Low-level lead exposure and mortality in US adults: a population-based cohort study – A Review

By LEAD Action News Contributor

24th July 2018

Reference:

Bruce P Lanphear, Stephen Rauch, Peggy Auinger, Ryan W Allen, Richard W Hornung

www.thelancet.com/public-health Published online March 12, 2018

http://dx.doi.org/10.1016/S2468-2667(18)30025-2

https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667(18)30025-2/fulltext?code=lancet-site

Introduction

The paper in The Lancet, Public Health journal by Professor Bruce Lanphear et al attracted a lot of attention. Titled "Low-level lead exposure and mortality in US adults: a population-based cohort study" it was published in March 2018 and was soon reported by over 70 news media.

The aim of this LEAD Action News article is to bridge the gap between the news media summaries and the Lanphear paper itself in an attempt to give a lay summary of the contents. Inevitably this means skimming over some of the background and details. It should be noted that the author is not an expert in statistics. Corrections or clarifications from readers would be welcomed.

For a proper understanding readers should study the full paper – and take a degree in statistics.

Summary

Basically, what this study did was to link two data sets to see if there was any correlation between elevated lead levels and some later fatal health problems. The study looked at all-cause mortality (death) together with cardiovascular disease (CVD) and ischaemic heart disease (IHD).

The most important output from the study was to report the hazard ratio for an increase in blood lead level (BLL). The hazard ratio was calculated for an increase from the level below which 10% of participants were found, to the level above which 10% of participants were found - from below 1·0 µg/dL to above 6·7 µg/dL. That is micrograms (1,000,000th of a gram) per decilitre (one tenth of a litre) – by any measure, very small amounts.

To put it another way, it answers the question "What is the risk of moving from the best 10% to the worst 10%?"

In summary, ignoring confidence intervals, the results presented were:

Outcome	Hazard Ratio	Population Attributable Fraction	Population Equivalent (USA)
All-cause mortality	1·37	18·0%	412 000
CVD	1·70	28·7%	256 000
IHD	2·08	37·4%	185 000

The meaning of these measures seems to be:

Hazard Ratio

This is the difference in the chance of an outcome occurring between one level of a measure and another.

In this case, an increase in BLL from below 1.0 to over 6.7 µg/dL would increase the risk of death by 37%, dying from CVD by 70% and dying from IHD by over 100%. That is over a period, on average, of 19.3 years. This is the time between the patient appearing on the first and second data sets. However, see the note below about confidence levels.

It should be added that µg/dL is the commonest measurement of BLL. The mean in this study was 2.71 µg/dL. In the USA 5 µg/dL is considered elevated (NIOSH, 2018 [1]), but in the UK the level would have to be 60 µg/dL for an adult to be suspended from work (UK Legislation, 2002 [2]).

Population Attributable Fraction (PAF)

The British Medical Journal (Mansournia and Altman, 2018 [3]) has a good description of what this is: "PAF is the estimated fraction of all cases that would not have occurred if there had been no exposure". In this case, the Lanphear et al study concludes that 18% of deaths would not have occurred had all subject's BLLs have been below 1.0 µg/dL. We know that death must occur at some time, so this would seem to mean the deaths that would be avoided over the mean period of follow-up, i.e. 19.3 years from the average age of subjects at the start which was 44.1 years. The same can be applied to CVD and IHD.

Population Equivalent

Having established the PAF in the sample of 14,289 subjects, this is then extrapolated to the whole population of the USA based on the number of all-cause deaths, and those caused by CVD and IHD.

This is how the headline figure of 412,000 deaths per year due to lead poisoning is derived.

Data Sets

The data analysis in the study uses two sources:

The Third National Health and Nutrition Examination Survey (NHANES III)

National Death Index (NDI)

The essential process was to match NHANES records with NDI records for each person in the study. NHANES provides the BLLs and NDI provides the cause of death, if deceased. This study does not therefore look at the lifelong impact of lead poisoning.

NHANES

NHANES III was conducted in the USA from 1988 to 1994 and included a total of 39,695 persons. There was a bias in the data collected with higher sample rates for unders 5s, over 60s, Mexican-American persons, and non-Hispanic black persons. However, this should not impact the results because the same individuals should be found in the NDI data.

The Lanphear et al study used data for those aged 20 or over in NHANES.

The NHANES data set includes wide variety of measures. Those chosen as the baseline in the Lanphear study are:

- Gender
- Ethnic origin
- Education
- Income
- Residence location
- Smoking
- Alcohol consumption
- Physical activity
- Hypertension
- Diabetes
- Healthy eating
- Body mass index

- Age
- HbA_{1c} – measures glucose concentration
- Serum cotinine – indicates smokers
- Urinary cadmium – measures kidney accumulation

NDI

The NDI data is a centralised database of death records from 1979. For this study individuals in NHANES are matched to the appropriate death record in NDI. It would appear that the cause of death is used to identify deaths from CVD or IHD.

Diseases

CVD

Cardiovascular disease is a general term for conditions affecting the heart or blood vessels. Around 1 in 4 deaths in the USA are caused by heart disease (CDC, [4]) and it is estimated that this costs $200billion per year (CDC Wonder [5]). If elevated lead levels are causing 28.7% of these deaths then this is costing over $57billion per year. This is not including the costs resulting from other non-fatal lead related conditions.

