

Neurobehavioral Effects of Dental Amalgam in Children

A Randomized Clinical Trial

Timothy A. DeRouen, PhD

Michael D. Martin, DMD, PhD

Brian G. Leroux, PhD

Brenda D. Townes, PhD

James S. Woods, PhD, MPH

Jorge Leitão, MD, MS

Alexandre Castro-Caldas, MD, PhD

Henrique Luis, MS

Mario Bernardo, DMD, PhD

Gail Rosenbaum, MS

Isabel P. Martins, MD, PhD

DENTAL AMALGAM, WHICH consists of approximately 50% elemental mercury, was thought for most of the 150 years it has been in use to be inert once it sets. Increasingly sensitive technology has recently demonstrated that some of the elemental mercury in amalgam is vaporized under pressure from mastication, and positive correlations have been found between urine, blood, and tissue mercury levels and the surface area or number of amalgam fillings.¹⁻³ Since high levels of mercury have been demonstrated to be toxic, the fact that dental amalgam induces some level of mercury exposure raised safety concerns. However, there is little or no evidence concerning health effects of low-level mercury exposure from amalgam, especially in children. A 2005 comprehensive review of evidence published since 1996 concluded that there still is not "sufficient evidence to sup-

Context Dental (silver) amalgam is a widely used restorative material containing 50% elemental mercury that emits small amounts of mercury vapor. No randomized clinical trials have determined whether there are significant health risks associated with this low-level mercury exposure.

Objective To assess the safety of dental amalgam restorations in children.

Design A randomized clinical trial in which children requiring dental restorative treatment were randomized to either amalgam for posterior restorations or resin composite instead of amalgam. Enrollment commenced February 1997, with annual follow-up for 7 years concluding in July 2005.

Setting and Participants A total of 507 children in Lisbon, Portugal, aged 8 to 10 years with at least 1 carious lesion on a permanent tooth, no previous exposure to amalgam, urinary mercury level <10 µg/L, blood lead level <15 µg/dL, Comprehensive Test of Nonverbal Intelligence IQ ≥67, and with no interfering health conditions.

Intervention Routine, standard-of-care dental treatment, with one group receiving amalgam restorations for posterior lesions (n=253) and the other group receiving resin composite restorations instead of amalgam (n=254).

Main Outcome Measures Neurobehavioral assessments of memory, attention/concentration, and motor/visuomotor domains, as well as nerve conduction velocities.

Results During the 7-year trial period, children had a mean of 18.7 tooth surfaces (median, 16) restored in the amalgam group and 21.3 (median, 18) restored in the composite group. Baseline mean creatinine-adjusted urinary mercury levels were 1.8 µg/g in the amalgam group and 1.9 µg/g in the composite group, but during follow-up were 1.0 to 1.5 µg/g higher in the amalgam group than in the composite group ($P<.001$). There were no statistically significant differences in measures of memory, attention, visuomotor function, or nerve conduction velocities (average z scores were very similar, near zero) for the amalgam and composite groups over all 7 years of follow-up, with no statistically significant differences observed at any time point (P values from .29 to .91). Starting at 5 years after initial treatment, the need for additional restorative treatment was approximately 50% higher in the composite group.

Conclusions In this study, children who received dental restorative treatment with amalgam did not, on average, have statistically significant differences in neurobehavioral assessments or in nerve conduction velocity when compared with children who received resin composite materials without amalgam. These findings, combined with the trend of higher treatment need later among those receiving composite, suggest that amalgam should remain a viable dental restorative option for children.

Trial Registration clinicaltrials.gov Identifier: NCT00066118

JAMA. 2006;295:1784-1792

www.jama.com

See also pp 1775 and 1835.

port a causal relationship between dental amalgam restorations and human health problems."⁴

Author Affiliations are listed at the end of this article.
Corresponding Author: Timothy A. DeRouen, PhD, School of Dentistry, University of Washington, Box 357480, Seattle, WA 98195 (derouen@u.washington.edu).

The use of dental amalgam for posterior restorations remains part of standard care in the United States and in most other countries. Although alternatives to amalgam have been developed (primarily resin composite material), available evidence suggests that they do not match the strength and durability of amalgam and are associated with more recurrent caries and higher failure rates.⁵⁻⁷ In addition, the composite restorations cost more, are more technique sensitive, and have not been assessed as far as related chemical exposures and their potential health effects. Given the cost-benefit dilemma associated with choosing between materials, it is important to determine any health risks associated with amalgam.

We report herein the results of a clinical trial comparing the health effects among children who had dental restoration performed using dental amalgam or resin composite materials.

METHODS

Study Design and Population

A detailed description of the study design and methods has been previously published.⁸ The objective of this clinical trial was to assess the safety of the use of mercury-containing amalgam in dental restorations in children. The hypothesis was that children exposed to low levels of mercury from amalgam may demonstrate less favorable health and development outcomes over time than children who received similar dental treatment without exposure to amalgam.

The targeted study population was students of the Casa Pia school system in Lisbon, Portugal, who were aged 8 to 10 years as of January 1, 1997. The Casa Pia system enrolls more than 4000 students from 7 campuses throughout Lisbon. This school system was selected because the University of Lisbon had prior collaborations with it and the students were known to have diverse backgrounds, high oral disease rates, limited prior dental treatment, and low rates of migration out of Lisbon. Retention over several years of fol-

low-up was thought to be of paramount importance, and this Portuguese school population offered the greatest promise for long-term follow-up.

Inclusion Criteria

Initially, all children born in 1986, 1987, or 1988 (8 to 10 years old as of January 1, 1997) enrolled in the Casa Pia school system were invited to participate. Over time, those who became 8 years old (born in 1989) were also included. The study protocol, approved by the institutional review boards at the University of Washington and the University of Lisbon, called for written informed consent to be obtained from parents or guardians, along with signed assent of the children.

