 mL of DDW extract was taken to the bubbler for the ethylation reaction, following the procedure described elsewhere (6,7). A flow chart for sample preparation is shown in Figure 1.

Results and Discussion

Homogenization of whole blood samples

Whole blood samples to be analyzed for mercury must be homogenized because about 90% of mercury is distributed in red cells. The use of anticoagulants such as heparin and shaking the sampling tube immediately after blood was drawn was found to be very helpful in obtaining homogeneous samples. If the sample received is not homogeneous, a large aliquot of the sample or even the whole sample should be taken for alkaline digestion, and then an appropriate aliquot of alkaline digestate is taken for solvent extraction. Accurate and precise results can be obtained only when homogenized samples are analyzed.

Elimination of contamination

All reagents, materials to be used for sample collection, storage, and analysis must be pre-analyzed for potential contamination. In this work, commercial heparinized tubes were checked for blanks and found that the MeHg was not detectable and the THg was less than 0.3 ng/tube. All materials to be in contact with the sample during sample preparation and analysis were acid cleaned. About 30 method blanks (MBs) have been analyzed for THg and MeHg during the analysis of 400 samples. It was found that MBs were less than 0.5 ng/g (method detection limit) for THg and undetectable for MeHg, indicating that contamination was not a critical problem using this method.

Matrix interference

The aqueous phase ethylation–GC separation–CVAFS detection is the most sensitive technique for determination of MeHg in various samples (5–7). However, matrix interference during the ethylation reaction was found to be very critical for many matrices especially blood samples making the use of the technique impossible for these complex matrices. Sample preparation can remove the interference making analysis possible.

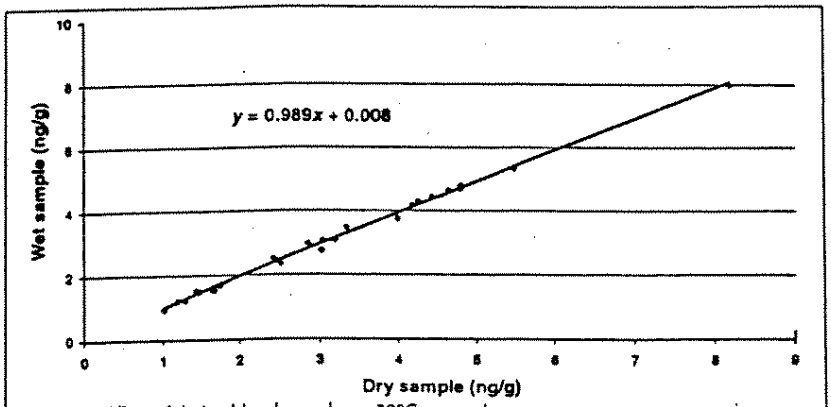


Figure 4. Effect of drying blood samples at 70°C on results.

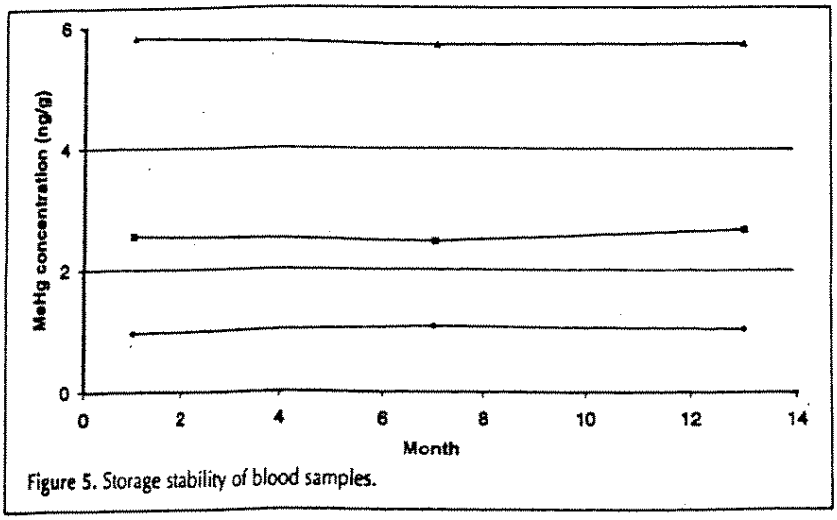


Figure 5. Storage stability of blood samples.

For analysis of biological samples using the method (5–7), the samples were digested with KOH/CH_3OH, and an aliquot of digestate was directly taken for the ethylation reaction. To reduce the matrix interference, only very small aliquots, typically several milligrams can be analyzed. Therefore only samples with high concentrations of MeHg can be analyzed directly using alkaline digestates. However, for many kinds of samples such as blood and milk, specific techniques can be applied for the isolation of MeHg from the matrix. The direct ethylation using alkaline digestates has been employed for the analysis of dolphin blood samples. It was found that recoveries were as low as 30% relative to alkaline digestion–solvent extraction as described in this paper.

A comparison of results using the technique of aqueous phase ethylation–GC separation–CVAFS detection using alkaline digestates with and without solvent extraction has been performed by analyzing 25 human blood samples, and the results are illustrated in Figure 2. The MeHg concentration obtained without solvent extraction had recoveries as low as 36% of the recovery obtained with solvent extraction.

It is worth noting that solvent extraction

can significantly reduce the matrix interference, but cannot eliminate the interference completely. Elimination of matrix interference is in a certain range of matrix concentration. To find the range, the matrix interference on results after using the technique of solvent extraction as a function of blood matrix mass was observed by analyzing three human blood samples with different MeHg concentrations (Figure 3). Matrix interference was observed with matrix mass above 50 mg. Compared to that without solvent extraction, the interference was observed with matrix mass above 5 mg (7), indicating the interference has been reduced by about a factor of 10 using solvent extraction. The concentrations of MeHg as Hg in blood samples collected from 150 children of 8–10-year-old in Portugese with nonoccupational and nonamalgamal exposure ranged from 0.47 to 9.12 ng/g. For the sample with lowest MeHg concentration, 50 mg blood contains about 24 pg of MeHg, which is about 40 times the absolute detection limit of the aqueous phase ethylation–GC–CVAFS technique (5,6). Therefore, blood samples can be easily analyzed for MeHg by the method after solvent extraction. According to the design of the procedure, 50 mg blood corresponds to about 15 mL DDW extract. Therefore, when less than 15 mL DDW extract is taken for the ethylation reaction, the result will be free from matrix interference.

Effects of drying blood samples on MeHg results

This improved solvent extraction method involves a procedure of drying blood samples at 70°C. An emulsion may form during solvent extraction when using the previous extraction procedure without drying. To solve this problem and ensure MeHg is quantitatively extracted, multiple CH_2Cl_2 extractions using a Teflon separation funnel have to be taken (8), which is time consuming and laborious. Drying blood prior to alkaline digestion can eliminate the emulsion; however, this may raise the question of wether MeHg is lost during sample drying. A comparison of results with and without drying samples prior to alkaline digestion for analysis of 25 blood samples (Figure 4) indicated that MeHg was not lost during drying at 70°C.

Stability of MeHg in storage of original samples and various extracts

Twelve blood samples were analyzed three times over a one-year period during storage in a conventional freezer. It was found that no significant change in concentration of MeHg occurred over this storage period. The results of three typical samples with high, middle, and low MeHg concentrations are shown in Figure 5. This indicated that blood samples could be stored in conventional freezers for at least one year with no change in MeHg concentration.

Using the method, MeHg was first extracted into alkaline solution, second extracted into CH_2Cl_2, and finally extracted into DDW. The stability of MeHg in various extracts was observed. It was found that MeHg in alkaline digestates was stable for at least two years, in CH_2Cl_2 for at least one month, and in DDW for one day. Therefore, MeHg in the DDW extract should be analyzed by ethylation–GC–CVAFS within 8 h after back-extraction is performed. It was found that the concentration of MeHg in DDW extract could decrease by 30% or more one day after back-extraction. The reason for this may be that some of MeHg was bound on the matrix co-extracted in DDW and was no longer available for the ethylation reaction. Moreover, although MeHg in CH_2Cl_2 was quite stable, the separation of CH_2Cl_2 phase from the extraction bottle needed to be conducted within 15 h. It was found that some of co-extracted matrix was absorbed on the wall of the bottle during evaporation of the solvent, which was helpful in reducing matrix interference further. However, if the separation of the solvent phase was conducted after 18 h of extraction, less or even no co-extracted matrix was absorbed on the wall of the bottle.

Table I. Summary of QC Data Generated from the Analysis of Whole Blood Samples Over a One and a Half Year Period for Methyl and Total Mercury

QC measurement	Parameters	Results	
Duplicate	Relative percent difference for methyl Hg	Range	0.0–9.7
		Mean ± SD	5.1 ± 2.4 (n = 50)
Pre-ethylation spike	Recovery for methyl Hg	Range	87.3–101.5%
		Mean ± SD	96.3 ± 4.8% (n = 22)
Pre-extraction spike	Recovery for methyl Hg	Range	79.1–95.3
		Mean ± SD	88.9 ± 6.1% (n = 25)
Certified reference materials	Recovery for methyl Hg	Range	85.7–101.0%
		Mean ± SD	93.7 ± 5.2% (n = 23)
Method blank	Methyl Hg	< 0.02 ng/g (n = 30)	
	Total Hg	< 0.5 ng/g (n = 30)	
Heparin tube blank	Methyl Hg	< 10 pg/tube (n = 5)	
	Total Hg	< 300 pg/tube (n = 5)	
Method detection limit	Methyl Hg	0.02 ng/g	
	Total Hg	0.5 ng/g	

Table II. Background Levels of Methyl, Inorganic, and Total Mercury in Whole Blood Samples of 150 8–10-Year-Old Portugese Children with Nonoccupational/Nonamalgamal Exposure

Mercury species	Concentration range (ng/g as Hg)	Mean ± SD
Methyl Hg	0.47–9.12	2.73 ± 1.52
Inorganic Hg	0.04–6.21	2.11 ± 1.41
Total Hg	1.03–12.0	4.82 ± 2.31
% Methyl Hg	56.6	

Evaluation of performance of the method

This improved solvent extraction method has been used for analysis of about 400 whole blood samples including human and animal blood over one and one half years. The samples were always analyzed with explicit quality control (QC) protocols and procedures. The QC data is summarized in Table I.

The evaluation of the accuracy of methods is best accomplished by the analysis of reference materials with similar matrices in quality and

quantity to the samples. However, because no blood samples certified for mercury were available, certified fish tissues were used for QC purposes. In addition, pre-extraction spike and pre-ethylation spike samples were prepared and analyzed to monitor extraction recovery and matrix interference as described.

It was noted that the mean recovery of pre-extraction spike samples was around 90%, which was lower than those obtained using previous extraction procedure (8). Using the previous procedure, after separation of two phases, the whole of solvent phase was separated using a Teflon funnel and collected into a Teflon bottle, and then residue in the funnel was washed several times with additional aliquots of the solvent. After washing, the solvent was combined into the bottle. Although recoveries close to 100% were obtained by washing residue, the washing procedure was too time consuming and laborious. The simplified procedure eliminated washing residue making the work much easier. Depending on project purposes, if necessary, results obtained using simplified procedure may be corrected with the mean of recoveries of spiked samples. At least three spiked samples should be prepared and analyzed for a sample preparation batch.

Application of the method

This technique has been used for the determination of MeHg in red blood protein of dolphins for the Mediterranean Sea Study and in whole blood for a result comparison in the Seychelles Study. Results (unpublished) obtained from these studies using the technique indicated that the method was promising.

The method has recently been used for the monitoring of methyl mercury in whole blood in the Children's Amalgam Trial Study. The study involves analyses of 300 blood samples collected from 150 Portugese children (8–10 years of age) for THg and MeHg. Background levels of total, methyl, and inorganic mercury are listed in Table II. The concentration of inorganic mercury in the table is the difference between THg and MeHg. The THg was determined using acid digestion, $SnCl_2$ reduction, gold sand trap collection, and CVAFS detection (9).

The method has also been used for analysis of blood samples from animal experiments for MeHg. Satisfactory results have been obtained.

Acknowledgments

This work was funded by Cebam Analytical, Inc. and supported by NIH grants P30 ES07033, ES04696, and DE11894.

Reference

1. T.W. Clarkson. The toxicology of mercury. *Crit. Rev. Clin. Lab. Sci.* 34(3): 369–403 (1997).
2. R.P. Junghans. A review of the toxicity of methylmercury comconds with application to occupational exposure associated with laboratory uses. *Environ. Res.* **31:** 1–31 (1983).
3. F. Bakir, S.F. Damluji, and L. Amin-Zaki. Methylmercury poisoning in Iraq. *Science* **181:** 230–241 (1973)
4. T.W. Clarkson. Methyl mercury and fish consumption: weighing the risks. *Can. Med. Assoc. J.* **158(11):** 1465–1466 (1998)
5. N.S. Bloom. Determination of picogram levels of methylmercury by aqueous phase ethylation, followed by cryogenic gas chromatography with cold vapor atomic fluorescence detection. *Can. J. Fish Aquat. Sci.* **46:** 1131–1140 (1989).
6. L. Liang, M. Horvat, and N.S. Bloom. An improved speciation method for mercury by GC/CVAFS after aqueous phase ethylation and room temperature precollection. *Talanta* **41:** 371–379 (1994).
7. L. Liang, N.S. Bloom, and M. Horvat. Simultaneous determination of mercury speciation in biological materials by GC/CVAFS after ethylation and room-temparature precollection. *Clin. Chem.* **40:** 602–607 (1994).
8. L. Liang, M. Horvat, E. Cernichiari, B. Gelein, and S. Balogh. Simple solvent extraction technique for elimination of matrix interferences in the determination of methylmercury in environmental and biological samples by ethylation–gas chromatography–cold vapor atomic fluorescence spectrometry. *Talanta* **43:** 1883–1888 (1996).
9. EPA Method 1631, Appendix A, Digestion procedures for the determination of total mercury in tissues, sludge, sediments, and soils and using acid digestion, $SnCl_2$ reduction, gold sand trap collection, and CVAFS detection, 1999.

Manuscript received October 8, 1999;
revision received December 28, 1999.

14. Appendix G. Progress on the Ancillary Study on Antibiotic Resistance

We have finished collecting the second year follow up oral and urine samples from 140 children. The number of children at each time point is given below in Table 1.

Table G.1. Children's samples collected at each time point.

