Original Article

Statistically Non-significant Papers in Environmental Health Studies included more Outcome Variables[*]

Pentti Nieminen[1,#], Khaled Abass[2,3,^], Kirsi Vähäkanga[4], and Arja Rautio[2]

1. Medical Informatics and Statistics Group, University of Oulu, PO Box 5000, 90014 Oulu, Finland; 2. Centre for Arctic Medicine, Thule Institute, University of Oulu, Finland; 3. Research Unit of Biomedicine, University of Oulu, Finland; 4. School of Pharmacy/Toxicology, Faculty of Health Sciences, University of Eastern Finland, Kuopio, Finland

Abstract

Objective   The number of analyzed outcome variables is important in the statistical analysis and interpretation of research findings. This study investigated published papers in the field of environmental health studies. We aimed to examine whether differences in the number of reported outcome variables exist between papers with non-significant findings compared to those with significant findings. Articles on the maternal exposure to mercury and child development were used as examples.

Methods   Articles published between 1995 and 2013 focusing on the relationships between maternal exposure to mercury and child development were collected from Medline and Scopus.

Results   Of 87 extracted papers, 73 used statistical significance testing and 38 (43.7%) of these reported 'non-significant' ($P>0.05$) findings. The median number of child development outcome variables in papers reporting 'significant' ($n=35$) and 'non-significant' ($n=38$) results was 4 versus 7, respectively (Mann-Whitney test P-value=0.014). An elevated number of outcome variables was especially found in papers reporting non-significant associations between maternal mercury and outcomes when mercury was the only analyzed exposure variable.

Conclusion   Authors often report analyzed health outcome variables based on their P-values rather than on stated primary research questions. Such a practice probably skews the research evidence.

Key words: Mercury exposure; Child development; Reporting bias; Literature review

INTRODUCTION

The existence of outcome reporting bias in medical research articles has been widely suspected for decades[1-6]. Selective outcome reporting occurs when a study measuring multiple outcomes reports those outcomes based on the nature or direction of their results[7]. For example, authors of published articles tend to report only statistically significant findings. Researchers are also less likely to submit manuscripts for publication of studies that were non-significant, negative, or neutral in outcome[3,7]. However, the consequences when the results are non-significant, negative, or neutral in outcome are of great interest. For example, if the findings are reported, how are they reported? If statistically non-significant results find their way into the published literature, do the authors report all outcomes including secondary outcomes and not only selected results?

Epidemiological environmental studies usually include several exposure variables and health-related outcomes[8]. In most cases, multivariable regression methods are applied to explore possible environmental prognostic variables with little or no prior information on which variables are the most important. Fitting multivariable regression models for several outcome variables provides challenges in data analysis and reporting[9]. Schriger et al.[10] recently demonstrated that readers are rarely shown the relationship between the outcome variables. This does not fully utilize the data generated or help other investigators decide whether the hypothesis is worth further study.

Analysis of the relationship between the outcomes of statistical tests and the number of reported outcome variables likely faces several potential confounding factors. The prestige and visibility of journals are potential confounding factors requiring adjustment. For example, studies with positive results may be more likely published in more visible journals than those with statistically non-significant results[7,11-12]. Furthermore, environmental studies emphasize current prevalent research methods differently; for example, some environmental journals may publish more cohort studies than cross-sectional or biological studies. Some authors have found that poor quality methodological reporting is associated with exaggerated positive findings[13-14]. There are also suspicions that a high number of coauthors means publishing ultra-thin salami slices with low quality and trivial findings[15].

We are interested in the reporting of health outcome variables in environmental health studies. We have previously reported about difficulties in meta-analysis in the field[16]. Here we explore the association between the number of reported outcome variables and statistical significance of the relationship between environmental mercury exposure and child development. To our knowledge, this is the first study to analyze this type of selective outcome reporting in medical journal articles.

METHODS

Set of Articles

The set of analyzed articles on mercury exposure and child development was obtained through Medline and Scopus. The literature was searched using the terms 'mercury' and 'child development', either as thesaurus terms, in titles, or as abstract terms with synonyms and closely related words. In summary, the selection criteria used to compose the list of publications subjected to analysis were as follows: publication type, original research article; publication language, English; year of publication, 1995-2013; topic, child development and mercury. The list of hits or search results from the article databases included 112 articles, all of which were then requested and supplied by the Medical Library of the authors' institute.