In addition to age, the inclusion criteria were (1) at least one carious lesion in a permanent tooth, (2) no previous exposure to amalgam, (3) urinary mercury level lower than 10 µg/L, (4) blood lead level lower than 15 µg/dL, (5) Comprehensive Test of Nonverbal Intelligence (CTONI) IQ of at least 67, and (6) no interfering health conditions. Data and specimens on inclusion measures were collected in Lisbon, shipped to Seattle for any laboratory analyses, and collated by the coordinating center in Seattle to determine eligibility, who then determined the randomized treatment assignment and transmitted it to Lisbon. Information on race was recorded by study staff based on the participant's appearance and was used only to evaluate demographic balance between randomized groups.

Intervention

The intervention was treatment for dental caries using amalgam for posterior restorations. The control condition was treatment for dental caries using resin composite material rather than amalgam. All dental treatment met existing standards of care in the United States and Portugal. Participants were randomized using stratification by the 7 schools in the system. In both groups, smaller and anterior restorations could be treated with other materials, selected from a list typical of use in the

United States and Portugal, but standardized to limit excess variability.⁸

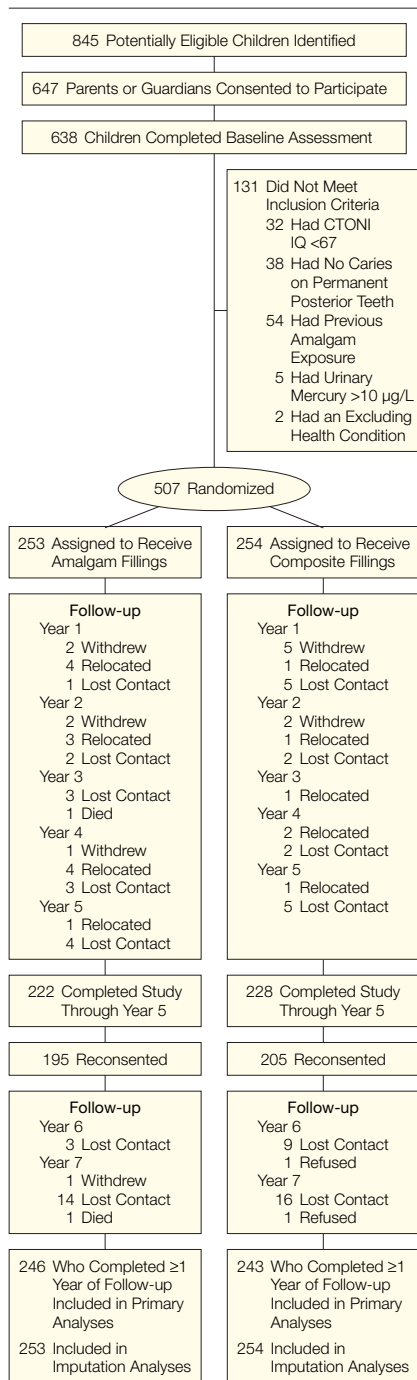
Primary Outcomes

Based on the toxicology of elemental mercury and information from studies of high-level exposure,⁹ the target organs for elemental mercury exposure from amalgam were identified to be the neurological and renal systems. The interdisciplinary investigation team prioritized 3 neurobehavioral domains most likely to be affected: memory, attention/concentration, and motor/visuomotor. Neurobehavioral tests in those domains were identified and described.¹⁰ The memory domain included Rey Auditory Verbal Learning and Visual Learning tests; the attention/concentration domain included Coding, Symbol Search, Digit Span, Finger Windows, Stroop, and Trails A and B; and the motor/visuomotor domain included Finger Tapping, Drawing, Matching, Peg-board, and Standard Reaction Time. Drawing was administered only through follow-up year 3; at the fourth and subsequent follow-up years adult versions of the following neurobehavioral tests were substituted for the child equivalents: Wechsler Memory Scale-R Visual Reproductions; Spatial Span from the Wechsler Memory Scale-III, Matrix Reasoning from the Wechsler Abbreviated Scale of Intelligence (WASI), and Symbol Search, Coding, and Digit Span subtests from the Wechsler Adult Intelligence Scale III (WAIS). Data analyses used US-derived norms. The primary outcome for each neurobehavioral domain was the combined *z* score for the tests in that domain. The fourth primary outcome identified was nerve conduction velocity, measured as the average of *z* scores for posterior tibial and ulnar nerve conduction velocities.

All primary outcomes were scheduled for annual assessments (initially planned for 5 years and extended mid-study to 7 years). Neurobehavioral tests were administered by a team of 3 psychometrists, each of whom was continually monitored and whose work was

calibrated over the 8.5-year testing period by review of videotaped testing sessions using ratings on a 136-item check-

Figure 1. Participant Flow Through the Study



Excluding health conditions include 1 case of diabetes and 1 case of neoplastic disease.

list (with 94.5% to 97.8% accuracy). Tests were double scored and data corrected when errors were identified (no severe violations of protocol were observed that required discarding data). Most nerve conduction velocity tests were performed by one technician, with trained substitutes used as necessary. Psychometrists and nerve conduction technicians had no reason to examine the children intraorally and were instructed not to in order to maintain blinding (although adherence could not be guaranteed). Participants could not be blinded due to the different appearance of the 2 kinds of materials.

Secondary Outcomes

For baseline screening, we used the CTONI because it is a nonverbal test developed to minimize the effects of language and culture on measures of intelligence. US norms for the CTONI are a mean (SD) of 100 (15), but international clinical experience with CTONI suggests that it underestimates IQ in other cultures by approximately 1 SD.¹¹ At the suggestion of the data and safety monitoring board to allow comparisons with a concurrent US trial for which IQ was the primary outcome,¹² we repeated the CTONI at year 7 and also included the WASI (performance subtests only). In the absence of Portuguese norms, we used US norms, recognizing that while there may be some cultural and language biases, they should be equally distributed between randomized groups. Single-void ("spot") urine samples were obtained at baseline (prior to any treatment) and at subsequent annual visits prior to any needed additional treatment. Urinary glutathione transferases (GST- α and GST- π) and porphyrins were monitored as indicators of renal responses to mercury (not necessarily permanent kidney damage)^{13,14} and will be reported separately. Renal glomerular function was monitored using creatinine-adjusted urinary albumin concentrations.