Initial	After Therapy	Year 1 Follow-up	Year 2 Follow-up
150	145	136	140

We now have approximately 16,000 aerobic isolates, 1,700 anaerobes, 300 black pigmented anaerobes and over 2,000 urine isolates collected from the 4 time points in the –80 ^{0}C freezer. We have phenotypically and genotypically characterized a minimum of 1 isolate from each specimen received that was positive for bacteria from initial, 3 month, 1 and 2 year follow-up visits. In many cases multiple isolates from each sample have been characterized. Thus for the oral samples we have a total of 897 oral isolates characterized for antibiotic and mercury phenotype and genotype from the 4 time points. Because not all urine samples had bacteria, fewer isolates from the 4 time points were characterized from the urine. The number of positive urine cultures for the 4 time periods are listed in Table 2.

Table G.2. Children's urine samples positive at each time point.

Initial	After Therapy	Year 1 Follow-up	Year 2 Follow-up
119 (79%)	134 (92%)	129(96%)	138 (99%)

Because we had so many negative urines (21%) we changed the timing of the urine shipments and also how we processed them. This allowed us to have 92-99% of the later urines with bacteria, which could be analyzed. We have done a minimum of 1 isolate from each urine sample with 562 urine isolates characterized for antibiotic and mercury phenotype and genotype from the 4 time points.

For characterization of isolates we did disk diffusion testing for mercury and antibiotic to determine susceptiblity to a standard set of antibiotics and mercury. For genotyping we used DNA-DNA hybridization with our developed Gram-positive oligonucletide probes. These were developed from known Gram-positive mercury resistance genes. The presence of Gram-positive mercury genes in 167 Gram-positive oral bacteria from this collection was presented at IADR March 2000 and currently we are writing a draft manuscript on the data. We have submitted 2 abstracts looking at a larger number of oral isolates for the presence of both the mercury resistance gene *merA* and the macrolide resistance gene *mef*(A) (AADR meeting 2001) or just macrolide resistance genes *mef*(A) and *erm* genes (Gram-positive meeting Dec 2000). In all cases, the isolates were not divided by time period collected or treatment group for these studies. However, the information will be used along with other data collected from other isolates for end point analysis at the end of the study. At that time all isolates that have been characterized during the years of the study will be divided by treatment group and time period collected for analysis to determine if one treatment group had more antibiotics and/or mercury resistant bacteria characterized than the other treatment group. We will also determine at the end of the study if

both groups had similar numbers of antibiotic and/or mercury resistant isolates prior to therapy and if that changed over time.

We have found from these studies that Gram-positive *erm*, *mef*(A) and *merA* genes are isolated in Gram-negative bacteria from both the oral and urine samples in the Casa Pia students. However, we now have similar results from isolates collected in other studies, primarily from the Seattle area, suggesting these findings are not unique to the Lisbon bacterial population, but are more widely found in the Gram-negative population. This is the first time these genes have been found in Gram-negative bacteria and is an important finding. Some of these findings on the *mef*(A) gene in Gram-negative bacteria have been published (Luna et al.) in Antimicrob. Agents Chemother. Sept 2000 (44:2503-2506). This publication used one isolate from the Casa Pia study and other isolates in the United States from both respiratory and genital sites.

We have cloned the *mef*(A) and *merA* genes from some of the Casa Pia isolates, as well as from US isolates and are working to determine if these two genes are linked in many of the isolates which co-transfer these 2 genes together. We have also set-up the PCR assay for the Gram-positive integrase which we will then clone as well. A major goal is to determine if these genes are on a single conjugative transposon, which is found in both Gram-positive and Gram-negative isolates.

Next month we will begin to collect year 3 follow-up samples. We will continue to characterize isolates so that we have 1-2 isolates from each oral site (A and B) and 2 isolates from each urine sample for each time point. This is a large amount of isolates, which we hope to complete within the next 12 months. We also plan to do year 2 analysis on the characterized isolates currently available and examine the number of isolates from each sample that grew on mercury-supplemented media, and the percentage of mercury resistant isolates in the population. In year 1 follow-up there was a trend for one treatment group to have higher percentage of mercury resistant isolates/sample than the other treatment group. We will see if this trend holds or not for the year 2 follow-up data.

Abstracts and publications associated with this ancillary study follow.

Dec 2000

RGPI002.doc
Resistant Gram-Positive Infections Abstract

DISTRIBUTION OF MACROLIDE RESISTANCE GENES IN GRAM-POSITIVE BACTERIA FROM NORMAL CHILDREN IN LISBON PORTUGAL- EXTENSION OF THEIR HOST RANGE

VA LUNA[1], HS LUIS[2], M BERNARDO[2], J LEITAO[2] and MC ROBERTS[1]. Univ. of Washington, Seattle, WA 98195[1] and Univ. of Lisbon, Lisbon Portugal[2]

There are currently 21 *erm* genes coding for rRNA methylases which confer macrolide-lincosamide-streptogramin B (MLS_B) resistance. The genes *erm*(B), (C) and (F) have the largest host ranges. The *mef*(A) gene codes for an efflux protein that confers resistance to macrolides and has the widest host range among other macrolide resistant mechanisms. In this study, we screened 718 randomly selected Gram-positive bacteria collected from oral and urine specimens from healthy children of which 364 (51 %) were macrolide resistant. Using DNA-DNA hybridization, polymerase chain reaction and sequencing, the *erm* genes, *erm*(A), (B), (C), and (F), were found in 168 (46%) of the isolates, 14 (4 %) isolates carried the *mef*(A) gene, and 8 (2 %) carried both *mef*(A) and one of the four *erm* genes. Additional 174 (47%) isolates did not hybridize with the 5 genes but may carry other *erm*, efflux or inactivating genes not examined, or have chromosomal mutations. The host range of the *mef*(A) gene was found in 2 new genera: *Lactobacillus* and *Staphylococcus*, while *erm*(A) and *erm*(B) was new in *Lactobacillus* and *erm*(F) was new to *Enterococcus* and *Staphylococcus*. This work expands the host ranges of *erm*(A), (B), (F) and *mef*(A) genes.

Macrolide-Lincosamide-Streptogramin resistance

AADR March 2001

Gram-positive Macrolide and Mecury Genes Transfer Together in Bacteria. VA LUNA, HS LUIS, M BERNARDO, J LEITAO and MC ROBERTS. (University of Washington, Seattle, WA and Univ. of Lisbon, Lisbon, Portugal).

The *mef*(A) gene codes for an efflux protein, confers resistance to macrolides, is often associated with mobile elements and has not been linked with other antibiotic resistance genes. We randomly screened 571 including 455 erythromycin resistant Gram-positive and 116 Gram-negative isolates identified in oral specimens collected from healthy children in the Casa Pia Schools, Lisbon, Portugal. Using DNA-DNA hybridization and polymerase chain reaction (PCR) these isolates were screened for the presence of *mef*(A) and a mercury reductase gene, *merA*. Ten Gram-positive (*Enterococcus*, *Streptococcus*, *Lactobacillus*, *Staphylococcus* spp.) and 17 Gram-negative (Acinetobacter, Klebsiella, *Enterobacter*, *Pseudomonas*, *Stenotrophomonas* spp.) carried both the *merA* and *mef*(A) genes. Two Gram-negative and 13 Gram-positive isolates were used for matings with *Enterococcus faecalis* JH2-2. Transconjugants were selected on 200 μM mercury or 10 μg/ml erythromycin at frequencies ranging from 10^{-6} to 10^{-10} per recipient. They carried both the *mef*(A) and *merA* genes regardless of the media used for selection. We have cloned the *mef*(A) and *merA* genes and are determining whether these two genes are on the same mobile element. This is the first report that the *mef*(A) gene may be linked to another gene conferring resistance. Supported by NIDCR Grant UO1-DE-11895 and Contract N01-DE-72623 and Training Grant T-32-DE-07063. vluna@u.washington.edu.

3819 Development of Gram-positive Probes for the Detection of Mercury Resistance in Oral Bacteria. M.C. ROBERTS*, K. JUDGE and K. YOUNG (University of Washington, Seattle WA. 98195).

Mercury (Hg) resistance genes have been described in Gram-positive staphylococci, enterococci and environmental bacteria such as bacilli. Using known DNA sequences from these three genera, we developed DNA oligonucleotide probes that would detect the structural gene for Hg resistance in characterized Hg resistant isolates. We examined Gram-positive species taken from the plaque and buccal epithelial cells of 8-11 year old children from Lisbon, Portugal. We examined by dot blot and DNA-DNA hybridization 9 *S. intermedius*, 163 viridians streptococci, 4 *S. aureus*, 8 *Staphylococcus* spp., 48 *Lactobacillus* spp., 2 *Enterococcus* spp and found that 5 (56%) *S. intermedius*, 85 (52%) viridians streptococci, 3 (75%) *S. aureus*, 5 (63%) *Staphylococcus* spp., 19 (39%) *Lactobacillus* spp., 0 *Enterococcus* spp. hybridized with the oligonucleotide probes specific for the Hg structural gene. Those that grow without blood were tested for phenotypic resistance to Hg and most of the positive isolates also could grow in the presence of 200 µM Hg and/or gave small zones when tested in a Hg disk assay. This suggests a correlation between expression of mercury resistance and the presence of the Hg gene related to those previously described in other Gram-positive species. Not all isolates which were phenotypically resistant to mercury hybridized with the probes suggesting that distinct mercury genes may be present in some of these isolates. Significant numbers of the oral bacteria examined hybridized with the Hg probes. In addition, this is the first time that lactobacilli have been shown to carry Hg resistance genes. These results suggest that other oral bacteria need to be examined for the presence of Hg resistance genes. This work was supported by a grant 01-DE11894 and contract 01-DE72623 from NIDCR.

SUBMITTED FOR PUBLICATION

Presence of *Porphyromonas gingivalis, Prevotella intermedia,* and *Prevotella nigrescens* from gingiva and buccal mucosa of children

Yasaman Sanai[1], G. Rutger Persson[1], Jacqueline R. Starr[2], Henrique S. Luis[3], Mario Bernardo[3], Jorge Leitao[3], and Marilyn C. Roberts[4]*

Department of Periodontics[1], Epidemiology[2], and Pathobiology[4], University of Washington, Seattle, Washington USA 98195, and Faculty of Medical Dentistry, University of Lisbon , Lisbon Portugal[3]

Short title: Presence of *P. gingivalis*, *P. intermedia*, and *P. nigrescens* in children

* Corresponding author: Department of Pathobiology, Box 357238, School of Public Health and Community Medicine, University of Washington, Seattle, WA 98195, e-mail: marilynr@u.washington.edu

ABSTRACT

Introduction. We cultured *Porphyromonas gingivalis*, *Prevotella intermedia*, and *P. nigrescens* from the gingiva and buccal mucosa of 47 of the 150 children examined.

Purpose. The objective of this study was to determine if 8-11 year-old children carried culturable *P. gingivalis*, *P. intermedia*, and *P. nigrescens*, in their gingiva and buccal mucosa over a one year time period.

Methods. Bacterial samples were collected from the gingiva and buccal mucosa of 150 Portuguese children at initial, after caries treatment, and at one-year follow-up. These samples were cultured and identified as *P. gingivalis*, *P. intermedia*, or *P. nigrescens* using specific 16S rRNA probes. Antibiotic resistant genes, *tet*(Q) and *erm* (F), were identified using specific DNA probes and used to separate the same species within a single specimen.

Results. A total of 116 isolates of black pigmented bacteria were cultured from 47 (31%) of 150 children. Five isolates were identified as *P. gingivalis*, 29 as *P. intermedia*, 33 as *P. nigrescens*, and 49 other species. In general, the bacteria were not culturable at more than one time period.

Conclusion. *P. gingivalis*, *P. intermedia*, and *P. nigrescens* are present on the gingiva and buccal mucosa of 8-11 year-old children. There were a few *P. gingivalis* and more *P. intermedia*, and *P. nigrescens* cultured from these oral sites.

ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, Sept. 2000, p. 2503–2506 Vol. 44, No. 9
0066-4804/00/$04.00+0
Copyright © 2000, American Society for Microbiology. All Rights Reserved.

Identification of the Conjugative *mef* Gene in Clinical *Acinetobacter junii* and *Neisseria gonorrhoeae* Isolates

VICKI A. LUNA,[1] SYDNEY COUSIN, JR.,[1] WILLIAM L. H. WHITTINGTON,[2] AND MARILYN C. ROBERTS[1]*

Departments of Pathobiology[1] and Medicine,[2] University of Washington, Seattle, Washington 98195

Received 28 December 1999/Returned for modification 28 March 2000/Accepted 1 June 2000

The *mef* gene, originally described for gram-positive organisms and coding for an efflux pump, has been identified in clinical isolates of *Acinetobacter junii* and *Neisseria gonorrhoeae*. These strains could transfer the *mef* gene at frequencies ranging from 10^{-6} to 10^{-9} into one or more of the following recipients: gram-negative *Moraxella catarrhalis*, *Neisseria perflava/sicca* and *Neisseria mucosa* and gram-positive *Enterococcus faecalis*. Three *Streptococcus pneumoniae* strains could transfer the *mef* gene into *Eikenella corrodens*, *Haemophilus influenzae*, *Kingella denitrificans*, *M. catarrhalis*, *Neisseria meningitidis*, *N. perflava/sicca*, and *N. mucosa* at similar frequencies. The *mef* gene can thus be transferred to and expressed in a variety of gram-negative recipients.

The *mef* gene encodes a membrane-bound efflux protein and confers resistance to macrolides but not to lincosamides or streptogramin B (24, 25). From geographically diverse areas, two gene variants with 90% nucleotide sequence identity, *mefA* and *mefE*, have been identified in gram-positive organisms, including *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Corynebacterium* spp., *Enterococcus* spp., *Micrococcus luteus*, and viridans group streptococci (2, 5, 9, 20, 23–25). Both genes have now been combined as *mef*(A) (20).

Recently, we have shown that both *Neisseria gonorrhoeae* and oral commensal *Neisseria* spp. carry known rRNA methylase genes (20). However, we have also identified several of these *Neisseria* spp. and oral gram-negative isolates that were macrolide resistant yet did not carry Erm determinants (ErmA, -B, -C, or -F) (21). We selected a gram-negative *Acinetobacter junii* strain and screened two groups of *N. gonorrhoeae* in order to determine if the *mef* gene was present. Because we have shown that the *mef* gene can be transferred by conjugation in various gram-positive genera, we also wanted to determine the mobility of the *mef* gene in these gram-negative isolates as well as the ability of gram-positive donors to move the *mef* gene to gram-negative species (9).