As the next step, all retrieved articles were screened to exclude those not actually analyzing the relationship between maternal mercury exposure and child development variables. Articles were also excluded if they included just a study protocol or technical report. No papers were deleted from the group based on overlap with other studies or a repetition of a previous study. The total number of included articles was 87.

Variables

A protocol for data collection was developed and employed. We read each paper and then completed the predesigned questionnaire. Where interpretation of the paper was ambiguous, the article was appraised by the other authors and conclusions were reconciled in group discussions.

To evaluate study characteristics of the reviewed articles, the following information was obtained:

Whether the relationship between maternal mercury exposure and child development was the primary outcome of the article (as stated in study objectives, abstract or introduction, labeled 'primary' in methods, or presented first in the results and central to the main conclusions);

Whether or not the authors used formal statistical analysis, and if so, whether they reported a statistically significant result (P-value<0.05);

Whether the primary research question or hypothesis was clearly stated in the report's introduction or methods section;

Whether the study design was a cohort or another observational study (cross-sectional survey or case-control study);

Whether mercury was the only measured explanatory compound or if other compounds were also assessed from the umbilical cord blood; the mother's serum, nail or hair; or estimated using other methods;

Whether sample size was low (<300 mother-child pairs) or high (≥300), the median value

of 300 was used as the cut-off point;

Whether assessment of correlation or association was the main strategy in the study analysis of the primary research question.

The number of outcome variables related to child development was counted from both the tables and text in each article. Health outcomes included the following domains of child development: general development; physical developmental milestones; cognitive, language, and intellectual development; behavioral outcomes; motor function; audition, visual, and visual-motor development; and neurodevelopmental disorders or other health outcomes[17]. Outcome assessment methods related to child development and health outcomes found in the included papers are listed in Appendix A (www.besjournal.com for the details).

To control for journal visibility, the impact factor for the publication year was recorded from the journal that published the article under study. The number of authors was also counted from each article.

Statistical Analysis

Statistical analyses aimed to explore possible differences in the reported number of child development outcome variables between articles reporting significant and non-significant results. Because the distribution of the number of child development outcome variables was skewed to the right, its distribution was summarized using medians and interquartile ranges. Scatterplots and Spearman's correlation coefficient were used to illustrate the correlation of the number of authors to the number of reported outcome variables. Mann-Whitney tests were used to analyze differences in the number of outcome variables between dichotomous categorical characteristics of the articles. Multivariable linear regression was applied to adjusted analysis to evaluate the independent effect of the significance of primary developmental mercury outcomes in mercury-exposed children on the square root of the number of reported outcome variables. A box plot was used to illustrate the distribution of the number of reported child development outcome variables. SPSS Statistics 22.0 was used for the analysis.

RESULTS

The statistical significance of the relationship between maternal mercury and child development in the article set is summarized in Table 1. All eligible articles were classified according to whether or not the authors had reported or decided that any of the relationships were statistically significant or non-significant. If no results of formal statistical significance testing or confidence interval estimation were included in reporting the finding, the article was classified as not evaluating statistical significance. We found that 73 (84.9%) articles reported a statistical test on the hypothesis of a relationship between mercury exposure and child development. Of these, 38 (52.1%) reported a statistically non-significant result. However, this proportion varied by study design (Table 1), with 28/46 (61%) statistically non-significant in cohort studies compared to 10/27 (37%) in others. The proportion of studies reporting non-significant or negative results was higher in the longitudinal cohort studies than in cross-sectional surveys or case-control studies. This was also seen in the analysis strategy: articles assessing strength of correlation or association more often reported non-significant findings (57.1%) than articles comparing groups (37.5%). Our literature review found no uniform evidence of a relationship between maternal mercury and child development.

The articles included in this study had a wide range of outcome variables. Figure 1 shows how the number of outcome variables varies by the statistical significance and number of authors. In three articles, the authors report findings from studies in which they had analyzed about 40 different child development variables[18-20]. There was a moderately positive correlation between the number of outcome variables and number of authors (Spearman's correlation coefficient=0.249, P=0.020). In multi-authored papers, the number of reported outcome variables tended to be higher, probably reflecting all those areas of child development in which the authors were interested.

