Prospective and longitudinal human studies of lead and cadmium exposure and the kidney

Johan Sommar

\begin{align*}
\ln(Y_{ij}) &= \mu + b_i + \epsilon_{ij} \\
\hat{\alpha}_b^2 / (\hat{\alpha}_b^2 + \hat{\alpha}_w^2) \\
Y_t &= c + \rho Y_{t-1} + \epsilon_t \\
\text{Var}(Y_t) &= \sigma^2 / (1 - \rho^2) \\
\text{Cov}(Y_{ij}, Y_{ik}) &= \sigma_b^2 + \sigma_w^2 \rho^{j-k} / (1 - \rho^2) \\
y_{it} - y_{oi} &= \beta_0 + \beta_1 y_{oi} + \beta_2 \ln(x_{oi}) + \epsilon_i \\
U_{ph,UF=1} &= U_{ph} \cdot UF_i^{-\beta_1} \\
\ln(Y_{ij}) &= \mu + b_i + \epsilon_{ij} \\
\hat{\alpha}_b^2 / (\hat{\alpha}_b^2 + \hat{\alpha}_w^2) \\
Y_t &= c + \rho Y_{t-1} + \epsilon_t \\
\text{Var}(Y_t) &= \sigma^2 / (1 - \rho^2) \\
\text{Cov}(Y_{ij}, Y_{ik}) &= \sigma_b^2 + \sigma_w^2 \rho^{j-k} / (1 - \rho^2) \\
y_{it} - y_{oi} &= \beta_0 + \beta_1 y_{oi} + \beta_2 \ln(x_{oi}) + \epsilon_i \\
U_{ph,UF=1} &= U_{ph} \cdot UF_i^{-\beta_1}
\end{align*}
Prospective and longitudinal human studies of lead and cadmium exposure and the kidney

Johan Sommar
If at first it doesn’t fit, fit, fit again.
--John McPhee
# Table of contents

Abstract ..................................................................................................... III  
Sammanfattning (Swedish summary) ...................................................... V  
Abbreviations ....................................................................................... VII  
List of papers ......................................................................................... VIII  
Introduction ............................................................................................ 1  
  One-way random effect and mixed models for investigation of Cd and Pb biomarker properties ........................................................................ 1  
  Kidney effects in relation to biomarkers of Cd and Pb ................................ 3  
  Matched case-control design for evaluation of risk of end-stage renal disease ................................................................. 3  
  The influence of metallothionein polymorphisms on Cd metabolism and kidney toxicity ................................................................. 5  
  Longitudinal studies on kidney recovery from Cd toxicity ................ 5  
Aims of the thesis .................................................................................... 6  
Methods .................................................................................................... 7  
  Study populations ................................................................................ 7  
    Paper I ............................................................................................... 7  
    Paper II ............................................................................................. 7  
    Paper III and IV ............................................................................... 9  
  Chemical analyses .............................................................................. 10  
    Paper I ............................................................................................... 10  
    Paper II ............................................................................................. 10  
    Paper III and IV ............................................................................... 10  
  Statistical analyses ............................................................................. 11  
    Paper I ............................................................................................... 11  
    Paper II ............................................................................................. 15  
    Paper III ........................................................................................... 16  
    Paper IV ........................................................................................... 17  
Results .................................................................................................... 20  
  Paper I ............................................................................................... 20  
  Paper II ............................................................................................... 23  
  Paper III ............................................................................................. 28  
  Paper IV ............................................................................................. 29  
Additional results .................................................................................. 33  
  Comparing different measures of change ........................................... 33  
  Variance components for Cd and kidney effect markers ................... 33  
Discussion .............................................................................................. 35  
  Study design and data analysis ......................................................... 35  
  Longitudinal study design ................................................................. 35  

Abstract

Cadmium and lead accumulate in humans and can have toxic effects. Exposure to cadmium is well known to cause kidney damage. Cadmium binds to metallothioneins, proteins that play a role in cadmium transport. Lead exposure’s main effect is on the central nervous system, but associations with kidney disease have also been found, although it is unknown if the latter is a causal association. The main source of both metals within the non-smoking population is from the diet.

This thesis aims to 1) compare the biomarkers lead and cadmium concentration in whole-blood, plasma and urine with regard to their ability to discriminate between individuals with different mean concentrations, and to describe the effect of urinary dilution, 2) estimate the association between end-stage renal disease and blood concentrations of cadmium, lead and mercury, using prospectively collected samples for exposure evaluation, 3) use longitudinal data on kidney function makers to evaluate kidney recovery after a substantial decrease in cadmium exposure, and 4) assess the influence of metallothionein polymorphisms (MT1A rs11076161, MT2A rs10636 and MT2A rs28366003) on cadmium-associated kidney toxicity and recovery due to a reduction in Cd exposure.

Repeated sampling of whole-blood, plasma and urine was conducted on 48 occupationally lead-exposed men and 20 individuals under normal environmental lead exposure, for estimation of the day-to-day and between individual-variation. Prospective samples were obtained for 118 cases that later in life developed end-stage renal disease, and 378 matched controls. Erythrocyte cadmium, lead, and mercury concentrations were determined and the risk of developing end-stage renal disease associated with metal concentrations was estimated. For evaluation of kidney recovery after a reduction in cadmium exposure and to test for gene-environment interactions, follow-up data on N-acetyl-β-d-glucosaminidase, β₂-microglobulin, albumin, and gene polymorphisms were obtained for 412 individuals within the Chinese population and the relation to blood and urinary cadmium was assessed.

The concentration of lead in blood was found to be the biomarker with the largest fraction of the total variance attributable to between-individual variation, and was therefore the biomarker with the best ability to discriminate between individuals with different mean concentrations,
both for individuals under occupational and normal environmental exposure (91 and 95%, respectively). Adjusting for urinary dilution had a great effect on the fraction of the total variance attributable to between-individual variation among individuals with normal lead exposure but only a minor effect among those who were occupationally exposed. Variance analysis showed that blood concentrations were also the best discriminating biomarker for cadmium.

Erythrocyte lead was, in a univariate model, associated with an increased risk of developing end-stage renal disease [odds ratio (OR) = 1.54 for an interquartile range increase, with a 95% confidence interval (CI) = 1.18-2.00], while erythrocyte mercury was negatively associated (OR = 0.75 for an interquartile range increase, with a 95% CI = 0.56-0.99). For erythrocyte cadmium, the OR was 1.15 with a 95% CI of 0.99-1.34. Associations with lead and cadmium were only seen among men. In the study on kidney recovery, the proportion of individuals with albumin level above the 95th percentile decreased between baseline and follow up, but no decrease was found for the tubular markers N-acetyl-β-d-glucosaminidase and β₂-microglobulin. Metallothionein polymorphisms modified cadmium-associated effects on N-acetyl-β-d-glucosaminidase and β₂-microglobulin levels but did not modify cadmium-associated change in any of the kidney function markers between baseline and follow up after a substantial decrease in exposure.

Blood concentrations of lead and cadmium are the biomarkers with the best ability to discriminate between individuals with different mean concentrations. Adjustment for urinary dilution has great influence on the fraction of the total variance attributed to between individual variation among urine samples with low lead concentrations, but only a small influence on samples with high lead concentrations. This suggests a difference in excretion. The association between end-stage renal disease and low-level lead exposure, as assessed through prospective erythrocyte samples, gives reason for concern, although further studies are needed to determine causality. A cadmium-associated increase in albumin is reversible after a substantial reduction in exposure, but this is not the case for the observed tubular effects. The tubular kidney effects of cadmium might be modified by the MT1A rs11076161 polymorphism.
Sammanfattning (Swedish summary)


Kadmium och bly ackumuleras i njure respektive ben och kan ha toxikologiska effekter. Det är välkänt att höga exponeringsnivåer av kadmium orsakar njurskada och även vid lägre exponeringsnivåer har studier funnit samband med markörer för njurfunktion. Exponering för bly påverkar i första hand det centrala nervsystemet. Studier har dock funnit samband mellan koncentrationer av bly i blod och njurens glomerulära filtrationshastighet. Det är oklart både om dessa associationer, vid låga exponeringsnivåer, är viktiga för hälsan och om de verkligen beror på att kadmium och bly orsakar njurskada. För att studera end-stage renal disease användes prospektiva kohorter där personer lämnat blodprov för forskning: Västerbottens interventionsprogram med prover som tagits vid Västerbottens hälsoundersökningar, MONICA-undersökningar i Norr- och Västerbotten, mammografundersökningarna i Västerbotten och Malmö kost cancer. Sammanlagt ingick över ett hundra tusen individer i dessa kohorter. Med hjälp av det Svenska njurregistret identifierades sedan 118 personer som senare i livet fått end-stage renal disease. Dessa jämfördes med 378 kontroller. För dessa 496 personer tinades blodprov (närmare bestämt röda blodkroppar) upp och analyserades för kadmium och bly. För att undersöka njurens förmåga till återhämtning studerades tre områden i Kina varav ett tidigare varit kadmiumexponerat. Erytrocytkoncentrationer av bly var, utan att ta hänsyn till några andra variabler, associerat med en ökad risk för att utveckla end-stage renal disease (med odds kvoten 1.54 för en interquartile range ökning av erytrocytbly, med ett 95% konfidensintervall 1.18-2.00). Sambanden kvarstod också efter att ha tagit hänsyn till övriga variabler. Förr erytrocytkadmium var odds kvoten 1.15 med 95% konfidensintervall 0.99-
och sambandet försvagades när hänsyn togs till andra variabler. Associationerna sågs bland män men inte bland kvinnor.


Åttioprocent av kadmium i celler är bundet till proteinet metallotheonin, vilket skyddar mot cellskada, men har också en roll i transporten av kadmium från levern till njurarna. En tidigare studie har visat att njurens känslighet för kadmiumexponering var associerad med genetiska skillnader i detta protein. För att studera genetiska associationer studerades de 412 personerna i den kinesiska studien [då också individernas genotyper av metallotheonin-polymorfierna MT1A rs11076161 (G/A), MT2A rs10636 (G/C) och MT2A rs28366003 (A/G) bestämdes]. Genetiken spelade roll för sambandet mellan förmåga att återta proteiner och kadmium men påverkade inte förändring av njurfunktion efter att man slutat äta kadmiumförorenat ris.

Kadmium- och blykonzentrationer i blod är de biomarkörer, av koncentrationer i blod, plasma och urin, med den bästa förmågan att skilja på individer med olika medelkonzentrationer. Justering för urinutspädning påverkade andelen av den totala variationen som kunde förklaras av skillnader mellan individer i stor utsträckning för individer med normal miljömässig exponering men inte bland yrkesexponerade, vilket tyder på en skillnad i hur utsöndringen går till. Associationen mellan end-stage renal disease och låg exponering för bly, uppknöt i prospektiva erytrocytprover, ger orsak till oro, men ytterligare studier behövs för att kunna utvärdera om detta är ett kausalt samband. En kadmiumrelaterad skada av den glomerulära filtrationen är reversibel efter en kraftig reducierung i exponering, men detta är inte fallet för tubulära skador. De tubulära njureffekterna av kadmiumexponering kan påverkas av metallotheonin-polymorfier.
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-Cd</td>
<td>Cadmium concentration in whole-blood</td>
</tr>
<tr>
<td>BMI</td>
<td>Body mass index</td>
</tr>
<tr>
<td>B-Pb</td>
<td>Lead concentration in whole-blood</td>
</tr>
<tr>
<td>Cd</td>
<td>Cadmium</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>Ery-Cd</td>
<td>Cadmium concentration in erythrocytes</td>
</tr>
<tr>
<td>Ery-Pb</td>
<td>Lead concentration in erythrocytes</td>
</tr>
<tr>
<td>Hg</td>
<td>Mercury</td>
</tr>
<tr>
<td>ICC</td>
<td>Intraclass correlation</td>
</tr>
<tr>
<td>MT</td>
<td>Metallothionein</td>
</tr>
<tr>
<td>NAG</td>
<td>N-acetyl-β-d-glucosaminidase</td>
</tr>
<tr>
<td>OR</td>
<td>Odds ratio</td>
</tr>
<tr>
<td>Pb</td>
<td>Lead</td>
</tr>
<tr>
<td>P-Cd</td>
<td>Cadmium concentration in plasma</td>
</tr>
<tr>
<td>P-Pb</td>
<td>Lead concentration in plasma</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>U-Cd</td>
<td>Cadmium concentration in urine</td>
</tr>
<tr>
<td>U-Pb</td>
<td>Lead concentration in urine</td>
</tr>
</tbody>
</table>
List of papers

This thesis is based on the following papers:


Reprints were made with the kind permission of the publishers.