Measures of Mercury Exposure

Urinary mercury analyses were performed according to methods

described by Corns et al¹⁵ using continuous cold-flow, cold-vapor atomic spectrofluorometry, using a PSA Merlin Mercury Analysis (Questron Corp, Mercerville, NJ). Urinary creatinine content was determined using a colorimetric determination assay kit (Sigma Chemical Co, St Louis, Mo). Creatinine-adjusted urinary mercury values were obtained by dividing the mercury concentration by the creatinine concentration. A cumulative measure of amalgam, in units of surface-years, was obtained from the number of amalgam restoration surfaces placed, weighted by the amount of time each restoration was in place.

Sentinel Adverse Health Events

As part of the safety monitoring plan, an attempt was made to identify children who experienced any "sentinel health events" during the study, defined as major disease diagnoses, hospitalizations, or death. The system depended on responses of parents or guardians to annual health history questionnaires, as well as reports from teachers. The study did not have the means or authority to obtain medical records to verify the reports, but no evidence surfaced that suggested the reports were inaccurate.

Statistical Analysis

Multivariate statistical analyses were performed using 2 different tests: the O'Brien test,¹⁶ extended to longitudinal data with interim annual testing¹⁷ to guard against subtle effects in all outcomes, no one of which might be significant by itself; and the Hotelling T^2 test, sensitive to detecting an effect in only one outcome. A 2-tailed approach was used, but with greater sensitivity toward detecting harmful effects of amalgam than composite. Because this was a longitudinal safety study, the test procedure was designed to consist of 7 annual tests. The overall significance level of .05 was divided between the Hotelling and O'Brien tests and allocated over the 7 interim analyses as specified in Table

4 of our design paper⁸ to adjust for the multiple comparisons. For illustration purposes, univariate methods were used to compute mean *z* scores for treatment groups for each primary outcome annually, with 95% confidence intervals surrounding each of those annual observed mean *z* scores. Annual comparisons between creatinine-adjusted albumin levels were made using the Wilcoxon rank sum test, and a *z* test based on a robust standard error¹⁸ was used to compare treatment groups on follow-up values of creatinine-adjusted urinary mercury.

The intent-to-treat principle was used for the analysis (all participants were retained in their assigned groups even if the treatment protocol was not followed), and all data available (whether complete or incomplete) on all randomized patients were included. Those who did not complete the 7 years of follow-up were considered censored at their last available follow-up. The potential effect of missing data was evaluated at the completion of the study by additional analyses conducted after multiple imputation¹⁹ and last observation carried forward methods were used to estimate missing data points. Data were analyzed using SAS version 9.1 (SAS Institute Inc, Cary, NC).

The sample size for the study was selected to ensure adequate power for detecting 2 potential scenarios. One was a small but near-uniform effect of 0.3 SD for the 3 neurobehavioral outcomes, and half of that (0.15 SD) for the nerve conduction outcome. The effect size of 0.3 SD represents a shift that would cause the proportion of abnormally low values in a normally distributed population to increase from 2.5% to 5.0%, thus doubling the proportion classified as abnormally low. For the other scenario, a potential effect in only 1 of the 4 outcomes was of interest, so an effect size of 0.5 SD in the nerve conduction outcome was used, with no effects in the others.

A sample size of 400 (200 in each group) through 5 years of follow-up provided adequate power (>97%) to

Table 1. Baseline Characteristics of Study Participants by Treatment Group

Variable	Amalgam Group (n = 253)	Composite Group (n = 254)
Sex, No. (%)		
Female	116 (46)	112 (44)
Male	137 (54)	142 (56)
Race, No. (%)		
White	178 (70)	181 (71)
Black	75 (30)	68 (27)
Asian	0	5 (2)
Age, mean (SD) [range], y	10.2 (1.0) [8.1-12.4]	10.1 (0.9) [8.3-12.0]
IQ on CTONI, mean (SD) [range]	85 (10) [67-118]	85 (10) [67-116]
Creatinine-adjusted urinary mercury concentration, mean (SD) [range], µg/g	1.8 (2.0) [0.1-23.5]	1.9 (1.8) [0.1-13.7]
Blood lead concentration, mean (SD) [range], µg/dL	4.7 (2.5) [1-16]	4.5 (2.2) [1-12]
Carious surfaces, mean (SD) [range], No.	15.6 (9.0) [0-52]	15.9 (10.2) [1-53]
Creatinine-adjusted albumin concentration, median (IQR), mg/g	8.6 (4.8-14.7)	8.3 (5.2-16.7)

Abbreviations: CTONI, Comprehensive Test of Nonverbal Intelligence; IQR, interquartile range. SI conversion factor: To convert lead to µmol/L, multiply values by 0.048.

detect both scenarios. To allow for those dropping out or otherwise lost to follow-up, enrollment of 500 was targeted, and 507 were actually enrolled. Midway through the trial, to enhance the power to detect an even smaller potential effect in only 1 outcome, follow-up was extended to 7 years.

RESULTS

Patient Characteristics and Treatment Groups

Of 845 children who were initially identified and whose parents/guardians were approached, consent was obtained for 647. Nine children did not return for some or all of the baseline screening measures, and 131 did not meet the inclusion criteria for the reasons given (FIGURE 1). A total of 507 children met inclusion criteria and were randomized: 253 to the amalgam treatment group and 254 to the composite treatment group. All 507 randomized participants were included in analyses, using available partial data for those without complete data, although the 18 children with no follow-up visits were included only in baseline comparisons.

The treatment groups were balanced on all baseline covariates (TABLE 1), which included sex, race,

age, IQ, creatinine-adjusted urinary mercury concentration, blood lead concentration, number of carious surfaces, and creatinine-adjusted albumin concentration.