Figure 2 illustrates how the number of child development outcome variables is distributed by the statistical significance of the article. Statistically non-significant articles reported more outcome variables, while the articles with significant findings concentrated on a limited number of outcomes.

As seen in the last row of Table 2, our analysis suggests that articles with a non-significant result are more likely to report several outcome variables (P=0.014). Without adjustment, this effect is also evident across reporting primary research questions, measured compounds, and the main analysis strategy. The multivariable linear regression analysis confirmed the independent association of a reported
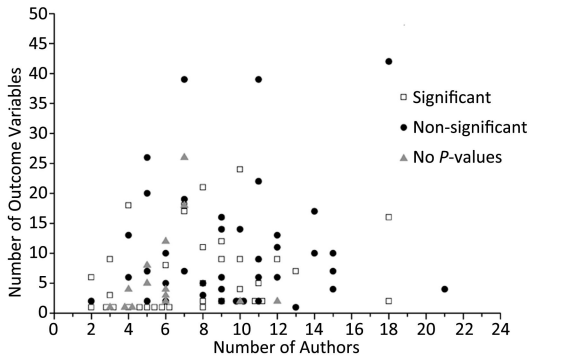
Table 1. Distribution of Statistical Significance of the Primary Outcome by Article Characteristics in 87 Original Articles Focusing on the Relationship Between Maternal Exposure to Mercury and Child Development

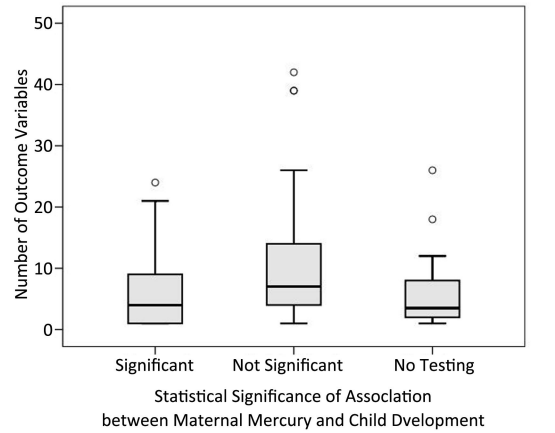| Item	Significant (P<0.05) n (%)	Not Significant (P≥0.05) n (%)	No Testing n (%)	Total n (%)
Primary research question stated				
Yes	26 (38.2)	32 (47.1)	10 (14.7)	68 (100)
No	9 (47.4)	6 (31.6)	4 (21.1)	19 (100)
Study design				
Cohort study	18 (32.7)	28 (50.9)	9 (16.4)	55 (100)
Other	17 (53.1)	10 (31.2)	5 (15.6)	32 (100)
Measured compounds				
Only mercury	21 (41.2)	21 (41.2)	9 (17.6)	51 (100)
Other included	14 (38.9)	17 (47.2)	5 (13.9)	36 (100)
Sample size [a]				
<300	20 (46.5)	17 (39.5)	6 (14.0)	43 (100)
300-9999	14 (35.0)	20 (50.0)	6 (15.0)	40 (100)
Main analysis strategy				
Assessment of correlation or association	25 (38.5)	32 (49.2)	8 (12.3)	65 (100)
Comparison of groups	10 (45.5)	6 (27.3)	6 (27.3)	22 (100)
All articles	35 (40.2)	38 (43.7)	14 (16.1)	87

Note. [a], Sample size was not reported in four articles.

non-significant outcome with a high number of analyzed outcome variables (*P*=0.026) when adjusted for article characteristics (number of authors, reporting of primary research question, study design, measured compounds, and main analysis strategy).

There was a low correlation between the number of outcome variables and publishing journal's impact factor (Spearman's correlation coefficient 0.13, *P*=0.232). The papers with statistically non-ignificant findings were published in journals with both low and

Figure 2. Number of outcome variables by the statistical significance in 87 original articles focusing on the relationship between maternal exposure to mercury and child development. The horizontal line in the middle of the box indicates the median value of citations and the lower (upper) boundary indicates 25th (75th) percentile. The box plot also displays outliers; cases with more than 1.5 box-length from the upper edge of the box are designated with a circle. The largest and smallest observed values that are not outliers are also shown. Lines are drawn from the ends of the box to those values.

Figure 1. Relationship between number of authors and number of reported outcome variables in 87 original articles focusing on the relationship between maternal exposure to mercury and child development.

