Article 25fa pilot End User Agreement

This publication is distributed under the terms of Article 25fa of the Dutch Copyright Act (Auteurswet) with explicit consent by the author. Dutch law entitles the maker of a short scientific work funded either wholly or partially by Dutch public funds to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU)'Article 25fa implementation' pilot project. In this pilot research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. Please note that you are not allowed to share this article on other platforms, but can link to it. All rights remain with the author(s) and/or copyrights owner(s) of this work. Any use of the publication or parts of it other than authorised under this licence or copyright law is prohibited. Neither Radboud University nor the authors of this publication are liable for any damage resulting from your (re)use of this publication.

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please contact the Library through email: copyright@ubn.ru.nl, or send a letter to:

University Library
Radboud University
Copyright Information Point
PO Box 9100
6500 HA Nijmegen

You will be contacted as soon as possible.

Application of A Causal Discovery Model to Study The Effect of Iron Supplementation in Children with Iron Deficiency Anemia

Fajar Agung Nugroho
Department of Informatics
Diponegoro University
Semarang, Indonesia
fajar@lecturer.undip.ac.id

Thomas H. A. Ederveen
Center for Molecular and Biomolecular
Informatics, Radboud Institute for Molecular
Life Sciences
Radboud University Medical Center
Nijmegen, The Netherlands
Tom.Ederveen@radboudumc.nl

Adi Wibowo
Department of Informatics
Diponegoro University
Semarang, Indonesia
bowo.Adi@live.undip.ac.id

Jos Boekhorst
NIZO
Ede, The Netherlands
Jos.Boekhorst@nizo.com

Marien I. de Jonge
Section Pediatric Infectious Diseases,
Laboratory of Medical Immunology
Radboud Center for Infectious Diseases
Radboud University Medical Center
Nijmegen, The Netherlands

Tom Heskes
Faculty of Science
Radboud University
Nijmegen, The Netherlands
Tom.Heskes@ru.nl

Abstract—Most clinical studies are descriptive, measuring different parameters that are associated with certain treatment effects, while causal relations cannot be confirmed. Computational models aid researchers to make predictions of causality and help to focus on the most relevant part of the data. This study used a computational model to find a causal link between iron supplementation effectiveness, RTI, and systemic inflammation parameters, and gut microbiome profiles.

We used a causal discovery algorithm on a randomized controlled trial dataset ($n=72$) of 6 month-old infants from Kenya. We also used correlation and partial correlation to determine the causal effect of any causal link. We found that (1) expression of the Transferrin Receptor (TfR) has a positive causal link with the serum TfR-Ferritin ratio, whereas serum Ferritin levels have a negative causal link to TfR expression. (2) C-Reactive Protein (CRP) together with IL-8 and IL-1 β have a positive causal relation with IL-6. (3) No causal link between iron supplementation and gut microbiome profile. The first and second result is in accordance with the current biological research findings. While the third result shows no causality model, the skeleton might give information for future studies on understanding the gut microbiome profile. Computer modeling helped to uncover causality between clinical parameters in iron deficiency anemia children with iron-micronutrient supplementation. This could lead to more focused studies to better understand the iron supplementation practice as well as the biological mechanism of RTI, gut microbiome alteration, and iron supplementation.

Keywords—causal discovery model, causality, iron deficiency anemia, iron supplementation, inflammation, gut microbiome

I. INTRODUCTION

Roughly, 31.15% of preschool children in Africa are categorized as iron deficiency anemia (IDA)[1][2]. In order

to counter that, a program was set for children to have a food supplement of Iron-containing micronutrient powders (MNP).

A study found that iron supplementation in Kenyan infants has an adverse impact on the gut microbiome [3][4], leading to gut inflammation and an abundance of pathogens. In addition, it also amplifies diarrhea and respiratory tract infections (RTIs) [5]. However, this clinical study could not reveal the causal relation between iron supplementation with the aforementioned adverse events. Thus, we propose computational models to aid predictions of causality and help to focus on the most relevant part of the data.

As an alternative method, we would like to explore any causal links within the dataset reported in [3][4] applying a causal discovery algorithm. This method calculates a bulk of statistical tests in order to attain, in case it is achievable, a model of causes and effects. We put in an application for causal modeling an algorithm that we employ from [6] called PC algorithm. The algorithm itself is the most influential constraint-based method that able to acquire completed partially directed acyclic graph (CPDAG) from a dataset.

The objective of this study is to make a causal model that could lead to more focused studies to better understand the iron supplementation practice as well as the biological mechanism of RTI, gut microbiome alteration, and iron supplementation. The causal model from this study then can be used as a guideline for a researcher to do experiments.

The paper is organized as follows. Section 2 describes the material and method that was used. Section 3 shows the resulting causal model. Sections 4 tells about the discussion related to the results.