*JN Sommar's name was until 2012-08-18 J Nilsson*
Introduction

Cadmium (Cd) and lead (Pb) have toxic effects in humans and accumulate in the kidneys and bones, respectively, with long biological half-lives of 10-30 years (1, 2). The main source of both Cd and Pb exposure among the general non-smoking population is from the diet. Dust and soil may also be important sources of Pb among children. Cd and Pb may cause kidney damage - this is well known for Cd and has been suggested for Pb by association studies with lead concentrations in whole blood (B-Pb), but causality is unclear and only a few prospective studies have been performed. Although a prominent effect from Cd exposure is kidney damage, Cd also has a toxic effect on bones. For Pb the main effect is on the central nervous system, with no safe exposure level (2).

Biomonitoring is widely used for exposure assessment (e.g., among lead workers) and in epidemiological studies. The epidemiology of Pb toxicity is mainly based on B-Pb, since Pb concentrations in urine (U-Pb) and plasma (P-Pb) are considered too imprecise (large variation between repeated measurements) (3). B-Pb is considered to mainly reflect the presence of Pb in soft tissue, i.e., current exposure, but it also reflects the long-term accumulation of Pb in bone. Most of the Pb in whole blood is present in cells, and therefore Pb concentrations in erythrocytes (Ery-Pb) will reflect the same time period of exposure as B-Pb. For Cd most studies on health effects are related to Cd concentrations in urine (U-Cd). U-Cd is primarily indicative of Cd levels in the kidney, which are determined by the long-term exposure, whereas Cd concentrations in whole-blood (B-Cd) reflect a combination of both current and historical exposure. Because almost all B-Cd is present in erythrocytes (Ery-Cd), B-Cd and Ery-Cd also reflect the same time period of exposure.

One-way random effect and mixed models for investigation of Cd and Pb biomarker properties

In addition to the kinetic aspects of different biomarkers it is desired that a biomarker has: 1) low variation among repeated measurements (low ‘day-to-day’ variation within individuals), and 2) high variation between individuals who differ in mean concentration. A biomarker with low variation between individual measurements and high variation between individuals with different mean levels is a biomarker with the ability to
discriminate between individuals with high and low levels. The biomarkers should reflect the individual exposure and body burden, and therefore a preferred biomarker has large variation between individuals with different mean levels (i.e., there is a close relation between the biomarker and exposure and body burden). A consequence is that the relation between these variances is of importance. One way to quantify this is to calculate the intra-class correlation coefficient (see below). Biomarkers with high variation between within-individual measurements are not preferred due to the large uncertainty associated with a single sample measurement. Exposure assessment based on a single sample is also the common practice within both biomonitoring of occupational exposure and within epidemiological research.

By utilizing repeated measurements within a one-way random effects model we are able to estimate variance components within and between individuals. As a joint measure of the biomarkers within individual variation (precision) and ability to discriminate between individuals with different mean levels, the fraction of the total variance that can be attributed to differences between individuals can be calculated. This is the definition of the intraclass correlation coefficient (ICC) within a one-way random effects model assuming independence between residuals within each individual. No previous study exists on within- and between-individual variance for Cd biomarkers, but one study exist on the ICC for B-Pb and P-Pb showing that a large fraction of the total variance could be attributed to differences between individuals for both B-Pb and P-Pb (4).

Besides comparing within- and between-individual variation among candidate biomarkers, within- and between individual variance may be estimated for deviations from the estimated association between each pair of biomarkers to investigate if certain individuals have measurements that deviate from the bivariate relation. This is of particular interest for the non-linear relationship between B-Pb and P-Pb. By studying the partition of Pb between cells and plasma within a one-way random effects model, estimating the fraction of the total variance of deviations (from their bivariate relations) that can be attributed to differences between individuals, we are able to assess if certain individuals have a greater deviation from the relation than others. If certain individuals have larger/smaller partition of Pb in plasma, then P-Pb measurements would contain information not contained within B-Pb measurements. Correspondingly, if the fraction of the total variance attributed to individual differences were large then this would mean that
the biomarkers give different information on the individual’s Pb exposure and body burden. However, if the partition of Pb between cells and plasma varies largely among within-individual measurements compared with between individuals, then P-Pb measurements would probably not give any additional insight with regards to an individual’s exposure and body burden than B-Pb. A previous study on Pb workers, with three repeated measurements taken over a single day, found that P-Pb measurements gave somewhat different information than B-Pb measurements (5).

Using mixed effects models we are able to use repeated measurements for estimation of the association between each pair of exposure biomarkers with higher accuracy of each individual’s biomarker concentration. The association between repeated measurements of B-Pb and P-Pb has previously been estimated (4), and in several studies using a single measurement for each individual (e.g., 6). However, there are no such studies on the bivariate relation between U-Pb and either B-Pb or P-Pb.

Mixed effects models may also be used to assess the association between biomarkers of effects and exposure biomarkers. There are few studies that have compared which of the biomarkers have the closest relation to exposure and health effects, but a closer correlation has been suggested for P-Pb compared with B-Pb (7, 8, Tian et al. submitted to Toxicology Letters).

**Kidney effects in relation to biomarkers of Cd and Pb**

**Matched case-control design for evaluation of risk of end-stage renal disease**

Individuals with end-stage renal disease have a loss of renal function that require permanent renal replacement therapy, i.e., dialysis or kidney transplantation. End-stage renal disease is usually a slowly progressing disease, and the disease affects blood concentrations of different substances, possibly also the biomarkers of metal exposure. Therefore prospectively collected samples are needed. For high efficiency estimation of dose-response effects, while minimizing the amount of sample required, a matched case-control design is appropriate for estimating the risk of developing the disease in relation to Cd, Pb, and mercury (Hg) exposure. The objective of such a case-control study is to compare
individuals that develop disease with those who do not, and to find the best model fit with regards to particular risk factors. The size of the effect can be interpreted in terms of the odds ratio (OR) for a unit increase in the risk factor. In epidemiological studies, with different study designs, the risk estimates are often expressed as the relative risk (RR) for a unit increase which may be more easily interpretable. However, for small disease probabilities the RR and the OR are asymptotically equivalent.

In Sweden the number of patients treated for end-stage renal disease each year has increased. This increase is probably due to an increasing prevalence of obesity in the Swedish population that has led to increased rates of diabetes and hypertension (9). Diabetes and nephrosclerosis due to hypertension are the two main causes of end-stage renal disease (10). Obesity has also been found to be an independent risk factor for end-stage renal disease both in diabetic and non-diabetics individuals (11-13). End-stage renal disease is the final stage of chronic kidney disease which is a progressive loss in renal function during a longer time period. Risk factors for chronic kidney disease within the general population are old age, proteinuria, smoking, and hypertension (14). The metals Cd, Pb and Hg are known to cause nephrotoxic effects among occupationally exposed (Cd: (15); Pb: (16, 17); Hg: (18)). Impaired kidney function has also been found to be associated with low-level Cd exposure in the general population (1), as indicated by e.g. increased excretion of urinary beta-microglobulin. For Pb, as previously mentioned, associations have been observed between decreased glomerular filtration rates and increasing B-Pb levels (2), and these associations have been observed even at low levels.

Several studies on end-stage renal disease have found an association with biomarkers of Cd and Pb (19-24), but these are all cross-sectional studies and this means that the difference in biomarker levels could have been caused by the disease itself. For Cd an indication of an association between end-stage renal disease and ecological exposure data on workers at and residents near a Cd-battery production facility has been found (25). Prospective studies on albuminuria and reduced estimated glomerular filtration rate have also found an association with low environmental Pb exposure (26, 27), but no association was found between risk of chronic kidney disease and Pb exposure in a study using occupational information (28). In a recent review the evidence for a causal effect of Pb has been challenged (29).
There are no previous studies on end-stage renal disease with a prospective design and use of relevant exposure biomarkers for Cd, Pb and Hg.

**The influence of metallothionen polymorphisms on Cd metabolism and kidney toxicity**

Genetic variations, especially polymorphisms of the low molecular weight metallothionein (MT) proteins, might affect Cd metabolism and susceptibility to Cd kidney toxicity. Cd binds to MTs, and MTs have a critical role in the transport of accumulated Cd from the liver to the kidneys (3). A genetic polymorphism is defined as a genetic sequence variation that occurs in >1% of the population, and single nucleotide polymorphisms (SNPs) are the most common of these sequence variations. A previous study found that the MT1A rs11076161 polymorphism modified Cd-associated kidney toxicity and levels of B-Cd at high levels of Cd exposure (30).

**Longitudinal studies on kidney recovery from Cd toxicity**

A follow-up study, in which the outcome is observed after a change in exposure, can be thought of as the simplest form of a longitudinal design. Because Cd is well known to cause kidney effects (3) a concern is if the Cd-related effect on kidney function is permanent or can be reduced by decreasing the exposure. Some studies have found that renal tubular damage among previously occupationally exposed to Cd was irreversible (31, 32). Renal dysfunction caused by environmental Cd exposure has also been found to be irreversible despite a 50% reduction in exposure (33). Subclinical kidney effects from low environmental exposure has, however, been found to be non-progressive (34).

The impact of MT-polymorphisms on Cd-associated change in kidney function has not previously been evaluated.
Aims of the thesis

- Implement one-way random effects models as a means of estimating the within- and between-individual variance components, in order to compare the biomarkers Pb and Cd concentration in whole-blood, plasma and urine with regard to their ability to discriminate between individuals with different mean concentrations. The same methodology will also be used to compare the effect of different methods of adjustment for urinary dilution.

- Estimate the association between end-stage renal disease and blood concentrations of Cd, Pb and Hg, using a prospective study design for exposure evaluation.

- Evaluate, using data from a longitudinal study design, kidney recovery after a substantial reduction in Cd exposure.

- To assess the genetic influence by metallothionein polymorphisms ($MT1A$ rs11076161, $MT2A$ rs10636 and $MT2A$ rs28366003) on recovery from Cd-associated kidney toxicity.
Methods

Study populations

Paper I

The study participants consisted of two groups; one group with 48 occupationally exposed individuals recycling batteries and one group consisting of 20 individuals under normal lead exposure. All individuals were male and the median age among occupationally exposed was 42 years (range 22-62) and among environmentally exposed 38 years (range 25-57). Blood and urine samples were taken every two to three months over two years from the occupationally exposed group and over one year from the environmentally exposed group. This yielded a total of 7-13 sampling occasions of the occupationally exposed individuals and 5 or 6 sampling occasions for the environmentally exposed individuals. However some samples were excluded due to difficulties with analysis, exclusion of hemolyzed plasma samples, and one occasion of samples being mixed between individuals.

All subjects gave informed written consent for their participation in the study, which was approved by the Ethical Committee of Lund University.

