Predicting the Prevalence of Soil Transmitted Helminths

PROJECT SCOPE

Ojo, C

Soil transmitted helminths (STH) are parasitic intestinal worms that cause intestinal nematode infections such as ascariasis, trichuriasis, and hookworm disease. According to the 2017 global burden of disease metrics, by The Lancet, soil transmitted helminths infections are

- The second most burdensome [4], after dengue: The 2017 all-age global burden level is 1920k (1200k to 2980k) disability adjusted life years.
- The most prevalent [3].

Therefore, STH infections are a focus of the World Health Organization (WHO). Soil transmitted helminths prevention, control, and elimination activities depend on knowledge of prevalence [1]. These prevalence metrics depend on field surveys, which are rather expensive to conduct. They are especially expensive when there are no measures to systematically/strategically direct field scientists to survey areas. In the absence of systematic/strategic direction, the field scientists have to survey as wide an area as possible, probably subject to constraints such as finances, accessibility, etc.

Hence, this project aims to develop a model that predicts the STH prevalence of an area, without depending on past or present field prevalence values. Instead, the model's independent variables will be based on domain expert understandings, recommendations, of the geospatial, demographic, and WASH (water, sanitation, and hygiene) variables that influence the transmission of STH.

Project Host & Core Team

The project host is the World Health Organization (WHO). Ojo, C is the postgraduate student, i.e., the project lead. Professor Antonio Montresor is the World Health Organization supervisor. Finally, Dr. Emanuele Giorgi of Centre for Health Informatics, Computing, and Statistics / Lancaster Medical School is the university supervisor.

I. Project Aim

The development a STH (Soil Transmitted Helminth) prevalence prediction model via **(a)** geospatial, **(b)** demographic, and **(c)** WASH (water, sanitation, and hygiene) **prognostic factors**.

II. Project Objectives

Hence, the project's objectives are to

- Investigate geospatial, demographic, and WASH prognostic factors of STH prevalence.

- Specifically determine whether (a) site level sanitation, and (b) population density – are prognostic factors of STH prevalence.
- Develop a geo-statistical STH prediction model.

III. Project Data

The project will rely on the

- The soil transmitted helminths' data sets of ESPEN (Expanded Special Project for Elimination of Neglected Tropical Diseases) – **for** ascariasis, hookworm infection, and trichuriasis examinations and cases data per site.
- The estimated WASH variables of IHME (Institute for Health Metrics and Evaluation) – **for** determining site level WASH variables values. The ESPEN estimates, which depend on the IHME methodology, are implementation level estimates.
- The National Oceanic & Atmospheric Administration (NOAA), WorldClim, and DIVA GIS – **for** historical geospatial and/or climatic variables values.

IV. Deliverables

The deliverables are

- STH Prevalence Prediction Model via GitHub
- Thesis Manuscript
- Datasheet [6]
- Model Card [5]

A. Approval of deliveries

Per each piece of work delivered

- Antonio Montresor shall provide written feedback.
- If more than one World Health Organization (WHO) stakeholder might provide feedback, then Antonio Montresor shall provide a written summary/consolidation of the WHO stakeholders' feedback.
- Antonio Montresor shall ensure that all WHO stakeholders that should provide feedback, do.

V. Timeline

	Notes
2022/06/20 - 2022/07/08	Addressing site level records identification issues (refer to the quality constraints section further below). The derivation of site level variable values; WASH and geospatial/climatic variables values. Data mapping, i.e.,
2022/07/12 - 2022/07/30	Exploratory data analysis. Exploration of Bayesian, and otherwise, geospatial binomial logistic models; binomial because prevalce values are proportions.
2022/08/01 - 2022/08/21	Final models development; focus on two models. Model validation. Model testing.
2022/08/22 - 2022/09/06	Manuscript writing.
2022/09/08 - 2022/09/11	Review model repository, model card, and data sheet.
2022/09/11 - 2022/09/15	Poster preparation.

VI. Out of Scope

The prediction of STH Prevalence for countries outwith the continent of Africa because the ESPEN data project focuses on African countries only, i.e., it only has the data of a set of African countries. Consequently, model development will be via the data of one or more African countries. Additionally, model validation, internal & external, will be via ESPEN countries only.

VII. Project Assumptions

The project makes a few key assumptions

- The soil transmitted helminth parasites of concern are roundworms (Ascaris lumbricoides), hookworms (Ancylostoma duodenale, Necator americanus), and whipworms (Trichuris trichiura); which cause ascariasis, hookworm infection, and trichuriasis.
- WHO is interested in **site level prevalence predictions**.
- The historical site level positive cases of ascariasis, hookworm infection, and trichuriasis, detailed by ESPEN, are dependable.
- Ascariasis, trichuriasis, and hookworm infection examinations, per site, are conducted within the same sample population of a site.
- Per site, the prevalence value of any soil transmitted helminth is via the prevalence formula of Silva & Hall [2], which depends on the prevalence of ascariasis, trichuriasis, and hookworm disease.

VIII. Project Constraints

	description
Project start date	6 June 2022
Project end date	9 September 2022
Deadlines	Masters Thesis: 9 September 2022 Poster Session: 16 September 2022 Model Repository: 12 September 2022 Data Sheet & Model Card: 12 September 2022

A. Time Constraints

- The final week of the masters modules overlaps with the first week of the masters project, therefore only a third of the week ending 10 June 2022 shall be spent on the masters project.
- Future planning may occasionally occur during week days, the time will be recovered during the evenings and/or weekends.

B. Budget Constraints

The project has no budget, i.e., no budget/payment for personnel or computational resources used; the student bears the cost of conducting the project.

C. Quality Constraints

The site level data of ESPEN has 3 core issues. Foremost, and per country, many records do not have a site level identification code. Second, records that have the same coordinates, do not usually have the same site identification code. Finally, records associated with the same site sometimes have slightly different coordinate values. Prior to exploratory analysis & modelling, these discrepancies will have to be addressed, i.e.,

- Each distinct site must have a unique site identifier and distinct coordinate values.
- records associated with the same site must have the same unique site identifier & coordinate values.

D. Equipment Constraints

No equipment has been provided for this project. The project's tasks will be, are, conducted via a personal computer. However, it is quite probable that the computer will stop working any time soon; the computer is quite old, and in recent times bits & pieces have stopped working. Hence, access to a fall-back laptop is critical.

References

1. A. Montresor, Helminth control in school-age children. (2011). https://www.who.int/publications/i/item/9789241548267.
2. N. de Silva & A. Hall, Using the prevalence of individual species of intestinal nematode worms to estimate the combined prevalence of any species. *PLOS Neglected Tropical Diseases*, **4** (2010) 1–7. https://doi.org/10.1371/journal.pntd.0000655.
3. G. B. of Diseases Collaborators, Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: A systematic analysis for the global burden of disease study 2017. *The Lancet*, **392** (2018) 1789–1858. https://doi.org/10.1016/S0140-6736(18)32279-7.
4. G. B. of Diseases Collaborators, Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: A systematic analysis for the global burden of disease study 2017. *journaltitle*, (n.d.).
5. M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, & T. Gebru, Model cards for model reporting. *CoRR*, **abs/1810.03993** (2018). https://doi.org/10.48550/arXiv.1810.03993.
6. T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. M. Wallach, H. D. III, & K. Crawford, Datasheets for datasets. *CoRR*, **abs/1803.09010** (2018). https://doi.org/10.48550/arXiv.1803.09010.