

UNIVERSIDAD NACIONAL DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO
POSGRADO REGIONAL EN CIENCIAS VETERINARIAS TROPICALES



Situación de las parasitosis intestinales en niños de zonas vulnerables de Costa Rica con base en doce años de datos (2008-2019), y modelado de sus determinantes predictivos utilizando aprendizaje automático

Alberto Solano Barquero

Universidad Nacional, Heredia, Noviembre de 2025

Tesis sometida a consideración del Tribunal Examinador del Posgrado Regional en Ciencias Veterinarias Tropicales, Maestría Académica en Enfermedades Tropicales para optar por el grado de *Magister Scientiae*

Situación de las parasitosis intestinales en niños de zonas vulnerables de Costa Rica con base en doce años de datos (2008-2019), y modelado de sus determinantes predictivos utilizando aprendizaje automático.

Alberto Solano Barquero

Universidad Nacional, Heredia, Septiembre de 2024

Tesis sometida a consideración del Tribunal Examinador del Posgrado Regional en Ciencias Veterinarias Tropicales, Maestría Académica en Enfermedades Tropicales para optar por el grado de *Magister Scientiae*

Miembros del Tribunal Examinador

Representante del Consejo Central del Posgrado

Dra. Julia Rodríguez Barahona
Coordinadora PCVET

Dr. Juan José Romero Zúñiga
Tutor

Dr. Bernardo Vargas Leitón
Lector

Dr. Javier Mora Rodríguez
Lector

Alberto Solano Barquero
Sustentante

RESUMEN GENERAL

Las parasitosis intestinales persisten como un problema de salud en las regiones más vulnerables del mundo, afectando a más de 1,4 mil millones de personas. Costa Rica, aun cuando ha presentado importantes avances en materia de salud pública desde la universalización del seguro social, sigue presentando perfiles epidemiológicos de enfermedades infecciosas muy similares a los que tenía antes de la década de los cuarenta en ciertas regiones muy vulnerables del país. En estas regiones, de difícil acceso y comunidades marginadas, se ubican focos ocultos de estas enfermedades infecciosas entre las que se incluyen las parasitosis intestinales.

Este trabajo es una investigación integrada acerca de la situación epidemiológica de las parasitosis intestinales en escolares de zonas vulnerables del país desde dos enfoques. En el primero, se hace un abordaje epidemiológico descriptivo basado en 4812 registros de diagnósticos realizados en escolares, recolectados por los proyectos ED-511 y TC-552 de la Universidad de Costa Rica, entre los años 2008 y 2019. En el segundo, se desarrollan modelos predictivos de los determinantes para estas infecciones, utilizando herramientas de aprendizaje automático (machine learning).

En el primer estudio se documenta que la proporción en general de infecciones parasitarias intestinales (tanto comensales como patógenos) alcanza un 56,4%, con 20,5% de protozoarios patógenos y 33,4% de helmintos transmitidos por el suelo. Estas cifras son considerablemente más altas que las que reportan las encuestas nacionales y los escasos estudios más recientes. Esta alta proporción de casos se concentró en escuelas de los distritos de Chirripó y Limoncito, constituyéndose en ‘puntos calientes’, tanto por protozoarios como por helmintos: prevalencias mayores que 30% y 15% por protozoarios y por helmintos, respectivamente. Estas infecciones afectan más al grupo etario entre los 7 y los 9 años, sin diferencias por sexo.

El análisis multivariado identificó agrupamientos regionales de las observaciones por escuela, mostrando un comportamiento muy homogéneo de las variables entre las escuelas de Chirripó y Limoncito. Asimismo, ciertos agrupamientos entre escuelas de otras regiones, aunque con mayor traslape. También, identificó, inicialmente, potenciales variables

altamente correlacionadas y explicativas de las parasitosis por protozoarios y por helmintos. Algunos ejemplos de estas variables altamente correlacionadas fueron la tenencia de agua de tubería dentro de las casas, muy correlacionada con la tenencia de alcantarillado sanitario, y negativamente correlacionada con la defecación al aire libre, las variables de analfabetismo en la población estaban también muy correlacionadas entre sí, el uso de letrinas sin alcantarillado sanitario o tanque séptico también estaban muy correlacionadas con la defecación al aire libre.

En el segundo estudio, se construyó un marco analítico para modelar los determinantes de infecciones por helmintos y protozoarios intestinales en escolares de Costa Rica. Para ello, se integraron en el análisis 68 variables independientes relacionadas con factores socioeconómicos, educativos, condiciones de vivienda, clima, configuración del terreno y dificultad de acceso geográfico.

La metodología incluyó, inicialmente, la reducción de dimensionalidad mediante análisis multivariado. Dicho análisis consistió en un análisis de componentes principales, el cual incluyó analizar si se formaban grupos (clústeres) entre las observaciones y la visualización de los vectores de residuos de las variables independientes en su relación con los vectores de los residuos de las variables objetivo. Seguidamente, la selección de variables mediante modelos secuenciales basados en aprendizaje automático o machine learning, tales como Random Forest Classifier y Elastic-Net. Por último, se incorporaron las variables seleccionadas a modelos lineales generalizados y modelos lineales generalizados mixtos para incorporar efectos aleatorios dados por la escuela y el distrito. Este enfoque permitió construir dos modelos predictivos: uno para helmintos y otro para protozoarios intestinales.

El modelo para los helmintos identificó 10 variables significativas que incluyeron la estacionalidad de las precipitaciones, las distancias entre las viviendas, la proporción de población indígena, la irregularidad topográfica, y la distancia entre la escuela y el centro de salud de nivel secundario o terciario más cercano. Con este perfil de variables, se explica cómo estas infecciones parecen estar determinadas por factores ambientales, de configuración territorial y acceso geográfico, y por una variable sociocultural que debe investigarse más a fondo. El modelo alcanzó una buena capacidad predictiva con un área bajo

la curva (AUC) de 0.74 y un R^2 condicional de 0.50 al incluir los efectos aleatorios, lo cual es notable para un fenómeno de tal complejidad.

En el caso de las infecciones por protozoarios, el modelo predictivo, aunque ofreció parámetros robustos y de buen ajuste, no resultó tan explicativo como el modelo para los helmintos. Aun así, alcanzó un R^2 condicional de 0.17, con una calibración adecuada. Para fenómenos multidimensionales como las infecciones por protozoarios, presenta una capacidad muy apreciable de explicación del fenómeno. En este caso, las variables más informativas fueron el índice de carencias críticas (necesidades básicas no satisfechas), el índice de vecinos más cercanos (NNIS), la diferencia entre temperatura máxima y promedio anual, y la cantidad total de lluvia anual.

Este trabajo proporciona tres aportes principales. En primer lugar, respalda, con datos duros, que las metodologías de estimación y las cifras nacionales sobre parasitosis intestinal en escolares de Costa Rica subestiman, en mucho, la frecuencia real de parasitosis intestinal en diversas regiones vulnerables de Costa Rica. En segundo término, proporciona evidencia robusta del carácter microgeográfico, estructural y multidimensional de los determinantes de estas infecciones. Finalmente, introduce, por primera vez en Costa Rica, una estrategia combinada de aprendizaje automático (machine learning) y estadística inferencial para predecir infecciones parasitarias intestinales en niños escolares. Ello deriva en su potencial valor en la salud pública de ser utilizado para mejorar los abordajes de estas infecciones y para incidir en políticas de salud más focalizadas.

Como conclusión, esta investigación confirma que el control percibido de las parasitosis intestinales en Costa Rica está sobreestimado para ciertas poblaciones vulnerables de nuestro país y muchos contextos más locales. Las comunidades más vulnerables siguen enfrentando un riesgo de enfermedad que, a los ojos de la salud pública, permanece un poco a la sombra; esto se refuerza por condiciones estructurales de exclusión.

A través de un abordaje innovador e integrador, este trabajo no solo brinda datos sobre esta realidad, sino que propone una estrategia y herramientas concretas para abordar, intervenir y mitigar los riesgos. Este conocimiento puede servir de base para una redefinición de las estrategias de vigilancia y control más justas, más precisas y mejor informadas.

AGRADECIMIENTOS

A mis tutores, Juan José Romero Zúñiga, Bernardo Vargas Leitón y Javier Mora Rodríguez, por su guía en este proceso, sus fundamentales aportes y retroalimentación.

A Dennis León, quien concibió inicialmente los proyectos de acción social que originaron estos datos, y a Javier Mora por su incansable trabajo impulsándolos y dándoles continuidad.

A Javier Mora y Alicia Rojas por su apoyo y por permanecer al tanto de los avances y ayudar con retroalimentación para mejorar y dar conclusión a este proceso.

A toda mi familia por su invaluable apoyo todos estos años. Mencionar en especial a mi esposa Laura y a mi hermana Melissa, también personas de ciencia y que admiro muchísimo: muchas gracias por sus consejos para este trabajo.

A los profesores y compañeros de la maestría por su entrega y sus conocimientos y experiencias compartidas.

Y a todas las personas que, a pesar de lo que pasa en este mundo de ahora, trabajan para que el conocimiento sea valorado y sea la guía para encontrar soluciones a los problemas.

DEDICATORIA

A mi madre, Norma Patricia Barquero Sanabria.

Por todo su esfuerzo dedicado a la salud pública de Costa Rica desde su puesto de trabajo durante muchos años, por su enorme sacrificio para conmigo y mis hermanos.

En tu honor, todos mis esfuerzos, Mamá.