Paper II

Baseline examinations

The Northern Sweden Health and Disease Study

The Northern Sweden Health and Disease Study (35) includes the population-based Västerbotten Intervention Project (36), the Northern Sweden Health WHO Monitoring of Trends and Cardiovascular Disease (MONICA) study, and the local Mammography Screening Project. Since the start of the Västerbotten Intervention Project in 1985, residents in Västerbotten County reaching 40, 50 or 60 years old (during some years those who were also 30 years old were also included) have been invited for a routine health screening. The MONICA study was also initiated in 1985 as a health and examination program for cardiovascular disease and diabetes in which a random selection of residents were invited to participate in a continuous health survey. In addition, since 1995 women in Västerbotten County have been encouraged to take part in the
Mammography Screening Project every 2 or 3 years. The first two studies include anthropometric, blood pressure, and lifestyle data, whereas the third study contains more limited information. Within these programs, participants are also asked to donate blood samples that are stored in a biobank at -70°C. Between the years 1985 and 2002, 74,000 individuals had donated a blood sample to the Northern Sweden Health and Disease Study. The participation rates during these years were 60% within the Västerbotten Intervention Project, 77% within MONICA, and 65% within the Mammography Screening Project.

The Malmö Diet and Cancer Study
The Malmö Diet and Cancer Study (37) contains baseline anthropometric, blood pressure, and lifestyle data for men and women aged 45-54 year who had been surveyed between the years 1991 and 1996 (n=39,447). Residents of the municipality of Malmö were randomly invited to participate in the study, but recruitment was also performed by advertisement and by recommendations from other participants. The only exclusion criterion for participation was lack of Swedish language skills. Participants who emigrated during the study period were censored from the study on the date they left the country. All participants were also asked to donate a blood sample to be stored in a biobank (-70°C). Forty-five percent of invited individuals participated in the study.

Follow-up
Since 1991, information on all individuals with end-stage renal disease (defined as a glomerular filtration rate <10-15 mL/min) starting renal replacement therapy, i.e. dialysis or transplantation, have been recorded in the Swedish Renal Registry [SRR; (38)]. All healthcare facilities performing renal replacement therapy report to the SRR (attaining information on >95% of patients). By linkage to the SRR, participants in the two cohorts that later developed end-stage renal disease (up to 31 December 2006) were identified. Each case of end-stage renal disease was then matched with three controls who were alive at the time of end-stage renal disease diagnosis. Matching was performed by cohort, age, gender, and time of sampling. A total of 139 cases and 382 controls were identified, and Ery-Cd and Ery-Pb levels were determined for 118 cases and 378 controls, and Ery-Hg levels were measured 87 cases and 288 controls. Complete sets of cases and controls with Ery-Cd and Ery-Pb measurements were obtained for 118 cases and 347 controls, and
complete Ery-Hg measurements were obtained for 86 cases and 244 controls.

All subjects gave informed written consent for their participation in the study, which was approved by the Ethical Committee of Umeå University.

Paper III and IV

In the Chinese Province of Zhejiang, three areas with different levels of Cd pollution were studied: Jiaoweibao (highly polluted), Nanbaixiang (moderately polluted), and Yantou (no particular pollution). The study areas and their participants have been described previously (39-42). Due to contamination from a Cd refinery, the water used to irrigate farming areas in the exposed areas contaminated the locally grown rice. The rice was consumed as the main food source in the highly polluted area up until 1996 when they stopped consuming the contaminated rice and started to consume commercially available rice. Even though consumption of locally grown vegetables continued, the Cd exposure from the vegetables was minimal compared with the previous exposure from the rice. In the moderately polluted area, consumption of locally grown rice also stopped in the 1990s due to rapid economic development in the area. The area with no particular Cd pollution was situated 40 kilometers from the contaminated areas and had the same living habits and socioeconomic conditions as the polluted areas.

Kidney function in these areas was assessed in 1995, 1998, and 2006. In 1998 a total of 790 individuals were surveyed, of which 497 were available at follow-up in 2006. Of the individuals lost to follow up, 60 had died, 9 were unavailable due to work outside the study area, 57 had moved out of the area, and 167 declined to participate. After excluding individuals >80 years old, those with systolic blood pressure >170 mmHg, and those who were taking any physician-prescribed medication, a total of 412 individuals remained.

The study subjects gave their informed written consent to take part in the study. The study was approved by the Ethical Committee of Public Health School at Fudan University, Shanghai, China.
Chemical analyses

**Paper I**

The Pb and Cd concentrations were determined by inductively coupled plasma-mass spectrometry (ICP-MS; Thermo X7, Thermo Elemental, Winsford, UK) in samples diluted with an alkaline solution according to (43). The analytical accuracy was checked against reference materials. For blood and urine Seronorm Trace Elements samples from SERO AS, Billingstad, Norway were used, and for plasma outdated plasma from blood donors spiked with 0.50 μL Cd and 1.0 μL Pb.

**Paper II**

Ery-Cd and Ery-Pb levels were determined by inductively coupled plasma-mass spectrometry (ICP-MS; Thermo X7, Thermo Elemental, Winsford, UK; (43)). The analytical accuracy was checked against reference materials from SERO AS, Billingstad, Norway (Seronorm Trace elements) and the Centre de Toxicologie du Quebec, Canada (human whole blood). Ery-Hg was determined by cold vapor atomic fluorescence spectrophotometry (44), and reference materials from Centre de Toxicologie du Quebec, Canada, were used for quality control.

**Paper III and IV**

B-Cd and U-Cd levels were determined by graphite-furnace atomic absorption spectrometry (GF-AAS) using standard addition as described by Jin et al. (2002) (41). β₂-microglobulin and albumin were measured by enzyme-linked immunoabsorbent assay (ELISA). Urinary N-acetyl-β-D-glucosaminidase (NAG) was determined as previously described by Tucker et al. (1975) (45). To adjust the urine spot samples for dilution, creatinine levels were determined by the Jaffe reaction method (46). Extraction of genomic DNA was performed utilizing QIAamp blood DNA mini kits (QIAGEN, Hilden, Germany). Thereafter Taqman allelic discrimination assay (ABI 7900; Applied Biosystems, Foster City, CA, USA) was used to separately analyze each SNP.
**Statistical analyses**

**Paper I**

To estimate the fraction of the total variance attributed to variance between individuals, a one-way random effect model with the natural logarithm of Pb concentrations as response and individual as a random factor was used to estimate the fraction of the total variance attributable to variance between individuals. The model consisted of one within- and one between-individual variance component that were assumed to be normally distributed and mutually independent. This one-way random effect model was defined as:

\[
\ln(Y_{ij}) = \mu + b_i + \varepsilon_{ij}, \quad i = 1, 2, ..., k \\
\quad j = 1, 2, ..., n_i
\]

where \( Y_{ij} \) is the exposure biomarker level for the \( i \)th person on the \( j \)th measurement, \( \mu \) is the true fixed mean of the natural logarithm of exposures, \( b_i \) represents the random effect for person \( i \), and \( \varepsilon_{ij} \) represents the random deviation of the natural logarithm of exposures \( \ln(Y_{ij}) \) from \( \mu \), i.e. \( \varepsilon_{ij} = \ln(Y_{ij}) - \mu \), where \( \mu_i = \mu + b_i \). \( b_i \) represents the random deviation of the \( i \)th person’s mean of the natural logarithm of exposure from \( \mu \) such that \( b_i = \mu_i - \mu \). For individual random effects the model assumes that \( b_i \) and \( b_j \) are independent for \( i \neq j \) and \( \{b_i, i \in [1,k]\} \sim N(0, \sigma_b^2) \).

When compound symmetry can be assumed, \( \varepsilon_{ij} \) and \( \varepsilon_{ik} \) are independent for \( \{(i,j) \neq (k,l)\} \) and \( \varepsilon_{ij} \sim N(0, \sigma_w^2) \) \( \forall i = 1, 2, ..., k; j = 1, 2, ..., n_i \). The total variation of logged exposures is, therefore, \( \sigma^2 = \sigma_w^2 + \sigma_b^2 \). For this model the covariance between each pair of logged exposures for individual \( i \) is equal to \( \sigma_w^2 \). The intraclass correlation between \( Y_{ij} \) and \( Y_{ik} \), where the class is the natural logarithm of exposures within individual \( i \), is equal to \( \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} \). Modeling first order autocorrelated residuals \( \varepsilon_{ij} \), where
\[ Y_{i,j} = c + \rho Y_{i,j-1} + \varepsilon_{i,j} \]

the covariance between \( Y_{ij} \) and \( Y_{i,k} \), where \( j \neq k \), was modeled as

\[ \sigma^2_b + \sigma^2_e \rho^{|j-k|} / (1 - \rho^2) \]

where \(|j-k|\) is the distance between within-individual measurements.

Model fits were performed using restricted maximum likelihood. Confidence intervals for the variance components and the amounts of the total variance attributable to differences between individuals were estimated using bootstrap samples. Mixed effects models were used to estimate the association between each pair of biomarkers and for estimating the associations between Pb concentrations and kidney effect markers. Statistical analyses were performed using the statistical package R, version 2.9.0 (47).

A revision of a previously established model of the relation between whole-blood and plasma Pb concentrations was used to investigate the partition of B-Pb between cells and plasma (48). In this model only one component for Pb binding in erythrocytes was used, where Pb binding constants were estimated from whole-blood and plasma Pb concentrations. A constant allowing for a nonzero intercept was also estimated from the whole-blood and plasma Pb concentrations and was added to the model. The relationship between whole-blood and plasma Pb concentrations is given by

\[
B_{Pb} = (1 - Ht)(P_{Pb} + c) + Ht \cdot (P_{Pb} + c) \left( 1 + \frac{BIND}{K_{bind} + (P_{Pb} + c)} \right),
\]

where \( B_{Pb} \) is the Pb concentration in whole blood, \( Ht \) is the hematocrit, \( P_{Pb} \) is the Pb concentration in plasma, \( BIND \) is the maximum binding capacity for the Pb-binding component, \( K_{bind} \) is the dissociation constant for the Pb binding of the component and \( c \) is a constant allowing for a nonzero intercept with the P-Pb axis.

Based on this model we calculated the ratio between measured P-Pb and predicted P-Pb from measured B-Pb:

\[
ratio = \frac{P_{Pb}}{P_{Pb}^*},
\]
where \( P_{Pb} \) is the measured P-Pb and \( P_{Pb}^* \) is the predicted P-Pb corresponding to measured B-Pb.

A ratio of 1 corresponds to a whole-blood sample in which Pb is distributed between cells and plasma according to the model above. A ratio larger than 1 indicates a larger proportion of Pb in plasma and a ratio less than 1 indicates a larger proportion of Pb in cells. Numerical methods were implemented using MATLAB 7.5 (The MathWorks Inc., Natick, MA, 2007, www.mathworks.com).

**Adjustments for U-Pb and U-Cd**

1. Using the sample densities \( x = (x_1, x_2, \ldots, x_n) \), density adjustment of U-Pb was performed by:

\[
\frac{\bar{x}_{g/ml}}{x_i \cdot g/ml} \cdot \frac{u_i \cdot \mu g/L}{L} = \frac{\bar{x} - 1}{x_i - 1} \cdot u_i \cdot \mu g \cdot Pb/L, \quad i = 1, 2, \ldots, n,
\]

where \( u_i \cdot \mu g/L \) is the Pb concentration for sample \( i \).

2. The amount of U-Pb per gram creatinine was calculated according to:

\[
\frac{u_i \cdot \mu g/L}{crea \cdot mmol/L} = \frac{u_i \cdot \mu g/L}{113.126 \cdot g/mol \cdot crea \cdot mmol/L} = \frac{1000 \cdot crea \cdot u_i \cdot \mu g \cdot Pb/g}{creatinine},
\]

where \( crea \cdot mmol/L \) is the creatinine content for sample \( i \).

