Heavy Metals in Human Health and Pregnancy: How Data Analysis, Mining, and Modeling Present a Solution

Emma Frennborn, Khalil Rust, Les Sztandera
Thomas Jefferson University, Jefferson College of Life Sciences (JCLS)
Philadelphia, Pennsylvania, United States of America
Emails: Emma.Frennborn@students.jefferson.edu, Khalil.Rust@students.jefferson.edu, Les.Sztandera@jefferson.edu

Abstract— Essential heavy metals, such as zinc, copper, iron, and manganese, play important roles in many biological processes. Other heavy metals, like lead, arsenic, cadmium, and mercury, can displace essential heavy metals and disrupt vital biological processes. Heavy metal exposure can be particularly detrimental to pregnant women and their developing fetuses; however, little is known about the combinatory impact that simultaneous exposure to multiple heavy metals may have on fetal development. Procuring a better understanding of how these metals influence fetal development is a critical first step to addressing this concerning lack of knowledge. There are numerous databases and datasets in existence storing data on heavy metal levels in maternal and fetal blood, and novel studies are being done to expand this collection of data. Data mining techniques present a tool that could be used to close this gap of knowledge by revealing patterns in data not previously discovered. Research at Thomas Jefferson University aims to aid in closing this gap of knowledge. In the study, data such as heavy metal blood concentrations, pregnancy complications, health outcomes, and demographic information will be collected from mothers and newborns. Data mining strategies will then be used to develop models capable of discovering data patterns. If this modeling is successful, such an approach can be utilized by healthcare providers in the future to assess patient risk and provide early intervention for at-risk pregnant patients.

Keywords- data mining; data modeling; databases; heavy metals; neonatal health; maternal health.

I. INTRODUCTION

First, this paper will describe the critical roles that essential heavy metals play in human health, as well as how toxic metals can disrupt these processes. Also, the effects that toxic heavy metal exposure can have during pregnancy and on fetal development will be emphasized.

A. Heavy Metals and Health

Certain heavy metals are considered essential, as they act as co-factors in important enzymes necessary for maintaining biological homeostasis. Essential heavy metals include zinc, iron, copper, and manganese [1]. Zinc finger motifs, which require zinc, are common protein structures in Deoxyribonucleic Acid (DNA) binding proteins like transcription factors [2]. Iron is found in hemoglobin, which is involved in oxygen transport in red blood cells [3]. Copper and manganese are often found in enzymes involved in oxidation reduction (redox) reactions and both can be found in superoxide dismutases. Superoxide dismutases neutralize superoxide radicals formed in the mitochondria and act as antioxidants [3]. Considering the breadth of roles that heavy metals play in maintaining human health, disrupting the levels of these beneficial heavy metals in humans can have detrimental effects.

There are also other heavy metals known to cause adverse health effects. Lead, mercury, cadmium, aluminum, arsenic, and cobalt are some of the most well studied heavy metals associated with causing heavy metal toxicity [1]. Many of these toxic heavy metals displace essential heavy metals in important enzymes and thus prevent these enzymes from functioning properly. For example, cadmium and mercury can displace zinc in zinc finger motifs and can prevent transcription factors from binding to DNA [2]. Also, if the heavy metals in superoxide dismutases are displaced, there is an accumulation of reactive oxygen species leading to oxidative stress, which if severe enough, triggers apoptosis [3]. Toxic heavy metals also react with thiol groups, which are found on the amino acid cysteine, so any exposed cysteine on a protein has the potential to be altered by a toxic heavy metal [4]. Considering that zinc-finger motifs and thiol groups are found in many proteins, there is the potential for many biological functions to be disrupted when toxic heavy metals enter human cells, highlighting the importance of studying heavy metal exposure [4].

Although heavy metal exposure can occur at work, like in the case of miners and welders, the majority of the population is exposed to heavy metals from pollution [1] [4]. Heavy metals in water runoff from industrial pollution can enter crops, cattle, fish, and drinking water [4]. If contaminated food or water is consumed, heavy metals can enter the bloodstream, and in the case of pregnant women, can also be transferred to developing offspring [5]. The effects of heavy metals on developing fetuses can be especially damaging. There is substantial cell proliferation and differentiation during development which requires a great amount of energy and resources. Increased oxidative stress due to dysfunctional superoxide dismutases or decreased gene expression due to disrupted zinc finger motifs in DNA binding proteins can thus greatly damage the growth and development of fetuses [2][3][5].