Table 2. The Median and Range of the Number of Outcome Variables by Article Characteristics for the Statistically Significant and Non-significant Articles Analyzing Links Between Maternal Mercury Exposure and the Primary Child Development Variable

Item	Significant ($P<0.05$)			Not Significant ($P\geq0.05$)			P-value of M-W[b] test
	Md	range	n	Md	Range	n	
Primary research question stated							
Yes	2	1-17	26	7	1-42	32	0.001
No	12	2-24	9	10	2-39	6	0.776
Study design							
Cohort study	5	1-21	18	8.5	2-42	28	0.114
Other	2	1-24	17	6	1-39	10	0.170
Measured compounds							
Only mercury	2	1-24	21	10	2-42	21	0.012
Other included	6	1-18	14	7	1-39	17	0.468
Sample size [a]							
<300	2	1-24	20	6	1-26	17	0.149
300-9999	5.5	1-21	14	12	2-42	20	0.129
Main analysis strategy							
Assessment of correlation or association	5	1-24	25	8	1-42	32	0.045
Comparison of groups	2	1-18	10	6.5	2-14	6	0.263
All articles using significance tests	4	1-24	35	7	1-42	38	0.014

Note. [a], Sample size was not reported in two articles; [b], Mann-Whitney test.

high visibility. Median values of the journal impact factor were as follows: 2.41 for 'significant papers', 2.46 for 'non-significant' papers, and 2.53 for papers not evaluating statistical significance.

DISCUSSION

This study investigated whether papers with statistically non-significant findings on the relationship between maternal mercury exposure and child development outcomes had a higher number of reported outcome variables assessing child development than the articles with significant findings. We found that the articles containing non-significant primary findings included more health outcome variables, an association not solely explained by the study characteristics. The influence of the outcome of statistical testing on the decision to report only a subset of findings is consistent with a misunderstanding of the relationship between statistical tests and scientific importance (that has been observed elsewhere in the medical literature)[21]. The practice of reporting only significant findings masks the information from studies showing that a specific characteristic is not a risk factor. Such a publication bias creates a false impression of the importance of the findings. Moreover, it skews the reality that may even reflect on practical decisions and also involves questions about research integrity (e.g., see Vähäkangas[22]).

Previous evidence concerning the extent of outcome reporting bias comes from findings that studies reporting relatively high effects (statistically significant findings) are more likely to be published than studies reporting lower effect sizes[7,23-24]. Our findings expand this observation to the number of reported effects in research papers. A research project may include several related outcome variables that are all analyzed. Our study suggests that when reporting significant findings, other related outcomes are omitted, but when reporting non-significant findings, all related outcomes are more often reported.

This is the first analysis of the number of outcome variables in relation to significance of results, and our hypothesis should be tested in further studies. We evaluated articles in one limited research area; results from other fields of clinical

epidemiology may differ. However, this example demonstrates that articles from observational studies may include misleading positive findings and are subject to reporting bias.

The health outcome concept 'child development' includes many dimensions: growth, motor functions, cognition, and behavior including intelligence and language skills as well as social interaction[17,25]. For example, McCarthy Scales of Children's Abilities were used to assess a child's level of development in both cognitive and motor abilities[26-28]. Intelligence can be assessed with several different scales, e.g., Crump et al. administered a battery of 26 psychological and scholastic tests[26]. One evaluated article in our set had analyzed 42 different neuropsychological outcome variables to study the relationship between mercury exposure from thimerosal (an organomercury compound also called thiomersal) and neuropsychological functioning[20]. These measurements usually form a correlated set of outcome variables. Coping with multiple outcomes, multiple models, and multiple comparisons is one of the biggest challenges in data analysis and reporting[21,29-30]. Our finding suggests the need to improve the reporting of studies with several outcome variables. It is important to fully report results for all primary responses including any non-significant findings[29]. However, it is not always necessary to present all descriptive statistics for the secondary outcomes, or they can be shown in an article supplement. The key is the recognition of 'close alternative' outcomes. This set of variables includes several inter-correlated health outcome scales or related measurements that were also analyzed. Variables selected for the final report are often interpreted as the only important outcome variables. This problem of interpretation can be reduced if the close alternatives to the selected primary outcomes are reported. The presentation of all variables originally analyzed in the study recovers information on the correlated set of outcomes variables and encourages readers to think in terms of sets of variables that together represent outcomes associated with the explanatory variables.