Dental Treatment and Mercury Exposure

The amount of treatment required was high initially because most children had a history of untreated caries. Both groups received the same amount of restorative treatment in the initial treatment year and over the next 4 years (TABLE 2). However, in treatment years 6, 7, and 8 approximately 50% more restorative treatment was needed in the composite group, consistent with previous findings.⁵⁻⁷ Children in the amalgam group had an average cumulative exposure of 50 surface-years, mostly due to the initial treatment. Two children in the composite group received amalgam fillings by mistake, but it had little effect on the statistical analysis (excluding them, or including them in the amalgam group, changed nonsignificant *P* values at year 7 by .02 less).

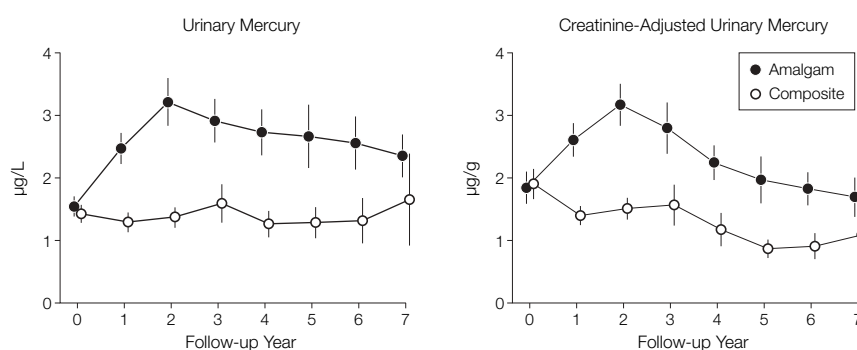
Urinary mercury concentrations increased following dental treatment in the amalgam group (FIGURE 2). The mean creatinine-adjusted mercury concentration was 1.8 µg/g at baseline, in-

Table 2. Dental Restorative Treatment by Treatment Group and Year

Year	Amalgam (n = 253)			Composite (n = 254)		
	Children With Surfaces Restored, No. (%)	Mean (SD)	Median (Range)	Children With Surfaces Restored, No. (%)	Mean (SD)	Median (Range)
1	249 (98)	10.1 (5.6)	10 (0-27)	248 (98)	9.9 (6.3)	9 (0-32)
2	53 (23)	0.9 (2.9)	0 (0-24)	58 (26)	0.9 (2.2)	0 (0-15)
3	61 (26)	1.0 (2.3)	0 (0-15)	67 (29)	1.2 (2.4)	0 (0-12)
4	63 (28)	1.2 (3.0)	0 (0-20)	77 (34)	1.5 (3.5)	0 (0-28)
5	81 (38)	1.7 (3.6)	0 (0-28)	91 (42)	1.5 (2.9)	0 (0-19)
6	80 (39)	1.8 (3.0)	0 (0-21)	106 (50)	3.0 (5.1)	1 (0-30)
7	83 (45)	2.1 (4.0)	0 (0-34)	94 (51)	3.0 (5.4)	1 (0-38)
8*	76 (45)	2.1 (3.3)	0 (0-20)	98 (58)	3.2 (3.9)	2 (0-21)
Cumulative surfaces restored, No.		18.7 (13.0)	16 (0-80)		21.3 (15.9)	18 (0-115)
Cumulative surfaces restored with amalgam, No.		16.1 (10.8)	14 (0-80)		0.1 (1.3)†	0 (0-20)
Cumulative amalgam exposure, surface-years		50.1 (37.2)	44.1 (0-248)		0.2 (2.6)	0 (0-41)
Restored surfaces at year 7, No.		13.8 (12.1)	11 (0-75)		15.7 (11.6)	13 (0-55)

*Year 8 refers to treatment provided after the final (follow-up year 7) testing. Eighth-year data complete as of December 5, 2005.

†Two children received amalgam fillings by mistake (20 surfaces and 1 surface, respectively).

Figure 2. Mean Urinary and Creatinine-Adjusted Urinary Mercury Concentrations by Treatment Group and Follow-up Year

Error bars indicate 95% confidence intervals.

creased to 3.2 µg/g by 2 years after baseline, subsequently leveled off, and then declined steadily from year 3 to year 7. Mercury levels were significantly higher in the amalgam group ($P < .001$) during follow-up, by approximately 1.5 µg/g in the first 3 years of follow-up, declining to approximately 1.0 µg/g later.

Retention

The percentage of children who remained in the study was 85% or greater through 5 years of follow-up, then declined to just under 70% through follow-up year 7 (TABLE 3).

One reason for loss of participants in the final 2 years was the need to reobtain informed consent for participation beyond 5 years. Of all data that could have been collected if all 507 children had remained in the study for 7 years of follow-up, 5% were missing because of missed visits or tests for children remaining in the study, and 13% were missing because of children lost to follow-up. The most complete data on primary outcome variables was for the neurobehavioral tests. Data completeness for nerve conduction tests was less than for neurobehavioral tests in the last 2 years due to logistical issues

in the scheduling and timing of the tests. Follow-up and data completeness percentages were similar in the 2 treatment groups (Table 3). Urinary mercury concentrations were obtained for at least 80% of children through year 5, and for years 6 and 7 lower rates of 73% and 65% merely reflect the declining retention rate. Measurement of urinary albumin was added to the protocol during the first year of the study and was therefore available for only 56% of the amalgam group and 57% of the composite group at baseline, but subsequent follow-up percentages were similar to those for urinary mercury. CTONI IQ was obtained for all at baseline screening, and at year 7 CTONI and WASI measures of IQ were available for 66% of the amalgam group and 63% of the composite group.