(This work was presented, in part, at the 39th Interscience Conference on Antimicrobial Agents and Chemotherapy, 23 to 27 September 1999, San Francisco, Calif.)

MATERIALS AND METHODS

Bacteria. Erythromycin-resistant (Ery^r) *Acinetobacter junii* strain 329 was isolated from the oral gumline of a child in Portugal (Table 1). The organism was identified by D. Stroman (Alcon Laboratories, Inc., Fort Worth, Tex.) by sequencing the 16S rRNA. Sixteen *N. gonorrhoeae* isolates with variable susceptibilities to erythromycin (MIC = 0.25 to 4 μg/ml) were collected from urethral specimens from adults in Seattle, Wash., and identified by the *Neisseria* Reference Laboratory at the University of Washington (Table 1). Thirteen *N. gonorrhoeae* isolates carrying the *tetM* 25.2-MDa plasmid were collected in Seattle and the eastern United States during 1983 through 1986 and have been previously described (7). Of the three Ery^r *S. pneumoniae* strains carrying the *mef* gene, two (n011 and 970146) were isolated in Washington State and the third (02J1048) was supplied by Pfizer Inc. (Groton, Conn.) (Table 1) (9, 26). Strains used as recipients are listed in Table 2. Antibiotic susceptibilities were determined using either disk diffusion or agar dilution as described by the NCCLS (13).

Media. *S. pneumoniae* isolates were grown on brucella blood agar (Difco, Detroit, Mich.) supplemented with 5% sheep red blood cells and incubated with 5% CO_2 at 36.5°C (8). The *Haemophilus influenzae* isolate was grown on brain heart infusion (BHI) agar (Difco) supplemented with 2 μg of NAD (Sigma Chemical Co., St. Louis, Mo.) per ml and 10 μg of hemin–L-histidine (Sigma Chemical Co.) per ml (16). *Eikenella corrodens*, *Kingella denitrificans*, *Moraxella catarrhalis*, *N. gonorrhoeae*, *Neisseria meningitidis*, *Neisseria perflava/sicca*, and *Neisseria mucosa* were grown on gonococcus (GC) medium agar (Difco) supplemented with Kellogg's supplement solution (0.22 M D-glucose, 0.03 M L-glutamine, 0.001 M ferric nitrate, and 0.02 M cocarboxylase) (Sigma-Aldrich Chemical Co., St. Louis, Mo.) as previously described (6, 15). *Enterococcus faecalis* JH2-2 was grown on BHI agar without supplements (9). For DNA dot blots and DNA extractions, the gram-positive bacteria were grown overnight at 36.5°C in BHI broth (Difco) supplemented with 0.03 M D-glucose and 0.04% DL-threonine (8, 9). The gram-negative bacteria, for DNA dot blots and DNA extractions, were grown in either BHI broth with no supplements (*A. junii*), BHI broth supplemented with 2 μg of NAD per ml and 10 μg of hemin–L-histidine per ml (*H. influenzae*), or GC broth supplemented with 1% sodium bicarbonate and Kellogg's supplement (see above) (*E. corrodens*, *K. denitrificans*, *M. catarrhalis*, and *Neisseria* spp.) as described previously (11, 12, 15).

Mating procedure. Donor and recipient bacteria were grown separately at 36.5°C overnight. Each isolate was suspended in 0.5 to 1.0 ml of BHI or GC broth (Difco) to a density of approximately 10^9 cells per ml (3 McFarland). Donor and recipient bacteria at ratios of 1:20 or 1:40 (donor to recipient) were mixed. The bacterial suspension was then plated directly onto brucella blood agar plates and incubated in CO_2 at 36.5°C for 48 h (8, 9). After incubation, the mating mixture was serially diluted onto antibiotic-supplemented plates as follows: for *H. influenzae* recipients, erythromycin (10 μg/ml) and rifampin (10 μg/ml); for *E. faecalis* recipients, erythromycin (10 μg/ml) and rifampin (10 μg/ml); for *M. catarrhalis*, *N. gonorrhoeae*, and *N. meningitidis* recipients, erythromycin (2 μg/ml) and rifampin (10 μg/ml); and for *E. corrodens*, *K. denitrificans*, *N. perflava/sicca*, and *N. mucosa* recipients, erythromycin (2 μg/ml) and tetracycline (10 μg/ml). Gram stains were prepared from all transconjugants, and all transconjugants were identified to species level by established methods (4, 14). Mating experiments were performed a minimum of two times. DNase I (1 mg/ml) (Sigma-Aldrich Chemical Co.) was added to the mating mixture in at least one of the two matings for comparison with mating without DNase to rule out transformation.

Extraction of whole-cell DNA. Whole-cell DNA was prepared from isolates and transconjugants as previously described (1, 8). DNA was separated on a 0.7% agarose gel and Southern blots were prepared (22).

Plasmid analysis of transconjugants. Plasmid extractions were prepared from *K. denitrificans*, *N. perflava/sicca*, and their respective transconjugants after 8 h of growth in 40 ml of GC peptone broth supplemented with Kellogg's supplement solution and 1% sodium bicarbonate as previously described (3, 15). DNA was separated on a 0.7% agarose gel in 0.5× Tris-borate-EDTA (TBE) buffer at 100 V for 1.5 h; Southern blots were prepared and hybridized with *mef*-specific ^{32}P-labeled oligonucleotide probe MF5 to examine the location of the *mef* gene.

Dot blots. Two milliliters of overnight bacterial growth in supplemented BHI or GC broth was placed into 1.5-ml sterile Eppendorf tubes and centrifuged at 8,000 × *g* for 2 min, and the supernatant was decanted. The bacterial pellet was resuspended with 1 ml of BHI broth to create a turbid suspension of 3 McFarland. Two hundred microliters of the suspension, in 25-μl aliquots, was spotted onto a GeneScreen Plus membrane (NEN Research, Boston, Mass.), dried, and treated with 0.5 M NaOH for 10 min, 1 M Tris-HCl for 3 min, and 1 M Tris-HCl with 1.5% NaCl, pH 7.5, for 10 min. The membrane was dried, washed in

* Corresponding author. Mailing address: Department of Pathobiology, Box 357238, School of Public Health and Community Medicine, University of Washington, Seattle, WA 98195-7238. Phone: (206) 543-8001. Fax: (206) 543-3873. E-mail: marilynr@u.washington.edu.

TABLE 1. Isolates carrying the *mef* gene

Isolate	Yr obtained	Location	Susceptibility pattern[a]	Plasmid size(s) (MDa)	Erm determinant(s)	Reference
S. pneumoniae 02J1048	Unknown	Unknown	Ery^r (*mef*) Rif^s Fus^s Str^s Tet^s	None	None	24
S. pneumoniae n011	1996	Seattle, Wash.	Ery^r (*mef*) Rif^s Fus^s Str^s Tet^s	None	None	9
S. pneumoniae 970146	1997	Seattle, Wash.	Ery^r (*mef*) Rif^s Fus^s Str^s Tet^r	None	None	9
A. junii 329	1999	Lisbon, Portugal	Ery^r (*mef*) Rif^s Fus^s Str^s Tet^s	None	None	This paper
N. gonorrhoeae 98.420	1998	Seattle, Wash.	Ery^r (*mef*) Pen^r Rif^s Str^s Tet^s	2.6	B, F	This paper
N. gonorrhoeae 98.560	1998	Seattle, Wash.	Ery^r (*mef*) Pen^r Rif^s Str^s Tet^s	2.6	B, C	This paper
N. gonorrhoeae 98.737	1998	Seattle, Wash.	Ery^r (*mef*) Pen^r Rif^s Str^s Tet^s	2.6	None	This paper
N. gonorrhoeae 85.022462	1985	DeKalb County, Ga.	Ery^r (*mef*) Tet^r	2.6, 25.2	B, F	7

[a] Ery, erythromycin; Fus, fusidic acid; Rif, rifampin; Str, streptomycin; Tet, tetracycline. *H. influenzae*, *N. meningitidis*, and *E. faecalis* were chromosomally resistant to the antibiotics.

chloroform-isoamyl alcohol (24:1), rinsed in water twice, washed in 1 M Tris-HCl with 1.5% NaCl for 10 min, and baked at 80°C for 1 h. The filters were stored at room temperature until hybridized with labeled oligonucleotide probes (8).

Labeled probes. MF5 (5′ GGT GCT GTG ATT GCA TCT ATT AC 3′) was used as the oligonucleotide probe for the *mef* gene (9). The oligonucleotide probes for the *erm* genes were the following: $ermB_F$ (5′ GAA AAG GTA CTC AAC CAA ATA 3′), $ermC_R$ (5′ GCT AAT ATT GTT TAA ATC GTC AAT 3′), and ermF1 (5′ CGG GTC AGC ACT TTA CTA TTG 3′) (19). A ^{32}P-labeled probe was used for whole-bacterial-cell dots as previously described (9).

PCR. PCR amplification used 40 ng of genomic DNA from *S. pneumoniae* 02J1048 as a positive control and 30 ng of genomic DNA as a template from the transconjugants and *A. junii*. Primers were MF4a (5′ ACC GAT TCT ATC AGC AAA G 3′) and MF6 (5′ GGA CCT GCC ATT GGT GTG 3′). Both primers are in the conserved regions of the *mef* gene. Each reaction mixture contained 2 U of *Taq* polymerase (Perkin-Elmer Cetus, Norwalk, Conn.), 200 μM deoxynucleoside triphosphates, 1× PCR buffer (1.5 mM $MgCl_2$), and 100 ng of each primer. Using a Perkin-Elmer Cetus thermal cycler, the reactions were carried out by denaturation at 94°C for 1 min, annealing at 37°C for 1 min, and elongation at 72°C for 2 min, for 35 cycles. The PCR products were dried, resuspended in 1/10 volume of sterile water, and separated on a 1.5% agarose gel with 0.5× TBE running buffer. The PCR bands were visualized by ethidium bromide staining, and Southern blots were prepared. The 940-bp PCR product hybridized with a labeled internal *mef* probe, MF5 (9). Positive and negative controls were run with each assay.

DNA-DNA hybridization. Southern blots using uncut whole-cell DNA or the PCR fragments were prepared on Magnagraph nylon (Micron Separation Inc., Westboro, Mass.) and hybridized with ^{32}P-labeled oligonucleotide MF5. DNA dot blots containing 30 to 300 μg of extracted whole-cell DNA from the donor strains and bacterial dot blots from transconjugants were placed on GeneScreen Plus membranes and hybridized with the radiolabeled oligonucleotide probe (8, 9).

Sequencing. Sequencing of selected isolates (*A. junii* 329 and *N. gonorrhoeae* 98.420 and 98.737) was performed using oligonucleotide primers MF7 (5′ ATG CAG ACC AAA AGC GCG AT 3′) and MF8 (5′ CGG TAT CTG TTC TGG TAG CG 3′). Both primers are in the conserved regions of the *mef* gene and within the generated PCR fragment. Each reaction mixture contained 2 U of *Taq* polymerase, 200 μM deoxynucleoside triphosphates, 1× PCR buffer (1.5 mM $MgCl_2$), and 100 ng of each primer. Using a Perkin-Elmer Cetus thermal cycler, the reactions were carried out by denaturation at 94°C for 1 min, annealing at 51°C for 1 min, and elongation at 72°C for 2 min, for 35 cycles. The PCR products were filtered by centrifugation with a 0.5-μm-pore-size separation filter (Micron) by following the manufacturer's directions. A small part of the PCR product obtained was dried and separated on a 1.5% agarose gel with 0.5× TBE running buffer. The 253-bp PCR product was then visualized by ethidium bromide staining. The remainder of the PCR product was then sequenced by following the Big Dye Terminator chemistry protocol (Perkin-Elmer Cetus, Foster City, Calif.). The sequence reaction mixture included 100 ng of the PCR product and 4 pmol of primer MF7, with a total volume of 12 μl. The *mef* sequence was compared to the *mef* sequence from *S. pneumoniae* (GenBank number U83667) using the Genetics Computer Group software (University of Wisconsin, Madison).

RESULTS

The *mef* gene in gram-negative bacteria. We examined 29 *N. gonorrhoeae* isolates for the presence of the *mef* gene using DNA-DNA hybridization of bacterial dot blots. Four (14%) were positive and verified as carrying the *mef* gene by hybridization of Southern blots and positive PCR assays. Three of the *mef*-positive isolates were from Seattle and one was from Georgia. The *mef* gene was also identified by the same means in an oral *A. junii* isolate from Portugal. The PCR fragments were sequenced from the *A. junii* isolate and from two of the *N. gonorrhoeae* isolates. The first 9 nucleotides of the PCR product could not be adequately sequenced; however, the three isolates had 97 to 98% nucleotide identity with the GenBank sequence of the pneumococcal *mef* gene over the remaining 244 nucleotides and had 98% amino acid identity over 81 amino acids.

All four *N. gonorrhoeae* isolates carried the gonococcal 2.6-MDa plasmid, and strain 85.022462 carried a 25.2-MDa plasmid as well. When the *mef* probe was hybridized with Southern blots containing both chromosomal and plasmid DNA, only the chromosomal band hybridized in all four isolates. The *A. junii* isolate had no visible plasmids, and the *mef* probe hybrid-

TABLE 2. Characterization of recipients used

Isolate	Location	Susceptibility pattern[a]	Reference
H. influenzae RD8	Laboratory	Ery^i Rif^{r}Fus^{r}Str^r	15
E. corrodens CTM 1.6	DeKalb County, Ga.	Ery^{s}Tet^{rb}	6
N. meningitidis NRL 9205[c]	Laboratory	Ery^s Rif^{r}Fus^{r}Str^r	16
N. perflava/sicca 86.000348	DeKalb County, Ga.	Ery^{s}Tetrd	6
N. mucosa CTM 1.1	DeKalb County, Ga.	Ery^{s}Tetrd	6
K. denitrificans 87.023461	DeKalb County, Ga.	Ery^{s}Tet^{rb}	6
M. catarrhalis NRL 30018	Laboratory	Ery^s Rif^r Str^r	15
E. faecalis JH2-2	Laboratory	Ery^s Rif^{r}Fus^r	19

[a] Ery, erythromycin; Fus, fusidic acid; Rif, rifampin; Str, streptomycin; Tet, tetracycline. *H. influenzae*, *N. meningitidis*, *M. catarrhalis*, and *E. faecalis* were chromosomally resistant to the antibiotics.