Acute or chronic exposure to heavy metals during pregnancy can result in severe morbidities in mothers as well as in offspring. Exposure to heavy metals during pregnancy can lead to impaired growth and development in children. Heavy metals exposure during pregnancy is also associated with asthma, obesity, and hypertension [6]-[9]. Exposure to multiple heavy metals can also lead to deficiency of nutritionally essential metals with long-term consequences [6][10][11]. Maternal heavy metal exposure is also associated with decreased maternal thyroid hormones [12]. Elevated levels of lead in umbilical cord blood have been correlated with increased risk of preterm delivery, low birth weight, small head circumference, and low birth length [5]. Mercury is known to cross the placenta and can have negative effects on the developing nervous system of the fetus leading to cognitive delays, reduced attention span, memory issues, and motor skill impairments [13]. Cadmium can impair placental function, reducing the transfer of oxygen and nutrients to the fetus [14]. Arsenic also interferes with brain development and can also increase the risk of miscarriage and of developing chronic diseases such as diabetes or heart disease later in life [15]. These are just a few of the heavy metals that can have a negative impact on fetal development but more exist.

Simultaneous exposure to multiple heavy metals may produce a toxic effect that is either additive, antagonistic, or synergistic. However, the literature is scarce regarding the combined toxicity of heavy metals. Understanding how toxic metals may interact is necessary for our ability to predict the health outcome of a developing fetus and hopefully provide early intervention. There are many available datasets that exist reporting on the levels of heavy metals in maternal and fetal blood, but to our knowledge, data mining techniques that could be utilized to reveal patterns and create predictions based on the current data that exists have not yet been applied to this problem.

In this paper, we first aim to summarize data mining techniques that can be applied to biomedical research questions, including the gap of knowledge about the combinatory impact that simultaneous metal exposure has on fetal and maternal health. We also introduce research to be conducted at Thomas Jefferson University to address the heavy metals in neonatal health problem utilizing statistics and datamining techniques.

B. *Data Mining Techniques*

Many research-affiliated hospitals in the United States have established biobanks storing patient samples such as cord blood and maternal blood with associated databases. The Mayo Clinic Umbilical Cord Blood BioBank, Baylor College of Medicine's PeriBank, and the Magee Obstetric Maternal and Infant (MOMI) Database and Biobank are a few examples of large biobanks with databases [16][17]. As the data stored in these databases is sensitive, these databases are generally private and require permission to access. There are also several papers that have already been published with datasets which were used for analyzing heavy metal levels in maternal and fetal blood samples [18]-[23].

Considering that there are large datasets which already exist containing heavy metal levels from maternal and fetal blood, data mining could be utilized to reveal patterns not previously described about pre-existing data or could be used to predict future outcomes. There are a wide array of data mining methods available, which when used in conjunction, can be used to create an accurate modeling system. Many of these methods involve machine learning, which can be unsupervised or supervised [24].

Supervised machine learning involves randomly dividing the data into three subsets. One set is used to train the machine to establish the parameters. Then, a validation set is used to refine the model, and the test set is used to ensure the model performs as expected [24]. A Bayesian Network is a supervised machine learning model that could be used in the context of the heavy metal problem to profound effect. Bayesian Networks are probabilistic graphical models that represent a set of variables as a Directed Acyclic Graph (DAG) [25]. Nodes in the model represent random variables while connections between nodes represent conditional probabilities. This model can be used to calculate the probability of a certain outcome given several different interacting variables [25]. An approach like this could help to elucidate the potential synergistic, additive and antagonist properties of multiple heavy metals, and predict outcomes based on the level of exposure of individual metals. Supervised machine learning requires that data is labeled and is an approach used to train a model [24]. This means there is a known output expected when analyzing the data. If the goal is to reveal patterns in the data, in which the output is unknown and undefined, unsupervised machine learning is the appropriate approach [24].

Unsupervised machine learning models, also known as descriptive models, are often used to find patterns that describe data and can be interpreted by humans [24]. These models achieve this by clustering data into categories based on the similarity between objects in the dataset. Unsupervised data mining is exploratory in nature and can lead to the discovery of unknown patterns or relationships in data [26]. An example of this mining technique would be an association rule-based analysis, also known as market based analysis. Association rules can discover correlations between items in substantial amounts of data, and this technology is often used in health care settings to determine associations between joint effects of disease risk factors and combinations of other risk factors [26]. This is done in a 2-step process: 1) all high frequency items in the data set are listed and 2) frequent association rules are generated based on these high frequency items [26]. Such an approach would lend itself to the heavy metal problem quite nicely. Determining association rules between the various heavy metals and their effect on human health or fetal development could clarify how their known effects interact with each other and even aid in the discovery of new combinatorial effects that were previously uncharacterized.