3. Sampling time for U-Pb was also considered by calculating the amount of Pb excreted per hour:

\[
\frac{u_i \cdot \mu g/L \cdot volume \cdot mL}{time \cdot h \cdot 1000} = u_i \cdot \mu g \cdot Pb/h
\]

The individuals provided morning urine and the time since the individuals last gave urine was recorded.
Adjustment methods 1-3 were also used for adjustment of U-Cd.

4. Adjustment of urinary dilution was performed utilizing a log-linear model:

   \[
   \ln(U_{pb_i}) = \beta_0 + \beta_1 \ln(UF_i) + \epsilon_i, 
   \]

   where \( U_{pb_i} \) is the Pb concentration in urine, \( UF_i \) is the urinary flow and \( \epsilon_i \in N(0,\sigma^2) \) is the error term for sample \( i \). Urinary flow was calculated as mL/h. This can be re-expressed as

   \[
   U_{pb_i} \cdot UF_i^{\hat{\beta}} = e^{\hat{\epsilon}_i}. 
   \]

   Using the estimated regression coefficients \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \) from (*) and letting the urinary flow equal one unit (i.e. \( UF_i = 1 \)) we obtain:

   \[
   \ln(U_{pb_i,UF=1}) = \hat{\beta}_0 + \hat{\beta}_1 \ln(1) + \epsilon_i = \hat{\beta}_0 + \epsilon_i 
   \]

   which can be re-written as \( U_{pb_i,UF=1} = e^{\hat{\epsilon}_i} \). Thereafter we solved

   \[
   \begin{cases} 
   U_{pb_i} \cdot UF_i^{\hat{\beta}} = e^{\hat{\epsilon}_i} \\
   U_{pb_i,UF=1} = e^{\hat{\epsilon}_i} 
   \end{cases} 
   \]

   for \( U_{pb_i,UF=1} \), and thus obtaining U-Pb at a urinary flow equal to one unit, \( UF_i = 1 \), that for each sample can be calculated as

   \[
   U_{pb_i,UF=1} = U_{pb_i} \cdot UF_i^{\hat{\beta}}. 
   \]
The association between end-stage renal disease and each of the potential risk factors Ery-Cd, Ery-Pb, Ery-Hg, BMI, diabetes, smoking, and hypertension, was calculated using conditional logistic regression based on the matched sets of cases and controls. The binary disease outcome was linked to a linear function of candidate risk factors:

\[
\logit(E[Y_i|\mathbf{X}_i]) = \mathbf{\beta} \mathbf{X}_i
\]

where \( Y_i \) is the disease outcome, \( \mathbf{X}_i \) is the vector of risk factors, and \( \mathbf{\beta} \) is the vector of regression parameters. The regression parameters were estimated by minimizing the likelihood function

\[
L = \prod_{i=1}^{k} \left( \frac{\prod_{j=1}^{n_i} \exp(\mathbf{\beta}^T \mathbf{x}_j)}{\sum_{j=1}^{n_i} \exp(\mathbf{\beta}^T \mathbf{x}_j)} \right)
\]

where \( k \) is the number of strata of matched case control sets, \( n_{i_1} \) is the number of cases in stratum \( i \), and \( n_{o_i} \) the number of controls in stratum \( i \), \( l_i \) is the set \( \left\{ \frac{n_{i_1} + n_{o_i}}{n_{i_1}} \right\} \) of possible choices of selecting \( n_{i_1} \) integers from the set \( \{1,2,\ldots,n_{i_1} + n_{o_i}\} \) (49).

Univariate association was estimated for each of the candidate risk factors. In addition to univariate analysis, Spearman’s correlation coefficient or odds ratio (OR) was calculated between each pair of candidate risk factors to be able to determine if each risk factor was a confounder, changed the effects of the metals, or was an independent risk factor. Finally, a multiple conditional logistic regression model was used to simultaneously estimate the effect of each candidate risk factor when also taking into consideration all other candidate risk factors changing the association between the metals and end-stage renal disease or independently predicting the risk of end-stage renal disease. A variable was considered to change the effect of Ery-Cd, Ery-Pb, or Ery-Hg if the regression coefficient for any of the metals changed by more than 10%.
Assessment of model assumptions were conducted using penalized cubic splines and by categorization of Ery-Cd, Ery-Pb and Ery-Hg by quartile limits. Sensitivity analyses were performed by excluding influential observations.

The estimates of the associations between end-stage renal disease and metal exposure were divided into subsets of time defined as quartile limits to determine if the time between sampling and end-stage renal disease was a confounding variable.

All statistical analyses were performed using the statistical package R, version 2.11.0 (47). p-values <0.05 were considered statistically significant.

Paper III

Impaired kidney function was defined as an individual having a kidney function marker (NAG, β2-microglobulin or albumin) above the 95th percentile of the same kidney function marker in residents from the non-exposed area. Because the age distribution in the non-polluted area differed from the age distributions in the polluted areas, the dataset containing all individuals was stratified into four age groups and weights for each age group in the non-polluted area were calculated as the total number of individuals in each age group divided by the number of individuals in the non-polluted area. Thereafter a dataset where individuals within each age group are used ten times the corresponding weight in that age group (the factor ten was used to increase the accuracy when rounding off the weights to integers) was constructed. In this way the number of individuals in every age group was the same in the non-polluted area and the polluted areas. The dataset was then used to construct the cut-off values defined as the 95th percentile for the kidney markers. Also, since the kidney function markers are related to age, different cut points were defined at baseline and follow up.

Tests between kidney marker and Cd biomarker levels between baseline and follow up were performed using Wilcoxon signed rank test. To evaluate the prevalence of impaired kidney function in relation to Cd biomarkers, the prevalence were calculated within exposure biomarker groups (0-2, 2-5, 5-10, 10-20 and >20 μg Pb/g creatinine). Log-linear models were used to model the association between prevalence of impaired kidney function and exposure groups (U-Cd in 1998). Pairwise
comparisons were made between the prevalence of kidney damage in 1998 and 2006 in each exposure group using paired t-tests. p-values <0.05 were considered statistically significant. All statistical analyses were performed using R version 2.9.0 (47).

**Paper IV**

For each SNP, deviations from Hardy-Weinberg equilibrium were tested using the Fisher-Freeman-Halton test. Differences in exposure biomarkers and kidney markers levels were estimated using univariate linear regression.

The association between each kidney marker and each exposure biomarker's interaction with the genetic variation within each SNP was modeled using a multiple log-linear regression model, and estimates were presented as the relative increase for a relative increase in exposure biomarker. The Shapiro-Wilk test was used to evaluate the assumption of normally distributed residuals. If deviations from normality were found, sensitivity analyses were performed by excluding outliers. Sensitivity analyses were also performed by excluding highly influential observations identified using Cook's distance. To achieve robust estimates, B-Cd measurements less than or equal to 0.01 μg/L were excluded in the 1998 cross-sectional analysis (n = 3 in the 1998 data).
Change in kidney marker levels between baseline and follow-up were modeled using robust multiple linear regression that was fitted using the MM-estimator. The change in kidney marker was evaluated by the difference in kidney marker levels between baseline and follow up:

\[ y_{it} - y_{ot} = \beta_0 + \beta_1 y_{oi} + \beta_2 \ln(x_{oi}) + \beta_3 \text{genotype}_{i1} + \beta_4 \text{genotype}_{i2} + \beta_5 \text{smoking}_{i} \ln(x_{oi}) + \beta_6 \text{gender}_{i} \ln(x_{oi}) + \beta_7 \text{age}_{i} + \varepsilon_i \]

where

\[ \varepsilon_i \in N(0, \sigma^2) \]

\( y_{oi} \) = kidney marker at baseline
\( y_{ot} \) = kidney marker at follow up
\( x_{oi} \) = exposure biomarker at baseline
\( \text{genotype}_{i1} \) = dummy variable for the heterozygote genotype
\( \text{genotype}_{i2} \) = dummy variable for the variant homozygote genotype

Estimates were given as the change in kidney marker level for a relative change in exposure biomarker level, and confidence intervals were estimated using bias-corrected non-parametric bootstrap samples (based on 5,000 bootstrapped samples).

Age, sex and smoking were considered as potential confounders in all multiple linear models and were included as covariates (modeling levels/change of kidney effect markers). A backwards stepwise procedure based in the Akaike information criterion was used for variable selection, where covariates that when excluded in the procedure changed the exposure biomarker parameters by more than 10% remained in the multiple model. To investigate possible multicollinearity Spearman’s correlation coefficient was calculated between each pair of explanatory variables.

Linearity between each kidney effect marker/change in each kidney effect marker and the exposure biomarkers was assessed using generalized additive models with penalized cubic splines, where the generalized cross-validation score were used as criteria for determining the amount of smoothing and the effective degrees of freedom.
All analyses were performed using R version 2.14.0 (47), and p-values less than 0.05 were considered statistically significant.
Results

Paper I

Geometric means of Pb concentrations among lead workers were 227 μg/L in whole blood, 0.57 μg/L in plasma and 23.7 μg/L in urine. The corresponding concentrations among normally exposed individuals were 20.7 μg/L, 0.09 μg/L, and 10.8 μg/L, respectively. Plots of Pb concentrations in whole blood, plasma, and urine for each individual with occupational exposure showed that all the biomarkers could fairly well discriminate between individuals with different mean levels (Figure 1a-c). For the individuals with normal environmental exposure, P-Pb was however too imprecise to be able to discriminate between individuals with different mean concentrations (Figure 1d-f).

Using one-way random effects modeling with individuals as the random effect, the estimated fraction of the total variance attributed to between individual variance among Pb workers was 91% for B-Pb, 78% P-Pb, and 82% for density adjusted U-Pb. After incorporating a first order autocorrelation structure, the results were 90% and 75% for B-Pb and P-Pb, respectively. The corresponding result for the individuals with normal lead exposure were 95% for B-Pb, 15% for P-Pb and 87% for creatinine-adjusted U-Pb. Adjustment for urinary dilution had only a marginal effect on the estimated fraction of the total variance attributed to between-individual variance among Pb workers, but it had a considerable effect among individuals with normal environmental exposure.

The relation between Pb concentrations in whole blood and plasma was modeled based on a single Pb dissociation constant ($K_{\text{bind}}$) of 0.8 μg/L and a binding capacity (BIND) of 1240 μg/L (Figure 2). Deviation from the model was evaluated using the ratio between observed P-Pb concentrations and model-estimated P-Pb (Figure 3). The fraction of the total variance attributed to between-individual variance for the ratios was 42%.
Figure 1. Pb concentrations in whole blood (a, d), plasma (b, e) and urine (c, f) from 48 occupationally exposed individuals collected during a 2-year period (a-c) and from 20 individuals not under occupational exposure (d-f) collected over one year. Each data point represents one sample. For comparability each vertical axis is scaled to the mean of the data points in the plot. Plot d is illustrative of a situation where the between-individual variance is large compared to the variance within individuals (‘day-to-day’ variation) and is indicative of a biomarker with the ability to discriminate between individuals with high and low levels. It is interesting to note that for the Pb-workers, U-Pb has a variance within individuals that is larger than that for whole blood (a), but at the same time the variance between individuals is also larger. (From paper I)
Figure 2. Whole-blood lead concentrations (B-Pb) versus plasma lead concentrations (P-Pb) together with the model describing the relationship between Pb concentration in whole blood and plasma. The model uses one Pb-binding component with a Pb binding capacity and dissociation constant estimated from the data for the individuals with occupational lead exposure. (From paper I)

Figure 3. Ratios of the measured P-Pb and predicted P-Pb among Pb workers as modeled from measured B-Pb using the model with one Pb-binding component. (From paper I)
Paper II

The geometric mean of Ery-Cd was 0.86 μg/L among cases and 0.66 μg/L among controls, and the geometric mean was 23% lower among controls with a 95% CI of 8.2–44.7%. The geometric mean of Ery-Pb was 66.2 μg/L among cases and 55.0 μg/L among controls, and the geometric mean was 17% lower among controls with a 95% CI of 7.8–29%. The arithmetic mean of Ery-Hg was 2.44 μg/L among cases and 3.06 μg/L among controls, and the difference between the two groups was 0.62 μg/L with a 95% CI of 0.17–1.08 μg/L.