ÍNDICE

RESUMEN GENERAL	IV
AGRADECIMIENTOS	V
DEDICATORIA	VI
ÍNDICE DE CUADROS	XI
ÍNDICE DE FIGURAS	XIII
ABREVIATURAS	XV
DESCRIPTORES	XVI
INTRODUCCIÓN GENERAL	17
1.1 Aspectos epidemiológicos generales	17
1.2 Protozoarios intestinales	17
1.3 Helminthos intestinales	21
1.4 Importancia de las parasitosis intestinales	23
1.5 Aspectos económicos	25
1.6 Antecedentes de la investigación sobre parásitos intestinales en Costa Rica	26
Referencias (Introducción General)	28
ARTÍCULO 1	33
(Formato: Revista Trop. Med. Int. Health) https://onlinelibrary.wiley.com/page/journal/13653156/homepage/forauthors.html	33
Beyond the Illusion of Control: Gastrointestinal Parasites in Costa Rica's Vulnerable Communities	33
Abstract	33
INTRODUCTION	34
METHODOLOGY	35
Sources of information on intestinal parasites in schoolchildren from Costa Rica	35
Study population and sampling methods	36
Additional data	37
Construction of Variable Scores	38
Data analysis	39
General statistical description	39
Logistic regression analysis	39
RESULTS	40
Overall gastrointestinal parasites proportions by district	41

Proportions of parasites by sex and age group	44
Proportions by parasite species	45
Logistic Regression analysis	47
DISCUSSION	48
Geographical patterns	49
STH species	49
IP species	50
Age and Sex influence	51
Logistic regression models	51
Consent to participate	54
References	54
ARTÍCULO 2	61
A Machine Learning-Based Framework for Identifying Multidimensional Determinants of Intestinal Parasites in Costa Rican School-Age Children	61
Abstract	61
Introduction	62
Methods	64
Data sources: coproparasitological results database	64
Sources of information on intestinal parasites in schoolchildren from Costa Rica	64
Study population and sampling methods	65
Sources of information on social, economic, environmental, and geographic data (potential predictors)	66
Data analysis process	66
Multivariate analysis	66
Machine learning tools for variable selection	67
Random Forest first selection	67
Final modelling	68
Generalized linear mixed model (GLMM)	68
Results	69
Multivariate analysis: Principal Component Analysis	69
PCA Biplot Interpretation	70
<i>Random Forest Analysis for Soil-Transmitted helminth infections</i>	72
Elastic-Net analysis for the soil-transmitted helminths determinants	74
Generalized Linear Mixed Model (GLMM) and random effects: school and district	76

<i>Soil-transmitted helminth infections model</i>	76
Model Calibration and Fitness	79
A formal predictive model for soil-transmitted infections in schoolchildren	80
Random Forest Analysis for potential predictors of intestinal protozoan infections	81
Variable Importance for Protozoan Intestinal Infections (IP) according to Random Forest Model	82
Elastic Net regression model for the selection of determinants of intestinal protozoan infections	84
Generalized Linear Mixed Model (GLMM) for Protozoan Intestinal Infections (IP)	85
Model calibration and fitness	88
A formal predictive model for intestinal protozoan infections in schoolchildren of Costa Rica	88
Discussion	89
PCA analysis	89
A machine-learning-based pipeline to find the most influential factors associated with intestinal parasitic infections:	91
Random Forest initial predictor selection	91
Elastic-Net analysis for further predictor selection	92
STH and IP final models.	96
References	99
Discusión y conclusión general	107
2.1 Recomendaciones y direcciones futuras	109
Supplementary material A	112
Supplementary material B	117
Supplementary material C	124
Supplementary material D	126

ÍNDICE DE CUADROS

Artículo 1.

Título	Página
Table 1. Distribution of records per district, age group, and sex.	42
Table SA-1. Overview of sampled schools' data.	112
Table SB-1. Socioeconomic variables explained with source references.	117
Table SB-2. Educational variables explained with source references.	118
Table SB-3. Housing conditions variables explained with source references.	119
Table SB-4. Climatic variables explained with source references.	120
Table SB-5. Topographic and geographic variables explained, and source references.	121
Table SB-6. Land configuration variables explained, and source references	123
Table D1. Land configuration variables explained and sources references.	126

Artículo 2.

Título	Página
Table 1. Performance metrics of the Random Forest Model for soil-transmitted helminths infection prediction across training, validation, and test datasets.	73
Table 2. Top 20 predictors of soil-transmitted helminth infections according to mean decrease in Accuracy (Random Forest Model).	74
Table 3. Non-zero coefficients of the selected variables according to Elastic-Net regression analysis in the process of STH infections determinants modelling.	76
Table 4. Results of a fitted generalized linear mixed model for the determinants of STH infections in schoolchildren of Costa Rica.	77
Table 5. Generalized model comparisons between a model with no random effects and the GLMM model with random effects for the STH infections outcome.	78

Título (Continuación)

	Página
Table 6. Performance metrics of the Random Forest Model for intestinal protozoan parasite infection prediction across training, validation, and test datasets.	81
Table 7. Top 20 most important variables influencing protozoan intestinal infections in schoolchildren from Costa Rica, determined by the Random Forest Analysis, based on Mean Decrease in Accuracy.	83
Table 8. Non-Zero Elastic Net Coefficients of the selected variables in the IP model.	85
Table 9. Comparative metrics of the generalized models applied to the selected predictors for the intestinal protozoan infections as outcome.	86
Table 10. Results of the GLMM model for each variable in the IP infections model.	86

ÍNDICE DE FIGURAS

Artículo 1.

Título	Página
Figure 1. Database curation and data selection process.	36
Figure 2. Map showing the locations of all participating schools and the planning regions where they belong.	38
Figure 3. Proportions per school and hotspots of intestinal parasites	43
Figure 4. Heatmap showing proportions of any intestinal parasite (AIP), soil-transmitted helminths (STH), and intestinal protozoans (IP) in schoolchildren of Costa Rica in schools from different regions and districts	44
Figure 5. Proportions and 95% confidence limits of intestinal parasites in schools from vulnerable zones of Costa Rica by age group (years).	45
Figure 6. Intestinal parasite species prevalence in schools from different regions and districts of Costa Rica in vulnerable places.	46
Figure 7. Forest plot visualization for logistic regression models' results. Odds ratios and 95% confidence limits.	47

Artículo 2.

Título	Páginas
Figure 1. Three-dimensional principal component analysis plot shows clustering of observations from schools in each region.	70
Figure 2. PCA loading-vector biplot of predictor variables with supplementary outcome vectors (STH, soil-transmitted helminth infection; IP, intestinal protozoan infection)	71
Figure 3. ROC-AUC curves for the training, validation, and testing datasets to observe the performance of the Random Forest Analysis for the response variable “Soil Transmitted Helminth infections”	73
Figure 4. Elastic Net model for STH infections determinants selection with $\alpha=0.85$ and λ 1.se.	75

Título (<i>Continuación</i>)	Página
Figure 5. Forest plot showing the adjusted odds ratios to see the association of each variable in the model with the outcome variable (STH infection =positive).	79
Figure 6. Calibration curve of the GLMM for STH infections.	80
Figure 7. ROC Curves and AUC for Random Forest Model Predicting Protozoan Intestinal Infections (PR) Across Training, Validation, and Test Sets.	82
Figure 8. Elastic Net model for selecting determinants of intestinal protozoan infections, with alpha=0.5 and the best lambda value (lambda.min).	84
Figure 9. Forest plot for the adjusted ORs and 95% intervals of each variable in the final GLMM model for intestinal protozoan infections in Costa Rican schoolchildren.	87
Figure 10. Calibration curve with LOESS for intestinal parasite infection (IP).	88

ABREVIATURAS

STH = Soil-transmitted helminths (geohelminths)

IP = Intestinal protozoan infections (infecciones por protozoarios intestinales)

AIP = Any intestinal parasite (presencia de cualquier parásito intestinal en la muestra)

DESCRIPTORES

- Parasitología
- Humanos
- Epidemiología
- Parasitosis intestinales
- Factores asociados
- Factores socioeconómicos
- Factores ambientales
- Clima
- Topografía
- Geografía
- Acceso
- Salud pública

INTRODUCCIÓN GENERAL

1.1 Aspectos epidemiológicos generales

Las parasitosis intestinales son infecciones causadas principalmente por protozoarios y helmintos. Estas infecciones se encuentran entre la lista de las enfermedades tropicales desatendidas de la Organización Mundial de la Salud (OMS) y forman parte de sus prioridades con miras a la eliminación hacia el año 2030.

Más de 1500 millones de personas padecen alguna enfermedad intestinal debida a parásitos intestinales. Estas son más comunes en poblaciones en condiciones de pobreza, falta de acceso a servicios de salud y pobre saneamiento ambiental básico (Chen et al., 2024). Las parasitosis intestinales causan además una considerable cantidad de daño a la población medida en años perdidos debido a discapacidad y por el evento de mortalidad prematura (Torgerson et al., 2015; Wong et al., 2020).

1.2 Protozoarios intestinales

Los protozoarios constituyen un grupo heterogéneo de organismos unicelulares y eucariotas, algunos de los cuales están asociados al intestino humano, ya sea como comensales o como patógenos (Chabé et al., 2017). Las especies reportadas con mayor frecuencia en el intestino humano incluyen *Blastocystis hominis*, *Entamoeba coli*, *Endolimax nana*, *Giardia intestinalis*, *Iodamoeba butschlii* y *Entamoeba histolytica* (Bahmani et al., 2017).

Con menor frecuencia, se identifican otras especies tales como *Entamoeba hartmanni*, *Cryptosporidium* spp., *Pentatrichomonas hominis*, *Chilomastix mesnili*, *Dientamoeba fragilis*, *Cystoisospora belli* y *Balantidium coli* (Chabé et al., 2017; da Silva et al., 2021a; Haque, 2007; Hotez, 2018).