Fig. 1. Stable Selection Framework

II. MATERIAL AND METHODS

A. Participants

We included 72 randomized control trial Kenyan infants that were studied in [3][4]. All subjects participated in the study should have the following scale:

- Age, between 6.5 to 9.5 months.
- Never had a chronic or acute disease.
- Never consumed vitamin and mineral in the past 8 weeks.
- Never had antibiotics in the past 10 weeks.

B. Data Description

We carefully included these few variables to be applied to causal discovery using a PC algorithm [6]. These variables represent the iron level in the blood (Zinc ProtoPorphyrin, transferrin receptor, sTFR-Ferritin index and serum ferritin), inflammation status (C-Reactive Protein, IL_1 β , IL_6, IL_8, and IL_17) as well as normal and pathogenic gut microbiome species (Bifidobacterium, Prevotella, Streptococcus, Lactobacillus, Bacteroides, Veillonella, Megasphaera, Faecalibacterium, Blautia, Clostridium, Collinsella, Megamonas, Sporacetigenium, Enterococcus, and Eubacterium).

C. Causal Modeling

One of the underlying mechanisms of causal discovery algorithm is Structural Equation Model (SEM). SEM able to generate a collection of functions that later on can become variables of particular data. As described in [7]. It works with the following formula:

$$X_i \leftarrow g_i(S_i, \epsilon_i), \quad i = 1, \dots, p, \quad (1)$$

The formula can be implemented very well by Maathuisin [7] to create a collection of random variables that embedding the causal graph in it.

$$\begin{aligned} P &\leftarrow g_1(M, \epsilon_P) \\ S &\leftarrow g_2(P, \epsilon_S) \\ R &\leftarrow g_3(M, \epsilon_R) \\ M &\leftarrow g_4(M, \epsilon_M) \end{aligned}$$

M denotes motivation for a prisoner to join a program (P) that can give them some social skills (S) that hopefully can lower the rearrests (R) level.

Reverse engineering data in order to find the corresponding SEM has been the main discussion of Pearl in [8]. Basically by employing a conditional (in)dependency test, we can have DAG(s) out from combination of three variables. The illustration of (in)dependence and its DAG(s) representation can be seen in Table I.

Constraint-based and score-based are the two main procedures in finding causal models from data [9]. The detail explanation of the score-based method can be found in [10], while the constraint-based is in [6]. Both models are accessible in R-package named `pcalg` [11].

TABLE I REPRESENTATION OF (IN)DEPENDENCY TO DAG(S)

(in)dependency	DAG(s)
$X \perp\!\!\!\perp Z \mid Y$	$X \rightarrow Y \rightarrow Z$ $X \leftarrow Y \leftarrow Z$ $X \leftarrow Y \rightarrow Z$
$X \not\perp\!\!\!\perp Z \mid Y$	$X \rightarrow Y \leftarrow Z$

Fig. 2. Output causal model executing full dataset (A), and the result of Stable Selection Framework (B).

We select PC Algorithm in this study on account of its advantage for being able to neglect the causal sufficiency assumption and the flexibility for us to tune its α (p values). Based on the previous study in [3] [4], we also put in the application for the value of α that is 0.05 as statistically worthy of attention.

Since the data that we obtained from [3] [4] can be categorized as relatively small sample size, thus we performed a stable selection framework, from Fig. 1, in order to validate the causal links that we found. We adopt this framework from [12]. We randomly select half of our data that would become a subset, then run PC algorithm on it so we can have a matrix causal model of it. After that, we repeat those actions into 500 times, until we have 500 matrices. Finally, we calculate them on average numbers.

III. RESULT

Performing PC algorithm, we computed a bulk of statistical tests in order to evaluate causal relations between variables. The graph of these relationships depicted in. There are two results in this figure. (A) is the causal model when we execute all obtained datasets ($n=72$). Whereas (B) is the result of a causal model implementing a stable selection framework. In these graphs we can observe two kinds of relation connecting between variables. There are (" $X \rightarrow Y$ ") and (" $X \leftrightarrow Y$ "), those are mean that X causes Y and there

might be common cause respectively. We also accommodate plus and minus (" $+ / -$ ") signs on our causal links as a manifestation of underlying positive or negative correlation on those particular two variables.

Derived from the model in Fig. 2, there are both Transferrin Receptor and Serum Ferritin have a direct causal link to sTFR_Ferritin. A positive sign means the two variables have a positive correlation (i.e. when the Transferrine Receptor increases then the sTFR_Ferritin will enhance), and a negative sign implicates those variables have a negative correlation.

We did not find any causal link in the group of gut microbiome variables, as depicted in Fig. 3, but we found the skeleton of it. This, however, could lead to further research looking for a correlation between them. We have confidence that when the sample size is by some means can be increased into two hundred records, we might find some causal links.