Primary Analysis of Group Differences

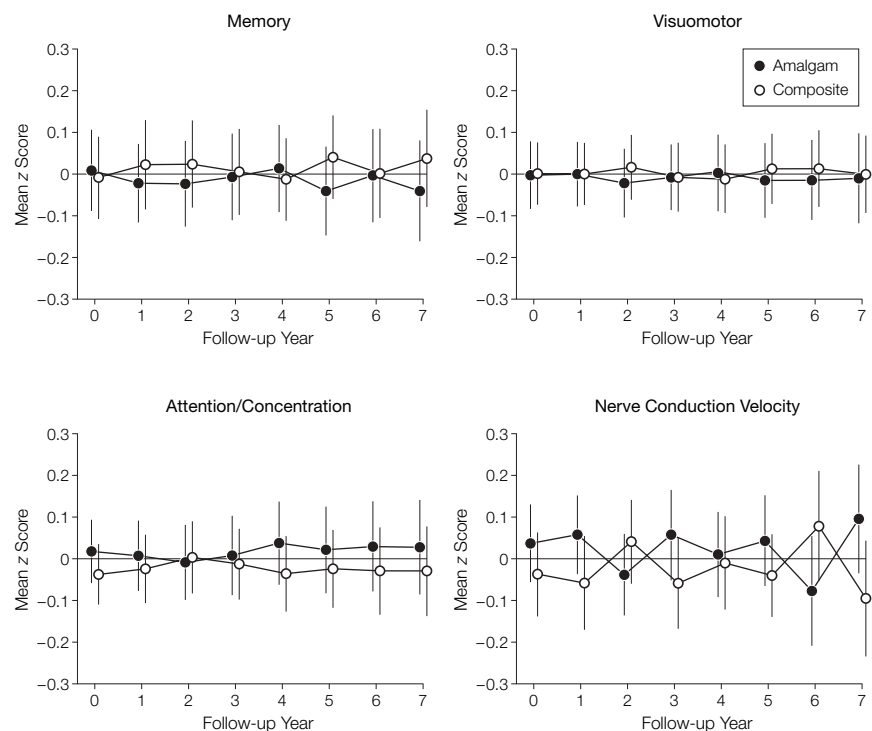
Annual interim analyses on the primary outcomes were performed and reported to the data and safety monitoring board using the O'Brien and Hotelling multivariate statistical tests previously described, with the .05 significance level spread over the 7 annual tests as specified in Table 4 in

Table 3. Children Within Each Treatment Group With Follow-up Data on Primary Outcome Variables by Follow-up Year

	No. (%) by Year						
	1	2	3	4	5	6	7
Neurobehavioral							
Amalgam group	241 (95)	237 (94)	228 (90)	218 (86)	212 (84)	187 (74)	172 (68)
Composite group	238 (94)	230 (91)	227 (89)	223 (88)	214 (84)	189 (74)	176 (69)
Nerve conduction velocity							
Amalgam group	230 (91)	229 (91)	205 (81)	205 (81)	201 (79)	141 (56)	140 (55)
Composite group	227 (89)	217 (85)	202 (80)	204 (80)	202 (80)	139 (55)	140 (55)
≥1 Primary outcome							
Amalgam group	242 (96)	237 (94)	233 (92)	222 (88)	214 (85)	190 (75)	173 (68)
Composite group	239 (94)	232 (91)	232 (91)	226 (89)	218 (86)	193 (76)	176 (69)

our design paper.⁸ For the final analysis in year 7, the prescribed significance levels were .011 for the Hotelling test, .024 for the O'Brien test to detect worse outcomes in the amalgam group, and .011 for the O'Brien test to detect better outcomes in the amalgam group. The final test statistic values were $F=0.60$ ($P=.66$) for the Hotelling test and $t=0.21$ (1-sided $P=.42$) for the O'Brien test. No evidence of group differences for primary outcome variables was found in any of the 7 annual interim analyses, with 2-sided P values for the Hotelling test ranging from .42 to .91, and 1-sided P values for the O'Brien test ranging from .29 to .48.

Univariate mean z scores and 95% confidence intervals for each primary outcome variable and each study year (unadjusted for multiple comparisons) are shown in FIGURE 3. Differences in mean z scores were small and not statistically significant at any year for the 3 neurobehavioral outcomes. The nerve conduction velocity outcome exhibited high variability, with inconsistent treatment effects over time and a treatment difference at year 7 that reached statistical significance at the (univariate) nominal .05 level of significance. Because of the inconsistency of the estimates, as illustrated by the treatment difference for year 6 that was in the opposite direction to year 7, and because of the large number of tests, this finding is not interpreted as evidence for a treatment effect (and in fact this finding is in the direction of more favorable results for the

Figure 3. Average Standardized z Scores by Treatment Group and Follow-up Year for Each Primary Outcome Variable

Error bars indicate 95% confidence intervals.

amalgam group, opposite to that hypothesized).

To illustrate the underlying data for the neurobehavioral test scores and nerve conduction velocities used in calculation of the primary outcome z scores, descriptive statistics are given for all measures at baseline and year 7 only (TABLE 4). The 2 groups were very similar at baseline and remained very similar at year 7, supporting the finding of no group differences from the primary analysis.

Children who did not complete 7 years of follow-up did not differ on baseline characteristics between the groups. To further assess any potential bias due to missing data, the primary analyses for year 7 were repeated with missing data replaced by estimated data from multiple imputation based on baseline characteristics and all other outcomes, as well as from the last value carried forward method. The multiple imputation method resulted in O'Brien test $t=-0.16$ ($P=.44$)

and Hotelling test $F=0.62$ ($P=.65$); the last observation carried forward method resulted in O'Brien test $t=-0.20$ ($P=.42$) and Hotelling test $F=0.58$ ($P=.68$), the lack of significance suggesting that missing data did not have much of an effect. Although not prespecified in the analysis protocol, a worst case subset comparison of all primary outcomes for the 20% with highest amalgam exposure at initial treatment (>13 surfaces) vs the composite group was not statistically significant either (O'Brien test $t=1.20$, $P=.12$; Hotelling test $F=0.91$, $P=.46$).