The number of outcome variables was also associated with the number of authors and stated primary research question. Collaboration between researchers has increased the proportion of multi-authored papers[31-32]. Coauthors often represent different subfields relevant to the research question under study. Multidisciplinary research may increase the number of interesting response variables in health studies. Although it is valuable for medical studies to evaluate several aspects of subjects' responses, it is important to identify a small set of primary outcome or response variables[21]. The importance of stating the purpose and a priori hypotheses of a research project in the report is obvious, but such a statement was missing in 20.5% (15/73) of papers. These papers also reported more outcome variables. In these cases, the results cannot be interpreted in light of a priori hypotheses. The large number of outcome variables may only indicate that the analyses are explanatory and speculative, and/or the study was conducted to create a hypothesis of potential association of an environmental factor and potential health effect. In any case, the purpose of study should be clearly expressed.

Emission sources of mercury are from both man-made and natural phenomena (e.g., volcanic eruptions). Humans are exposed to mercury mostly through fish consumption and to a lesser extent from medical or cosmetic compounds and dental amalgam[33]. Mercury is a highly toxic metal, has no benefits for human health, is one of the most neurotoxic compounds known, and exposure to it can cause adverse effects during any period of development[34]. However, the lack of sufficient and consistent scientific findings on its association with defects in neurological development in early life exposure has raised public concerns[35-36]. In Davidson et al. [35] review on the association between exposure to mercury and child development outcomes, the insufficient data on mercury neurotoxicity at low doses, discrepancies in data interpretation, and the limitation of dose-response data on neurodevelopment were criticized. Therefore, it is recommended to report all studied outcomes despite their significances and P-values.

Why do the studies on child development with non-significant results (significant) include a high number of outcome variables? First, child development is a broad area of study comprising several aspects worthy of exploration. Investigators may try to seek links from many sources and include everything, as they should, when no significant findings are found. Similar diligence in reporting both the study's purpose and all findings would be needed when positive findings are reported. The number of related outcome variables and their significance may then either validate the findings or cast doubt on their biological significance.

Furthermore, medical literature shows a strong tendency to accentuate significance testing, specifically statistically significant outcomes. Small p-values have a hypnotic effect on the editors and readers alike. Altman[37] stated over two decades ago that in all medical fields 'P-values and P<0.05 rule most', and this remains true. Significant outcomes are more likely to be reported and published[7,38]. Our analysis suggests that authors reporting significant associations have selected a subset of outcomes conveniently describing with low P-values the relationship between maternal mercury and a set of child development measures. Our results reinforce increasing concern with the misuse of significance testing in interpreting medical data[39].

In summary, we empirically investigated the relationship between the number of outcome variables in a research paper and the statistical significance of its primary outcome. After adjustment for likely confounders, articles with non-significant primary findings had substantially more outcome variables. We conclude that the motivation of authors to report outcomes is often associated with the size of the outcomes' P-value rather than their intrinsic scientific information. This does not help future investigators to know which outcomes are redundant, which provide unique information, and which are most responsive to changes in the exposure variable. Editors and referees should be aware of the consequences and likely presence of an outcome reporting bias. To combat this bias, the improved education of researchers, registration of all studies, and their systematic inclusion in meta-analyses should be encouraged by the academic community.

REFERENCES

1. Chalmers I. Underreporting research is scientific misconduct. JAMA, 1990; 263, 1405-8.

2. Hahn S, Williamson PR, Hutton JL. Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee. Journal of Evaluation in Clinical Practice, 2002; 8, 353-9.

3. Chan AW, Hrobjartsson A, Haahr MT, et al. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA, 2004; 291, 2457-65.

4. Ioannidis JP. Why most published research findings are false. PLoS Med, 2005; 2, e124.

5. McGauran N, Wieseler B, Kreis J, et al. Reporting bias in medical research - a narrative review. Trials, 2010; 11, 37.

6. Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol Sci, 2011; 22, 1359-66.

7. Song F, Parekh S, Hooper L, et al. Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess, 2010; 14, 9-193.

8. Thomas DC. Statistical methods in environmental epidemiology. Oxford: Oxford University Press 2009.

9. Royston P, Sauerbrei W. Multivariable model-building: A Pragmatic Approach to Regression Anaylsis based on Fractional Polynomials for Modelling Continuous Variables. Chichester: John Wiley & Sons 2008.