En el Cuadro 1 se presentan los principales protozoarios intestinales, su clasificación taxonómica, su carácter patógeno o comensal, el sitio anatómico del tracto digestivo humano en el que se localizan y las patologías más comúnmente asociadas a cada especie (Fitri et al., 2022; Li et al., 2020). Aunque las especies comensales no causan enfermedad, su presencia

indica contaminación fecal, refleja la ingestión previa de alimentos o agua contaminados, o alguna otra forma de exposición oral a heces humanas.

Entre las especies patógenas más frecuentemente identificadas en el intestino humano se encuentran *Giardia intestinalis*, *Entamoeba coli* y *Cryptosporidium* spp. (Chabé et al., 2017; Haque, 2007).

Giardia intestinalis es un protozoo flagelado altamente adaptado a la transmisión hídrica. Sus quistes, producidos en gran cantidad, pueden contaminar fuentes acuáticas y transmitirse con facilidad por la ingestión de agua contaminada (Adam, 2021).

Entamoeba histolytica es una ameba que habita en la región ileocecal y el colon de humanos y animales domésticos como perros y gatos. Su transmisión ocurre mediante la ingestión de quistes presentes en las heces de individuos o animales infectados. Estos quistes pueden sobrevivir en el ambiente por un tiempo prolongado, facilitando la contaminación de manos, alimentos y agua. *E. histolytica* puede causar diversas patologías, siendo la disentería amebiana —caracterizada por diarrea con sangre y moco, resultado de un daño severo a la mucosa intestinal— la más conocida (Baxt & Singh, 2008).

Cabe destacar que esta especie es morfológicamente indistinguible de la ameba comensal no patógena *Entamoeba dispar* cuando se utilizan técnicas convencionales de microscopía de luz. Por esta razón, en los exámenes coproparasitológicos se suele reportar la presencia de *Entamoeba histolytica/E. dispar*, sin diferenciación específica. Esta limitación representa un desafío diagnóstico y terapéutico, ya que la única forma de distinguir entre ambas especies es mediante técnicas de PCR y secuenciación, procedimientos que resultan costosos y poco prácticos para su implementación rutinaria en laboratorios clínicos (Cui et al., 2019; Silva et al., 2021b).

Cryptosporidium spp. es un protozoo apicomplejo cuya transmisión se da mediante la ingestión de ooquistes altamente resistentes, presentes en heces que contaminan el suelo y el agua. Estos ooquistes pueden eventualmente contaminar manos, alimentos y fuentes de agua potable. Su diagnóstico presenta desafíos, y en personas inmunocompetentes, la infección suele ser asintomática, lo que contribuye a su subdiagnóstico. Sin embargo, en niños pequeños con sistemas inmunológicos inmaduros, puede causar cuadros severos de diarrea.

Un metaanálisis realizado por Khalil et al. (2018) identificó a *Cryptosporidium* como la quinta causa más frecuente de diarrea en niños menores de cinco años, atribuyéndosele aproximadamente 48 000 muertes y una elevada carga de morbilidad. Además, la criptosporidiosis se asoció con retrasos importantes en el crecimiento vinculados a episodios diarreicos (Khalil et al., 2018).

Balantidium coli es el único protozoario del filo *Ciliophora* (ciliados) reportado como parásito intestinal en humanos. Se localiza comúnmente en el intestino de los cerdos, y la interacción estrecha entre humanos y estos animales puede facilitar su transmisión. Aunque su presencia en humanos es infrecuente, siendo más común en Asia y América del Sur, es capaz de causar patología intestinal, incluyendo diarrea disentérica. Su transmisión ocurre mediante la ingestión de quistes que sobreviven en el ambiente y contaminan fuentes de agua, alimentos o las manos humanas (da Silva et al., 2021a).

Cuadro 1. Especies de parásitos protozoarios y helmintos asociados con el intestino humano según su clasificación, sitio anatómico, si son patógenos o comensales y patologías específicas asociadas.

Parásito	Sitio anatómico	Clasificación	Patologías asociadas
Protozoarios			
<i>Entamoeba histolytica</i>	CC	P	Disentería amebiana, abscesos hepáticos y pulmonares
<i>Entamoeba dispar</i>	CC	C	Ninguna
<i>Entamoeba hartmanni</i>	CC	C	Ninguna
<i>Entamoeba coli</i>	CC	C	Ninguna
<i>Endolimax nana</i>	CC	C	Ninguna
<i>Iodamoeba bütschlii</i>	CC	C	Ninguna
<i>Cryptosporidium hominis</i>	ID	P	Enteritis, diarrea severa
<i>Cryptosporidium parvum</i>	ID	P	Enteritis, diarrea severa
<i>Cyclospora cayetanensis</i>	CC	P	Diarrea severa
<i>Cystoisospora belli</i>	CC	P	Diarrea severa
<i>Sarcocystis hominis</i>	ID	C	Enteritis
<i>Toxoplasma gondii</i>	QT	p	Enteritis
<i>Blastocystis</i> spp.	ID	C/P	Ninguna, enteritis
<i>Dientamoeba fragilis</i>	CC	P	Enteritis
<i>Pentatrichomonas hominis</i>	C	C	Ninguna
<i>Trichomonas hominis</i>	C	P	Enteritis Mayoría casos leves

Cuadro 1. Continuación			
Parásito	Sitio anatómico	Clasificación	Patologías asociadas
<i>Trichomonas tenax</i>	B	C	Ninguna
<i>Trichomonas gingivalis</i>	B	C	Ninguna
<i>Balantidium coli</i>	CC	P	Colitis severa, diarrea sanguinolenta
<i>Enterocytozoon bieneusi</i>	ID	P	Enteritis, diarrea acuosa profusa
<i>Retortamonas intestinalis</i>	CC	C	Ninguna
<i>Enteromonas hominis</i>	CC	C	Ninguna
<i>Giardia intestinalis</i>	ID	P	Enteritis, diarrea y esteatorrea, síndrome de malabsorción intestinal
Helmintos			
<i>Ascaris lumbricoides</i>	ID	P	Enteritis, obstrucción intestinal, apendicitis, diarrea
<i>Trichuris trichiura</i>	CS, R	P	Colitis, sangrado intestinal bajo, proctitis, prolapso rectal, diarrea
<i>Necator americanus</i>	ID	P	Enteritis, anemia ferropénica, hipoproteinemia, diarrea
<i>Ancylostoma duodenale</i>	ID	P	Enteritis, anemia ferropénica, hipoproteinemia, diarrea
<i>Strongyloides stercoralis</i>	ID, CC	P	Enterocolitis, síndrome de hiperinfección, diarrea
<i>Enterobius vermicularis</i>	CC	P	Colitis, apendicitis, proctitis
<i>Taenia saginata</i>	ID	P	Enterocolitis, diarrea, malabsorción intestinal
<i>Taenia solium</i>	ID, C, QT	P	Enterocolitis, diarrea, malabsorción intestinal
<i>Dipylidium caninum</i>	ID	P	Enterocolitis, diarrea, malabsorción intestinal
<i>Dibothriocephalus latus</i>	ID	P	Enterocolitis, malabsorción, diarrea
<i>Hymenolepis nana</i>	ID	P	Enterocolitis, diarrea
<i>Hymenolepis diminuta</i>	ID	P	Pocos síntomas, dolor abdominal, diarrea

Abreviaturas: B: boca; CC: ciego y colon; CS: colon sigmoides; ID: intestino delgado; QT: quistes tisulares; R: recto. P: patógeno; C: comensal

Los protozoarios comensales más frecuentes en el intestino humano son *Blastocystis hominis*, *Endolimax nana* y *Entamoeba coli* (Amin, 2002). *Blastocystis hominis* habita en el tracto intestinal humano y presenta una notable heterogeneidad genética, con diversos subtipos identificados hasta la fecha (Asghari et al., 2021; Cifre et al., 2018). Para algunos de estos subtipos se ha documentado evidencia de potencial patogenicidad.

De los 17 subtipos conocidos de *Blastocystis*, los subtipos 1 al 9 han sido hallados en humanos, siendo el subtipo 3 el más prevalente. Se ha observado que hasta en un 11 % de los casos de infección intestinal por *Blastocystis*, los pacientes pueden presentar manifestaciones

clínicas tanto intestinales como cutáneas (Seguí et al., 2018). En el caso de *E. nana* y *E. coli*, no se han reportado casos en que se hayan relacionado efectos patógenos con su presencia en el intestino.

1.3 Helmintos intestinales

En el Cuadro 1 se presenta una lista de las principales especies de helmintos intestinales que afectan al ser humano. Las helmintiasis transmitidas por el suelo, también conocidas como geohelmintiasis —entre ellas, la ascariasis causada por *Ascaris lumbricoides*, la tricuriasis por *Trichuris trichiura*, la uncinariasis por *Necator americanus* o *Ancylostoma duodenale*, y la estrongiloidiasis por *Strongyloides stercoralis*— constituyen, con diferencia, las helmintiasis más prevalentes a nivel mundial (Figura 1). Estas parasitosis se caracterizan por su amplia distribución geográfica y su alta capacidad de transmisión al ser humano (Jourdan et al., 2018).

Más de una cuarta parte de la población mundial está en riesgo de adquirir una geohelmintiasis. La Organización Mundial de la Salud (OMS) estimó que, para el año 2010, aproximadamente 875 millones de niños requerían tratamiento periódico contra estas infecciones. Desde entonces, no se ha actualizado el análisis global del impacto de las geohelmintiasis (Chen et al., 2024; Pullan et al., 2014).

Las regiones tropicales, con elevados índices de pobreza y con deficiente acceso a servicios de salud y saneamiento básico, son las más afectadas por este tipo de infecciones. Esta distribución geográfica se asocia con una serie de factores de riesgo presentes en las comunidades afectadas, tales como elevados índices de pobreza, bajos niveles de desarrollo social, acceso limitado al agua potable, restricción en los servicios de salud y deficiente saneamiento ambiental (Jourdan et al., 2018).