[b] Carries a 25.2-MDa plasmid with the *tetM* gene (16).

[c] Serotype A.

[d] The *tetM* gene is carried on the chromosome (6).

TABLE 3. Conjugal transfer of the *mef* gene[a]

Donor	Recipient	Frequency (per donor cell)
A. junii 329	*E. faecalis* JH2-2	9.5×10^{-6}
	M. catarrhalis NRL 30018	5.0×10^{-8}
	N. mucosa CTM 1.1	8.5×10^{-8}
	N. perflava/sicca 86.000348	1.0×10^{-9}
N. gonorrhoeae 98.737	*E. faecalis* JH2-2	4.1×10^{-9}
	M. catarrhalis NRL 30018	1.6×10^{-8}
	N. mucosa CTM 1.1	1.2×10^{-8}
S. pneumoniae 02J1048	*N. meningitidis* 9205	1.3×10^{-8}
	N. perflava/sicca 86.000348	1.4×10^{-8}
	E. corrodens CTM 1.6	1.7×10^{-6}
S. pneumoniae n011	*N. mucosa* CTM 1.1	4.0×10^{-8}
	K. denitrificans 87.923461	1.0×10^{-7}
	H. influenzae RD8	3.0×10^{-7}
S. pneumoniae 970146	*H. influenzae* RD8	1.6×10^{-8}
	M. catarrhalis NRL 30018	1.3×10^{-9}

[a] Mating experiments were performed a minimum of two times.

ized with the chromosomal band in this isolate (data not shown).

We examined the five isolates for the presence of *erm* genes, which code for rRNA methylase genes (20). Of the four *Neisseria* isolates, 98.420 and 85.022462 hybridized with probes for *erm*(B) and *erm*(F) while 98.560 hybridized with the probes for *erm*(B) and *erm*(C). *N. gonorrhoeae* 98.737 and *A. junii* 329 did not hybridize with any of the *erm* probes examined (Table 1). In addition, *A. junii* 329 was susceptible to clindamycin, suggesting that this isolate did not carry an Erm determinant.

In mixed cultures, we were able to transfer the *mef* gene from *A. junii* 329 and *N. gonorrhoeae* 98.737 to *E. faecalis* at frequencies of 10^{-6} and 10^{-9} per donor cell, respectively, and to *M. catarrhalis* and *N. mucosa* at a frequency of 10^{-8} per donor cell (Table 3). We were able to transfer the *mef* gene by conjugation from *A. junii* to *N. perflava/sicca* at a frequency of 10^{-9} per donor cell (Table 3). DNA-DNA hybridization of bacterial dot and Southern blots of cellular DNA and amplification of PCR products confirmed the presence of the *mef* gene in transconjugants. The recipients did not carry plasmids and the *mef* gene probe hybridized with the chromosomal band in these transconjugants, although the presence of very large plasmids could not be ruled out. The transconjugants grew on media supplemented with 2 to 10 μg of erythromycin per ml, depending upon the recipient, whereas the original recipients did not grow on media supplemented with erythromycin. This corresponds to a six- to eightfold increase in MICs (0.03 μg/ml versus 2 to 10 μg/ml). All of the transconjugants remained susceptible to clindamycin. As in our previous work, no other antibiotic resistance determinants were transferred with the *mef* gene.

Conjugation of the *S. pneumoniae mef* gene with gram-negative bacteria. To examine whether gram-positive *S. pneumoniae* could transfer *mef* to gram-negative recipients, we used three pneumococcal donors and obtained frequencies of transfer of *mef* to *E. corrodens*, *H. influenzae*, *K. denitrificans*, *M. catarrhalis*, *N. meningitidis*, *N. mucosa*, and *N. perflava/sicca* that were similar those obtained with *A. junii* as the donor (Table 3). Although the *K. denitrificans* and *E. corrodens* transconjugants carried the 25.2-MDa *tetM* plasmid, the *mef* probe hybridized with the chromosomal region but not with the 25.2-MDa plasmid.

DISCUSSION

This is the first time the *mef* gene has been identified in clinical gram-negative isolates. These isolates were collected over a 13-year time span and from different geographical areas and ecological niches (oral and genital). We showed that these isolates transfer the *mef* gene to both gram-negative and gram-positive recipients (Table 3). In all cases, the *mef* gene appeared to be localized to the chromosome rather than to indigenous plasmids, although plasmids larger than 25.2 MDa cannot entirely be ruled out. Previous work with the *mef* gene in gram-positive donors and recipients has demonstrated not a specific location in the chromosome but potentially multiple random sites (9, 10). We are currently examining whether this is also true with the gram-negative recipients used in this study. However, we did not find the *mef* gene inserted into the 25.2-MDa plasmid. The *mef* sequence in the *A. junii* and *N. gonorrhoeae* isolates had 97 to 98% nucleotide identity to the *mef* gene found in *S. pneumoniae*, suggesting that these genes originally came from a gram-positive donor. A similar picture has emerged for the *tetM*, *tetO*, and *erm* genes found in the gram-negative flora (18–21).

To examine whether the *mef* genes located in *S. pneumoniae* could be transferred into gram-negative recipients, we examined three randomly chosen isolates and documented their ability to transfer *mef* into a variety of different oral recipient species (Table 3). Our data suggest that the *mef* gene host range may extend to a larger number of species and genera than the *Acinetobacter* and *Neisseria* organisms in this study. A survey of a variety of gram-negative species and genera from different ecological niches and locations should be done to better define the host range of the *mef* gene and its potential impact on macrolide therapy.

ACKNOWLEDGMENTS

The study was supported in part by NIH grants U01-DE-1189 and AJ-131448 and contract N01-DE-72623. V. Luna was supported in part by Oral Biology Training Grant T-32-DE-07063, and S. Cousin was supported by the Poncin Research Award Scholarship Fund.

We gratefully acknowledge the valuable technical assistance of Karen Winterscheid, Maria Strejac, and Kathleen Judge.

REFERENCES

1. **Anderson, D. G., and L. L. McKay.** 1983. Simple and rapid method for isolating large plasmid DNA from lactic streptococci. Appl. Environ. Microbiol. 46:549–552.
2. **Clancy, J., J. Petitpas, F. Dib-Hajj, W. Yuan, M. Cronan, A. Kamath, J. Bergeron, and J. Retsema.** 1996. Molecular cloning and functional analysis of a novel macrolide-resistance determinant *mefA* from *Streptococcus pyogenes*. Mol. Microbiol. 22:867–879.
3. **Elwell, L. P., M. Roberts, L. W. Mayer, and S. Falkow.** 1977. Plasmid-mediated beta-lactamase production in *Neisseria gonorrhoeae*. Antimicrob. Agents Chemother. 11:528–533.
4. **Facklam, R. R., and F. Sahm.** 1995. *Enterococcus*, p. 308–314. *In* P. R. Murray, E. J. Baron, M. A. Pfaller, F. C. Tenover, and R. H. Yolken (ed.), Manual of clinical microbiology, 6th ed. American Society for Microbiology, Washington, D.C.
5. **Kataja, J., P. Huovinen, M. Skurnik, The Finnish Study Group for Antimicrobial Resistance, and H. Seppälä.** 1999. Erythromycin resistance genes in group A streptococci in Finland. Antimicrob. Agents Chemother. 43:48–52.
6. **Knapp, J. S., S. R. Johnson, J. M. Zenilman, M. C. Roberts, and S. A. Morse.** 1988. High-level tetracycline resistance resulting from TetM in strains of *Neisseria* spp., *Kingella denitrificans*, and *Eikenella corrodens*. Antimicrob. Agents Chemother. 32:765–767.
7. **Knapp, J. S., J. M. Zeilina, J. W. Biddle, G. H. Perkins, W. D. DeWitt, M. L. Thomas, S. R. Johnson, and S. A. Morse.** 1988. Frequency and distribution in the United States of strains of *N. gonorrhoeae* with plasmid-mediated high-level resistance to tetracycline. J. Infect. Dis. 155:819–822.

8. **Luna, V. A., and M. C. Roberts.** 1998. The presence of the *tetO* gene in a variety of tetracycline resistant *Streptococcus pneumoniae* serotypes from Washington State. J. Antimicrob. Chemother. 42:613–619.
9. **Luna, V. A., P. Coates, J. H. Cove, A. E. Eady, T. T. H. Nguyen, and M. C. Roberts.** 1999. A variety of gram-positive bacteria carry mobile *mef* genes. J. Antimicrob. Chemother. 44:19–25.
10. **Luna, V. A., D. B. Jernigan, A. Tice, J. D. Kellner, and M. C. Roberts.** 2000. A novel multiresistant *Streptococcus pneumoniae* serogroup 19 clone from Washington State identified by pulsed-field gel electrophoresis and restriction fragment length patterns. J. Clin. Microbiol. 38:1575–1580.
11. **Mayer, L. W., K. K. Holmes, and S. Falkow.** 1974. Characterization of plasmid deoxyribonucleic acid from *Neisseria gonorrhoeae*. Infect. Immun. 10:712–717.
12. **Morse, S. A., S. R. Johnson, J. W. Biddle, and M. C. Roberts.** 1986. High-level tetracycline resistance in *Neisseria gonorrhoeae* is result of acquisition of streptococcal *tetM* determinant. Antimicrob. Agents Chemother. 30:664–670.
13. **National Committee for Clinical Laboratory Standards.** 1995. Approved standard. Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically. NCCLS document M7-A3. National Committee for Clinical Laboratory Standards, Villanova, Pa.
14. **Pickett, M. J., D. G. Hollis, and E. J. Bottone.** 1991. Miscellaneous gram-negative bacteria, p. 410–428. *In* A. Balows et al. (ed.), Manual of clinical microbiology, 5th ed. American Society for Microbiology, Washington, D.C.
15. **Roberts, M., L. P. Elwell, and S. Falkow.** 1977. Molecular characterization of two beta-lactamase-specifying plasmids isolated from *Neisseria gonorrhoeae*. J. Bacteriol. 131:557–563.
16. **Roberts, M. C., C. D. Swenson, L. M. Owens, and A. L. Smith.** 1980. Characterization of chloramphenicol-resistant *Haemophilus influenzae*. Antimicrob. Agents Chemother. 18:610–615.
17. **Roberts, M. C., and J. S. Knapp.** 1988. Host range of the conjugative 25.2-megadalton tetracycline resistance plasmid from *Neisseria gonorrhoeae* and related species. Antimicrob. Agents Chemother. 32:488–491.
18. **Roberts, M. C.** 1996. Tetracycline resistance determinants: mechanisms of action, regulation of expression, genetic mobility, and distribution. FEMS Microbiol. Rev. 19:1–24.
19. **Roberts, M. C., W. O. Chung, D. Roe, M. Xia, C. Marquez, G. Borthagaray, W. L. Whittington, and K. K. Holmes.** 1999. Erythromycin-resistant *Neisseria gonorrhoeae* and oral commensal *Neisseria* spp. carry known rRNA methylase genes. Antimicrob. Agents Chemother. 43:1367–1372.
20. **Roberts, M. C., J. Sutcliffe, P. Courvalin, L. B. Jensen, J. Rood, and H. Seppala.** 1999. Nomenclature for macrolide and macrolide-lincosamide-streptogramin B resistance determinants. Antimicrob. Agents Chemother. 43:2823–2830.
21. **Roe, D. E., P. H. Braham, A. Weinberg, and M. C. Roberts.** 1995. Characterization of tetracycline resistance in *Actinobacillus actinomycetemcomitans*. Oral Microbiol. Immunol. 10:227–232.
22. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed., p. 5.31–5.32. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
23. **Seppala, H., A. Nissinen, Q. Yu, and P. Huovinen.** 1993. Three different phenotypes of erythromycin-resistant *Streptococcus pyogenes* in Finland. J. Antimicrob. Chemother. 32:885–891.
24. **Shortridge, V. D., R. K. Flamm, N. Ramer, J. Beyer, and S. K. Tanaka.** 1996. Novel mechanism of macrolide resistance in *Streptococcus pneumoniae*. Diagn. Microbiol. Infect. Dis. 26:73–78.
25. **Sutcliffe, J., A. Tait-Kamradt, and L. Wondrack.** 1996. *Streptococcus pneumoniae* and *Streptococcus pyogenes* resistant to macrolides but sensitive to clindamycin: a common resistance pattern mediated by an efflux system. Antimicrob. Agents Chemother. 40:1817–1824.
26. **Sutcliffe, J., T. Grebe, A. Tait-Kamradt, and L. Wondrack.** 1996. Detection of erythromycin-resistant determinants by PCR. Antimicrob. Agents Chemother. 40:2562–2566.
27. **Tait-Kamradt, A., J. Clancy, M. Cronan, F. Dib-Hajj, L. Wondrack, W. Yuan, and J. Sutcliffe.** 1997. *mefE* is necessary for the erythromycin-resistant M phenotype in *Streptococcus pneumoniae*. Antimicrob. Agents Chemother. 41:2251–2255.

EXAMINATION OF DIETARY METHYLMERCURY EXPOSURE IN THE CASA PIA STUDY OF THE HEALTH EFFECTS OF DENTAL AMAGAMS IN CHILDREN

Carina C. Evens[1], *Michael D. Martin[1], James S. Woods[1],

Henrique Luis Soares2, Mario Bernardo[2], Jorge Leitäo[2],

P. Lynne Simmonds[1], Lian Liang[3], and Timothy DeRouen[1]

[1]Univeristy of Washington, Seattle, WA

[2]Univeristy of Lisbon, Lisbon, Portugal

[3]Cebam Analytical, Seattle, WA

Running title: Methylmercury exposure in the children's amalgam study

*Please address all correspondence to: Michael D. Martin, DMD, MSD, Ph.D.