10. Schriger DL, Cooper RJ, Lopez-O'Sullivan A, et al. Is the relationship among outcome variables shown in randomized trials? Trials, 2015; 16, 584.

11. Nieminen P, Rucker G, Miettunen J, et al. Statistically significant papers in psychiatry were cited more often than others. J Clin Epidemiol, 2007; 60, 939-46.

12. Brembs B, Button K, Munafo M. Deep impact: unintended consequences of journal rank. Front Hum Neurosci, 2013; 7, 291.

13. Altman DG. Poor-quality medical research: what can journals do? JAMA, 2002; 287, 2765-7.

14. Fanelli D, Ioannidis JP. US studies may overestimate effect sizes in softer research. Proc Natl Acad Sci USA, 2013; 110, 15031-6.

15. Ioannidis JP, Tatsioni A, Karassa FB. Who is afraid of reviewers' comments? Or, why anything can be published and anything can be cited. Eur J Clin Invest, 2010; 40, 285-7.

16. Nieminen P, Lehtiniemi H, Vähäkangas K, et al. Standardised regression coefficient as an effect size index in summarising findings in epidemiological studies. Epidemiology Biostatistics and Public Health 10[4], e8854. 2013.

17. Ewen JB, Shapiro BK. Disorders of attention or learning in neurodevelopmental disorders. Semin Pediatr Neurol, 2005; 12, 229-41.

18. Cao Y, Chen A, Jones RL, et al. Does background postnatal methyl mercury exposure in toddlers affect cognition and behavior? NeuroToxicology, 2010; 31, 1-9.

19. Murata K, Sakamoto M, Nakai K, et al. Effects of methylmercury on neurodevelopment in Japanese children in relation to the Madeiran study. International Archives of Occupational and Environmental Health, 2004; 77, 571-9.

20. Thompson WW, Price C, Goodson B, et al. Early thimerosal exposure and neuropsychological outcomes at 7 to 10 years. New Engl J Med, 2007; 357, 1281-92.

21. Nieminen P, Miettunen J, Koponen H, et al. Statistical methodologies in psychopharmacology: a review. Hum Psychopharmacol Clin Exp, 2006; 21, 195-203.

22. Vähäkangas K. Research ethics in the post-genomic era. Environ Mol Mutagen, 2013; 54, 599-610.

23. Song F, Eastwood AJ, Gilbody S, et al. Publication and related biases. Health Technology Assessment 4[10]. 2000.

24. Hopewell S, Loudon K, Clarke MJ, et al. Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database of Systematic Reviews, 2009; 1, MR000006.

25. Tervo RC. Identifying patterns of developmental delays can help diagnose neurodevelopmental disorders. Clin Pediatr (Phila), 2006; 45, 509-17.

26. Crump KS, Kjellstrom T, Shipp AM, et al. Influence of prenatal mercury exposure upon scholastic and psychological test performance: benchmark analysis of a New Zealand cohort. Risk Anal, 1998; 18, 701-13.

27. Palumbo DR, Cox C, Davidson PW, et al. Association between prenatal exposure to methylmercury and cognitive functioning

in Seychellois children: a reanalysis of the McCarthy Scales of Children's Ability from the main cohort study. Environmental Research, 2000; 84, 81-8.

28. Freire C, Ramos R, Lopez-Espinosa MJ, et al. Hair mercury levels, fish consumption, and cognitive development in preschool children from Granada, Spain. Environmental Research, 2010; 110, 96-104.

29. Lang T, Secic M. How to report statistics in medicine. Philadelphia: American College of Physicians, 1997.

30. Keselman HJ, Cribbie RA, Holland B. Pairwise multiple comparison test procedures: an update for clinical child and adolescent psychologists. J Clin Child Adolesc Psychol, 2004; 33, 623-45.

31. Hammad FT, Shaban S, Abu-Zidan F. Multiple authorship and article type in journals of urology across the Atlantic: trends over the past six decades. Med Princ Pract, 2012; 21, 435-41.

32. Geminiani A, Ercoli C, Feng C, et al. Bibliometrics study on authorship trends in periodontal literature from 1995 to 2010. J Periodontol, 2014; 85, e136-e143.