Univariate analysis showed that the risk of developing end-stage renal disease was associated with Ery-Pb (OR = 1.010 per μg/L increase with a 95% CI of 1.004-1.016) and Ery-Hg (OR = 0.86 per μg/L increase with a 95% CI of 0.74-0.996). The association was however not statistically significant for Ery-Cd (OR = 1.17 per μg/L increase with a 95% CI of 0.99-1.38). Using multiple conditional logistic regression taking into consideration all other statistically significant covariates or covariates that changed the effect of the metals, both Ery-Pb and Ery-Hg were statistically significantly related to the development of end-stage renal disease (Table 1). Ery-Cd was not statistically significant in the multiple model (OR = 1.11; 95% CI 0.78-1.57; not shown in table).

To investigate the shape of the associations, analyses were also performed by categorizing Ery-Cd, Ery-Pb, and Ery-Hg by quartile limits among cases (Figure 4a-c). More elaborate analyses were performed on the shape of the associations using penalized cubic splines (Figure 5). Time between sampling and end-stage renal disease as a possible confounder was assess in subsets of time between sampling and end-stage renal disease diagnosis. (Figure 6).

Gender-stratified analyses, taking into consideration all statistically significant covariates and covariates that change the metal estimates, indicated a stronger dose-response relation among men than among women for Ery-Cd (men: OR = 1.39, 95% CI = 0.96–2.00, N = 151; women: OR = 0.79, 95% CI = 0.51–1.23, N = 93) and Ery-Pb (men: OR = 1.014, 95% CI = 1.0023–1.026, N = 151; women: OR = 1.00, 95% CI = 0.98–1.022, N = 93). A negative association was found for Ery-Hg among women that was not apparent among men (men: OR = 0.89, 95% CI = 0.71–1.13, N = 151; women: OR = 0.68, 95% CI = 0.47–0.98, N = 93).
Table 1. Multiple modeling of the associations between end-stage renal disease and erythrocyte concentrations of Pb and Hg, taking into consideration the covariates at baseline that either were statistically significant or that changed any of the metal estimates. (From paper II)

<table>
<thead>
<tr>
<th>Multiple model</th>
<th>OR</th>
<th>95% CI</th>
<th>IQR</th>
<th>ORIQR 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of individuals = 244</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ery-Pb (µg/L)</td>
<td>1.013</td>
<td>(1.003, 1.023)</td>
<td>0.37-1.3</td>
<td>1.74 (1.13, 2.68)</td>
</tr>
<tr>
<td>Ery-Hg (µg/L)</td>
<td>0.81</td>
<td>(0.66, 0.99)</td>
<td>39.8-83.5</td>
<td>0.66 (0.44, 0.98)</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>1.13</td>
<td>(1.02, 1.25)</td>
<td>1.53-3.5</td>
<td>1.75 (1.10, 2.79)</td>
</tr>
<tr>
<td>Diabetes (y/n)</td>
<td>35.8</td>
<td>(4.37, 294)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertension* (y/n)</td>
<td>4.23</td>
<td>(1.97, 9.11)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OR = odds ratio for a unit increase; CI = confidence interval; IQR = interquartile range; ORIQR = odds ratio for an interquartile range increase; BMI = Body Mass Index.

*Hypertension was defined as systolic blood pressure >140 or diastolic blood pressure >90 mm Hg, or as taking antihypertensive medication at baseline.
Figure 4. Univariate association between erythrocyte concentrations of cadmium (Ery-Cd) (a), lead (Ery-Pb) (b), and mercury (Ery-Hg) (c) and the risk of developing end-stage renal disease. Quartile concentration limits were calculated from cases. (From paper II)
Figure 5. Conditional logistic regression using penalized cubic splines was used to investigate linearity between log-odds and exposure variables. The solid line is the estimated association and the dashed lines represent the corresponding confidence interval. Cases and controls are indicated on the metal axis. Log-odds curves are for 1 μg/L increases. (From paper II)
Figure 6. Effect modification of end-stage renal disease by the time between sampling and diagnosis as determined by the odds ratio (OR) for a 1 µg/L increase in Ery-Cd (a), Ery-Pb (b), and Ery-Hg (c). The estimates are shown as ORs first for all data and then for subsets of cases with time between sampling and end-stage renal disease above the quartile limits. The ORs for Ery-Pb and Ery-Hg are adjusted for the other metal (i.e., Ery-Hg and Ery-Pb, respectively), BMI, and blood pressure. Ery-Cd was adjusted for the other two metals, BMI, blood pressure, and smoking. Due to the limited number of cases within each subset, diabetes was not included as a covariate in these analyses. (From paper II)
Paper III

At baseline, urinary albumin, $\beta_2$-microglobulin, and NAG levels were significantly higher in residents of the highly polluted area compared with residents of the control area. Between baseline and follow-up, B-Cd decreased among individuals in all three areas whereas U-Cd decreased only in the highly polluted area. NAG and $\beta_2$-microglobulin levels increased in all areas ($p<0.01$), whereas albumin levels had a statistically significant decline in the highly polluted area ($p<0.01$).

The predictivity of kidney marker levels, on an individual basis, was evaluated by calculating the probability of having an elevated level (above the 95th percentile in the control area) at follow-up given that the marker was also elevated at baseline. For $\beta_2$-microglobulin, the OR of an elevated level at follow-up was 24.8 with a 95% of CI 11.2-55.3, comparing normal with elevated level at baseline. Much weaker predictivity was found for NAG (OR = 2.6, 95% CI = 1.6-4.4) and albumin (OR 3.0 95% CI 1.2- 7.5).

Comparison of the prevalence curves for kidney dysfunction at baseline and follow-up in relation to U-Cd at baseline showed that elevated albumin levels at follow-up were not related to U-Cd at baseline despite a clear association at baseline. No such associations were found for NAG or $\beta_2$-microglobulin.
Between baseline and follow-up, NAG and \( \beta_2 \)-microglobulin increased statistically significantly with geometric means of 5.05 U/g creatinine and 0.11 mg/g creatinine, respectively. In contrast, albumin decreased statistically significantly with a geometric mean of 1.38 mg/g creatinine. The exposure biomarkers B-Cd and U-Cd both decreased with geometric means of 2.28 \( \mu \)g/L and 0.24 \( \mu \)g/g creatinine, respectively. Only the decrease in B-Cd was, however, statistically significant.

Statistically significantly higher B-Cd were found among \( MT1A \) rs11076161 AA individuals compared with GG individuals both at baseline and follow-up (the differences in geometric means were 2.45 \( \mu \)g/L at baseline and 1.15 \( \mu \)g/L at follow up). At follow-up, higher levels of NAG, \( \beta_2 \)-microglobulin, and albumin were also found among AA individuals compared with GG individuals (2.63 U/g creatinine, 0.17 mg/g creatinine, and 1.8 mg/g creatinine, respectively).

Modification of Cd-related kidney dysfunction, and changes in kidney function, by MT polymorphisms were evaluated. Firstly, modification of cross-sectional associations between each kidney marker and exposure biomarker by the MT polymorphisms were estimated both at baseline and at follow-up. Secondly, modification of the association between changes in kidney function, between baseline and follow up, and exposure biomarkers were evaluated. Figure 7 shows the associations between kidney markers and biomarkers of cadmium exposure for each SNP that showed a statistically significant genetic and exposure biomarker interaction.

At baseline, \( MT2A \) rs28366003 was found to modify the U-Cd-associated effect on \( \beta_2 \)-microglobulin, where individuals with the GG genotype increased by a factor of 2.56 when U-Cd increased to the two-fold, compared with 1.30 for AA (\( p = 0.030 \)). The \( MT2A \) rs28366003 polymorphism also modified the association between NAG and B-Cd, NAG increased by a factor of 1.65 for individuals with the AG genotype when B-Cd increased to the two fold, compared with 1.30 for AA (\( p = 0.026 \)). Also, the B-Cd associated effect on \( \beta_2 \)-microglobulin was modified by \( MT1A \) rs11076161, individuals with the AG genotype increased by a factor of 1.45 compared with 1.21 among GG when B-Cd increased to the two-fold (\( p = 0.021 \)).
At follow-up, the U-Cd-associated effect on β₂-microglobulin was found to be modified by the MT1A rs11076161 polymorphism, β₂-microglobulin levels increased by a factor of 1.45 among individuals with the AG genotype and by a factor of 1.65 among those with the AA genotype when U-Cd increased to the two-fold, compared with 1.22 among those with GG genotype (p = 0.038 and 0.018, respectively). Also, because B-Cd mainly reflects current exposure, B-Cd at follow-up was not able to discriminate between individuals who had different exposure biomarker levels at baseline. Therefore, the kidney function markers at follow-up were also evaluated in relation to B-Cd at baseline. The baseline B-Cd-associated effect on β₂-microglobulin was found to be modified by MT1A rs11076161, individuals with the AG genotype β₂-microglobulin increased by a factor 1.42 compared with 1.26 among GG individuals when B-Cd increased to the two-fold.

Comparing the Cd-associated effects on NAG, β₂-microglobulin and albumin at baseline and follow up, confidence intervals for the estimated associations were overlapping and therefore effects at baseline and follow up cannot be concluded statistically significant. Modeling the difference in kidney effect markers no statistically significant modification of the Cd-associated effect was found for any of SNPs.

The reversibility of the Cd-associated effects on kidney function was also assessed without considering the effect of genetic interactions. Standardized regression coefficients for the association between kidney markers at baseline and U-Cd at baseline were 0.37 (95% CI = 0.28-0.47) for NAG, 0.39 (95% CI = 0.30-0.48) for β₂-microglobulin, and 0.18 (95% CI = 0.08-0.28) for albumin, compared with 0.22 (95% CI = 0.12-0.31), 0.31 (95% CI = 0.22-0.41), and 0.10 (95% CI = 0.00-0.19) for the associations between kidney markers at follow-up and U-Cd at baseline, respectively. Confidence intervals for baseline and follow-up estimates were all overlapping and therefore the Cd-associated effects were not concluded statistically significantly different.

On an individual basis, the predictivity of the kidney marker levels at follow-up, given the levels at baseline, was higher for β₂-microglobulin than both NAG and albumin. Taking into consideration age, gender, and smoking, the level of NAG at follow-up increased by 0.0064 (95% CI = 0.0019–0.011) U/g creatinine for each U/g creatinine increase at baseline, the level of β₂-microglobulin increased by 0.58 (95% CI = 0.47–0.69) mg/g creatinine, and the level of albumin increased by 0.015 (95% CI = 0.008–0.021) mg/g creatinine.
CI = 0.0047–0.026) mg/g creatinine for each mg/g creatinine increase at baseline (standardized coefficients were 0.014 for NAG, 0.60 for β2-microglobulin, and 0.014 for albumin).
Figure 7. Associations between the kidney effect markers and Cd exposure biomarkers that were found to be statistically significantly modified by polymorphisms in MT1A and MT2A. (From paper IV)
Additional results

Comparing different measures of change

The change in kidney function was measured in paper III as the change in a binary variable and in paper IV as the difference between levels at baseline and follow-up. The change in binary levels measured only transitions across the cut-off value (defined as the 95th percentile of levels in the control area), giving the event space of no damage, continuous damage, recovery and impairment. In contrast, paper IV utilized continuous scales. The relation between the two measures is shown in Figure 8a-c. These show that the transition outcomes were able to discriminate between differences in kidney markers with minimal overlap. Individuals that were classified as having a kidney recovery also had the largest decline in kidney marker levels, and individuals with impaired kidney function had the largest increase. Thus, transitions above or below the cut-off value were not due to minor changes in kidney marker levels.