Department of Oral Medicine, Box 356370

University of Washington

Seattle, WA 98195

Centro de Estudos Egas Moniz, Faculty of Medicine* and

School of Dental Medicine*** University of Lisbon, Portugal

Department of Psychiatry and Behavioral Sciences** and

Department of Dental Health***, University of Washington, USA

Isabel Pavão Martins, MD PhD; Department of Neurology, Centro de Estudos Egas Moniz, Hospital de Sta Maria 1600 Lisboa, Portugal; Telephone/Fax: 351 21 7934480; e-mail: labling@mail.telepac.pt

ABSTRACT

This study examined methylmercury concentrations in blood of children participating in the Casa Pia Study of the Health Effects of Dental Amalgams in Children over a one year period and related them to their diets in terms of fish and other seafood consumption. One hundred fifty children between the ages of 8 and 10 who were residents of the Casa Pia School system of Lisbon, Portugal, participated. Parents or caregivers completed a food frequency questionnaire designed specifically for this study at baseline. Children provided urinary and blood samples for mercury determinations at baseline and at one year following placement of dental tooth fillings. Mercury levels in fish samples from children's diets were also obtained. Mercury determinations in urine, blood and fish were performed using cold vapor atomic fluorescence spectroscopy. The mean value of baseline methylmercury concentrations in blood increased as the report of seafood consumption increased, although not statistically significantly. However, blood methylmercury and total mercury concentrations were significantly lower at one year follow-up than at baseline. Sixty one percent of parents/caregivers reported that their children consumed fish on a weekly basis. The fish offered at a sample of the schools contained low levels of methylmercury (range 13.9-23.6 ng/g). Thus, children participating in the Casa Pia dental amalgam study are exposed to low dietary levels of methylmercury by way of fish consumption, and this finding was reflected in the low mean blood methylmercury concentrations observed. The present findings indicate that dietary methylmercury is not a significant source of mercury exposure and is not likely to confound the association of dental amalgam mercury with potential health effects in the present study.

INTRODUCTION

This study focused on determining exposure to dietary methylmercury in a sample of 150 children aged 8-10 years who are participants in the Casa Pia Study of the Health Effects of Dental Amalgams in Children (a component of the Children's Amalgam Trial). Blood methylmercury (HgB) assessments were made at baseline, (prior to initial placement of fillings), and at one year followups. Consumption of fish and other seafood (crustaceans, mollusks, etc) in meals was obtained through the use of a food frequency questionnaire completed by either the parent or caregiver. Urinary mercury concentrations and mercury levels in fish samples were also examined.

Mercury exists naturally in three forms: metallic or elemental, inorganic, usually in combination with other elements such as chlorine, and organic including methylmercury. The latter is predominantly found in marine mammals, crustacea, mollusks and fish (Bjerregaard, 1996; Langworth et al, 1991). It is estimated that 86% of mercury in the diet is derived from fish consumption (USDHHS, 1994). Concentrations of methylmercury are highest in predatory fish such as pike, tuna, swordfish, walleye, and bass (Svensson et al., 1992; Cernichiari, et al., 1995; Marsh et al., 1995; Peterson et al., 1994). People who consume fish frequently have increased exposure to methylmercury. (Ratcliffe et al, 1996) It is believed that among persons not exposed to mercury in occupational settings, the most important source of mercury is through the diet, followed by exposure through the environment (air and water) and amalgam dental fillings (USDHHS, 1994; Loftenius et al, 1998; Nadarajah et al, 1996). Because about 90% of methylmercury in found in erythrocytes, analysis of mercury concentrations in blood is

one of the best ways to determine recent exposure (Gothe et al, 1985; Ratcliffe et al, 1996). Hair samples have also been used to assess cumulative organic mercury levels (Cernichiari, et al., 1995; March et al., 1995), although this approach is less reliable due to possible environmental contamination. Urine samples are predominantly used to assess recent exposure to inorganic mercury (USDHHS, 1994).

The health effects of methylmercury have been clearly demonstrated in two epidemics involving exposure of human populations to methylmercury-contamination - Minamata, Japan through the ingestion of contaminated fish and Iraq through the ingestion of contaminated grains (Clarkson, 1992; Ratcliffe et al, 1996). From these episodes, it has been determined that methylmercury predominantly affects the central nervous system (brain), immune function and kidneys (Aschner and Aschner, 1990; ATSDR, 1992; Kosuda et al, 1994; Loftenius et al, 1998). It is believed that exposure to methylmercury is more hazardous for children than adults due to the greater sensitivity of children's developing brains. Some of the effects associated with methylmercury intoxication include motor disturbances, sensory dysfunction, paresthesia, tremor, and incoordination (Gerstner and Huff, 1977; Girard et al., 1996). Methylmercury is more neurotoxic than inorganic mercury because of its greater ability to cross the blood brain barrier.

The specific aims of the present study were to (1) measure exposure to methylmercury at baseline and at one year following placement of dental fillings in a group of children residing within the Casa Pia school system , (2) examine the potential relationship present between number of fish/seafood meals eaten and organic mercury

levels in blood, and (3) describe the consumption of fish/seafood of these children using a food frequency questionnaire and menus acquired from the various schools.

METHODS

Study Population: The population of this study was comprised of a subset of 150 children enrolled in the Casa Pia Study of Health Effects of Dental Amalgams in Children (NIDCR U01 11894), a randomized clinical trial designed to evaluate the potential health effects associated with elemental mercury exposure derived from dental amalgam tooth fillings. Approximately 500 children aged 8-10 years at the beginning of enrollment on January 1, 1997 from the Casa Pia schools of Lisbon, Portugal, constitute the parent study population. The same recruitment and enrollment procedures as well as inclusion and exclusion criteria used in the parent Casa Pia study applied to this investigation. Overall, children who were born in 1986-88 were eligible. Other criteria included: at least one carious lesion in posterior permanent teeth, no prior amalgam exposure, IQ $\geq$ 67, blood lead $\leq$ 10 μg/dl, baseline urinary mercury $\leq$10 μg/l and no existing interfering health conditions. Of 647 children for whom screening and consent procedures were complete, 32 were excluded due to having no caries, 38 for prior exposure to mercury amalgam, 2 for a pre-existing health condition, and 5 due to having a urinary Hg level greater than 10 ug/l.

Dietary Mercury Exposure by Modified Food Frequency Questionnaire: Diet assessment is difficult in adults and even more challenging in children. Various methods have been used to assess nutrient values in children, such as the 24-hour dietary recall

form, food frequency questionnaire, and dietary records. This investigation employed a modified food frequency questionnaire designed specifically for this study and used questions similar to those asked by Peterson et al (1994) in a study that evaluated the fish consumption patterns of Wisconsin Chippewa Indians and blood mercury levels. The modified food frequency questionnaire was added to an already existing health screening questionnaire that either parents or caretakers of the children completed at baseline. Questions related to consumption of fish and seafood were asked. For the purpose of this study, "fish" was defined as free-swimming vertebrates and "seafood" as all other edible marine life (mollusks, crustaceans, etc.) Additional questions related to bread products, fat, sugar, vegetable and fruit consumption were added to reduce bias by disguising the focus of this study. Choices for level of consumption included: never, rarely, once/week, 2-3 times/week, once/day, 2-3 times/day, or 4 or more times/day. For analytical purposes these categories were grouped into never/rarely, weekly or daily consumption. Although it has been shown that selenium interacts with mercury to decrease the bioavailability of mercury (Hansen et al, 1984), specific questions related to selenium were not asked other than those about fish, since fish is a substantial source of selenium. Moreover, the region of Portugal in which Lisbon is located is considered neither a selenium-deficient nor selenium-excess area; thus, there was no reason to believe that the selenium content of the food or water ingested by the children would act as a potential confounder in this study. It has been noted that alcohol (USDHHS, 1994) also interacts with mercury. Given the young age of these children, alcohol consumption was assumed to be none. Descriptions about the children's diets were made by using the food frequency questionnaire and menus from the various schools.

Blood Collection Protocol: Sample collection, transportation and storage were performed as previously described (Liang et al., 2000). Briefly, five ml blood samples were collected by venipuncture for each child in a heparin-containing vacutainer. The sample was frozen and double-bagged in plastic and then shipped to Cebam Analytical, Inc., Seattle, for analysis. All baseline blood samples were collected between December, 1997 and April, 1998, whereas one-year follow-up samples were collected and shipped one year later. Explicit quality control protocols (Liang et al., 2000) were employed throughout the collection, shipment and analysis procedures to prevent mercury contamination of samples.

Blood Mercury Analysis: Blood mercury concentrations were employed as an estimate of methylmercury exposure. Total and organic mercury concentrations in blood samples were directly measured by ethylation–GC-CVAFS after alkaline digestion-solvent extraction, as described by Liang et al, (2000). The detection limit was found to be approximately 0.02ng/g by this method. Inorganic mercury was calculated as the difference between total and organic mercury.

Urinary Mercury Measurement: Urinary mercury analysis was performed by the method of Corns et al (1994) by continuous cold-flow, cold-vapor atomic spectrofluorometry, using a PSA Merlin Mercury Analysis (Questron Corp., Mercerville, NJ), as previously described (Woods et al, 1998)

Fish Mercury Measurement: Samples of cooked fish from meals served during a one week period during the one year followup (2/23/99 to 2/26/99) at four different Casa Pia schools were analyzed for total, inorganic, and methylmercury. Five fish samples from each school were analyzed as follows: The total and inorganic mercury

concentrations in fish samples were analyzed by CVAFS following preparation of tissues as described by Atallah and Kalman (1993). The organic mercury content of tissue samples was determined by the difference in the total and inorganic mercury concentrations.

Statistical Analyses: Descriptive statistics were used to characterize the children in this study. Analysis of variance was used to compare children within fish/seafood consumption groups and paired t-tests were used to compare mercury values over time. The .05 level of significance was used in hypothesis testing, and analyses were based on the assumption of normality.

RESULTS

Baseline Findings: The demographic characteristics of the study population are presented in Table 1. The majority of children were 10 years of age at baseline and of Caucasian descent with an almost even split of males and females. The mean blood methylmercury concentration at baseline was 2.71 ng/g and the mean total blood mercury concentration was 5.2 ng/g (Table 2). Thus, methylmercury accounts for 53% of the total blood mercury. The mean baseline urinary inorganic mercury concentration was 1.3 μg/l. Overall, there were more reports of fish consumption than seafood consumption, with the majority of parents/caregiver reporting their children eating fish on a weekly basis (Table 3). The mean value of blood methylmercury increased as the report of fish and seafood meals consumed increased (Figure 1: 2.03 ± 1.41 ng/g, none; 2.72 ± 1.50 ng/g, weekly; and 2.85 ± 1.49 ng/g, daily). However, these differences were not statistically significant.

Almost 700 menus were evaluated from the various schools within the Casa Pia school system. Most provided lunch only, but one of the schools provided both lunch and dinner. The overall average of fish/seafood meals offered in this school was 2.4 per week. This value is lower than the reported average from the questionnaire (4.3 fish/seafood meals per week), but closer to the mode of fish/seafood consumption from the questionnaire (2.5 fish/seafood meals per week).

One-Year Follow-Up Findings: Of the 150 children present at baseline, 132 were available at one year following placement of dental fillings. Of the 18 children who had missing data at one-year follow-up, 7 dropped out of the main Casa Pia study, 1 moved away, and 10 refused to give blood for a second time. The demographics of the remaining 132 children were quite similar to those of the total sample at baseline (Table 1). As can be seen from comparisons of Table 4 with Table 2, there were significant differences between baseline and one-year follow-up mercury levels. Both methylmercury and total mercury values were significantly lower at one-year follow-up than at baseline, whereas the mean urinary mercury concentration increased significantly as expected since approximately half of the children had dental amalgam fillings placed in the interim.

Mercury Analysis in Fish: The form of mercury present in the dietary fish samples was mostly organic, as expected, with only a small amount (approximately 10%) present as inorganic mercury. Results of analyses presented in Table 5 indicate low levels of total mercury present in these fish samples, consistent with the blood mercury results from the children.

DISCUSSION

The mean blood methylmercury concentrations found in children participating in the Casa Pia study were much lower than those reported for the children in either the recent Seychelles Child Development Study (Davidson et al., 1998) or the Faroe Islands study (Grandjean et al., 1997). Seven year-old children in the Faroe Islands study had hair methylmercury values ranging from 1.7-6.1 ppm, with a mean of 2.99 ppm, corresponding to a blood mercury level of 12 ng/g (Clarkson et al., 1986). Sixty-six month-old children from the Seychelles Study had hair methylmercury values ranging from 0.9-25.8 ppm , with a mean of 6.5, corresponding to a mean blood methylmercury value of 26 ng/g. Both of these means are well above the average blood methylmercury values (2.4 – 2.7 ng/g) found in the present Casa Pia study. It has been suggested that the high methylmercury values in the Seychelle children may be due to frequent fish consumption (average 12 fish meals per week) rather than to consuming fish with high levels of methylmercury. This again is substantially higher than the overall average 4.3 fish/seafood meals per week reported consumed by children in the present study.

The reason for the drop in mean Hg values from baseline to one year follow-up is unclear. The quality control measures assured that this drop was not due to a measurement error. Possible contributors include a continued drop in environmental exposure as air pollution levels from fossil fuels in the study location changed; and the children have increased in mass during the study period. Assuming even a steady state exposure, the concentration in blood and urine may be lowered for this reason.

While a direct comparison between the Portuguese children in this study and children in the U.S. is difficult, it appears that Portuguese children consume fish and

seafood more frequently than children in the United States. The U.S. Department of Agriculture's Continuing Surveys of Food Intakes by Individuals (CSFII) 1989-1991 three-day report indicates that 24.3% of males and 21.4% of females aged 6-11 years reported eating fish/seafood during this time period. This estimate corresponds to a fish/seafood consumption rate among U.S. children of approximately one meal/week. This estimate may be low, however, because reports did not include fish sandwiches made with tuna or fish patties.

Several limitations of the present investigation should be noted. There was concern that when parents/caretakers were completing the dietary questionnaire, they gave answers that they believed were desired by the interviewer, such as presenting an underestimate of the intake of "unhealthy foods" or overestimate the intake of "healthy food" such as fish. While efforts were made to minimize this bias in the design of the dietary questions, some misclassification might have taken place. However, this misclassification should have been non-differential or had the effect of biasing the results towards the null. Some error in assessment of measurement was also introduced by missing information on the dietary questionnaire.

In summary, children in the Casa Pia study of the health effects of dental amalgam fillings appear to be exposed to low levels of methylmercury from their diets despite relatively moderate intake of fish and seafood. Blood methylmercury levels derived from dietary intake are an order of magnitude less than those found in studies in which significant neurobehavioral effects were suggested. These findings indicate that dietary methylmercury is not a significant source of mercury exposure and is not likely to

confound the association of dental amalgam mercury with potential health effects in the parent study.

REFERENCES

Mercury Toxicity. Agency for Toxic Substances and Disease Registry. 1992. *Am. Fam. Physician.* 46(6):1731-1741.

Aschner, M., Aschner, J. 1990. Mercury neurotoxicity: mechanisms of blood-brain barrier transport. *Neurosci. Behav. Rev.*14:169-176.