33. Castoldi AF, Johansson C, Onishchenko N, et al. Human developmental neurotoxicity of methylmercury: impact of variables and risk modifiers. Regul Toxicol Pharmacol, 2008; 51, 201-14.

34. Bose-O'Reilly S, McCarty KM, Steckling N, et al. Mercury exposure and children's health. Curr Probl Pediatr Adolesc Health Care, 2010; 40, 186-215.

35. Davidson PW, Myers GJ, Weiss B. Mercury exposure and child development outcomes. Pediatrics, 2004; 113, 1023-9.

36. Dorne JL, Kass GE, Bordajandi LR, et al. Human risk assessment of heavy metals: principles and applications. Met Ions Life Sci, 2011; 8, 27-60.

37. Altman DG. Statistics in medical journals: developments in the 1980s. Stat Med, 1991; 10, 1897-913.

38. Dickersin K, Chalmers I. Recognizing, investigating and dealing with incomplete and biased reporting of clinical research: from Francis Bacon to the WHO. J R Soc Med, 2011; 104, 532-8.

39. Kim JS, Kim DK, Hong SJ. Assessment of errors and misused statistics in dental research. Int Dent J, 2011; 61, 163-7.

Appendix A. List of Variables (assessment tests, disorders, or measures) Related to Child Development and Health Outcomes Used in the Included 87 Papers

Domain of Child Development	Test Instruments, Disorders, or Measures
General development	Test instruments: Comprehensive Developmental Inventory for Infants and Toddlers, Gesell Developmental Schedules
Physical developmental milestones	Test instruments: Physical Developmental Index Measures: age of sitting, age of standing, age of talking, age of walking, birth head circumference, birth length, birth weight, creeping, gestational age.
Cognitive, language, and intellectual development	Test instruments: Bayley Scales of Infant Development (Bailey III), Boston Naming Test, Burt Word Recognition Test, California Verbal Learning Test, Children's Memory Scale, Clay Diagnostic Survey, Confrontational Naming Task, Fagan Test Of Infant Intelligence, Kaufman Brief Intelligence Test, Key Mathematics Test, McArthur-Bates Communicative Development Inventory, McCarthy Scales Of Children's Abilities, Peabody Picture Vocabulary Test, Preschool Language Scales, Raven Colored Progressive Matrices, SACMEQ School Achievement Test, Spoken Language Quotient, Stanford-Binet Intelligence Scale, Tactual Performance, Test of Language Development, Wechsler Intelligence Scale For Children (WISC-R WISC-III,WISC-IV), Woodcock-Johnson Tests of Cognitive Abilities.
Behavioral outcomes	Test instruments: Bayley Scales of Infant Development (Bailey III), Behavior Rating Inventory of Executive Function, Child Behavior Checklist, Conners' Rating Scales, Continuous Performance Test, Disruptive Behavior Disorders Rating Scale, Gordon Diagnostic System Neonatal Behavioral Assessment Scale, Disorders: level of auto-aggressive behavior, level of disruptive behavior.
Motor function	Bender-Gestalt Test, Bruininks-Oseretsky Test, Finger·Tapping Test, Grooved Pegboard Test, Ear-Hand Coordination Test, Eye-Hand Coordination Test, Fine Motor Coordination Test, Postural Sway Test, Trail Making Test, Tremor Test Measures: tics, reaction time.
Audition, visual, and visual-motor development	Test instruments: Beery-Buktenica Developmental Test Of Visual-Motor Integration, Bender-Gestalt Test, Drawing Test (visual-motor), Haptic Matching Tests, Matching Test (visual spatial), Pegboard Test (fine motor skills), Teller Visual Acuity Test, Visual Recognition Memory. Measures: Audiometric tests.
Neurodevelopmental disorders and other health outcomes	Test instruments: Bayley Scales Of Infant Development (Bailey II, Mental Developmental Index MDI and Psychological Developmental Index PDI), Denver Developmental Screening Test, Neurological Development Tests, Profile of Mood States (Bipolar version), Disorders: Attention deficit disorder ADD, attention deficit hyperactivity disorder ADHD, autism and autism spectrum disorders, congenital anomalies, developmental disorder, emotional disturbance, failure to thrive, learning disability, pervasive developmental disorder not otherwise specified (PDD-NOS), pneumonia. Measures: long chain polysaturated fatty acids, sex hormones.