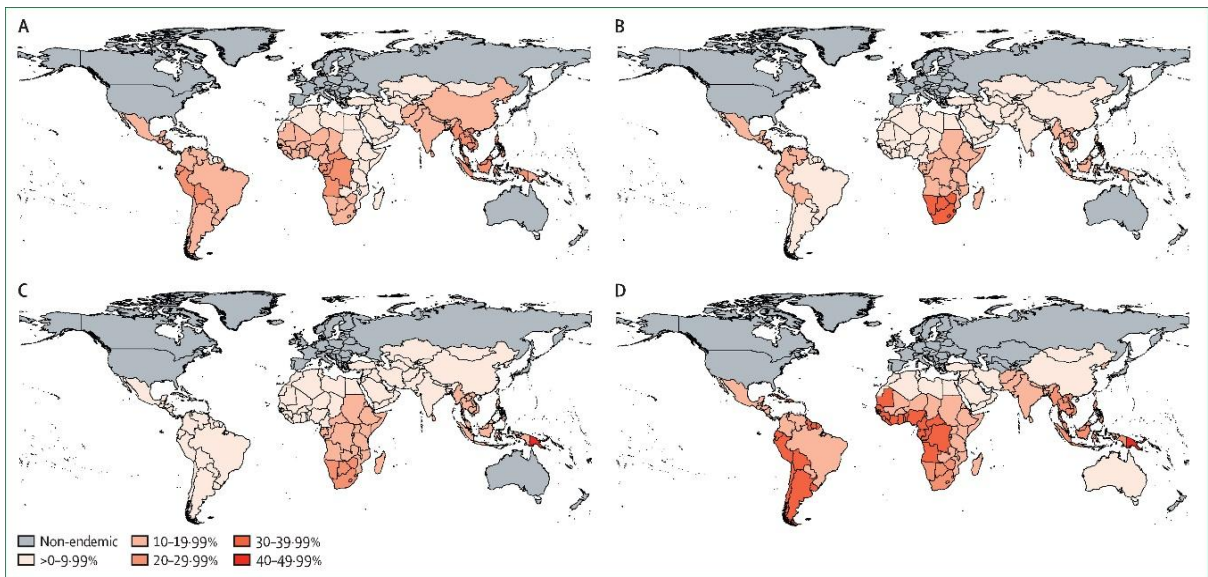


Figura 1. Prevalencias estimadas globales de (A) ascariasis, (B) tricuriasis, (C) uncinarias y (D) estrongiloidiasis. Modificado de OMS, 2015.

En cuanto a los métodos de diagnóstico utilizados en el laboratorio clínico, el frotis directo con solución salina y Lugol, así como la técnica modificada de Kato son los más comunes para la detección de helmintiasis. Se debe considerar que en muchas regiones incluso de nuestro país no se cuenta con el método de Kato lo cual limita significativamente la posibilidad de detectar las parasitosis causadas por helmintos.

Por otra parte, para algunas helmintiasis específicas como la uncinariasis, causada por *Strongyloides stercoralis*, la sensibilidad del frotis directo y el Kato son limitadas, y requieren de la realización del método de Baerman que permite la detección de larvas en las heces (Wong et al., 2020). Debido a la limitada disponibilidad y uso rutinario de esta técnica, la verdadera prevalencia de *S. stercoralis* continúa siendo desconocida en la mayoría de los países, especialmente en aquellos con escasos recursos, lo que representa un vacío significativo en la investigación epidemiológica (Fleitas et al., 2020).

Así mismo, helmintos como *Enterobius vermicularis* requieren métodos de detección especializados. *E. vermicularis* es un nematodo intestinal humano que, por lo general, provoca infecciones con escasa sintomatología. El prurito anal, predominantemente nocturno, es el signo más sugestivo de la infección, aunque no necesariamente el más frecuente, y además constituye un síntoma inespecífico. Las hembras de este helminto

depositan sus huevos en la región perianal, lo que reduce significativamente la probabilidad de detectar el parásito en muestras de heces, contribuyendo así a su subregistro en los estudios coproparasitológicos (Lashaki et al., 2023).

1.4 Importancia de las parasitosis intestinales

Entamoeba histolytica

Se estima que, anualmente, se producen aproximadamente 100 millones de casos de disentería amebiana, colitis o abscesos hepáticos asociados a *Entamoeba histolytica*, con una mortalidad cercana a las 100 000 personas por año (Cui et al., 2019). *E. histolytica* posee diversos factores de virulencia que le permiten adherirse a la mucosa colónica, degradarla y posteriormente invadirla. Esta capacidad invasiva le posibilita, en ciertos casos, migrar a través de la circulación linfática o venosa hacia otros órganos como el hígado, los pulmones o el cerebro, donde puede inducir procesos inflamatorios y lesiones severas (Baxt & Singh, 2008; Kantor et al., 2018).

Desde el punto de vista morfológico, *Entamoeba histolytica* es indistinguible de *E. dispar* y *E. moshkovskii* cuando se utilizan técnicas convencionales de microscopía, lo cual ha conducido a una posible sobreestimación de su prevalencia en estudios epidemiológicos. A diferencia de *E. histolytica*, que puede digerir tejido y sangre, *E. dispar* y *E. moshkovskii* son especies comensales que se alimentan exclusivamente de bacterias intestinales (da Silva et al., 2021b).

Giardia intestinalis

Clínicamente, la giardiasis se manifiesta con mayor intensidad en niños pequeños, pacientes inmunocomprometidos y personas con fibrosis quística. *Giardia intestinalis* (también conocida como *G. lamblia* o *G. intestinalis*) se clasifica en seis grupos genéticos (assemblages); de estos, los subtipos A y B se encuentran tanto en humanos como en animales, mientras que los restantes (C a H) se presentan principalmente en animales como caninos, felinos, roedores y focas (Heyworth, 2016).

Se ha estimado que la ingestión de tan solo 10 a 25 quistes es suficiente para desencadenar una infección sintomática. No obstante, en individuos inmunocompetentes mayores de cinco

años, muchas infecciones pueden resolverse espontáneamente, aunque en algunos casos los síntomas leves pueden persistir durante varias semanas (Allain & Buret, 2020).

G. intestinalis es un protozoo flagelado no invasivo. Tras la ingestión de los quistes, el parásito se reproduce rápidamente de forma asexual, lo que le permite colonizar con eficacia el intestino delgado. Se adhiere mediante un disco adhesivo a las microvellosidades intestinales, provocando daño estructural y alteraciones significativas en la absorción de nutrientes. Estas alteraciones parecen estar asociadas tanto a la respuesta inmunitaria local como a las secreciones del propio parásito, además de la obstrucción mecánica de las superficies absorbentes. Como consecuencia, puede producirse diarrea con alto contenido de grasa y cuadros clínicos de malabsorción (Adam, 2021).

Al igual que ocurre con otros patógenos unicelulares, diferentes cepas de *G. intestinalis* parecen asociarse con diferentes grados de severidad en la interacción entre el parásito, el epitelio intestinal y el sistema inmunológico. Existen cada vez más evidencias que sugieren que ciertos subtipos poseen factores de virulencia que potencian el daño a las microvellosidades intestinales, provocando consecuencias clínicas más significativas. En estos casos, los síntomas pueden persistir durante semanas o incluso meses después de la eliminación completa del parásito (Adam, 2021; Allain & Buret, 2020; Halliez & Buret, 2013).

Ascaris lumbricoides

Comúnmente conocida como lombriz intestinal, es uno de los helmintos parásitos más frecuentes a nivel mundial, posiblemente el más prevalente (Pullan et al., 2014). Las manifestaciones clínicas de la infección por *A. lumbricoides* pueden variar desde formas leves —la mayoría de los casos son asintomáticos— hasta cuadros moderados o severos, dependiendo principalmente de factores del hospedero, como la edad, el estado nutricional y la carga parasitaria.

Durante las distintas etapas de la infección, se produce una compleja interacción entre factores propios del parásito y del hospedero. Durante la migración larval —que sigue la ruta intestino–hígado–pulmón–corazón— pueden presentarse síntomas gastrointestinales, daño hepático leve y signos respiratorios, cuya severidad depende, en gran medida, de la intensidad de la infección.

Las formas graves de ascariasis se asocian, por lo general, con cargas parasitarias extremadamente altas. Estas infecciones pueden desencadenar complicaciones severas, incluyendo obstrucción intestinal, perforación, pancreatitis o colangitis, con riesgo significativo para la vida del paciente (Herbstein et al., 2020).

1.5 Aspectos económicos

Dos de las principales parasitosis intestinales, la criptosporidiosis (infección por *Cryptosporidium* spp.) y la amebiasis (principalmente causada por *Entamoeba histolytica*), fueron responsables de aproximadamente 10,6 millones de DALY's (años de vida ajustados por discapacidad, por sus siglas en inglés) a nivel mundial, según las estimaciones más recientes sobre la carga global de enfermedad por parásitos intestinales, correspondientes al año 2010. Como se muestra en la Figura 2, las geohelmintiasis representaron aproximadamente 5,18 millones de DALY's en ese mismo periodo (Turkeltaub et al., 2015).