Atallah, R. H. and Kalman, D.A. 1993. Selective determination of inorganic mercury and methylmercury in tissues by continuous flow and cold vapor atomic absorption spectrometry. *J Analyt. Toxicol.* 17: 87-92.

Bjerregaard, P. 1996. Cardiovascular disease and environmental pollutants: the arctic aspect. *Arctic Med.l Res.* 55 (Suppl 1):25-31.

Cernichiari, E., Toribara, T., Liang, L. et al. 1995. The biological monitoring of mercury in the Seychelles study. *Neuro. Toxicol.*. 16(4):613-628

Clarkson, T. W., Hursh, J.B., Sager, P.R. and Syversen, T.L.M. 1986. Mercury. In *Biological Monitoringof Metals* (Clarkson, T.W., Friberg, L., Nordberg, G.F. and Sager, P.R. Eds) Plenum Press, NY, pp199-246.

Clarkson, T.W. 1992. Mercury: major issues in environmental health. *Environ. Health Perspect.*100:31-38.

Corns,W.T., Stockwell, P.B. and Jameel, M. 1994. Rapid method for the determination of total mercury in urine samples using cold vapour atomic fluoroescence spectrometry. *Analyst.* 119(11):2481-2484.

Davidson, P.W., Myers, G.J., Cox, C., Axtell, C., Shamlaye, C, Sloane-Reeves, J., Cherichiari, E., Needham, L., Choi, A, Yang, Y. et al. 1998. Effects of prenatal and postnatal methylmercury exposure from fish consumption on neurodevelopmental outcomes at 66 months of age in the Seychelles Development Study. JAMA 280: 701-707.

Gerstner, H. and Huff, J. 1977. Clinical toxicology of mercury. *J. Toxicol. Environ. Health* 2:491-526.

Girard, M., Noel, F. and Dumont, C. 1996. Varying mercury exposure with varying food source in a James Bay Cree community. *Arctic Med. Res.* 55:69-74.

Gothe, C., Langworth, S., Carlseon, R. et al. 1985. Biological monitoring of exposure to metallic mercury. *Clin. Toxicol.* 23:381-389.

Grandjean, P., Weihe, P., Whire, R.F., Debes, F., Araki, S., Yokoyama, K., Marata, K., Sorensen, N., Dahl, R. and Jorgensen, P.J. 1997. Cognitive deficit in7-year old children with prenatal exposure to methylmercury. *Neurotoxicol. Teratol.* 19(6):417-428.

Hansen, F., Kromann, N., Wulf, H. et al. 1984. Selenium and its interaction with mercury in whole blood and hair in an east Greenlandic population. *Sci. Total Environ.* 38:33-40.

Knobeloch,L., Ziarnik, M., Anderson, H. 1995. Imported seabass as a source of mercury exposure: a Wisconsin case study. *Environ. Health Perspec.* 103:604-606.

.Langworth S, Elinder CG, Gothe CJ, Vesterberg O. 1991. Biological monitoring of environmental and occupational exposure to mercury. *Internatl. Arch.Occ.-Environ. Heath.* 63:161-167.

Liang, L., Evens, C., Lazoff, S., Woods, J.S., Cherichiari, E., Martin, M.D., and DeRouen, T. 2000. Determination of methyl mercury in whole blood by ethylation-GC-CVAFS after alkyline digestion-solvent extraction. *J. Analytic. Toxicol.* 24:328-332.

Marsh, D., Clarkson, R., Myers, G. et al. 1995. The Seychelles study of fetal methylmercury exposure and child development: introduction. *Neuro Toxicology* 16(4):583-596.

Peterson, D., Kanarek, M., Kuykendall, M. et al. 1994. Fish consumption patterns and blood mercury levels in Wisconsin Chippewa Indians. *Arch. Environ.l Health.* 49:53-74.

Svensson, B., Schetz, A., Milsson, A. et al. 1992. Fish as a source of exposure to mercury and selenium. *Sci. Total Environ.* 126:61-74.

U.S. Department of Health and Human Services. 1994. Toxicological Profile for Mercury (update) TP-93/10.

Woods, J.S., Martin, M.D. and Leroux, B.G. 1998. Validity of spot urine samples as a surrogate measure of 24-hour prophyrin excretion rates. Evaluation of diurnal variations in porphyrin, mercury, and creatinine concentration among subjects with very low occupational mercury exposure. *J. Occup. Environ. Health* 40:1090-1101.

ACKNOWLEDGEMENTS

This research was funded through a cooperative agreement with the National Institute of Dental and Craniofacial Research (NIDCR) (DE11894), NIDCR contract (NO1DE72623), NIDCR training grant (T32DE07227) and with additional support from University of Washington Superfund Program Project Grant from the National Institute of Environmental Health Sciences (2 P42 ES04696).

Issues in Design and Analysis of a Randomized Clinical Trial to Assess the Safety of Dental Amalgam Restorations in Children

Running Title: Safety of Dental Amalgam Restorations

Timothy A. DeRouen[1,2], Ph.D., Brian G. Leroux[1,2], Ph.D.,

Michael D. Martin[3,4], D.M.D., Ph.D., Brenda D. Townes[5], Ph.D.,

James S. Woods[6,7], Ph.D., M.P.H., Jorge Leitão[8], D.M.D., M.S.,

Alexandre Castro-Caldas[9], M.D., Norman Braveman[10], Ph.D.

Departments of Dental Public Health Sciences[1], Biostatistics[2], Oral Medicine[3], Epidemiology[4], Psychiatry and Behavioral Sciences[5], and Environmental Health[6], University of Washington, Seattle, Washington

Battelle Centers for Public Health Research and Evaluation[7], Seattle, Washington

Faculties of Dental Medicine[8] and Medicine[9], University of Lisbon, Lisbon, Portugal

National Institute of Dental and Craniofacial Research[10], NIH, Bethesda, Maryland

Funded by the National Institute of Dental and Craniofacial Research Cooperative Agreement U01 DE11894
Corresponding author: Timothy A. DeRouen, University of Washington Box 357475, Seattle, WA 98195

Abstract

The Casa Pia Study of the Health Effects of Dental Amalgams in Children is a randomized clinical trial designed to assess the safety of low-level mercury exposure from dental amalgam restorations in children. It is being carried out in 507 students (eight to twelve years of age at enrollment) of the Casa Pia school system in Lisbon, Portugal, by an interdisciplinary collaborative research team from the University of Washington (Seattle) and the University of Lisbon, with funding from the National Institute of Dental and Craniofacial Research. Since the goal of the trial is to assess the safety of a treatment currently in use, rather than the efficacy of an experimental treatment, unique design issues come into play. The requirements to identify as participants children who have extensive unmet dental treatment needs and who can be followed for seven years after treatment are somewhat in conflict, since those with the most treatment needs are usually in lower socioeconomic categories and more difficult to track. The identification of a primary study outcome measure around which to design the trial is problematic, since there is little evidence to indicate how health effects from such low-level exposure would be manifested. The solution involves the use of multiple outcomes. Since there are concerns about safety, multiple interim comparisons over time between treatment groups are called for which, in conjunction with the use of multiple outcomes, requires an extension of statistical methodology to meet this requirement. Ethical questions which have to be addressed include whether assent of the children participating is required or appropriate, and whether the Director of the school system, who is the legal guardian for approximately 20% of the students who are wards of the state and live in school residences, should provide consent for such a large number of children. Approaches taken to address these and other design issues are described.

Key Words: *safety, dental amalgam, mercury, multiple endpoints, interim analyses, children, neurobehavioral test*

Introduction and Background

Dental amalgam is the most widely used dental restorative material to repair cavities in teeth throughout the world. Approximately 100 million persons have amalgam restorations in the United States alone [1]. In addition to silver and other metallic components, amalgam contains approximately 50% elemental mercury, a known toxic substance that is now banned from most medical uses, and its use in other areas is under increased scrutiny to minimize environmental and occupational exposures. Mercury exists naturally in three forms: metallic or elemental (the form in dental amalgam); inorganic, usually in combination with other elements such as chlorine; and organic including methylmercury. While industrial use is often the source for high-level occupational exposures, mercury is relatively ubiquitous, and almost everyone has some background level of mercury exposure through the environment (air and water) or through their diet (methylmercury, primarily from consumption of fish). It is in the presence of these other background mercury exposures that any potential health effects of low-level mercury exposure from dental amalgam must be evaluated.

Dental amalgam has been in regular use in dentistry for more than 100 years, and while its safety has periodically been questioned, the lack of valid scientific evidence regarding its safety has made evidence-based discussions difficult at best. For many years, amalgam was thought to be relatively inert once it hardened. However, it contains approximately 50% elemental mercury, which readily vaporizes under pressure. Recent studies [2,3] have identified positive correlations between blood and tissue mercury levels with the number and total surface area of amalgam fillings. It is postulated that stress on the amalgam surface, such as that produced by chewing, grinding of teeth, or tooth brushing, causes the breakdown of a surface barrier and the release of mercury vapor into the mouth. When this release coincides with

inhalation through the mouth, inhaled mercury vapor readily diffuses across the alveolar membranes and is distributed to the central nervous system and other target organs mainly in red blood cells.

At times mercury from dental amalgams has been suggested as the cause of disease manifestations which range from mild dermatologic conditions, to chronic debilitating neuromuscular diseases, and even acute, life-threatening health outcomes [4-7]. The question of the safety of dental amalgam as a restorative material was addressed by a Food and Drug Administration Panel in 1991. While this panel recommended the continued use of dental amalgam, it acknowledged the lack of solid research into the issue, and voted unanimously for further research concerning the safety of mercury-containing dental amalgam. This need for a research program to examine the potential long-term biological effects of dental amalgam was echoed in a 1993 final report of the Committee to Coordinate Health and Related Programs, Public Health Service, which had been directed by the Assistant Secretary for Health to examine the question of amalgam safety [8].

If there were alternative restorative materials that were entirely equivalent to amalgam in terms of strength, longevity, cost, and ease of use, then the known risk for high-level exposure to mercury suggests that it might be prudent to eliminate low-level exposure from amalgam, even in the absence of documented health effects. However, it is generally recognized that no other material is as easy to apply and, once in place, lasts as long. The use of alternative materials, especially in the absence of ideal conditions and equipment, may result in more secondary decay around restored teeth and the need for more frequent replacement of restorations. Also, the alternative materials have not been thoroughly tested for health effects other than localized allergic reactions, and they may have their own harmful systemic effects. The elimination of

dental amalgam as a restorative material would have widespread public health and economic consequences, likely resulting in higher costs for restorations and reduced efficacy of alternatives. Thus, the safety of dental amalgam has become an important issue.

In an effort to assess the health effects of dental amalgam, the National Institute of Dental and Craniofacial Research (NIDCR) issued a Request for Applications for a randomized clinical trial to be conducted in children, who are the group receiving most restorations and are thought to be most vulnerable to the effects of mercury during their growth and development. This paper describes issues in design and analysis that had to be addressed in one such trial, called the Casa Pia Study of the Health Effects of Dental Amalgams in Children.

Design and Analysis Issues for a RCT on Treatment Safety

The question of safety of a treatment currently in widespread clinical use is an unusual question for which to design a randomized clinical trial. Moreover, given the previous long-term use of the treatment without objectively documented adverse health effects, it was expected that any health effects, if present, would be subtle and therefore difficult to detect. Thus, there were several design and analysis issues that had to be addressed in an unusual manner. This was not to be a treatment efficacy study. The dental treatments that would eventually be compared were assumed because of their long-term use to be effective as far as their use in successfully repairing carious dental lesions. There was some question as to which treatment would end up having greater longevity, but that question was entirely secondary.

The primary question was whether the use of dental amalgam was safe, so this study had to be designed as a safety study. In general, the study was proposed to be a randomized clinical trial in which half of the participants would receive treatment that included the use of dental

amalgam as appropriate in the standard of care (amalgam used for large posterior restorations in molars or premolars, but tooth-colored composite material used in front teeth), and the other half of the participants would receive treatment that included only alternative (non-amalgam) materials, specifically composite material consisting of resin (i.e., methacrylates) and inorganic filler (quartz, colloidal silica or silica glass) which is currently in use as an alternative to amalgam. Both treatment approaches are considered acceptable care, so neither group would be receiving treatment considered experimental. In the context of a safety trial, design and analysis issues that required special consideration included: the population/site in which to conduct the trial, the choice of study outcome measure(s), the types and frequency of any tests or examinations, the length of follow-up, the dental materials to use in treatment, the measures of mercury exposure to use, the statistical test(s) to be employed, and based on that, the sample size to be required.

The Population/Site in Which to Conduct the Trial

Issues considered in the identification of an appropriate study population included the need to enroll several hundred children who had a considerable amount of untreated dental caries in order to provide enough exposure to amalgam through treatment that any health effects, if present, might be manifested. At the same time, the children should not have had prior exposure to amalgam. To monitor for health effects subsequent to treatment, the ability to retain the children in the study for several (five to seven) years was important. There are no published data that would indicate the time course over which adverse health effects associated with the use of amalgam, if any, might occur.

Initial consideration was given to sites and populations within the U.S. Experience with outreach activities in the Pacific Northwest and a survey of oral health of third grade students in

the State of Washington indicated that there were children with untreated disease and no prior amalgam exposure, and such children were likely to be from low income and ethnic minority populations. One of the lessons learned from our outreach activities in those populations was that long-term follow up and retention could be problematic because of frequent movement and other difficulties in maintaining contact. Instead, an affiliation with the Faculty of Dental Medicine at the University of Lisbon in Portugal, and experience through them with the Casa Pia school system in Lisbon, led us to identify that as an ideal site in which to conduct the study.

The Casa Pia school system in Lisbon includes more than 4000 students in elementary and high school grades in seven schools. It was founded more than 200 years ago to provide an education for orphaned and homeless children in Lisbon, and it has since evolved into a large school system that uses a combination of private and public support to educate children from a much broader variety of backgrounds. Consistent with the original mission of Casa Pia, the school is the official guardian for approximately 20% of the students who are wards of the state and live in residences operated by the school. The other 80% live at home, including many from middle- to upper-socioeconomic backgrounds whose families pay tuition for their enrollment in a school system that is perceived as providing a good education.