Secondary Analyses

Average WASI IQ scores at year 7 were similar in the 2 groups and not significantly different, and the follow-up CTONI IQ scores were similar in the 2 groups (Table 4). There were no significant group differences in median values of creatinine-adjusted albumin over the 7 years of follow-up (TABLE 5), with each of the annual comparisons at $P>.14$ and the observed median values (6.5-9.9 mg/g) in the normal range. In monitoring urinary GST concentrations, we did not find any extremely high observed values that might indi-

cate kidney damage, and concentrations of urinary porphyrins did not suggest substantial mercury accumulation in the kidneys.

Adverse Sentinel Health Events

There were 4 adverse sentinel health events reported in the amalgam group (1 death due to unintentional gunshot, 1 death due to hepatitis, a brain aneurysm, and a case of kidney stones) and 5 events reported in the composite group (2 diagnosed cases of epilepsy, 1 case of hyperthyroidism, 1 case of asthma, and 1 psychiatric hospital-

Table 4. Neurobehavioral Test Scores, Nerve Conduction Velocities, and Intelligence Measures at Baseline and Year 7

Baseline, Mean (SD)			Year 7, Mean (SD)		
Test	Amalgam Group	Composite Group	Test*	Amalgam Group	Composite Group
Memory					
RAVLT Memory†	8.36 (2.91)	8.1 (3.07)		9.65 (2.86)	9.73 (2.79)
RAVLT Total Learning†	39.09 (9.98)	37.95 (9.61)		46.06 (9.09)	47.36 (9.48)
WRAML Visual Memory†	6.52 (3.12)	6.56 (3.04)	WMS-III Reproductions (delayed)†	33.02 (6.24)	32.98 (6.01)
WRAML Visual Learning‡	7.83 (2.64)	8.14 (2.75)	WMS-III Reproductions (immediate)†	35.15 (4.47)	35.79 (3.68)
Attention/Concentration					
Coding‡	9.04 (3.14)	8.64 (3.08)	WAIS-III Digit Symbol‡	9.45 (2.86)	9.42 (2.98)
Symbol Search‡	9.39 (2.69)	9.41 (2.59)	WAIS-III Symbol Search‡	9.77 (3.08)	9.40 (2.85)
Digit Span‡	7.4 (2.73)	7.37 (2.53)	WAIS-III Digit Span‡	7.70 (2.21)	7.64 (2.17)
Finger Windows‡	7.32 (2.35)	7.28 (2.47)	WAIS-III Spatial Span‡	9.34 (2.99)	9.03 (2.96)
Trails A, seconds§	27.95 (12.74)	27.69 (13.05)	Adult Trails A, seconds§	28.72 (11.26)	28.94 (12.06)
Trails B, seconds§	65.25 (34.41)	65.1 (35.61)	Adult Trails B, seconds§	65.34 (25.07)	63.84 (25.5)
Stroop Word‡	42.18 (6.56)	41.54 (6.39)		41.41 (8.04)	41.7 (8.09)
Stroop Color‡	44.15 (6.01)	43.03 (5.62)		42.67 (8.14)	41.59 (8.16)
Stroop Color-Word‡	44.17 (6.93)	43.3 (6.84)		48.42 (9.41)	46.99 (9.71)
Visuomotor					
WRAVMA Drawing‡	101.06 (12.27)	101.71 (10.79)	Test dropped		
WRAVMA Matching‡	95.57 (13.72)	96.19 (12.4)	WASI Matrices†	24.83 (5.02)	24.44 (5.33)
WRAVMA Pegs (dominant)‡	101.94 (16.87)	103.04 (16.68)		119.51 (17.82)	119.76 (18.67)
WRAVMA Pegs (nondominant)‡	106.18 (14.64)	106.81 (15.03)		119.01 (15.55)	119.38 (15.83)
Standard Reaction Time, mean§	0.9 (0.2)	0.9 (0.2)		0.77 (0.15)	0.76 (0.14)
Finger Tapping (dominant)†	36.66 (6.17)	36.29 (6.05)		50.51 (6.56)	50.5 (6.35)
Finger Tapping (nondominant)†	32.02 (5.34)	31.33 (5.37)		44.48 (6.34)	44.49 (6.33)
Nerve Conduction Velocity					
Tibial, m/s§	51.12 (5.29)	51 (5.58)		50.78 (5.07)	50.15 (5.09)
Ulnar, m/s§	59.57 (6.39)	58.75 (6.51)		59.26 (6.41)	57.58 (6.52)
Intelligence					
CTONI	85 (10)	85 (10)		81 (12)	81 (12)
WASI	NA	NA		94 (14)	92 (13)

Abbreviations: CTONI, Comprehensive Test of Nonverbal Intelligence; NA, not available; RAVLT, Rey Auditory Verbal/Visual Learning Test; WAIS-III, Wechsler Adult Intelligence Scale III; WASI, Wechsler Abbreviated Scale of Intelligence; WMS, Wechsler Memory Scale; WRAML, Wide Range Assessment of Memory and Learning; WRAVMA, Wide Range Assessment of Visual Motor Ability.

*Some of the tests were replaced in year 4 to account for the aging of the children. Blank cells indicate that the tests remained the same.

†Raw test score.

‡Scaled test score.

§Lower values represent better performance. For all other tests, higher values represent better performance.

||The drawing test was dropped after year 3 and was not replaced.

ization). The variety of events observed in the 2 groups does not suggest a discernable pattern.

COMMENT

The 7 years of longitudinal data on these children provide extensive evidence concerning the relative safety of amalgam in dental treatment. Substantial amalgam exposure did lead to creatinine-adjusted urinary mercury levels that were higher in the amalgam group. However, the amount of the increase over the composite group leveled off to approximately 1.0 µg/g over time, and all the average levels remained within the range of 0 to 4 µg/L usually cited as background levels.^{20,21} Despite group differences in mercury levels, we found no statistically significant differences in measures of memory, attention, visuo-motor function, or nerve conduction velocities. This remained the case after adjusting for baseline covariates and after imputing values for missing data. A total of 9 sentinel adverse health events were observed, but with no discernible pattern of differences between the groups.