Variance components for Cd and kidney effect markers

For B-Cd 93% of the total variance could be attributed to differences between individuals (within- and between-variance components were 0.072 and 0.90 respectively). Almost the same fraction was found for Cd in creatinine-adjusted urine (89% with within- and between-individual variance components were 0.086 and 0.67, respectively). For P-Cd, the result was 61% (within- and between-variance components were 0.080 and 0.13, respectively).
Figure 8. Changes in the kidney function markers NAG (a), β₂-microglobulin (b), and albumin (c) between baseline and follow-up in relation to U-Cd.
Discussion

Study design and data analysis

Longitudinal study design

A longitudinal study is an observational study in which repeated measurements are taken over time. In typically epidemiological studies repeated measurements are recorded on several individuals. This type of study design has been used in papers I, III and IV. In paper I repeated measurements were taken on Pb concentrations in whole blood, plasma and urine for comparison of their ability to distinguish between individuals with different mean concentrations. By estimating the within- and between-individual variance components, we found that whole blood was the superior Pb biomarker with respect to the fraction of the total variance attributed to between individual variance. As the joint measure of within- and between-individual variance the fraction of the total variance attributed to between individual variance was chosen, which is the definition of the intra-class correlation in a model assuming no autocorrelation between individual measurements (50). Another suggested measure is the ratio between the within- and between-individual variance components (51), however for the purpose of comparing different biomarkers, the boundedness of the fraction of the total variance to the interval 0 to 1 may be more easily interpretable than the upper unboundedness of the ratio between the within- and between-individual variance components taking values between 0 and infinity.

In most situations it is reasonable to model each individual response as a function of some covariates (fixed effects), a between individual random effect that capture between-individual effects that we do not aim to estimate, and a residual term incorporating measurement errors and unobserved covariates. Because there is error in the measurements of the metal biomarkers, modeling dose-response relations in this type of typical model the relation will be attenuated, i.e., the regression coefficient will diminish towards the null. The attenuation of the regression coefficient in this type of linear model is \( 1 + \frac{\lambda}{n} \), where

\[ \lambda = \frac{\sigma_w^2}{\sigma_b^2} \]. If estimates of the within- and between-individual variance
components are available, we can estimate \( \lambda \) as \( \hat{\lambda} = \frac{\hat{\sigma}_w^2}{\hat{\sigma}_b^2} \) and the regression parameters will be reduced in expectation by the factor \( \left(1 + \frac{\hat{\lambda}}{n}\right) \). Rappaport et al. (1995), and several other studies, have compared the attenuation using environmental monitoring with the attenuation using biological monitoring (51-55). In a balanced repeated measurement design, the number of repeated measurements needed to keep attenuation below a certain percentage can be calculated (51, 52, 54-57). By incorporating fixed effects, mixed models have been used to compare variations within and between individuals with different jobs (58). Using mixed effects modeling of repeated measurement of exposure may increase the validity of the estimate but reduce precision for fixed effects estimates compared with single measurements (59). In this work (paper 1) one-way random effect modeling was employed to estimate the amount of the total variance attributed to between-individual variance without considering any fixed effects. Including covariates may affect variation within and between individuals and reduce the attenuation of the fixed effect that we aim to estimate (60). Smith et al. (2002) compared the intraclass correlation for Pb in blood and plasma and found that, at least for plasma, considering age, season of sample collection, and blood hemolysis at sampling as fixed effects did not affect the within-individual variation of P-Pb (4).

We can also estimate the change in outcome for a given change in a predictor variable that is subject to some uncertainty, where we are interested in the unattenuated association between the change in outcome and the true predictor variable. The uncertainty may be day-to-day within-individual variation (61, 62) or the predictor variable may be a random sample from a population. Consider the case of simple linear regression modeling where both the predictor variable and the random error are normally distributed, the outcome \( y \) and predictor \( x \) are modeled as

\[
y_i = \alpha^* + \beta^* x_i + \epsilon_i, \quad \epsilon_i \in \mathcal{N}(0, \sigma^2) \\
\]

\[
w_i = x_i + u_i, \quad x_i \in \mathcal{N}(0, \sigma^2) \quad \text{and} \quad u_i \in \mathcal{N}(0, \sigma^2) \\
\]

\[
36
\]
Then, if \( \varepsilon_i, x_i \) and \( u_i \) are independently distributed we know that the relation between \( y_i \) and \( u_i \) can be modeled as

\[
y_i = \alpha + \beta u_i + \delta_i, \quad \delta_i \sim N(0, \phi^2)
\]

where \( \beta \) is given by

\[
\beta = \beta^* \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}
\]

Under this model, the uncertainty in the predictor variable \( x \) will reduce the regression coefficient in expectation by the factor \( \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} \) (63).

Modeling longitudinal data often requires incorporation of some autocorrelation structure because there is usually a (positive) correlation between repeated measurements. Assuming compound symmetry for a sequence of autocorrelated data results in biased estimates of the standard error and, therefore, leads to invalid inference about the data. Therefore, the statistically significant autocorrelation for blood and plasma Pb levels was also modeled with a first order autocorrelation structure, but the estimated variance components were almost identical. Because incorporation of a first order autocorrelation structure gives different correlation estimates for different lag times, and because adding this autocorrelation structure did not significantly change the estimates for blood and plasma, the estimates assuming compound symmetry were used for comparison with U-Pb levels.

In this thesis, data from a follow-up study were modeled with changes in kidney marker outcomes in relation to baseline levels of Cd exposure biomarkers. Paper III used dichotomization of kidney function, obtaining a binary outcome for kidney function with a cut-off value defined as the 95th percentile of the levels found in normally exposed individuals. Thereafter the change in prevalence of individuals above this cut-off between baseline and follow-up was studied in relation to baseline U-Cd. This dichotomization may give more easily illustrated data, but it substantially reduces information about variability in kidney function. Therefore, when studying interaction effects in paper IV, the differences in continuous kidney marker levels were used. Other possible measures of
change are, for instance, relative change and the ratio between measurements taken at baseline and at follow-up. Using a dichotomized approach utilizing a cut-off to evaluate change in kidney function, individuals with kidney marker levels close to the cut-off value can with a small change be re-categorized compared with individuals with levels far from the cut-off. Also, there is no generally accepted method for the choice of cut-off value. The arbitrary method calls for sensitivity analyses. Utilizing sensitivity analyses we also need to take into consideration that we make multiple testing, at least if these analyses results in choosing a different cut-off.

In our study of the change in kidney function markers between baseline and follow up, individuals with higher levels at baseline had a greater probability of decline compared with individuals with low values at baseline. This could partially be due to the regression to the mean phenomenon where individuals with a high level at baseline are more likely to have a decreased level at follow-up compared to the others irrespective of changes in any predictor variables (64-67). It is imperative, therefore, to consider the association between change and the level at the previous time point. This was however further complicated by the association between baseline kidney marker levels and U-Cd levels, because including baseline levels of the kidney marker may explain a U-Cd-associated effect. The choice of including the baseline kidney marker level as an independent variable is not trivial (68, 69) because a causal U-Cd-associated effect may be biased both by inclusion of the baseline kidney marker levels due to overadjustment and by omitting the baseline kidney marker levels due to confounding. Therefore, the analyses of the interaction of MT polymorphisms with the Cd biomarker's effect on changes in kidney markers were also assessed without considering the baseline kidney marker levels. The results of this analysis suggested the same interactions as found previously (data not shown).

Considering baseline kidney marker levels as an independent variable in the regression analysis is also important due to the properties of the measure of chance. The difference between kidney markers at baseline and follow-up give large weight to individuals with high levels because they also generally have a larger absolute change, whereas the ratio between baseline and follow-up give large weight to individuals with low levels since they generally have a larger ratio between follow up and baseline levels. There is also a boundedness issue for the measure of kidney function since this measure is non-negative, and therefore
individuals with a low level at baseline cannot decrease in the same extent as those with high levels at baseline. We would prefer a function with known distribution \( f \) of the baseline level \( y_0 \) and follow up \( y_1 \) that minimizes the correlation between \( f(y_0, y_1) \) and \( y_0 \), however minimizing the association between this function and the kidney marker level at baseline was not a possible method in this case because the risk factor of interest was associated with baseline kidney marker levels.

**Robust estimates**

The normal distribution of change in kidney markers contained many outliers and the linear model between change in kidney markers and exposure biomarkers contained several over-influential observations. Therefore robust estimation methods were used for estimation of the associations between change in kidney markers and biomarkers of exposure.

Even though the mean is the optimal estimator of the location parameter of a normal distribution, it is not the optimal estimator in the presence of outliers (70). These outliers can be excluded, for instance if they are considered false, or they can be down-weighted using a robust estimator. Rejecting outliers that are correct observations results in underestimated variance for the normal distribution and leads to false inference.

To increase the robustness in the baseline cross-sectional analysis of the association between kidney markers and B-Cd in paper IV, three influential observations with deviated values of the independent variable (close to zero) were excluded to increase the robustness of the estimate.

**Matched case-control design**

The case-control study design is more efficient than cohort studies when evaluating the association between rare diseases such as end-stage renal disease and possible risk factors, and it is appropriate for studying slowly progressing diseases. Such a study design can also allow for a quick and cost-effective study (71). Because the study design assumes that the probability of being included in the study is independent of the risk factors, the design is sensitive to selection bias. Compared to a longitudinal study, another possible source of bias may arise in paper II due to the lack of measurements of kidney function at baseline because the individuals may have already had reduced kidney function before
entering the study. If the disease causes increased levels of the exposure biomarker, the regression coefficient will be overestimated due to reversed causality. The matched case-control design can also be inefficient when estimating the association between disease and common exposures if there is not a particularly large attributable risk percentage from these exposures. Efficiency increases when matching variables are closely associated with the outcome and only weakly associated with the risk factors. Efficiency decreases, however, if the matching variables are closely associated with the risk factors. In paper II the risk of end-stage renal disease was found to be closely correlated with age (increasing efficiency), but age also correlated with metal concentrations (reducing efficiency). The situation where matching factors are closely associated with the risk factors is sometimes referred to as over-matching because cases and controls have been matched to such an extent that they have the same level of the risk factors that we aim to estimate (72).

Compared to logistic regression, we have less ability to evaluate model fits using conditional logistic regression. However, some goodness-of-fit methods do exist (73-75). In the case where the models give similar results, the Hosmer-Lemeshow goodness-of-fit test can be performed with matching variables included as covariates in the logistic regression model. Estimation by unconditional maximum likelihood is suitable when the number of parameters is small compared to the number of individuals, but for a large number of parameters compared to the number of individuals the conditional maximum likelihood method is preferred. Therefore, using logistic regression for estimation of associations based on m matched strata of cases and controls is not suitable because the model acquires m-1 parameters to account for the matching as well as additional parameters for risk factors and covariates. Ignoring the matching in the statistical analysis results in biased estimates (76), where the bias is towards the null (77).