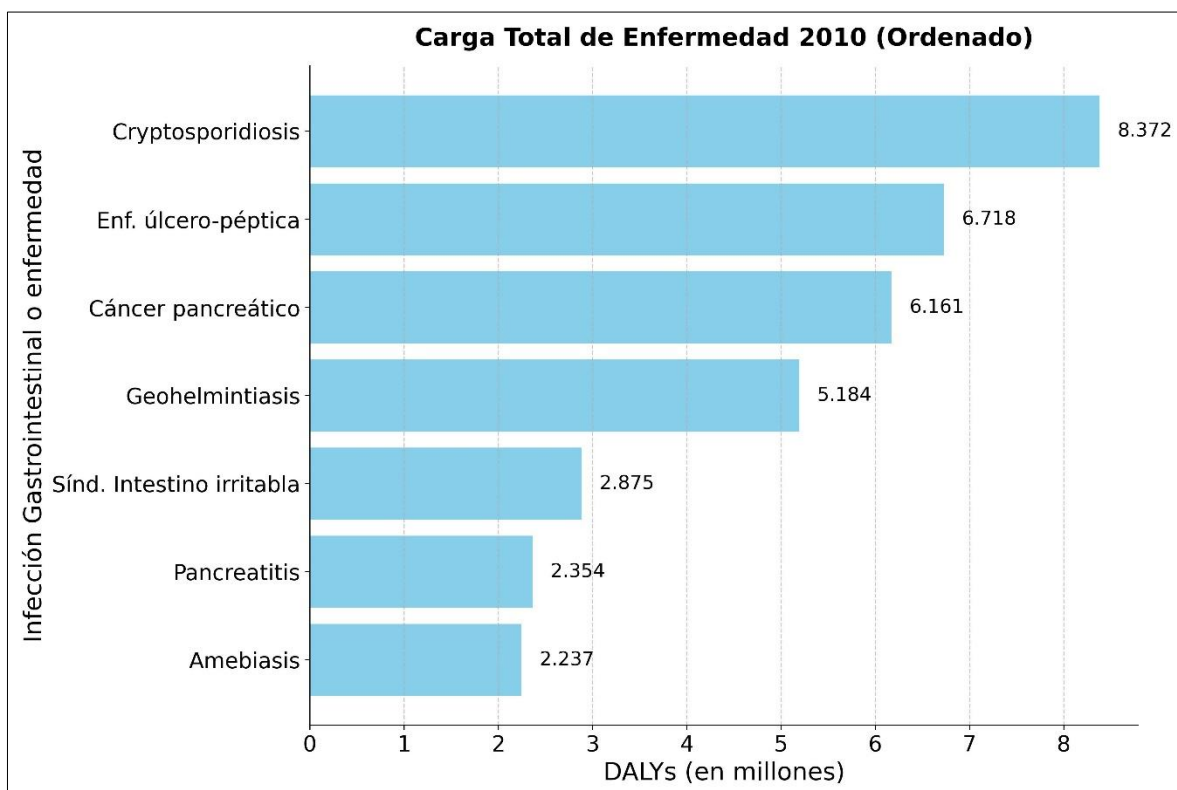


Figura 2. Años perdidos debido a muerte prematura o discapacidad (DALYs por sus siglas en inglés) causados por enfermedades gastrointestinales para el año 2010. Modificado de Turkeltaub et al. (2015).

Estos datos se basan en la detección de estos parásitos mediante exámenes coproparasitológicos convencionales, utilizando microscopía de luz y técnicas habituales como el frotis directo de heces y la técnica modificada de Kato. Sin embargo, se estima que existe un subregistro significativo, dado que no se emplean rutinariamente métodos de mayor sensibilidad, como la reacción en cadena de la polimerasa (PCR), debido a su elevado costo técnico y económico, especialmente en los países donde estas parasitosis son más prevalentes (Turkeltaub et al., 2015).

1.6 Antecedentes de la investigación sobre parásitos intestinales en Costa Rica

Desde inicios del siglo XX, Costa Rica ha incorporado la vigilancia de enfermedades infecciosas en contextos escolares, comenzando con la creación del Departamento de Anquilostomiasis en 1914 (Salas-Conejo & Meneses-Castro, 1977). No obstante, los datos generados sobre parasitosis intestinal han sido históricamente dispersos y escasamente sistematizados. Aunque se llevaron a cabo esfuerzos significativos por parte del Ministerio de Salud y sus instituciones predecesoras, rara vez se desarrollaron investigaciones analíticas que permitieran un aprovechamiento pleno de dicha información (Rodríguez-Herrera, 2005).

Entre las décadas de 1950 y 1980, diversos estudios documentaron altas prevalencias de helmintiasis y protozoos intestinales en poblaciones escolares, pacientes hospitalarios y universitarios (Lizano & Abate, 1953; Jiménez-Quirós, 1958; Ruiz & Vinocour, 1966; Vásquez-Rojas & Zumbado, 1980). Estos trabajos reportaron frecuencias de infección de hasta un 80 %, con predominio de *Trichuris trichiura*, *Giardia intestinalis* y *Entamoeba coli*. Aunque relevantes, estos estudios carecieron de continuidad, cobertura nacional y capacidad para analizar factores de riesgo de forma integral.

Las décadas posteriores aportaron algunas encuestas puntuales y estudios hospitalarios que sugirieron una disminución en la prevalencia de estas infecciones (Bolaños & Freer, 1997; Agüero et al., 2009). Sin embargo, sus limitaciones metodológicas impidieron extrapolar los hallazgos a nivel nacional o distinguir entre contextos rurales y urbanos. Por ejemplo, la Encuesta Nacional de Nutrición 2008-2009 reportó prevalencias de ascariasis y tricuriasis inferiores al 2%, aunque no logró identificar con precisión zonas vulnerables ni potenciales focos de transmisión activa.

La ausencia de estudios actualizados con diseños epidemiológicos robustos genera tres consecuencias críticas: (1) la desparasitación masiva se implementa sin criterios basados en evidencia, lo que podría inducir resistencia antiparasitaria o efectos adversos innecesarios; (2) se desconoce la localización de focos activos de infección y zonas sin datos disponibles; y (3) se produce una falsa percepción de control, que debilita tanto la educación preventiva como la formación continua del personal de salud en el tema de las parasitosis intestinales.

Además de su impacto directo sobre la salud, las parasitosis intestinales pueden afectar negativamente la nutrición, el desarrollo cognitivo y el rendimiento escolar, especialmente en la población infantil. Desde un enfoque ecológico, los parásitos interactúan con la microbiota intestinal, influyendo en procesos inmunológicos, neurológicos y metabólicos (Zaiss & Harris, 2016). Estudios recientes han demostrado que las infecciones intensas por *Giardia* reducen la diversidad microbiana intestinal, mientras que la presencia de helmintos en cargas moderadas podría incrementarla (Mejía et al., 2020).

En Costa Rica, la ausencia de una vigilancia activa y la baja precisión diagnóstica en los sistemas de salud constituyen limitantes importantes para la detección efectiva de estas infecciones. La notificación de estas patologías depende, en su mayoría, de diagnósticos clínicos presuntivos, rara vez confirmados por coproscopía. Además, la formación en parasitología clínica ha disminuido drásticamente desde 2002, y las técnicas diagnósticas más utilizadas —como el frotis directo— presentan baja sensibilidad, especialmente frente a infecciones leves o por especies como *Strongyloides stercoralis* y *Enterobius vermicularis*.

En este contexto, el análisis de una base de datos amplia y anonimizada sobre infecciones parasitarias intestinales en escolares de comunidades vulnerables representa una oportunidad valiosa para caracterizar epidemiológicamente la situación actual en Costa Rica. Asimismo, permite explorar asociaciones con factores sociales y ambientales que podrían incorporarse en modelos predictivos útiles para la toma de decisiones en salud pública y el fortalecimiento de la vigilancia epidemiológica de estas infecciones.

Referencias (Introducción General)

- Adam, R. D. (2021). *Giardia duodenalis*: biology and pathogenesis. *Clin. Microbiol. Rev.*, 34(4), e00019–24. <https://doi.org/10.1128/CMR>
- Agüero, M. L. Á., & de Salud, M. (2009). Parásitos intestinales. In Asencio-Rivera et al. (Eds). Encuesta nacional de nutrición, Costa Rica, 2008-2009. Ministerio de Salud, Instituto Costarricense de Investigación y Enseñanza en Nutrición y Salud, Caja Costarricense de Seguro Social, Instituto Nacional de Estadística y Censos, Instituto Costarricense sobre Drogas. San José, Ministerio de Salud. 79p.
- Allain, T., & Buret, A. G. (2020). Pathogenesis and post-infectious complications in giardiasis. In M. Ortega-Pierres (Ed.), *Adv. Parasitol.*, 107, 173–199. Academic Press. <https://doi.org/10.1016/bs.apar.2019.12.001>
- Amin, O. M. (2002). Seasonal prevalence of intestinal parasites in the United States during 2000. *Am. J. Trop. Med. Hyg.*, 66(6), 799–803. <https://doi.org/10.4269/ajtmh.2002.66.799>
- Asghari, A., Sadeghipour, Z., Hassanipour, S., Abbasali, Z., Hasan Ebrahimzadeh-Parikhani, & Hashemzaei, M., Alimardani, V., Hatam, G., Ebrahimzadeh-Parikhani, H., & Ir, V. A. (2021). Association between *Blastocystis* sp. infection and immunocompromised patients: a systematic review and meta-analysis. *Environ. Sci. Pollut. Res.*, 28, 60308–60328. <https://doi.org/10.1007/s11356-021-16187-1/Published>
- Bahmani, P., Maleki, A., Sadeghi, S., Shahmoradi, B., & Ghahremani, E. (2017). Prevalence of Intestinal Protozoa Infections and Associated Risk Factors among Schoolchildren in Sanandaj City, Iran. *Iran.J. Parasitol.*, 12(1), 108–116. <http://ijpa.tums.ac.ir>
- Baxt, L. A., & Singh, U. (2008). New insights into *Entamoeba histolytica* pathogenesis. *Curr. Opin. Infect. Dis.*, 21(5), 489–494. <https://doi.org/10.1097/QCO.0b013e32830ce75f>
- Bolaños, N., & Freer, E. (1997). Prevalencia de parásitos intestinales en las áreas de salud de Acosta y Coronado. *Rev. Costarric. Cienc. Méd.*, 18(4), 41-49.
- Chabé, M., Lokmer, A., & Ségurel, L. (2017). Gut Protozoa: Friends or Foes of the Human Gut Microbiota? *Trends Parasitol.*, 33(12), 925–934. <https://doi.org/10.1016/j.pt.2017.08.005>