Included in CVD are:

- Coronary heart disease – when blood flow to the heart is blocked, or reduced, causing angina, heart attacks and heart failure
- Strokes and transient ischaemic attack – when blood supply to the brain is cut off permanently or temporarily.
- Peripheral arterial disease – when there is blockage in the arteries to the limbs
- Aortic disease – such as aortic aneurysm, bulging of the main artery carrying blood from the heart

IHD

This is another name for coronary heart disease. It is probably selected for this study because it accounts for around 370,000 deaths in the USA of which about 146,000 occurred before age 75. These are data from the National Center for Health Statistics. Underlying Cause of Death 1999-2016 on CDC WONDER Online Database, released December, 2017. Data are from the Multiple Cause of Death Files, 1999-2016, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program (CDC, [6]).

Statistical Methods

The paper lists several statistical methods used in the data analysis. Most of these are highly complex and sophisticated. They include:

Cox proportional hazards models

> This appears to be a way to calculate the impact of different values of a variable on survival rates. In this case, differing values of blood lead (SDAHT, [7])

Schoenfeld residuals

> The Cox proportional hazard model assumes that the hazard ratio is constant over time. This is validated using the Schoenfeld residuals method. In the case of the Lanphear study none of the models violated this assumption using this test.
>
> (StataCorp, [8])

SUDAAN (version 10.0.1)

> SUDAAN is a software package for analysing correlated data. The software includes features to correctly account for complex design features, as found in the NHANES data, including weighting (RTI, [9]).

Taylor linearisation

> One of the features of SUDAAN is support for Taylor linearization. This appears to be used to find an approximate linear relationship at a point in the complex NHANES data (RTI, [9]).

Five-knot restricted cubic splines

> Because the relationship between blood lead levels and survivability may not be linear, i.e. not a straight line on a chart, splines are used to investigate what shape the line might be (Wikipedia, [10]).

Confidence Intervals

The findings in the paper are expressed with 95% confidence intervals. What this means is that we can be 95% confident that, based on an extrapolation of the sample data, that the true population value lies between the lower and upper confidence intervals.

For example, the all-cause mortality hazard ratio was found, in the sample, to be 1·37. For this result, the 95% confidence intervals are given as 1·17–1·60. Therefore, the true hazard ration is probably somewhere between 1.17 and 1.60.

Another example is the population attributable fraction for all-cause mortality. This was found to be 18% with confidence intervals of 10·9–26·1. So it is probable that the actual percentage of deaths attributable to lead is between just below 11%, to over 26%. Converting this into numbers of deaths we can say that between 249,000 and 597,000 deaths each year are caused by death.

Another type of interval is used for the follow-up time. That is the time from the subject being included in the NHANES data to their appearance in the NDI data. This is stated as 19·3 years with an interquartile range of 17·6–21·0. This is the range of values where the middle 50% of the follow-up times lie. This helps describe the range of values or the shape of the histogram. For this we have actual, not extrapolated, data, so can say for sure that 50% of the follow-up times lie within about 2 years below and 2 years above the average follow-up time.

Limitations

The paper describes some limitations. Interpretation of these could include:

> Baseline measurements of lead exposure may not reflect lifelong cumulative lead loading. Some might be the result of incidental rather than environmental exposure or continued environmental exposure could result in greater cumulative lead loading.
>
> Cause of death can be misreported and impact the analysis. Data quality is always a factor.
>
> Other risk factors are adjusted for where possible. However, some, e.g. air pollutants and arsenic, were not adjusted for and could have an impact.
>
> Although the NHANES data contains many parameters there may be some not included that impact the results. These are called confounders and may include additional socioeconomic and occupational factors.

Conclusion

> It could be said that death is unavoidable so can we say that this study really measures avoidable deaths? What we can say is that we die from the first thing that kills us. This study suggests that, it is probable, that, currently, between 249,000 and 597,000 Americans are dying sooner than they would have because they had elevated blood lead levels around 20 years previously.

References

1. The National Institute for Occupational Safety and Health (NIOSH). Adult Blood Lead Epidemiology and Surveillance (ABLES) https://www.cdc.gov/niosh/topics/ables/description.html

2. UK Legislation. The Control of Lead at Work Regulations 2002 http://www.legislation.gov.uk/uksi/2002/2676/regulation/10/made

3. Mansournia M. A and Altman D. G. Population attributable fraction *BMJ 2018; 360* https://www.bmj.com/content/360/bmj.k757

4. CDC, Heart Disease Facts https://www.cdc.gov/heartdisease/facts.htm

5. CDC, Wonder, About Multiple Cause of Death, 1999-2016 http://wonder.cdc.gov/mcd-icd10.html

6. CDC, Wonder, About Underlying Cause of Death, 1999-2016 http://wonder.cdc.gov/ucd-icd10.html

7. STDHA, Statistical tools for high-throughput data analysis, Cox Proportional-Hazards Model http://www.sthda.com/english/wiki/cox-proportional-hazards-model

8. StataCorp, stcox PH-assumption tests — Tests of proportional-hazards assumption https://www.stata.com/manuals13/ststcoxph-assumptiontests.pdf

9. SUDAAN® Statistical Software for Analyzing Correlated Data https://www.rti.org/impact/sudaan-statistical-software-analyzing-correlated-data

10. Wikipedia, Spline (mathematics) https://en.wikipedia.org/wiki/Spline_(mathematics)