An advantage of using matching as a method to eliminate or reduce confounding by for example age, being a strong risk factor for end-stage renal disease, is that matching compared to including age as a covariate does not assume a functional form. Matching also reduces possible loss of efficiency when the empirically measured age distribution deviates between cases and controls. Matching should only be performed with respect to risk factors that confound the association but do not need to be estimated (78).
Shape of dose-response

When the risk factors are on continuous scales and there is a linear relationship with the outcome (e.g., log odds of disease) we gain more information than if we would perform categorization, for instance by ordered categories based on quartile limits (79). This type of categorization could, however, be useful if the assumed linear model suffers from outliers or deviations from linearity that prove to be difficult to capture. Another argument for categorization is that such a model produces estimates that are easily interpretable and do not depend on the scale of the underlying variable in the sense that it does not give an estimate for a unit increase, giving for instance odds ratios for each category compared to a reference. Such models might, however, be sensitive to the choice of limits for categorization and calls for sensitivity analyses by choosing different categorizations to assure robustness. In the case of a case-control study the categorization might be based on the distribution among both cases and controls (obtaining an equal number of individuals in each category, distributing cases among categories in an intuitive dose-response fashion, and representing the levels among all studied individuals), among controls only (that might, for instance, represent the general population if there is a clear dose-response), or among cases (distributing cases equally among categories to obtain narrow confidence intervals for each category estimate). In this thesis we first suggest to model each continuous variable using penalized cubic splines for investigation of the shape of dose-response, and revealing overly influential observations.

On the bases of these benefits and limitations inherent in the modeling, we in this thesis compare smoothing splines, categorization and continuous models on the original scale.

Variable selection for modeling dose-response effects

Besides determining the shape of the dose-response for each variable in the model, there is also the issue of selecting the variables and variable interactions to be included in the model (79-82). Every strategy for selection of variables should first assess multicollinearity between the covariates and identify candidate variables that are related to the exposure risk factor that we aim to estimate or to possible disease outcomes, because these may inappropriately confound the associations between risk factors and disease. If a risk factor is negatively associated
with some confounder, we might estimate a positive association between disease outcome and the risk factor and a negative association between disease and the confounder where no effect on disease would remain if the effects were added together. Positive association between the risk factor and the confounder means that we are not able to determine if the effect on disease is caused by the risk factor or the confounder.

Secondly we implemented a variable selection algorithm for inclusion and exclusion of candidate variables that biases the risk factor estimates or affects the robustness of these estimates. We aimed to find the set of variables that confounds the association between the risk factor and disease, that by inclusion changes/modifies the effect of the risk factor or that are independently associated with the disease. The reason for including variables that are only associated with disease is conservatism. Even if the variable’s effect on the risk factor is very limited we choose to take this variable into consideration because it is statistically significantly associated with the outcome. For instance, these variables are often those that are known to be strong predictors of disease. This prior information about candidate variables can also be beneficial to consider in the variable selection algorithm, although forcing the inclusion of a known prior risk factor may heavily bias the risk factor estimate if the prior information is, in fact, not valid (79). Because the purpose of the models presented in this thesis is to obtain robust estimates for the risk factors and not to provide the model with the best predictive ability, variable selection algorithms ought to be evaluated with regards to their ability to produce valid estimates and tests of these estimates. A suitable variable selection algorithm is very sensitive to finding confounding effects and has high specificity for exclusion of variables that do not confound or modify the association between risk factors and disease.

One algorithm for variable selection is stepwise regression in which we first choose a set of variables that we would like always to remain in the model, and thereafter stepwise select variables to the model with the most statistically significant coefficient and excludes included variables with least statistically significant coefficients. The algorithm can be performed with both forward and backwards selection or with either forward or backwards selection alone. The forward selection algorithm starts with the model including only pre-selected variables and the backwards algorithm starts with all variables included. The algorithms continue until all variables are statistically significant or the number of iterations reaches a pre-determined limit. The risk factor of interest will always be a
pre-selected variable because otherwise the probability greatly increases for not selecting the set of variables that are confounders but are other independent risk factors of disease.

Another algorithm is the change-in-estimate algorithm in which variables are selected or excluded based on the resulting change to the risk factor of interest. Variables could, for instance, be entered into the model if the coefficient for the risk factor changed more than 10% and could be excluded if the coefficient changed by less than 10%.

In this thesis a backwards stepwise procedure has been implemented in which candidate variables were excluded if the variable were not statistically significantly associated with the outcome and did not change the effect of any of the risk factors in the model that we aimed to study by more than 10%. These variable selection algorithms are used because the total number of possible subsets of variables is large (2^number of variables). None of these methods for variable selection, however, are guaranteed to include all confounder variables. For this reason, and because the algorithm may perform poorly on a particular dataset, it is essential to record variable estimates and standard errors at each step.

The effect of within- and between-individual variation when using biomarkers in epidemiological studies

This section will discuss attenuation of dose-response relations and how this attenuation is affected by the choice of exposure biomarker. Based on the variance components estimates in paper 1, the attenuation of the regression coefficient in a dose-response model can be calculated for each of the Pb exposure biomarkers. As a consequence of the between- and within-individual variances, the least attenuation would be obtained by using B-Pb as the exposure biomarker (\( \hat{\lambda} =0.11, 0.33, \) and 0.23 for B-Pb, P-Pb, and density adjusted U-Pb, respectively). The results for P-Pb and B-Pb are in agreement with a previous study on 17 women who had blood samples taken weekly for 4 consecutive weeks and 14 women who had blood samples taken every 1 to 2 months over a period of 9 months (4). The results of that study showed that 83% of the total variance was attributable to differences between individuals for B-Pb and 78% of the total variance was attributable to differences between individuals for P-Pb. These values are comparable to the values of 91% and 78% for B-Pb and P-Pb, respectively, in our study of occupationally exposed individuals. Estimates of variance components were not given in the
previous study, and therefore attenuation cannot be calculated, but this corresponds to greater attenuation for P-Pb compared to B-Pb. A consequence of these differences in variances between and within individuals is that of the biomarkers studied in paper I, B-Pb should be chosen if the purpose is to minimize the attenuation in a dose-response model. However, it is possible that, for example, P-Pb correlate stronger than B-Pb to e.g. neurological effects because it is the Pb ions in the plasma that have access to the neurons, not the Pb inside the erythrocytes.

For cadmium, estimates of the variance components show that using B-Cd as the exposure biomarker for estimation of a dose-response relationship gives less attenuation than using P-Cd or U-Cd ($\hat{\gamma} = 0.08$, 0.70, and 0.13, for B-Cd, P-Cd, and density adjusted U-Cd, respectively). To the best of our knowledge, no previous study on variance components for cadmium exists. Because the estimated attenuation was almost the same when using B-Cd and U-Cd, the choice of biomarker may depend on the exposure timespan reflected by the biomarker. B-Cd is primarily indicative of current exposure levels whereas U-Cd reflects the long-term exposure that may be more closely associated with health effects in the general population.

**Comparing Pb biomarkers in relation to health effects**

The main endpoint of interest is the exposure biomarkers relation to health effects. Besides describing within- and between-individual variance and attenuation, the one-way random effects model was used to evaluate within- and between-individual variations of the partitioning of Pb between cells and plasma. Variance components estimates for deviations from the B-Pb and P-Pb model indicated that part of the variation could be attributed to differences between individuals, suggesting that B-Pb and P-Pb gave somewhat different information regarding an individual’s exposure. This has also been noted in a previous study (5). Because B-Pb saturates at higher levels, B-Pb has a non-linear relation with exposure, and also with health effects. P-Pb, however, does not have this property, and P-Pb is also believed to have a higher biological availability (83, 84). P-Pb is, therefore, thought to have a closer relation to health effects, but only a few studies exist comparing the exposure biomarkers correlation with exposure and health effects (7, 8). One study (Tian et al. submitted to Toxicology Letters) provides preliminary data showing that P-Pb might have a slightly closer
correlation than B-Pb with blood hemoglobin, blood zinc protoporphyrin, urinary δ-aminolevulinic acid, urinary albumin, urinary β<sub>2</sub>-microglobulin, urinary N-acetyl-β-D-glucosaminidase, and sensory nerve conduction velocity. The differences, however, were small and were not statistically significant. Because the attenuation of the relationship between effect markers and exposure biomarker would be greater for P-Pb than for B-Pb, this strengthens the possibility of a closer relation to P-Pb than B-Pb.
What does this thesis add to the knowledge about the relation between kidney effects and metals?

In this thesis, kidney function has been studied in two cohorts. In paper II the relationship between end-stage renal disease and Cd, Pb, and Hg was investigated in a nested case-control study within a prospective population-based cohort with low exposure levels. Papers III and IV studied the relationship between biomarkers of kidney effects and higher Cd-exposure levels within a Chinese population in a longitudinal study design with an 8-year follow up. In addition, the variance analyses in paper I generated some information about how Pb is excreted through urine.

Effects of metals on the kidney

Paper II showed that the risk of developing end-stage renal disease increased with increasing Ery-Pb, and perhaps also with Ery-Cd, but decreased with increasing Ery-Hg. The study design in this paper, however, did not permit conclusions to be drawn regarding causality. Because end-stage renal disease is a slowly progressing disease, some of the cases may have already had reduced kidney function at baseline. Potentially this may affect erythrocyte metal concentrations. However, analyses in subsets of time between sampling and end-stage renal disease indicated that the time from sampling to diagnosis had only a limited effect on the association between end-stage renal disease and Ery-Pb. For Ery-Cd, however, the association was mainly driven by the individuals with a short time period between sampling and diagnosis. Therefore, reversed causality for Pb appears less likely than for Cd, though it cannot be ruled out. Another observation that cast doubt on a cause-effect association is the gender difference in which a dose-response relation for Cd and Pb was only seen among men. No mechanisms or confounders that explain this observation can be given. If there were a causal association between end-stage renal disease and Ery-Pb, it would be hard to accept that such a large effect could be caused by Pb alone at these low exposure levels. It is possible that Pb affects the kidney’s ability to resist other factors and that this makes the kidney more susceptible.

Previous cross-sectional studies have shown higher levels of Pb among end-stage renal disease cases than in controls (19, 21, 23, 24);
albuminuria and reduced estimated glomerular filtration rate have been found associated with B-Pb (20, 85); and renal dysfunction and hyperuricemia have also been found to correlate with B-Pb (22). In prospective studies, progression of renal function has been found to be associated with low-level environmental exposure (26, 27). The use of prospective occupational information to assess Pb exposure through a job exposure matrix, however, showed no relation to the development of severe chronic kidney disease (28). Two longitudinal studies within the general population measured changes in serum creatinine levels. One found an association with B-Pb (86), but the other did not find an association with blood, tibia, or patella Pb concentrations (87). An association between changes in serum creatinine levels and B-Pb has also been found in a study of Pb workers (88). In a longitudinal study of 537 Pb workers, associations were found between creatinine clearance and serum creatinine, and B-Pb, among men and between B-Pb and blood urea nitrogen levels and B-Pb in women. However, that study concluded that further longitudinal studies are important due to confounding issues and other factors that can affect the estimated Pb-associated effects (89). In intervention studies, progression of renal insufficiency declined after Pb chelation therapy (27, 90). A recent review concluded that the evidence was limited for a causative effect of Pb exposure on deteriorating glomerular filtration rates, chronic kidney disease, and end-stage renal disease (29). Although the need for longitudinal studies and experimental evidence remain to establish causality, paper II, being the first prospective study with use of exposure biomarkers, gives reasons for concern.