- Chen, J., Gong, Y., Chen, Q., Li, S., & Zhou, Y. (2024). Global burden of soil-transmitted helminth infections, 1990–2021. *Infect. Dis. Poverty.*, 13(1). <https://doi.org/10.1186/s40249-024-01238-9>
- Cifre, S., Gozalbo, M., Ortiz, V., Soriano, J. M., Merino, J. F., & Trelis, M. (2018). *Blastocystis* subtypes and their association with Irritable Bowel Syndrome. *Med. Hypotheses*, 116, 4–9. <https://doi.org/10.1016/j.mehy.2018.04.006>
- Cui, Z., Li, J., Chen, Y., & Zhang, L. (2019). Molecular epidemiology, evolution, and phylogeny of *Entamoeba* spp. *Infect. Genet. Evol.*, 75, 104018. <https://doi.org/10.1016/j.meegid.2019.104018>
- da Silva, R. K. M., Dib, L. V., Amendoeira, M. R., Class, C. C., Pinheiro, J. L., Fonseca, A. B. M., & Barbosa, A. da S. (2021a). Balantidiasis in humans: A systematic review and meta-analysis. *Acta Trop.*, 223, 106069. <https://doi.org/10.1016/j.actatropica.2021.106069>
- da Silva, C. A. V., de Oliveira, I. M. C., Cruz, R. E., Silva Prado, G. K., Santos, F. V., Neves, N. C. V., Gomes, M. A., Silva Oliveira, F. M., & Caliari, M. V. (2021b). South American *Entamoeba dispar* strains produce amoebic liver abscesses with different pathogenicities and evolutionary kinetics. *Acta Trop.* 224, 106114. <https://doi.org/10.1016/j.actatropica.2021.106114>
- Fitri, L. E., Candradikusuma, D., Setia, Y. D., Wibawa, P. A., Iskandar, A., Winaris, N., & Pawestri, A. R. (2022). Diagnostic Methods of Common Intestinal Protozoa: Current and Future Immunological and Molecular Methods. *Trop. Med. Infect. Dis.*, 7(10), 253–265. <https://doi.org/10.3390/tropicalmed7100253>
- Fleitas, P. E., Travacio, M., Martí-Soler, H., Socías, M. E., Lopez, W. R., & Krolewiecki, A. J. (2020). The *Strongyloides stercoralis*-hookworms association as a path to the estimation of the global burden of strongyloidiasis: A systematic review. *PLoS Negl. Trop. Dis.*, 14(4), 1–13. <https://doi.org/10.1371/journal.pntd.0008184>
- Halliez, M. C. M., & Buret, A. G. (2013). Extra-intestinal and long term consequences of *Giardia duodenalis* infections. *World J. Gastroenterol.*, 19(47), 8974–8985. <https://doi.org/10.3748/wjg.v19.i47.8974>

- Haque, R. (2007). Human intestinal parasites. *JHPN*, 25(4), 387–391.
- Herbstein, J. A., Breglia, G. A., Rodrigues Paquete, C. A., & Vega, A. (2020). Death by asphyxiation caused by *Ascaris lumbricoides* case report. *Forensic Sci. Int. Rep.*, 2, 100084. <https://doi.org/10.1016/j.fsir.2020.100084>
- Heyworth, M. F. (2016). *Giardia duodenalis* genetic assemblages and hosts. *Parasite*, 23, 1–5. <https://doi.org/10.1051/parasite/2016013>
- Hotez, P. J. (2018). Human Parasitology and Parasitic Diseases: Heading Towards 2050. *Adv. Parasitol.*, 100, 29–38. <https://doi.org/10.1016/bs.apar.2018.03.002>
- Jiménez-Quirós, O., & Brenes, R. R. (1958). Parasitosis intestinal en el universitario costarricense I. Helmintiasis. *Rev. Biol. Trop.*, 6(1), 113-122.
- Jourdan, P. M., Lamberton, P. H. L., Fenwick, A., & Addiss, D. G. (2018). Soil-transmitted helminth infections. *The Lancet*, 391(10117), 252–265. [https://doi.org/10.1016/S0140-6736\(17\)31930-X](https://doi.org/10.1016/S0140-6736(17)31930-X)
- Kantor, M., Abrantes, A., Estevez, A., Schiller, A., Torrent, J., Gascon, J., Hernandez, R., & Ochner, C. (2018). *Entamoeba histolytica*: Updates in clinical manifestation, pathogenesis, and vaccine development. *Can. J. Gastroenterol. Hepatol.*, 2018(1), 4601420. <https://doi.org/10.1155/2018/4601420>
- Khalil, I. A., Troeger, C., Rao, P. C., Blacker, B. F., Brown, A., Brewer, T. G., Colombara, D. V., De Hostos, E. L., Engmann, C., Guerrant, R. L., Haque, R., Houpt, E. R., Kang, G., Korpe, P. S., Kotloff, K. L., Lima, A. A. M., Petri, W. A., Platts-Mills, J. A., Shoultz, D. A., ... Mokdad, A. H. (2018). Morbidity, mortality, and long-term consequences associated with diarrhoea from *Cryptosporidium* infection in children younger than 5 years: a meta-analysis study. *Lancet Glob. Health*, 6(7), e758–e768. [https://doi.org/10.1016/S2214-109X\(18\)30283-3](https://doi.org/10.1016/S2214-109X(18)30283-3)
- Lashaki, E. K., Mizani, A., Hosseini, S. A., Habibi, B., Taherkhani, K., Javadi, A., Taremiha, A. R., & Dodangeh, S. (2023). Global prevalence of enterobiasis in young children over the past 20 years: a systematic review and meta-analysis. *Osong Public Health and Res. Perspect.*, 14(6), 441–450. <https://doi.org/10.24171/j.phrp.2023.0204>

- Li, J., Wang, Z., Karim, M. R., & Zhang, L. (2020). Detection of human intestinal protozoan parasites in vegetables and fruits: a review. *Parasit. vectors*, 13(1), 1–19. <https://doi.org/10.1186/s13071-020-04255-3>
- Lizano, C., & Abate, J. D. (1953). Incidencia de parásitos intestinales en los niños de la sección de pediatría del Hospital San Juan de Dios. *Rev. Biol. Trop.*, 1(2), 223-333. <https://doi.org/10.15517/rev.biol.trop.1953.28782>
- Mejía, R., Damania, A., Jeun R., Bryan P.E., Vargas P., Juarez M., Cajal P.S., Nasser J., Krolewiecki A., Lefoulon E., Long, C., Drake, E., Cimino R.O. & Slatko B. (2020). Impact of intestinal parasites on microbiota and cobalamin gene sequences: a pilot study. *Parasit. Vectors*, 13(1), 200. <https://doi.org/10.1186/s13071-020-04073-7>
- Pullan, R. L., Smith, J. L., Jasrasaria, R., & Brooker, S. J. (2014). Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. *Parasit. vectors*, 7(1), 1–19. <https://doi.org/10.1186/1756-3305-7-37>
- Rodríguez-Herrera, A. (2005). La reforma de salud en Costa Rica. Santiago, CEPAL. 50p.
- Ruiz, A., & Vinocour, E. (1966). Protozoosis intestinales en universitarios costarricenses. *Acta Méd. Costarric.*, 9(3), 157-164.
- Salas-Conejo, R., & Meneses-Castro, R. (1977). Memoria del cincuentenario del ministerio de salud y de la salud pública en Costa Rica. Unidad de Planificación, Ministerio de Salud de Costa Rica.
- Seguí, R., Muñoz-Antoli, C., Klisiowicz, D. R., Oishi, C. Y., Köster, P. C., de Lucio, A., Hernández-de-Mingo, M., Puente, P., Toledo, R., Esteban, J. G., & Carmena, D. (2018). Prevalence of intestinal parasites, with emphasis on the molecular epidemiology of *Giardia duodenalis* and *Blastocystis* sp., in the Paranaguá Bay, Brazil: a community survey. *Parasit. vectors*, 11(1), 490. <https://doi.org/10.1186/s13071-018-3054-7>
- Torgerson, P. R., Devleeschauwer, B., Praet, N., Speybroeck, N., Willingham, A. L., Kasuga, F., Rokni, M. B., Zhou, X.-N., Fèvre, E. M., Sripa, B., Gargouri, N., Fürst, T., Budke, C. M., Carabin, H., Kirk, M. D., Angulo, F. J., Havelaar, A., & de Silva, N. (2015). World Health Organization Estimates of the Global and Regional Disease Burden of

11 Foodborne Parasitic Diseases, 2010: A Data Synthesis. *PLOS Med.*, 12(12), e1001920. <https://doi.org/10.1371/journal.pmed.1001920>

Turkeltaub, J. A., McCarty, T. R., & Hotez, P. J. (2015). The intestinal protozoa: Emerging impact on global health and development. *Curr. Opin. Gastroenterol.*, 31(1), 38–44. <https://doi.org/10.1097/MOG.0000000000000135>

Vásquez-Rojas, A. R., & Zumbado, H. (1980). Parasitosis intestinal en Cartago y Limón. *Acta Méd. Costarric.*, 23(3), 245-251.