Section 2 of this paper will describe a proposed study to be performed at Thomas Jefferson University in regard to this topic. Specifically, section 2A will describe the data collection process, while section 2B will go over the proposed analysis strategy. Section 3 and the acknowledgement section will conclude the paper.

II. PROPOSED RESEARCH

The following section will describe the data collection process.

A. Data Collection

In an Institutional Review Board (IRB) approved study at Thomas Jefferson University, 107 pregnant women will be enrolled. Consent will occur both in the outpatient setting prior to delivery and on admission to Labor and Delivery (L & D). L&D is a specialized unit that provides care to pregnant women during labor, childbirth, and the immediate postpartum period. Maternal blood samples will be collected on admission to L & D and the cord blood at delivery. Blood samples will be centrifuged, serum collected, and frozen at -80ºC. Serum samples will be analyzed by mass spectrometry for 25 metals (Na, K, Mg, Ca, Zn, Se, Cu, Li, Co, NI, Tl, Al, Cr, Sr, Cd, Ba, Be, V, Fe, As, Mo, Pb, Ag, Mn, and U). The following clinical data will be collected: maternal age, race/ethnicity, insurance, zip code, Body Mass Index (BMI), medical morbidities (asthma, hypertension, diabetes), pregnancy complications (preeclampsia, preterm birth, fetal growth restriction), maternal anemia, prior full term or preterm delivery, delivery outcomes (delivery mode, gestational age, birthweight, neonatal sex) and neonatal outcomes (duration of hospital stay, admission to the Neonatal Intensive Care Unit (NICU). The World Health Organization guidelines will be adhered to for postnatal care, including routine postpartum evaluation of all women and infant pairs at 3 days, 1-2 weeks, 6 weeks, and 12 weeks postpartum. The next section will detail the methodology that will be used for data analysis.

B. Data Analysis

Data will be analyzed through International Machines Corporation (IBM) Analytics software based on the project outline that is divided into distinct, but interlocked research goals. First, we aim to develop a database of blood samples from umbilical cords and mothers that reflect the maternal toxic elements and their potential transfer, as well as mother and newborn nutritional status. Secondly, we aim to develop a data analytics model to discover and prioritize data patterns. IBM Analytics software will be utilized to calculate power analysis for all statistical analysis to be undertaken in the newborn and mothers' blood samples study. We estimate that, for testing whether the mother's blood at the delivery will predict the toxicity level transferred to a newborn as well as whether the relation is being affected through the metals' mixture composition for the toxic materials, in order to achieve the value of power 0.95 as well as the medium size effect in regression analysis, the sample size required to be selected is around 107. Thus, it can be inferred that Power Analysis for the method of regression will help in stating the exact size of the sample on the basis of the research questions to demonstrate statistical significance. IBM Analytics software allows for multiple analysis of mean group differences and variance. The Variance Components procedure, for mixed-effects models, estimates the contribution of each random effect to the variance of the dependent variable. By calculating variance components, we will determine where to focus attention to reduce the variance in the computational models. We intend to explore four different methods for estimating the variance components: minimum norm quadratic unbiased estimator (MINQUE), analysis of variance (ANOVA), Maximum Likelihood (ML), and Restricted Maximum Likelihood (REML). If the ML method or the REML method is used, an asymptotic covariance matrix table is also displayed. Other available output includes an ANOVA table and expected mean squares for the ANOVA method and an iteration history for the ML and REML methods. WLS Weight will allow us to specify a variable used to give observations different weights for a weighted analysis to compensate for variation differences. ANOVA and MINQUE do not require normality assumptions. ML and REML require the model parameter and the residual term to be normally distributed. In terms of Data Management and Quality Control Mechanisms, three standard Jefferson security procedures based on Jefferson Information Systems and technology (IS&T) / Information Security questionnaires will be utilized to review data hosting to assure compliance with applicable security controls. The details of these internal procedures cannot be disclosed publicly as to protect proprietary information belonging to Thomas Jefferson University. Thomas Jefferson University recognizes interoperability as crucial to the sharing of research data and resources to promote efficiency in research, and utilizes standards articulated by the Jefferson Research Integrity, Conduct and Compliance Office. Software applications are hosted on servers with networked storage located at a data center, providing data security and disaster recovery services. Jefferson employs best practices with regard to data privacy and security, complying with the Common Rule, HIPAA, as well as state regulations.