For Ery-Cd, the increased risk of end-stage renal disease was close to statistically significant only in the univariate analysis. This Cd associated risk was reduced in the multiple conditional logistic regression model. Together with the fact that these associations were mainly driven by individuals with a short time between sampling and disease (as mentioned above), the observations are not strong indicators of a cause-effect relationship. Indications of an association have been found between end-stage renal disease and ecological exposure data on workers and residents near a cadmium-battery production facility, that is after higher exposure (25). Estimations of dietary intake of Cd found no association with either all-cause mortality or with mortality due to renal failure (91). Cases of end-stage renal disease have, however, been found to have higher levels of B-Cd than corresponding controls (19, 21, 23, 24). Also, less prominent kidney effects have been found to be associated with Cd
exposure (3) and correlations with albuminuria and reduced estimated glomerular filtration rates have been reported (85). Therefore, we do not consider the lack of association in the current study to be in conflict with previous studies. It is possible that there is an increased risk associated with Cd that could not be identified with the limited number of cases available for the current work in this thesis.

A decreased risk of end-stage renal disease associated with Hg is difficult to explain, but confounding by unobserved variables should be considered for all metal estimates. Hg can be considered to be indicative of lifestyle (92) and is associated with socio-economic status (28, 93, 94). A protective effect of long-chain omega-3 polyunsaturated fatty acids from fish that is correlated with erythrocyte levels of mercury (95) is possible, and these fatty acids have also been suggested as treatments for kidney diseases such as IgA-nephritis (96, 97). If we speculate that the association between end-stage renal disease and Hg is confounded by these factors, then this confounding should also be considered when interpreting the association of end-stage renal disease with Pb. Occupational exposure and fish consumption (98) are possibly correlated with B-Pb levels. A possible explanation for the association with fish consumption may be correlation with dietary Pb intake through leafy vegetables. Fish consumption is correlated with a healthy life style, and confounding by fish consumption, therefore, should decrease the Ery-Pb–associated risk rather than increasing it. Therefore the negative association between end-stage renal disease and Hg does not appear to further question the causality of the Pb-associated effect. Confounding by occupation could only explain a minor fraction of the cases in this population because occupational exposure is rare.

In papers III and IV, U-Cd–associated effects on the kidney function markers NAG, β2-microglobulin, and albumin were found both at baseline and at follow-up. These effects are well known (1, 3). A novelty is however the observed gene-environment interactions between Cd biomarkers and MT polymorphisms in relation to kidney effect markers. Two SNPs (MT1A rs11076161 and MT2A rs28366003) that modify Cd-associated renal toxicity were identified. At baseline in 1998, increased B-Cd corresponded to a larger increase in NAG among MT2A rs28366003 AG individuals compared with AA individuals. In addition, MT1A rs11076161 was found to modify the relation between B-Cd and β2-microglobulin, increased B-Cd corresponded to a greater increase in β2-microglobulin for AG individuals compared to GG individuals. At follow-
up the same SNP also modified the association between U-Cd and \( \beta_2 \)-microglobulin, and both AG and AA individuals had a greater effect on \( \beta_2 \)-microglobulin with increased U-Cd than GG individuals. A cross-sectional study on 512 individuals (including the individuals studied in paper IV) available at follow-up in the same population also found genetic effects on Cd metabolism and Cd-associated kidney toxicity (30). The purpose of paper IV, however, was to evaluate whether the observed genetic effects on Cd-associated kidney toxicity may be a result of a genetic effect on kidney recovery. The previous study found that \( MT1A \) rs11076161 modified the association between \( \beta_2 \)-microglobulin and B-Cd. In paper IV this association was only found for U-Cd at follow-up, but it was also seen for B-Cd at baseline. Because B-Cd mainly reflects current exposure and U-Cd is a better indicator of long-term exposure, it is expected that U-Cd would have a closer relation to effects at follow-up than B-Cd. They also found that \( MT1A \) rs11076161 modified the association between NAG and B-Cd, and \( MT2A \) rs28366003 the association between NAG and U-Cd. These associations were also suggested at baseline in our study but were not statistically significant.

The differences in findings between our study and the previous might be due to the restrictions on age and blood pressure used in papers III and IV, or to a difference between those who were followed up and those who were recruited at follow-up. These hypotheses were evaluated using bootstrapped samples with a sample size equal to 80% of the 512 individuals with non-missing data who were available at follow-up. No differences were found due to the restrictions on age and blood pressure, but differences were found between those who were followed up and those who were recruited at follow-up. The additional individuals in Lei et al. (2012), that were not included in paper IV, were not invited to participate but still wanted to be examined. It is difficult to speculate how this could have affected the results. The common finding between the studies was that \( MT1A \) rs11076161 might predict Cd-associated kidney dysfunction.

**Kidney recovery**

Paper III showed that the prevalence of individuals having elevated albumin levels (above the 95th percentile for individuals in the control area) at baseline was associated with U-Cd at baseline. No such association was found at follow-up, however, and this suggests that the Cd-associated effect on albumin excretion was reversed upon a decrease
in Cd exposure. Corresponding prevalence curves for elevated levels of NAG at follow-up showed a slightly lower prevalence at follow-up compared with baseline levels, but this difference was not statistically significant. No difference was found between prevalence curves for β₂-microglobulin. A 3-year follow-up study found a decline in Cd-associated kidney effects for β₂-microglobulin and albumin among individuals within this population with U-Cd <10 μg/g after reduced cadmium exposure (99). Using continuous markers of kidney function instead of only the 95th percentile cut-off, we found in paper IV that point estimates for U-Cd-associated effects were reduced for all of the kidney markers. This is suggestive of some recovery also for the tubular function after a reduction in Cd exposure, but the 95% confidence intervals for the effects at baseline and follow-up were overlapping and therefore the estimates cannot be concluded statistically significantly different. The levels of the tubular markers NAG and β₂-microglobulin increased between baseline and follow-up, whereas albumin as a marker of glomerular function decreased. Both β₂-microglobulin and NAG increase with age and this could be a possible explanation for our results. Increased levels of NAG and β₂-microglobulin despite a large reduction in exposure are consistent with previous studies in which the results were interpreted as evidence of irreversible kidney damage (33, 100-102). A decrease in albumin has previously been found, but only among men, in a 5-year follow-up study in Belgium after a reduction in Cd exposure (34).

In paper IV genetic modifications of Cd-related kidney toxicity were found both at baseline and follow-up, but no modifications in Cd-related changes in kidney function were found. This suggests that genetics affect Cd sensitivity but not kidney recovery from Cd-associated kidney toxicity. The genetic effect on Cd-associated kidney toxicity is however too limited to have any clinical applications, but the finding could possibly be used in epidemiological studies to target the most susceptible part of a population and may also provide further indications of what mechanisms are involved in Cd nephrotoxicity.

**The effect of adjustment for urinary dilution**

For U-Pb, adjustment for dilution greatly affected the fraction of the total variance that was attributable to between-individual variance among individuals under normal environmental Pb exposure. This effect was not seen among individuals with occupational Pb exposure. The general opinion is that Pb is excreted through glomerular filtration. If the
excretion is through straight forward glomerular filtration, adjustment for urinary dilution would decrease the within-individual variation of U-Pb. Since this was not noted among lead workers it suggests a difference in excretion between lead workers and the general population. This has also been suggested in older literature (Vostal 1966 cited in (103)). U-Pb has been reported to be correlated with the urinary flow rate even after adjustment for dilution (e.g., by measuring creatinine levels) (104, 105). This relation to urinary flow, however, was not seen in paper I. Two major differences between these two studies are that our study has a longer time between sampling (every 2 to 3 months over 2 years compared to 4 times a day on 6 consecutive days) and lower U-Pb levels (ranging from 1 to 295 μg/L compared to 220 to 590 μg/L).
Conclusions

- B-Pb has a better ability to discriminate between individuals with different mean concentration compared to Pb concentration in plasma and urine. Pb in plasma and urine also showed acceptable ability among Pb workers. This was evident from variance analyses using a one-way random effect model, estimating within- and between-individual variance components.
- The mechanism of Pb excretion in urine is different among Pb workers compared to individuals under normal Pb exposure, as adjustment for urinary dilution greatly affects the between-individual fraction of the total variance among individuals under normal Pb exposure but not among Pb workers.
- Pb concentrations in blood, and possibly also Cd concentrations, are predictors of future risk for end-stage renal disease, although further studies are needed to evaluate causality.
- Cd-associated albuminuria is reversible upon a substantial reduction in Cd exposure, while Cd-associated effects on the tubular markers N-Acetylglucosamine and β2-microglobulin remain.
- Analysis of prospective longitudinal data does not suggest that metallothionein polymorphisms affect kidney recovery from Cd toxicity. Therefore, the observed genetic association with Cd-related kidney effects appears to be due to an effect on Cd toxicity rather than on recovery related to decreased Cd exposure.
Acknowledgements

There are many people that I would like to thank for their contributions to this work and for supporting me in different ways. First I want to thank my main supervisor Ingvar Bergdahl for sharing his wisdom with great enthusiasm. I have learned a lot from you during these years, and I am very grateful. Secondly I would like to thank my co-supervisors Leif Nilsson, for advice when it is most needed and for encouragement in the most intense writing phase, and Ulf Strömberg for inspiring lectures and discussions at PHIME meetings as well as giving feedback when needed.

I would also like to say special thanks to some of my coauthors: Staffan Skerfving, from whom I have learned many things during lectures and while working on our papers, Thomas Lundh for going the extra mile to help me and finding answers to my questions, Maria Svensson and Bodil Björ for discussions and excellent work on the end-stage renal disease paper, Karin Broberg for introducing me to and guiding me through the world of genetics, Yihuai Liang for always being kind and helpful, and Taiyi Jin and Gunnar Nordberg for interesting discussions.

Additionally I would like to thank Anna Oudin, Ingrid Liljelind and Lars Barregård for reading my manuscripts and providing me with useful feedback at my mid-seminar, and Marie Eriksson and Bengt Järnholm at my before dissertation seminar. Additional thanks to Bengt Järnholm for advice during the writing of this thesis.

I would like to thank all participants of PHIME for joyful times and interesting work.

I have also been privileged with the best co-workers! Thank you all for your support and the nice working atmosphere which means a lot. I would like to thank my fellow PhD-students and Lars Modig for moral support, and additional thanks to David Olsson, Andreas Tornevi, Kadri Meister, Christofer Åström and Daniel Åström for always making time for discussions. Another thanks to David for his own way of giving support and encouragement.

For great company, and always making time when I’ve needed, I would also like to thank my former roommates Joacim Rocklöv and Peter Boman.
I would also like to thank someone with a special place in my heart, my wife Louise, for her support and understanding during the many 'off-office hours' I have spent writing this book.

I thank my family, extended family, and friends for supporting me during these years. Joel for being there since the first day of college, and Tommy my 'magister' in all subjects in life.

For access to samples I would like to acknowledge the Malmö Diet and Cancer study and the Northern Sweden Health and Disease Study (NSHDS), including samples from the Västerbotten Intervention Project (VIP), Northern Sweden WHO Monitoring of Trends and Cardiovascular Disease (MONICA) and Mammography Screening Project in Västerbotten.
References


10. Swedish Renal Registry [www.snronline.se].


48. Bergdahl IA, Sheveleva M, Schutz A, Artamonova VG, Skerfving S. Plasma and blood lead in humans: capacity-limited binding


57. Johannesson S, Rappaport SM, Sallsten G. Variability of environmental exposure to fine particles, black smoke, and trace


61. Riggs DS, Guarnieri JA, Addelman S. Fitting straight lines when both variables are subject to error. Life Sci. 1978 Apr 3-17;22(13-15):1305-60.


67. Blomqvist N. ON THE BIAS CAUSED BY REGRESSION TOWARD THE MEAN IN STUDYING THE RELATION BETWEEN

68. Cologne JB. Re: "When is baseline adjustment useful in analyses of change? An example with education and cognitive change". American Journal of Epidemiology. 2006 Dec 1;164(11).