Because study participants were Portuguese children, the question of study generalizability may be a concern. The use of a randomized clinical trial study design with treatment groups identical at baseline should mitigate some of those concerns, since the question addressed by the study is whether the groups differed as a result of treatment, not whether performance on any specific test was representative of children in the United States or other countries. If the results here are not generalizable, it would mean that amalgam may have different effects on the development of children in different cultures; ie, that neurotoxicity of mercury depends on the cultural context, which seems unlikely.

It is important to note what kinds of effects this study was, and was not, designed to detect. The hypothesis was that children exposed to constant low levels of mercury from dental amalgam would over time perform worse,

on average, on neurobehavioral and nerve conduction measures than children not exposed to dental amalgam. The evidence from this study does not support that hypothesis. This study does suggest that children treated with dental amalgam will experience slightly higher urinary levels of mercury, but those levels are likely to remain in the general range of background levels. Furthermore, it suggests that, on average, children exposed to dental amalgam will not show any effects on their neurobehavioral or neurological development through adolescence, at least for those measures addressed in this study.

This study was not designed to detect whether a very small fraction of children may have genetic predispositions to sequester elemental mercury at an extraordinarily high rate, or have rare allergic or other kinds of adverse reactions to elemental mercury. While we monitored for unusual individual responses and did not observe any, we are not able to definitively rule out the possibility of such occurrences if the rate of occurrence is 1:100 or smaller. However, given these findings on average response, it does suggest that any future research should focus on the possibility of rare outcomes.

This study also was not designed to evaluate the safety of alternative dental materials, specifically the resin composite material used. While we did perform a 2-tailed comparison of the treatments, the outcomes for this study were specifically selected to be sensitive to effects of elemental mercury. The absence of a worse effect for composite in outcomes sensitive to elemental mercury does not reveal much about the safety of composite. After this study was under way, initial reports surfaced that chemicals in the composites may serve as endocrine disruptors,²² and there is evidence that there are at least short-term exposures to some of these chemicals from the placement of composite restorations.²³ However, these findings are preliminary and were not available in time to allow inclusion of outcomes sensitive to potential health effects of composites. This point is im-

Table 5. Creatinine-Adjusted Urinary Albumin Levels by Treatment Group and Follow-up Year*

Year	Albumin Level, Median (IQR) mg/g of Creatinine	
	Amalgam	Composite
1	7.7 (3.1-11.5)	7.4 (4.2-12.5)
2	8.6 (5.5-13.4)	9.4 (5.3-16.1)
3	9.0 (5.5-17.9)	9.9 (6.8-16.7)
4	8.7 (5.6-14.5)	9.2 (5.8-20.8)
5	8.0 (5.4-12.5)	8.2 (5.1-14.3)
6	7.3 (4.8-14)	7.5 (4.8-14.3)
7	6.5 (4.3-12.3)	6.8 (4.4-13.7)

Abbreviation: IQR, interquartile range.

*Differences between groups were not statistically significant at any time point.

portant in discussions of the risks and benefits of the use of amalgam compared with alternative materials in dental restorations.

One potential conclusion from this study might be that there is no need to advocate removal of existing amalgams in children since there is no evidence of risk, but as a precaution future use of amalgam should be avoided since it does involve some (albeit low) level of mercury exposure. While this trial provides detailed information on exposures and potential risks associated with dental amalgam restorations, there is no comparable information available on the exposures and risks associated with resin composite restorations, the most commonly used alternative to amalgam.

CONCLUSIONS

In summary, this trial showed that children treated with dental amalgam did not, over a 7-year follow-up period, demonstrate statistically significant differences in neurobehavioral and neurological test results compared with similar children treated with other dental materials. These findings, especially in light of the observed higher treatment need in the composite group 5 or more years after initial treatment, suggest that amalgam should remain a viable clinical option in dental restorative treatment.

Author Affiliations: Departments of Dental Public Health Sciences (Drs DeRouen and Leroux), Biostatistics (Drs DeRouen and Leroux), Oral Medicine (Dr Martin), Epidemiology (Dr Martin), Psychiatry and Behavioral Sciences (Dr Townes and Ms Rosenbaum), and Environmental and Occupational Health Sci-

ences (Dr Woods), University of Washington, Seattle; Battelle Centers for Public Health Research and Evaluation, Seattle, Wash (Dr Woods); Faculty of Dental Medicine (Drs Bernardo and Leitão and Mr Luis) and Faculty of Medicine (Drs Castro-Caldas and Martins), Universidade de Lisboa, Lisbon, Portugal; and Institute of Health Sciences, Universidade Catolica Portuguesa, Lisbon, Portugal (Dr Castro-Caldas).

Author Contributions: Dr DeRouen had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: DeRouen, Martin, Leroux, Townes, Woods, Leitão, Bernardo.

Acquisition of data: DeRouen, Martin, Townes, Woods, Leitão, Castro-Caldas, Luis, Bernardo, Rosenbaum, Martins.

Analysis and interpretation of data: DeRouen, Martin, Leroux, Woods.

Drafting of the manuscript: DeRouen, Martin, Woods, Martins.

Critical revision of the manuscript for important intellectual content: DeRouen, Martin, Leroux, Townes, Woods, Leitão, Castro-Caldas, Luis, Bernardo, Rosenbaum, Martins.

Statistical analysis: DeRouen, Martin, Leroux, Woods.

Obtained funding: DeRouen, Martin, Woods.

Administrative, technical, or material support: DeRouen, Martin, Townes, Woods, Leitão, Luis, Bernardo.

Study supervision: DeRouen, Martin, Townes, Woods, Castro-Caldas, Rosenbaum, Martins.