In terms of Data Analytics, we aim at developing models to discover and prioritize data patterns to provide information and actionable knowledge to both medical practitioners as well as public health policy decision makers. Data Analytics will be used: 1. To explore data to find new patterns and relationships (data mining); 2. To evaluate and test previous decisions (randomized controlled experiments, multivariate testing); 3. To explain why a certain outcome

 ISBN: 978-1-68558-247-0

happened (statistical analysis, descriptive analysis); and 4. To venture into the future (forecast) results (predictive modeling, predictive analytics). All four research avenues capture very well the significance and impact of Data Analytics. It could be summarized as a leading theme for the whole research proposal: "In God we trust, all others bring data, especially in maternity health care, and its impact on future healthy cities."

The data from the proposed research will help in identifying pregnant women at risk for developing heavy metals toxicity and the deficiency of essential nutrients. Identifying deficiency of nutritionally essential metals in pregnant women and their newborn and supplementation may improve pregnancy outcomes, as well as improve growth and development in children and prevent long-term morbidities. The significance of the proposed research is accentuated by environmental impact on maternal/child health through measuring maternal heavy metal exposure and fetal transfer. Although the acute and chronic effects are known for some metals, little is known about the health impact of mixtures of toxic elements.

III. CONCLUSION

Heavy metals play a complex role in human health. Harmful metals are known to interact with essential metals in a competitive way, but their combinatorial effects have not been sufficiently studied. An understanding of these combinatorial effects is important as individuals are often exposed to many of these harmful metals at once, due to the impact of industrial pollution. Fortunately, there is a growing collection of data concerning the impact of these metals on pregnant mothers and newborns. Data mining technologies present an efficient and productive method to utilize these databases. Proper use of datamining techniques used in conjunction with the collected data on these metals could elucidate the additive, synergistic, and antagonistic effects of these metals, thereby filling a gap of knowledge. As such, exploration of this topic would be of immense importance. At Thomas Jefferson University, we aim to collect data on maternal and neonatal heavy metal blood levels and health outcomes to add to the data already in existence. Then, data mining techniques will be used to develop data modeling systems capable of revealing patterns in this data not previously reported. Understanding the connection between health outcomes and heavy metal blood levels in both mothers and developing fetuses could allow clinicians to better predict pregnancy and delivery complications and thus could provide early intervention, if needed, to prevent complications. These healthcare goals can only be achieved once the data available on heavy metal levels during pregnancy/delivery are analyzed and modeled.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Les Sztandera of Thomas Jefferson University for his mentorship and guidance. Also, the authors gratefully acknowledge the travel funding received from the Jefferson Global Health Initiatives Council (GHIC) Global Office, which made presenting this work at the IARIA AI Health 2025 conference possible.

REFERENCES

[1] P. B. Tchounwou, C. G. Yedjou, A. K. Patlolla, and D. J. Sutton, "Heavy metal toxicity and the environment," *Exp. Suppl.*, vol. 101, pp. 133-164, 2012.

[2] A. Hartwig, "Zinc finger proteins as potential targets for toxic metal ions: differential effects on structure and function," *Antioxid. Redox Signal.*, vol. 3, no. 4, pp. 625-634, Aug. 2001.

[3] M. Jaishankar, T. Tseten, N. Anbalagan, B. B. Mathew, and K. N. Beeregowda, "Toxicity, mechanism and health effects of some heavy metals," *Interdiscip. Toxicol.*, vol. 7, no. 2, pp. 60-72, Jun. 2014.

[4] D. Witkowska, J. Słowik, and K. Chilicka, "Heavy metals and human health: Possible exposure pathways and the competition for protein binding sites," *Molecules*, vol. 26, no. 19, p. 6060, 2021.

[5] S. S. Zinia, K. H. Yang, E. J. Lee, M. N. Lim, J. Kim, W. J. Kim, and the Ko-CHENS Study Group, "Effects of heavy metal exposure during pregnancy on birth outcomes," *Sci. Rep.*, vol. 13, no. 1, p. 18990, Nov. 2023

[6] I. Karakis, D. Landau, M. Yitshak-Sade, et al., "Exposure to metals and congenital anomalies: a biomonitoring study of pregnant Bedouin-Arab women," *Sci. Total Environ.*, vol. 517, pp. 106-112, 2015.

[7] J. M. Madrigal, V. Persky, A. Pappalardo, and M. Argos, "Association of heavy metals with measures of pulmonary function in children and youth: Results from the National Health and Nutrition Examination Survey (NHANES)," *Environ. Int.*, vol. 121, pt. 1, pp. 871-878, 2018.