Wong, L. W., Ong, K. S., Khoo, J. R., Goh, C. B. S., Hor, J. W., & Lee, S. M. (2020). Human intestinal parasitic infections: a narrative review on global prevalence and epidemiological insights on preventive, therapeutic and diagnostic strategies for future perspectives. *Expert Rev. Gastroenterol. Hepatol.*, 14(11), 1093–1105. <https://doi.org/10.1080/17474124.2020.1806711>

Zaiss, M.M. & Harris N.L. (2016). Interactions between the intestinal microbiome and helminth parasites. *Parasite Immunol.*, 38(1), 5-15. <https://doi.org/10.1111/pim.12274>

ARTÍCULO 1

(Formato: Revista Trop. Med. Int. Health)

<https://onlinelibrary.wiley.com/page/journal/13653156/homepage/forauthors.html>

Beyond the Illusion of Control: Gastrointestinal Parasites in Costa Rica's Vulnerable Communities.

Alberto Solano-Barquero^{1,2}, Juan José Romero-Zúñiga³, Bernardo Vargas-Leitón³, Alicia Rojas^{1,2}, Javier Mora^{1,2}.

¹Laboratorio de Helmintología, Faculty of Microbiology. University of Costa Rica. San José, Costa Rica.

²Centro de Investigación en Enfermedades Tropicales. University of Costa Rica. San José, Costa Rica.

³Programa de Investigación en Medicina Poblacional, Escuela de Medicina Veterinaria. Universidad Nacional. Heredia, Costa Rica.

Corresponding author: Alberto Solano-Barquero. albesol@ucr.ac.cr

Abstract

Introduction: Intestinal parasites are prevalent worldwide, having an especially high burden on poor tropical countries. Although Costa Rica's health system has helped to reduce the frequency of these infections in some central urbanized and rural areas, the problem in vulnerable zones has not been epidemiologically characterized over the last two decades, and numerous reports from these places exist. **Methods:** A curated database of fecal test results from children in vulnerable zones of Costa Rica between 2008 and 2019 was organized. Sociodemographic, climatic, land configuration, educational, housing conditions, and topographic variables for each observation were added from public datasets. Descriptive epidemiological and comparative analyses were performed. Using QGIS, maps of the prevalence of protozoan and helminth infections were elaborated. **Results:** A high prevalence of intestinal parasites (56.4%) was found among schoolchildren from vulnerable districts, with protozoan and helminth main hotspots in Chirripó and Limoncito. Logistic regression revealed that social, geographic, and environmental variables significantly influenced infection risk. **Conclusions:** Data demonstrates the persistence of high rates of intestinal parasites in vulnerable communities of Costa Rica, highlighting potential associations that reveal the multiple dimensions of the problem.

Keywords: intestinal parasites, schoolchildren, Costa Rica, protozoans, soil-transmitted helminths, epidemiology

INTRODUCTION

Intestinal parasites are one of the most common infections worldwide, affecting more than 1.4 billion people worldwide [1,2]. Schoolchildren are the most vulnerable to these infections. A substantial body of evidence indicates that intestinal parasites pose a significant burden to children, impeding their physical and cognitive development and overall well-being [3-5].

Intestinal parasitic infections in children are a complex phenomenon and have many determinants, as demonstrated by a large body of evidence around the world [6-9]. Many of those determinants directly relate to poverty and difficult access to general health services and environmental sanitation. These elements are not the only ones involved, and, possibly, several idiosyncrasies in each regional or small-scale setting can be contributing to parasite infections [10,11].

Costa Rica is a small country in Central America where social security improvements in the 1940s to 1980s helped to make remarkable changes in the general health indicators for the country, including a substantial rise in the population's general longevity [12]. Epidemiological information on parasitic intestinal infections in Costa Rica depends almost entirely on a passive reporting system of incident cases [13], making it challenging to research the phenomena of risk factors linked to these infections in our nation [14].

Available public data have shown that the epidemiology of infectious and chronic diseases in this country changed radically since social security improvements in the early 1940's, towards a higher prevalence of chronic diseases and cancer and an important lowering of prevalences and incidences of infectious diseases, including intestinal parasitic infections, at least in some specific urbanized regions [15]. This was demonstrated by subsequent official nutritional and intestinal parasite surveys over the years, which showed a decrease in the prevalence of parasitic infections, particularly in soil-transmitted helminths. However, some specific locations in Costa Rica are poorly represented in those official reports, mainly those hard-to-access locations with certain other characteristics that made them particularly vulnerable to intestinal parasitic infections [16, 17].

Since 2008, Universidad de Costa Rica's Social Affairs office has supported the Faculty of Microbiology in developing projects to increase awareness of intestinal parasite infections in vulnerable areas of the country by providing medical and educational training on appropriate treatment methods. These projects have been focused on schoolchildren living in these vulnerable communities around the country [18]. Data on the results of these diagnostic procedures and the population characteristics from which these samples came have accumulated over more than a decade. Such an important amount of information constitutes an excellent opportunity to improve the knowledge about intestinal parasitic infections in schoolchildren from these vulnerable communities when combined with official data on the population's environmental, social, demographic, and educational aspects.

This study aims to provide information on the basic epidemiology of intestinal parasitic infection in schoolchildren from some vulnerable zones in Costa Rica and perform an initial approach to the complexity of the relations between intestinal parasitic infections and their potential determinants.

METHODOLOGY

Sources of information on intestinal parasites in schoolchildren from Costa Rica

Thirty-six databases containing schoolchildren's fecal test results were revised and anonymized to create a curated database. Data came from the University of Costa Rica's official results reports on intestinal parasites in schoolchildren from seven regions of Costa Rica and obtained information on parasite species, children's age and sex, school, and educational level for each record. Inclusion criteria were age between 4 and 16 years old, having complete information in the record, including: sex, school, age, and official result of the coprological test. Records that could not be linked to the official reports, records with incomplete information, and records from students who were sampled more than once were among the exclusion criteria. A total of 4812 records were included in the study. A flow diagram of the selection process can be seen in Figure 1.

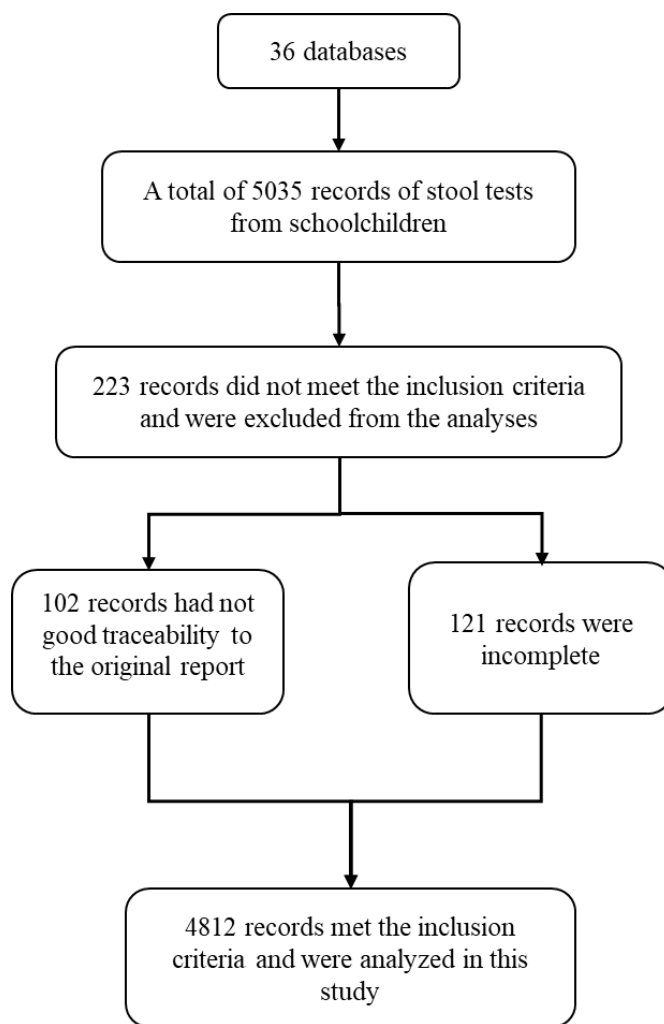


Figure 1. Database curation and data selection process.

Study population and sampling methods

The database's observations came from the results of coproparasitological tests conducted on schoolchildren aged 4 to 16 who attended schools in one of 24 districts located throughout the nation's seven designated health planning zones or health regions (Central Norte, Central Sur, Central Este, Huetar Caribe, Huetar Norte, Chorotega, and Brunca).

All the children were enrolled in their local school, and sample collection was part of official University of Costa Rica community projects (TC-552 and ED-511) that aimed to improve the diagnosis and treatment of intestinal parasite infections in the nation's impoverished communities. A map showing the locations of all participating schools and the regions to which they belong is displayed in Figure 2.

Every child in every school was allowed to provide one fecal sample. All collected samples were transported to the Universidad de Costa Rica's Faculty of Microbiology's Laboratorio de Helminología for processing. Each sample was examined using a fecal smear (for intestinal protozoans) and a modified Kato-Katz technique, which involved spreading 60 mg of feces across a large slide, covering it with cellophane impregnated with Kato's solution, and pressing it against a hard surface. For ten minutes, each sample was screened for helminth eggs. Only the presence of parasite species is reported in this modification; eggs are not counted. A fecal sample was obtained from a varying number of students in each school; Table SA-1 in Supplementary Materials A shows all the information regarding the sample fraction for each school.

Additional data

Additional data on several types of variables were included in the database. A climatic analysis, obtained by remote sensing climate stations, was provided by the Instituto Meteorológico Nacional (IMN). These stations were mapped and the information from the nearest station to each school was used. Monthly and yearly average and max temperatures, humidity, ultraviolet radiation, days with rain, and total rain measures were obtained. The Instituto Nacional de Estadística y Censos (INEC: [National Institute of Statistics and Census]) provided additional information on socioeconomic, educational, and demographic data to use as variables for the study [19]. Supplementary material B (Tables SB-1 through SB-11) has details on those variables.