IV. DISCUSSION

In the current study, we present an alternative approach of data analysis for finding causal links between variables in IDA dataset. Finding causality by clinical research is to warrant the biological plausibility. However, it might demand huge resources which most of the time is scarce.

Fig. 3. The output of the causal model on gut microbiome variables.

Thus, the causal discovery model plays a role to give a hint for further clinical research.

We found that variables representing iron availability: transferrin receptor (TfR) and serum TfR-Ferritin ratio; serum Ferritin and TfR—have a positive and negative causal link, respectively. This causality is compatible with the iron physiology in human body—when the expression of TfR is increased, the serum TfR-Ferritin will also increase, indicating low level or high demand for serum iron[13][14]. On the other hand, when the serum ferritin is low, the expression of TfR will increase[14][15]. Hence, the compatibility of the finding and the biological process, show that the model is robust.

The causality amongst CRP, IL-1 β , and IL-8 with IL-6 might be explained by the fact that all those variables have inflammatory properties. We also found an interesting finding that there is no causal link between iron supplementation and gut microbiota. In the clinical trial, they discovered that iron supplementation adversely affects the normal flora and increased the abundance of pathogens in the gut. This could be partially explained by the fact that we use only a small data set.

One of the strengths of this study is that we use a very small sample set of data since a study of causal discovery normally works with 200 or more sample size. However, we found some causal links out of it.

In conclusion, implementation of causal discovery employing PC algorithm in Kenyan clinical trial of children with iron deficiency anemia found that (1) expression of the Transferrin Receptor (TfR) has a positive causal link with the serum TfR-Ferritin ratio, whereas serum Ferritin levels have a

negative causal link to TfR expression. (2) C-Reactive Protein (CRP) together with IL-8 and IL-1 β have a positive causal relation with IL-6. (3) No causal link between iron supplementation and gut microbiome profile.

ACKNOWLEDGMENT

The microbiota dataset is described in detail in publication [3][4] and data was kindly provided by the authors.

REFERENCES

- [1] WHO, The global prevalence of anemia in 2011, Geneva: World Health Organization, 2015.
- [2] N. Kassebaum, R. Jasrasaria, M. Naghavi and e. al, "A systematic analysis of global Anemia burden from 1990 to 2010," vol. 123, p. 615–24, 2014.
- [3] T. Jaeggi, G. Kortman, D. Moretti and et al, *Gut*, vol. 64, pp. 731-724, 2015.
- [4] D. Paganini, M. Uyoga, G. Kortman and e. al, *Gut*, vol. 66, p. 1956–1967, 2017.
- [5] S. Zlotkin, S. Newton, A. Aimone and e. al, "Effect of iron fortification on malaria incidence in infants and young children in Ghana: a randomized trial," *JAMA*, vol. 310, p. 938–47, 2013.
- [6] P. Spirtes, C. Glymour and R. Scheines, *Causation, prediction, and search*, MIT Press, Cambridge, second edition, 2000.
- [7] K. Bollen, *Structural Equations with Latent Variables*, New York: John Wiley, 1989.
- [8] J. Pearl, *Causality: Models, reasoning and inference*, Cambridge: Cambridge Univ Press, 2000.
- [9] R. Daly, Q. Shen and S. Aitken, "Learning Bayesian networks: Approaches and issues," *Knowl Eng Rev* 26(2), p. 99–157, 2011.
- [10] D. Chickering, "Optimal structure identification with greedy search,"

Mach. Learn. Res., 3, p. 507–554, 2003.

- [11] M. Kalisch, M. Machler, D. Colombo, M. Maathuis and P. Buhlmann, "Causal inference using graphical models with the R package pcalg," *Stat. Softw.*, 47(11), p. 1–26, 2012.
- [12] R. Rahmadi, P. Groot, M. Heins, H. Knoop and T. Heskes, "Causality on cross-sectional data: Stable specification search in constrained structural equation modeling," *Applied Soft Computing Journal*, vol. 52(May), p. 687–698, 2017.
- [13] K. Punnonen, K. Irjala and A. Rajamäki, "Serum transferrin receptor and its ratio to serum ferritin in the diagnosis of iron deficiency," *Blood*, vol. 89, no. 3, pp. 1052-1057, 1997.
- [14] K. Suega, Y. Kandarini and J. Tubung, "Role of Soluble Transferrin Receptor and Transferrin Receptor-Ferritin Index to Detect Iron Deficiency Anemia in Regular Hemodialysis Patients," *Open Access Maced J Med Sci*, vol. 7, no. 1, pp. 97-102, 2019.
- [15] E. J. Lee, E.-J. Oh, Y.-J. Park, H. K. Lee and B. K. Kim, "Soluble Transferrin Receptor (sTfR), Ferritin, and sTfR/Log Ferritin Index in Anemic Patients with Nonhematologic Malignancy and Chronic Inflammation," *Clin Chem*, vol. 48, no. 7, pp. 1118-21, 2002.