[8] I. Al-Saleh, N. Shinwari, A. Mashhour, and A. Rabah, "Birth outcome measures and maternal exposure to heavy metals (lead, cadmium, and mercury) in Saudi Arabian population," *Int. J. Hyg. Environ. Health*, vol. 217, no. 2-3, pp. 205-218, 2014.

[9] S. F. Farzan, C. G. Howe, Y. Chen, et al., "Prenatal lead exposure and elevated blood pressure in children," *Environ. Int.*, vol. 121, pt. 2, pp. 1289-1296, 2018.

[10] I. Karakis, D. Landau, R. Gat, et al., "Maternal metal concentration during gestation and pediatric morbidity in children: an exploratory analysis," *Environ. Health Prev. Med.*, vol. 26, no. 1, p. 40, 2021.

[11] G. O. Arinola, A. Dutta, O. Oluwole, C. O. Olopade, "Household air pollution, levels of micronutrients and heavy metals in cord and maternal blood, and pregnancy outcomes," *Int. J. Environ. Res. Public Health*, vol. 15, no. 12, 2018.

[12] X. Sun, W. Liu, B. Zhang, et al., "Maternal heavy metal exposure, thyroid hormones, and birth outcomes: A prospective cohort study," *J. Clin. Endocrinol. Metab.*, vol. 104, no. 11, pp. 5043-5052, 2019.

[13] C. S. Oliveira, P. A. Nogara, D. M. P. Ardisson-Araújo, M. Aschner, J. B. T. Rocha, and J. G. Dórea, "Neurodevelopmental effects of mercury," *Adv. Neurotoxicol.*, vol. 2, pp. 27-86, 2018. doi: 10.1016/bs.ant.2018.03.005. Epub May 24, 2018.

[14] J. L. Young and L. Cai, "Implications for prenatal cadmium exposure and adverse health outcomes in adulthood," *Toxicol. Appl. Pharmacol.*, vol. 403, p. 115161, Sep. 2020.

[15] N. Y. Ortiz-Garcia, et al., "Maternal exposure to arsenic and its impact on maternal and fetal health: A review," *Cureus*, vol. 15, no. 11, p. e49177, Nov. 2023.

[16] S. N. Caritis, "Magee obstetric maternal & infant (MOMI) database and Biobank: Magee-Womens Research Institute & Foundation," *Magee*..

[17] K. M. Antony, et al., "Generation and validation of a universal perinatal database and biospecimen repository: PeriBank," *J. Perinatol.*, vol. 36, no. 11, pp. 921-929, Nov. 2016.

[18] I. B. Goldberg, et al., "Early pregnancy metal levels in maternal blood and pregnancy outcomes," *Sci. Rep.*, vol. 14, no. 1, p. 27866, Nov. 2024

[19] K. Kot, et al., "Interactions between 14 elements in the human placenta, fetal membrane and umbilical cord," *Int. J. Environ. Res. Public Health*, vol. 16, no. 9, p. 1615, 2019.

[20] I. Al-Saleh, N. Shinwari, A. Mashhour, G. E. D. Mohamed, A. Rabah, "Heavy metals (lead, cadmium, and mercury) in maternal, cord blood and placenta of healthy women," *Int. J. Hyg. Environ. Health*, vol. 214, no. 2, pp. 79-101, 2011.

[21] T. Takatani, et al., "Association between maternal blood or cord blood metal concentrations and catch-up growth in children born small for gestational age: an analysis by the Japan environment and children's study," *Environ. Health*, vol. 23, no. 1, p. 18, Feb. 2024.

[22] J. Grzesik-Gąsior, J. Sawicki, A. Pieczykolan, and A. Bień, "Content of selected heavy metals in the umbilical cord blood and anthropometric data of mothers and newborns in Poland: preliminary data," *Sci. Rep.*, vol. 13, no. 1, p. 14077, Aug. 2023.

[23] H. Ma, et al., "Maternal and cord blood levels of metals and fetal liver function," *Environ. Pollut.*, vol. 363, pt. 2, p. 125305, 2024.

[24] W. T. Wu, Y. J. Li, A. Z. Feng, et al., "Data mining in clinical big data: the frequently used databases, steps, and methodological models," *Mil. Med. Res.*, vol. 8, p. 44, 2021.

[25] J. Puga, M. Krzywinski, and N. Altman, "Bayesian networks," *Nat. Methods*, vol. 12, pp. 799–800, 2015.

[26] I. Yoo, et al., "Data mining in healthcare and biomedicine: a survey of the literature," *J. Med. Syst.*, vol. 36, no. 4, pp. 2431-2448, Aug. 2012. Epub May 3, 2011.