There are several reasons why the Casa Pia school system was thought to be an ideal setting in which to conduct this kind of study. First, past experience suggested that there was an enormous unmet need for dental restorative treatment. This was confirmed in a small feasibility study in which we examined 41 children in our targeted range of 8-10 year olds and observed an average of 6.1 carious teeth, with 4 of those carious permanent teeth. By contrast, the average level of disease noted in the US population for a similar age group, as documented in the NHANES-III study [9] is 1.57 decayed or filled teeth (with 0.56 of those in permanent teeth). At

the same time we confirmed that prior exposure to amalgam was not likely to be an issue, since among the 41 children examined we saw only one amalgam restoration. Also, urinary mercury samples from the children indicated low background levels of mercury from other sources (median urinary mercury level of 2.0 micrograms/liter), which also was important in trying to assess the mercury exposure levels and health effects due solely to amalgams.

Another characteristic that made this setting ideal was that the University of Lisbon had previously worked with the Casa Pia schools in having dental hygienists provide instruction in oral hygiene to students, which led school administrators to be enthusiastic about cooperating in a study that could provide needed dental care for a large number of their students. This cooperation was essential in carrying out the logistics of the study, since it allowed students to be transported, tested, and treated during school hours. The final characteristic that made this population so attractive for conducting this kind of study was that the school system, in its admissions process, developed an extensive social network of contacts for each student that would prove to be quite valuable for maintaining contact and follow-up for enrollees, especially for those who leave the Casa Pia school system.

The decision to conduct this trial in a foreign country was not made without regard to ethical concerns or whether the results would be generalizable to the U.S. Trials are sometimes conducted in foreign countries to minimize costs. In this case, although the cost of delivering dental care to 500 children in Portugal was less than it would have been in the U.S., the additional cost for investigators to travel between Seattle and Lisbon to supervise and coordinate the study somewhat offsets those savings, so that cost was not an overriding factor in the decision. For all the reasons described above, it was felt that the scientific quality of the study would benefit from the use of this study population. The main concern was the question of

whether a study conducted in another country could be generalized to the U.S. population. In order to maximize the generalizability of the study to the U.S., care was taken to ensure that the dental practices and materials in the study were not only those used in Portugal, but were also typical of those in the U.S. It was established that the racial composition of the Casa Pia population was not markedly different from the general U.S. population, and it was also recognized that if the study were conducted in urban settings in the U.S., the known distribution of caries among children in the U.S. suggests that many potential participants would come from families recently immigrated to the U.S.

The Choice of Study Outcome Measure(s)

Mercury's toxic effects on human beings can vary with the form, extent, and route of exposure. The fate of mercury in the body, although extensively investigated, remains somewhat unknown. Target organs for elemental mercury include the kidneys and the central nervous system. Information about the deleterious effects of mercury in human beings generally comes from occupational or environmental disasters that have produced very high exposure levels. The study of health effects of mercury exposure from dental amalgam requires the extrapolation of information obtained in such very high exposure levels to low-level exposure where there is little data.

Place Figure 1 here.

Consideration of study outcome measures in a safety trial such as this is particularly difficult when it is not clear how or when any health effects will be manifested, if there are any. What is known about signs and symptoms of mercury toxicity is summarized in Figure 1, which indicates the urinary mercury levels at which the various signs and symptoms have been documented. Most of this knowledge comes from environmental and occupational studies in

which exposure levels result from shorter duration exposures and are much higher than would be expected in a study of dental amalgams. At the time the study was being designed, the lowest exposure level at which sub-clinical changes in behavior had been detected was 36 μg/liter of urine [10], with most of the clinical signs and symptoms documented at levels of 100-1000 μg/liter. In a study of exposure from amalgam, the urinary mercury levels were expected to be less than 10, and probably mostly in the range of 2–5 μg/liter. Thus, identification of one or more primary outcome measures required extrapolation to low levels for which there was no information. (Information in Figure 1 about motor function (hand steadiness) changes at 26μg/liter [11] and potential pre-clinical effects on symptoms, mood, motor function, and cognition related to exposures in the range of 0-4 μg/liter [12] were published after the study was underway). Based on the toxicology of elemental mercury and information from the studies of high-level exposures, the target organs were felt to be the renal and neurologic systems. Discussions among an interdisciplinary team of experts in toxicology, neuropsychology, neurology, nephrology, epidemiology, and biostatistics, led to identification of a constellation of potential primary outcome measures, consisting of test performance in three neurobehavioral domains thought to be potentially affected by mercury (memory, attention/concentration, and motor/visual motor), nerve conduction velocity (posterior tibial and ulnar nerves), and urinary glutathione transferase levels (GST-α and GST-π).

The neurobehavioral tests were chosen on the basis of their appropriateness for the age of the children, documented sensitivity to the integrity of brain functions, cross-cultural applicability, and appropriateness for measuring abilities in each of three functional domains plus intelligence (Table 1). The Comprehensive Test of Nonverbal Intelligence [13] was administered as a pre-enrollment measure of IQ. The IQ itself was considered to be too gross a

measure of neurobehavioral abilities, and was therefore considered a secondary outcome measure. The focus in the choice of primary outcome measures was to assess attention, memory, and motor/visual motor functions in depth in order to identify possible subtle changes that might occur with mercury exposure. The neurobehavioral battery included: Rey Auditory Verbal Learning Test [14], Visual Learning and Finger Windows subtests from the Wide Range Assessment of Memory and Learning [15], Trail Making Test [16], Stroop Color Word Test [17], Standard Reaction Time [18], Finger Oscillation Test [19], Pegboard, Matching, and Drawing tests from the Wide Range Assessment of Visual Motor Abilities [20], and Digit Span, Coding and Symbol Search subtests from the Wechsler Intelligence Scale for Children-III [21].

Subsequent discussions concerning the GST enzymes, which are indicators of damage to specific regions of the renal nephron, led to the conclusion that the enzymes are not only highly elevated when there is permanent kidney damage, but their levels are also likely increased in response to low-level mercury exposure when no permanent damage has occurred. Moreover, there was no information available which indicated the levels of the GST enzymes that differentiated permanent renal damage from reversible renal response. Nor was there any information relating the level of GST to clinical symptoms associated with continuous levels of renal function and/or malfunction. Had GST levels been included as a primary outcome measure under those circumstances, the validity of the study could have been jeopardized by the observation of a significantly increased GST level among those exposed to amalgam that may have merely been a confirmation of exposure rather than an indication of any serious renal damage. Once this issue concerning GST levels was fully understood, the GST enzymes were designated as secondary outcome measures, subject to exploratory analyses, along with another renal measure of response to mercury, urinary porphyrin levels. Unfortunately, no renal measure

was identified that was sensitive enough to serve as a primary outcome measure, since anything as definitive as kidney failure was not expected to occur as a result of such low-level mercury exposure.

Although exceedingly rare, mercury-induced autoimmune reactions that can lead to glomerular nephritis have been reported [22]. Therefore, we are measuring urinary albumin to monitor for such potential immunologic reactions associated with mercury exposure, but this is a secondary outcome measure because it is expected to be so rare.

Further discussion of the remaining four potential primary outcome measures did not resolve the issue of which of those measures was most likely to be affected by and sensitive to low-level mercury exposure. Although most clinical trials concerning treatment efficacy are designed around one definitive outcome measure, one consequence of designing a safety trial, as in this case, may be the need to be cautious and include several primary outcome measures. In this case, the decision was made to use four primary outcome measures, as defined in Table 1. Each of these measures is calculated as an average of Z-scores (standardized to have mean 0 and SD 1 across the sample) for tests within a particular domain. The secondary outcome measures are given in Table 2.

Place Table 1 here.

Place Table 2 here.

The Length of Follow-up and Frequency of Tests

Although there was a possibility that any health effects of low-level mercury exposure could be manifested in a relatively short period of time, the expectation was that it would likely take several years to demonstrate any differences in the development of the children. While follow-up until adulthood was perhaps the ideal, that would have required at least 8 to 10 years of follow-up, which did not seem feasible financially and could mean risking a non-evidence-based solution to a potentially volatile question prior to completion of the trial. The compromise was to require at least 5 years, and hopefully 7 years, of follow-up. The study was initially proposed for 5 years of follow-up after treatment, and recently approved renewal funding extended it to 7 years.

The question of the frequency of testing and examination was resolved by the nature of the tests and practicality. The neurobehavioral tests could not be repeated too often without the risk of learning effects. On the other hand, they were affected by development of the children in school, and it was thought important to test the children at approximately the same time of the school year to avoid introducing more variability in the test scores. The children needed to have a dental exam at least annually and, if there was disease progression, be treated. All of these requirements combined led to the development of an annual testing/treatment cycle as optimal.

The Dental Materials to Use

Since the purpose of the study was not to test the efficacy of any specific brands of dental materials, but to assess whether dental amalgam as used in practice is safe, materials selected for

use in the study were chosen to be typical of what is employed in practice in Portugal and in the U.S. The selection of materials typical of U.S. practice was made in order to ensure that the results could be generalized to the U.S. In order to avoid introducing any unnecessary variability in exposures and therefore potentially in outcome measures, it was desirable to standardize the brands of materials available to the dentists who were providing treatment. The most important material, of course, was amalgam. Although there are several manufacturers of amalgam, the basic composition of the different amalgams, in terms of the percent of mercury, is approximately the same (50%). The choice of amalgam then was the brand most widely used in the U.S. (Dispersalloy by Dentsply). Similarly, the alternative material to be used instead of amalgam, posterior composite, was selected because it is the most widely used composite in the U.S. (Z100 MPq by 3M). The other materials that the dentists could use in treatment, including composite adhesive, glass ionomer base, compomer, and compomer adhesive, were selected to be representative of what is used in practice in the U.S., and are given in Table 3. In the protocol, the children in the amalgam group were to be treated with amalgam as appropriate in the standard of care, which basically means in large restorations in posterior teeth, but they could at the same time be treated with other materials (composite, glass ionomer, or compomer) for restorations not appropriate for amalgam. Thus, the two treatment groups differed in exposure to amalgam, but they could overlap in exposure to the other materials.

Place Table 3 here.

Measures of Mercury Exposure

While the trial merely randomizes patients to one of two treatment regimens (one including dental amalgam, the other excluding dental amalgam), there is a need in secondary analyses to document the amount of mercury exposure produced by the amalgam treatment, and

to allow investigation of any dose-response effects. The best measure of mercury exposure depends on the form of the mercury. For example, mercury exposure from diet is predominantly methylmercury, primarily from fish consumption. Because about 90% of methylmercury is found in erythrocytes, analysis of mercury concentrations in blood is one of the best ways to determine recent exposure to methylmercury [23]. Hair samples have also been used to assess cumulative organic mercury levels, [24, 25] although it is considered less reliable due to possible environmental contamination. Single void ("spot") urine samples are predominantly used to assess recent exposure to elemental mercury [26, 27].

The main assessment of exposure to elemental mercury from amalgams used in this study is analysis of urinary mercury, determined from samples obtained at baseline (prior to treatment), and at all subsequent annual visits prior to re-treatment. Measurements of urinary mercury concentrations made at this point in time should involve the least amount of variability, since exposure has been of sufficient length of time since the previous treatment to allow body compartments to achieve steady-state mercury levels [28]. Urinary mercury analysis is performed by the method of Corns et al [29] by continuous cold-flow, cold-vapor atomic spectrofluorometry, using a PSA Merlin Mercury Analysis (Questron Corp., Mercerville, NJ). Urinary creatinine content (used to also provide creatinine-adjusted urinary mercury levels) is determined using the colorimetric determination assay kit obtained from Sigma Chemical Co. (St. Louis, MO).

All dental care providers in the study are trained to current stand-of-care in placing amalgam and composite restorations. As a result, no meaningful variation in the quality of care is expected. Also, individual treatment appointments are essentially randomly assigned to the various providers in the study, which should distribute any very unlikely provider effects equally

between the two treatment groups. Any problematic treatments (such as restorations which were significantly "high" upon placement) would be very few in number and would very quickly result in re-treatment for any of a number of clinical reasons including discomfort to the patient, cracking, etc. None of these issues are expected to have any impact on the levels of mercury exposure experienced by study participants.

In addition to urinary mercury concentrations, the other measure of mercury exposure is obtained from clinical records of the amalgam restoration surfaces placed. Given the date of the placement of each surface, one can calculate the number of months of exposure contributed by each surface at a given point in time, and the total number of surface-months of exposure contributed by all restorations in a person as a cumulative measure of exposure at a given point in time. To aid in further refining the surface-months of exposure to take into account the size of each restoration, if needed, each restoration is categorized into one of three categories of size, small, medium, or large, and these categories could potentially be differentially weighted in determining cumulative weighted surface-months of exposure. While it is unlikely that study participants would receive dental treatment outside of the study, at annual re-examinations of the subjects notations are made of all restorations present in the mouth. With this data, it is possible to discern whether or not an amalgam has been placed which was not placed as part of the study, and if so, take it into account in measuring amalgam exposure.

While exposure to elemental mercury from amalgam is of primary interest in the study, there was concern that dietary exposure to methylmercury could potentially act as a confounder and mask any effects of amalgam exposure. Therefore, in a sample of 150 of the participants, a dietary questionnaire was administered to assess intake of seafood, and blood samples were taken to measure recent methylmercury exposure. For this, total and organic mercury

concentrations in blood samples were directly measured by ethylation-GC-CVAFS after alkaline digestion-solvent extraction [30].

The Statistical Tests Employed

There were several issues that had to be taken into account in determining the statistical approach to be used. The first two were issues resulting directly from fundamental decisions made about study design: that there would be multiple (four) primary outcome measures, and that the study would be longitudinal, with follow-up and annual testing for seven years.

Third, there was a need to have a two-sided alternative hypothesis, although the two sides were not considered to be of equal importance. Our expectation was that if the two treatment groups differed, the difference would be in the direction of amalgam being harmful. This was particularly true since the outcome measures were specifically selected to be sensitive to the effects of amalgam. Reports released as the protocol was being finalized suggested there may be low-level exposure to estrogenic-like compounds from composites, so it was desirable to incorporate a legitimate test for harmful effects of the composite material. However, the reports on composites were too preliminary and came too late in the process to allow incorporation of outcome measures specifically selected to detect harmful effects of composites. Another reason for emphasizing the detection of adverse effects of amalgam over adverse effects of composite was the clear difference between treatment groups in the levels of mercury exposure, whereas the difference in exposure to composite material was a matter of degree (making any observed treatment effect in this direction difficult to interpret).

Fourth, there was a need for annual interim analyses. Although the dental treatments given were considered standard of care, from an alternative point of view we were exposing

children to low levels of a substance that in higher doses has been shown to have health effects, and it was felt important to closely monitor the effects of the trial and assess each year whether the trial should be continued.