Financial Disclosures: Dr Leitão reports being principal investigator for a clinical research project at the Biomaterials Laboratory of the University of Lisbon, of which he is the director, supported by the Ivoclar Vivadent Company to test a novel indirect veneer material. Ivoclar Vivadent manufactures both composite resins and dental amalgams, but none of the products used in that investigation are discussed in this article. Dr Leitão reports not personally receiving any direct compensation from this contract between Ivoclar Vivadent and the University of Lisbon. Dr Bernardo reports receiving financial support from Dentsply DeTry Company, a manufacturer of dental instruments and amalgam, to travel to the University of Washington, Seattle, to participate in professional development in research methods. No other authors reported financial disclosures.

Funding/Support: This work was funded by Cooperative Agreement U01 DE11894 from the National Institute of Dental and Craniofacial Research (NIDCR) of the National Institutes of Health.

Role of the Sponsor: The U01 is an assistance funding mechanism in which federal officials play a facilitative role in the conduct of the trial, but scientific decision making is the responsibility of the principal investigator. The NIDCR had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript.

Acknowledgment: We wish to acknowledge the many contributions by others to various aspects of this study. Key study personnel who received financial support for their work include: Lurdes Vaz, RDH, Helena Amaral, BS, Gorety Ribeiro, MS, Pedro Rodrigues, MS, Susana Rodrigues, BS, Helena Nazareth, BS, Isabel Morgado, BS, Patricia Santos (dental assistant), Teresa Guerreiro (dental assistant), Victoria Lopes, BS, Mamede Carvalho, MD, PhD, Jaime Portugal, DDS, Margarida Patrocinio, DDS, from the University of Lisbon; and Lynne Simmonds, MS, John Kushleika, MS, Tessa Rue, MS, Ying Huang, MS, Tonya Benton, MS, from the University of Washington. We also acknowledge the oversight provided by the independent data and safety monitoring board, and the extraordinary cooperation of the students, teachers, and administrators of the Casa Pia de Lisboa school system.

REFERENCES

1. Dye BA, Schober SE, Dillon CF, et al. Urinary mercury concentrations associated with dental restorations in adult women aged 16-49 years: United States, 1999-2000. *Occup Environ Med*. 2005;62:368-375.
2. Ritchie KA, Burke FJ, Gilmour WH, et al. Mercury vapour levels in dental practices and body mercury levels of dentists and controls. *Br Dent J*. 2004;197:625-632.
3. Mackert JR Jr, Berglund A. Mercury exposure from dental amalgam fillings: absorbed dose and the potential for adverse health effects. *Crit Rev Oral Biol Med*. 1997;8:410-436.
4. Brownawell AM, Berent S, Brent RL, et al. The potential adverse health effects of dental amalgam. *Toxicol Rev*. 2005;24:1-10.
5. Van Nieuwenhuysen JP, D'Hoore W, Carvalho J, Qvist V. Long-term evaluation of extensive restorations in permanent teeth. *J Dent*. 2003;31:395-405.
6. Sjogren P, Halling A. Survival time of class II molar restorations in relation to patient and dental health insurance costs for treatment. *Swed Dent J*. 2002;26:59-66.
7. Mjor IA, Dahl JE, Moorhead JE. Placement and replacement of restorations in primary teeth. *Acta Odontol Scand*. 2002;60:25-28.
8. DeRouen TA, Leroux BG, Martin MD, et al. Issues in design and analysis of a randomized clinical trial to assess the safety of dental amalgam restorations in children. *Control Clin Trials*. 2002;23:301-320.
9. Friberg L, Nordberg G. Inorganic mercury: a toxicological and epidemiological appraisal. In: Miller W, Clarkson T, eds. *Mercury, Mercurials and Mercaptans*. Springfield, Ill: Charles C Thomas; 1973.
10. Martins IP, Castro-Caldas A, Townes BD, et al. Age and sex difference in neurobehavioral performance: a study of Portuguese elementary school children. *Int J Neurosci*. 2005;115:1687-1709.
11. Townes BD, Rosenbaum JG, Martins IP, Castro-Caldas A. Neurobehavioral assessment of children: a cross-cultural perspective. *Psychologica*. 2003;34:177-185.
12. Children's Amalgam Trial Study Group. The Children's Amalgam Trial: design and methods. *Control Clin Trials*. 2003;24:795-814.
13. Harrison DJ, Kharbanda R, Scott Cunningham D, McLellan LI, Hayes JD. Distribution of glutathione S-transferase isoenzymes in human kidney: basis for possible biomarkers of renal injury. *J Clin Pathol*. 1989;42:624-628.
14. Bowers MA, Aicher LD, Woods JS. Quantitative determination of porphyrins in rat and human urine and evaluation of urinary porphyrin profiles during mercury and lead exposures. *J Lab Clin Med*. 1992;120:272-280.
15. Corns WT, Stockwell PB, Jameel M. Rapid method for the determination of total mercury in urine samples using cold vapour atomic fluorescence spectrometry. *Analyst*. 1994;119:2481-2484.
16. O'Brien PC. Procedures for comparing samples with multiple endpoints. *Biometrics*. 1984;40:1079-1087.
17. Leroux BG, Mand LA, DeRouen TA. Group sequential testing in dental clinical trials with longitudinal data on multiple outcome variables. *Stat Methods Med Res*. 2005;14:591-602.
18. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73:13-22.
19. Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res*. 1999;8:3-15.
20. *Criteria EH, 118: Inorganic Mercury*. Geneva, Switzerland: World Health Organization; 1991.
21. *Toxicological Profile for Mercury*. Atlanta, Ga: Agency for Toxic Substances and Disease Registry; 1999.
22. Wada H, Tarumi H, Imazato S, Narimatsu M, Ebisu S. In vitro estrogenicity of resin composites. *J Dent Res*. 2004;83:222-226.
23. Martin MD, Bajet D, Woods JS, Dills RL, Poulsen EJ. Detection of dental composite and sealant resin components in urine. *Oral Surg Oral Med Oral Pathol Oral Radiol Endodontol*. 2005;99:429.