Fifth, there was the question of how any adverse effects of the amalgam, if present, would be manifested among the primary outcome measures. The adverse result considered most likely was consistent but subtle effects in all or most of the outcome measures, no one of which was large enough to be statistically significant by itself. If the four outcomes were equally affected by mercury exposure, and they were consistent in demonstrating small harmful effects from the low-level mercury exposure, we would want the statistical test to be sensitive to these consistent subtle effects even though no one outcome measure might stand on its own as being statistically significant. On the other hand, if it turned out that one of the four primary outcomes was much more sensitive to low-level mercury exposure than the other three, we would want the statistical test to be able to detect an effect manifested in only one outcome measure. As is indicated later, these two different scenarios for the way any harmful effects could be manifested (and the lack of information related to which was more likely) would lead to the use of two different tests.

The final issue that needed to be taken into account in the statistical approach was that care should be taken to avoid unnecessary risks of obtaining false positive findings. Since amalgam is used worldwide, is less expensive (it takes less dentist time), and is generally thought to be more durable, especially when employed under less-than-ideal working conditions (it is less technique sensitive), a false positive finding of a harmful effect of amalgam could lead to an unnecessary ban on its use with severe economic and public health consequences. Of course, if there truly are harmful effects of amalgam, then we want to detect them, and the benefit of

avoiding the associated health effects would likely outweigh the economic and public health consequences of a ban. However, the temptation in trying to detect a small harmful effect is to increase the size of the rejection region (significance level), and therefore the Type I error rate. For this study, it is important to use a test procedure that is sensitive to potential harmful effects of mercury, but to balance the need for sensitivity by maintaining a reasonable overall Type I error rate.

The use of multiple outcome measures led to the adoption of an O'Brien [31] test as a means of combining the multiple endpoints. However, the test was originally proposed for a one-time analysis at the end of a study, and its use in this study with longitudinal data and interim annual testing required the extension of the test for that purpose [32]. The O'Brien-type test is particularly sensitive to possible consistent small effects in all outcome measures, but it is not sensitive to a possible effect manifested in only one outcome. To protect against that possibility, the Hotelling T-squared multivariate test was also adopted as part of the overall testing procedure. The overall significance level of the entire statistical procedure was set at 0.05. A two-sided testing procedure was adopted to allow legitimate testing for harmful effects of both restorative materials. However, unequal significance levels were used for the two sides in recognition of the fact that the study was designed and outcome measures were selected to be sensitive primarily to harmful effects of amalgam. The overall 0.05 significance level was divided into a level of 0.03 for an O'Brien-type test sensitive to uniform harmful effects of amalgam in all outcome measures, a level of 0.01 for the O'Brien-type test sensitive to uniform harmful effects of composite in all outcome measures, and a level of 0.01 for the Hotelling T-square test sensitive to a large effect (in either direction) in just one outcome measure. Furthermore, since the longitudinal nature of the study required interim annual testing, the

overall significance level of 0.05 was spent over seven interim tests, with the overall significance level for each annual analysis given in Table 4. As is shown, the spending function we used did not distribute the alpha levels uniformly over all years. Because we felt caution was called for in the first year of the trial to quickly detect a potential early harmful effect (i.e., an incremental effect manifested in the first year), we spent a higher proportion of the alpha level (0.0125) in the first test. If the effect were not incremental in nature, then we assumed it would be cumulative and linear over time, and we could spend the remaining alpha equally (0.0015) over years 2 through 6, while saving the majority of the alpha (0.03) for the final analysis at year 7.

Place Table 4 here.

Sample Size Determinations

Using the statistical procedures described above, the sample size required for the study had to be determined based on the power of the procedure for detecting potential harmful effects that might be observed. Given the infinite number of possible scenarios, especially with four outcome measures and two different statistical tests involved, we decided to focus on what we thought might be typical effects which we would want high power to detect. We looked at two scenarios: one with small but (relatively) consistent effects in all outcome measures, and the other with a pronounced effect in only one outcome measure. Scenario 1 consisted of effect sizes of 0.3 standard deviations in each of the three neurobehavioral outcome measures (memory, attention/concentration, and visual motor) and an effect size of 0.15 in nerve conduction velocity. The postulated effect size of 0.3 standard deviations is abstract and somewhat arbitrary, but in the absence of any evidence as to what kind of difference between the two groups might be of concern, this effect size does have a descriptive interpretation based on a presumed normal distribution. If the population mean shifts by 0.3 standard deviations, the

proportion that was originally identified as the lower 2.5% of the population distribution (and perhaps classified as "abnormally low") would increase to 5%, resulting in an approximate doubling of the "abnormally low" proportion of the population. We felt this to be a reasonably small effect that we would want to detect. Scenario 2 consisted of a larger effect (an effect size of 0.5 SD) in only one outcome measure, nerve conduction velocity, with no effect in the other outcome measures.

Exploration of power for these scenarios with varying sample sizes resulted in the conclusion that a total sample size of 400 (200 in each group) with complete data at the end of the study would provide adequate power (greater than 97%) to detect the two scenarios. In order to account for loss to follow-up and ensure that we would end up with adequate power, we set an enrollment minimum of 450 with a target of 500. We have continued to monitor the power of the study based on our experience with recruitment and follow-up of subjects (see details on implementation below). In the competing renewal proposal recently reviewed and recommended for funding, it was proposed that the follow-up period be extended to 7 years after initial treatment to allow for sufficient power to detect Scenario 2. The annual attrition rate was assumed to be 5% in the first year and 3% subsequently. The results showed that the study will have sufficient power (94%) with just a five-year follow-up period to detect small effects that are roughly uniform across outcome measures as in Scenario 1. The power to detect similar adverse effects of composite material is smaller because smaller significance levels are used for these tests, but is still adequate (87%). Adequate power (94-95%) exists with 5-year follow-up to detect a linear effect in just one outcome measure equal to 0.1 per year over five years. However, five-year follow-up is inadequate to detect a linear effect in just one outcome measure equal to 0.06 per year (cumulative effect size 0.3 over 5 years or 0.42 over 7 years). The power is just

49% with five years of follow-up, but extension of the follow-up period to seven years increases the power to a reasonable level of 85%.

Implementation

Enrollment commenced in February, 1997, when final authorization to proceed came from our Data Safety Monitoring Board and NIDCR. As is shown in Figure 2, we had projected an enrollment period of approximately 10 months (December 1997) to enroll the minimum number of 450. Because of a slight delay at the beginning in getting started, and because the enrollment rate was slightly slower than anticipated, we reached the minimum enrollment number of 450 in May 1998. However, because the goal enrollment of 500 seemed within reach, we continued enrolling until October 1998, when enrollment was stopped at 507, with 253 randomized to one treatment group and 254 to the other. As is indicated in Table 5, the 507 enrollees were obtained from a pool of 845 students in the appropriate age range who were contacted. Of those contacted, 647 (76%) provided consent and were screened. Among those screened and declared ineligible, most either had previous exposure to amalgam, or did not have at least one carious lesion on posterior permanent teeth. Implementation of enrollment and follow-up raised additional design issues that then had to be addressed.

Ethical Issues in Informed Consent

Since participants in the study are minors, consent had to be obtained from parents or guardians for their participation. The parental consent form was reviewed and approved by IRB's at the University of Washington and the University of Lisbon. However, the issue was raised as to whether we also were required to obtain the assent of the child. The technical answer to the question was that we were not required to obtain assent from the children because

the treatments being provided in both groups were considered standard of care, not experimental. However, we decided that it would likely lead to better cooperation by the children if we included them in the consent process, so we also asked for their written assent.

One issue raised by the Data Safety and Monitoring Board (DSMB) about the informed consent process concerned the 20% of the children in the schools who were wards of the state and who lived in Casa Pia housing. For those children, the Director of the Casa Pia school system was their legal guardian. The ethical issue raised by the DSMB was whether it was appropriate for one person, the Director, to have consent responsibility for such a large group of students. One concern was whether it was a conflict of interest for the senior administrator of the school system to make such a decision, given that the school system was responsible for the general welfare of these children, and there could be a financial benefit to the school system to have all of these children receive free dental care through participation in the study. On the other hand, access to dental care for all children in the Casa Pia school system was quite limited, regardless of whether they lived at home or at school. There was just one part-time dentist for the entire school system who only treated emergencies, and very few of the families sought or had access to dental care in private or public health settings. Thus, there seemed to be no more of an incentive for the Director of Casa Pia to consent to students who were residents of Casa Pia to participate, than there was for parents of individual students to consent. In both cases, participation brought with it access to dental care. A related ethical concern about this was whether the offer of free dental care was too great a financial incentive, and constituted a form of coercion. The counter argument was that the "exposure" in the trial consisted of dental treatment, and since the targeted population was not seeking dental treatment, they probably would not have participated if the dental treatment had not been free.

When the DSMB questioned the appropriateness of one person providing consent for the many children (approximately 100) who were residents of the schools, it created an issue with the potential to affect the cooperation from the Casa Pia schools that had been (and continues to be) so important to the study. The Director of Casa Pia is an individual who is highly respected in the community and who takes his responsibilities as legal guardian of the resident children very seriously. Addressing the DSMB concern by requiring additional consent by someone other than the Director would have meant imposing a consent process different from what is legally required in Portugal, it might have been interpreted as questioning the integrity of the Director, and therefore it could have caused problems. The solution, as suggested by the NIH Office of Protection from Research Risks, was to ask the IRB at the University of Lisbon to revisit the consent process for the resident Casa Pia children to determine if the provision of consent by the Director met local community ethical standards. They met, reviewed the process, and confirmed its appropriateness.

Sentinel Health Events Monitoring System

While the study is designed to evaluate the safety of treatment with dental amalgam, there is no guarantee that any adverse effects will only occur among the primary or secondary outcome measures we have specified. In reviewing our follow-up on patients post treatment, the DSMB asked us to also monitor for any unexpected medical diagnoses or health events that might occur during follow-up. The only mechanism initially in place to detect such occurrences was the routine questions asked by the treating dentist about any recent illnesses. Since the patients were children, that process was not thought to be sufficient to ensure that any significant illnesses would be identified. We implemented a short health history questionnaire sent to parents and guardians annually, but the DSMB thought our questions were too open ended and

lacked sufficient detail, while we resisted developing a lengthy annual questionnaire to send to parents. It was questionable whether our informed consent covered administration of a subsequently developed annual questionnaire, and (more importantly) investigators from the University of Lisbon doubted if we would get a good return rate from parents and thought we might run the risk of alienating them. In order to provide reasonable assurance that we could identify any serious illnesses developed by the study participants during follow-up, we instituted a multifaceted sentinel health events monitoring system. The system consists of three components. One is the annual short health history questionnaire sent to parents. The second was the development of a structured set of health history questions (asking about a variety of symptoms) that a dental hygienist administers to each child at the time of the annual exam. The third component consists of consultation with teachers and school officials, including the school physician, to systematically monitor the children in the study for reported diagnoses or extended absences that could be related to illness. While any reported illness identified through this system may not be completely reliable, if sufficient numbers of illnesses of any specific type are reported to be of concern, additional efforts can (and will have to) be made to get proper consent for access to medical records and appropriate confirmation.

Besides annual interim statistical analyses to compare group responses on the primary endpoints of the study, and the sentinel health events monitoring system to identify unexpected medical diagnoses or events, part of the overall monitoring of the study involves identifying and assessing individual extremes in neurobehavioral performance each year to identify any children who may require referral for treatment. This process involves using the Reliable Change Index [33,34] which takes into account the standard error of measurement as well as the correlation between performances on a test at two time points to identify children whose performance falls

outside a one-sided 95% confidence interval for their predicted value. Those children identified by the RCI method as having extreme (poor) scores on several of the individual neurobehavioral tests have their overall performances and clinical records examined by the study neuropsychologist (blinded to treatment group identification) as clinical case studies to determine if the patterns of test performances and clinical neurological exam findings indicate a diagnosis requiring referral for treatment.

Baseline Characteristics

As a result of the randomization process, patients were assigned to one of two treatment groups (designated as Groups A and B or 1 and 2 to maintain blinding). As is shown in Table 6, the two groups were balanced as far as demographic characteristics, with approximately 54% male and 70% Caucasian in both groups, with only slight variations in year of birth. The treatment needs of the two groups at baseline were also comparable, with one group averaging 15.6 decayed surfaces (primary and permanent teeth combined), and the other 15.8 decayed surfaces. The background levels of urinary mercury in the two groups are shown in Figure 3, with only slight group differences (one group averaging 1.42 micrograms of mercury per liter of urine, the other group averaging 1.54). Both group averages are at the lower end of what has been observed in populations with no known overt sources of mercury exposure.

Place Table 6 here.

Place Figure 3 here.

Baseline measures for the two groups in the primary outcome measures are shown in Figure 4 (for the three neurobehavioral domains of memory, motor/visual motor, and attention/concentration, as well as for nerve conduction velocities). Again, the two groups are balanced at baseline in the primary outcome measures, which will allow for subsequent group comparisons to be made without having to adjust for baseline differences. However, the use of regression adjustment for baseline scores (or alternatively a change score) will be considered if it provides higher precision, and hence higher power.

Place Figure 4 here.

Summary

The need to examine the safety in children of a treatment (dental amalgam) already in widespread use presented some unique design challenges. It required identifying a population with extensive dental disease that could be followed up long term, identifying study outcome measures in the absence of hard evidence concerning how health effects of low-level mercury exposure would be manifested, and the development of statistical methods that allowed group comparisons on multiple outcome measures at multiple interim time points. The Casa Pia Study addressed those issues and others, and is currently underway. Follow up will continue until seven years of follow up after treatment has accrued, which will occur in 2005.

Acknowledgments

The authors wish to acknowledge the many contributions by others on the study team to various aspects of the study design and baseline results presented here. These include: (from the University of Lisbon) Henrique Luís, Dr. Mario Bernardo, Lurdes Vaz, Helena Amaral, Dr. Isabelle Pavao, Dr. Mamede Carvalho, Goretty Ferreira, Pedro Rodrigues, Helena Nazareth, Dr. Paula Marques; (from the University of Washington) Gail Rosenbaum, Natalie Hawkins, Aimee Sparks, Dr. Jacqueline Farwell, Dr. Harvey Sarnat, Dr. Sandra Watkins, Lynne Simmonds, Dr. Glen Johnson; and (from NIDCR) Dr. Maryann Redford.