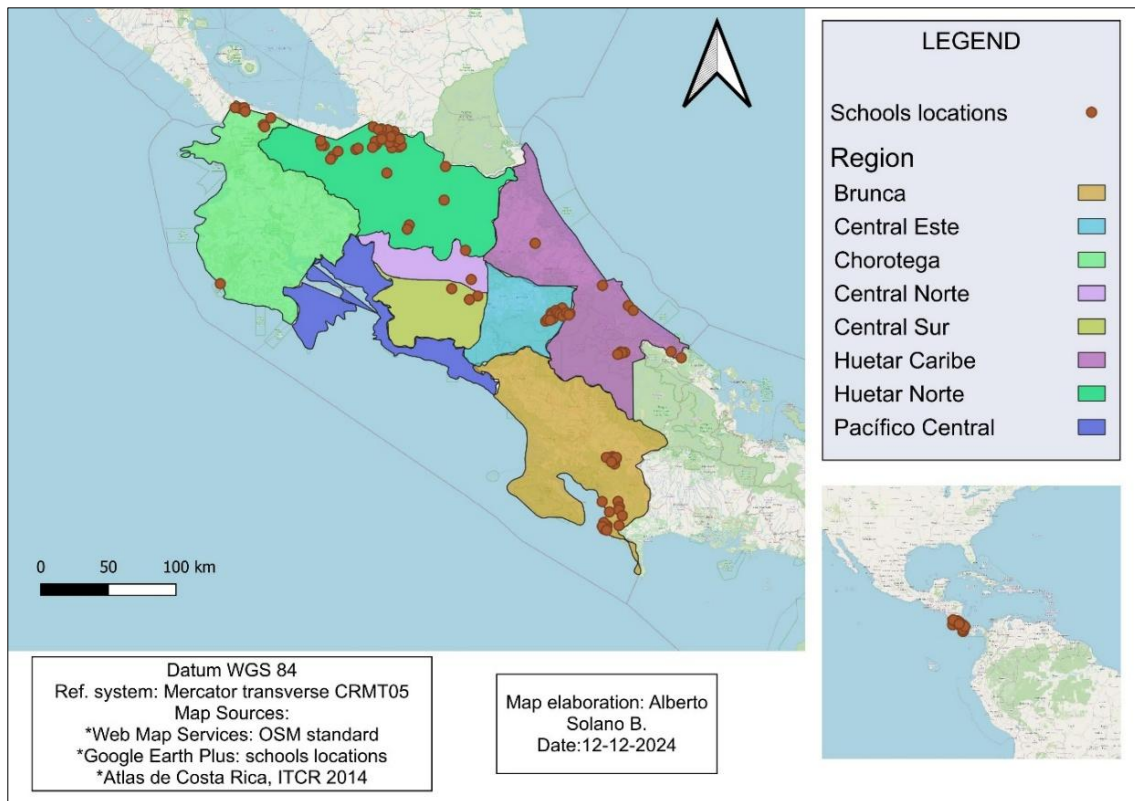


Figure 2. Map showing the locations of all participating schools and the planning regions where they belong. No records from Pacifico Central were available. There were schools from all regions except the Pacifico Central Region.

GIS systems (Google Earth Pro, QGIS 3.4) and satellite images provided by the National Geographic Institute were used to obtain topographic and land configuration data variables defined for this study [20]. All variables and their codification can be seen in Supplementary Material A (Table SA-1).

Construction of Variable Scores

With the additional data previously mentioned, scores of socio-economic (SEC), education (SED), housing condition (SHC), land configuration (SLC), climatic (SCLIM), and a difficulty of access index (DAI) were constructed based on the data mentioned above. Other variables include school, sex, age group, year of the record, and educational level (preschool or primary school). Scores were constructed in a way that higher values mean better or theoretically protective conditions against intestinal parasites (socioeconomic, housing, schooling, land configuration, and access). The climatic score was constructed,

giving more weight to humidity, rainfall, and higher temperatures. Details on the construction of the scores can be read in Supplementary Material B.

Data analysis

General statistical description

The proportions of children infected with intestinal protozoa (IP), soil-transmitted helminths (STH), and any intestinal parasite (AIP) were calculated. These proportions do not represent actual prevalences, either at the district level or at the regional level, because the data did not come from a randomized probabilistic design, where population size was neither previously calculated nor reached those levels. Even so, data was grouped by district and region to facilitate comprehension of the location of the sampled schools.

As substantial differences in sample sizes across schools and subgroups (sex, age, schools, and regions) were detected, bootstrap resampling procedures were implemented to compare the proportions among different subgroupings. To do this, we first calculated the 25th percentile of the sample sizes across the interest groupings and used this value as the target sample size for each school during resampling.

For each iteration, we randomly selected (with replacement) a subsample of this fixed size from each school, ensuring that all schools contributed equally to the estimation process, regardless of their original sample size. We repeated this procedure 10,000 times and calculated the distribution of the difference in proportions in each comparison. Statistical significance of differences in frequencies estimated between sex, age group, and regions was calculated through a chi-square test or Fisher's exact test if necessary; all calculations were made using R software (rsample, base, and tidy models packages) [21].

Using QGIS 3.4, maps of frequencies by school were constructed to reveal parasite hotspots among the regions studied for IP and STH infections [22].

Logistic regression analysis

Predictor variables were made numeric (a few categorical variables were converted to dummy variables) and standardized by the max-min method. General variables Region, Sex, Level (preschool or primary), Age group and constructed scores: Socioeconomic (SEC), Educational (SED), climatic (SCLIM), Housing conditions (SHC), Land configuration

(SLC), difficulty of access index (DAI), and the official Index of Social Development (SDI) were used as explanatory variables for a Logistic Regression Analysis.

This study aimed to explore how different groups of variables affect the outcomes of intestinal protozoan infections (IP) and soil-transmitted helminth infections (STH). Odds ratios and 95% confidence intervals were obtained and visualized as a forest plot.

Two distinct models were developed using response variables: one for any intestinal protozoan infection (IP) and one for STH infections (any helminthic infection). Age group, sex, educational attainment, and region all needed reference definitions for categorical variables. For the age group, the reference group was 7 to 9 years old, and the categories were 4 to 6, 7 to 9, 10 to 12, and >12 years old. Since some of the "Central zone" regions lacked sufficient data, we combined all the central regions into one category, "Central," which served as the reference category. Primary was the reference category for level, while male was the reference group for sex.

All statistical analyses were conducted in R statistics software, R (4.4.2), by using the packages base, rsample, stats, broom, car, tidymodels and dplyr. Visualizations were made with R's ggplot package or with the matplotlib package of Python [23].

RESULTS

A total of 4812 fecal test results were selected based on the inclusion and exclusion criteria. Table 1 shows the overall population characteristics from which those records were gathered. Districts with larger samples were Chirripó, Limoncito, Pavones, Los Chiles, Sixaola, Telire, Yolillal, and Delicias. The least represented districts are Valle la Estrella, Florencia, Amparo, and Caño Negro. There is a well-balanced representation among sexes, while by age groups, the least represented group is the >12 years old, with 393 records (Table 1).

Supplementary material A (Table SA-1) can be consulted for other detailed information on the schools and sampling fractions by school and region. The general proportion of any intestinal parasite AIP was 56.4% (95% CI = 55.0-57.8%). Besides, the overall proportion of pathogenic protozoans was 20.5% (95% CI = 19.4-21.6%). A map of the general proportions

of intestinal protozoans and soil-transmitted helminths can be seen in Figure 3, which also shows the location of each sampled school.

Available data show at least five protozoan infection hotspots and three for helminth infections; these were defined as regions where the prevalence of these infections in schoolchildren is above 15.0% (in helminthic infections) or above 30.0% in the case of protozoan infections (Figure 3).

Overall gastrointestinal parasites proportions by district

In Figure 4, a heatmap shows the mean proportions of AIP, STH, and IP infections for schools of each district within each region. The highest proportions of AIP-infected schoolchildren are found in schools from Limoncito (73.2%; 95% CI = 70.1-76.2%) and Chirripó (71.5%; 95% CI = 69.3-73.7%).

Schools from the same districts had the highest proportions of STH infections in schoolchildren, 39.0% (95% CI=34.8-41.4%) for Limoncito, and 36.6% (95% CI = 34.3-39.0%) for Chirripó. IP infections were also highly prevalent in schools from these districts.

Still, schools from all the districts had prevalences above 10.0%, and many of them above 20.0%, particularly Telire and Río Blanco in Huetar Atlántica region, also Río Azul, Pital, and Delicias in Huetar Norte region (Figure 4). Supplemental material Table SC-2 contains all the prevalences by school for AIP, IP, and STH infections, while Table SC-3 contains all the prevalences by school for each parasite species.

The proportion of IP infections showed greater variation both between and within regions compared to AIP and STH infections. On the other hand, the proportion of helminth infections was mainly below 10%, with many cases at 5% or less (Figure 4).

Differences in prevalences, both by region and by district, for all response variables (AIP, STH, IP) were statistically significant (chi-square test $p < 0.0001$ in all cases).

Table 1. Distribution of records per district, age group and sex.

District	4 to 6 yo			7 to 9 yo			10 to 12 yo			>12 yo			TOTAL n (%)
	Female n	Male n	Total n	Female n	Male n	Total n	Female n	Male n	Total n	Female n	Male n	Total n	
Chirripó	150	160	310	384	314	698	223	228	451	95	87	182	1641 (34.1)
Limoncito	115	126	241	157	162	319	105	115	220	23	35	58	838 (17.4)
Los Chiles	33	42	75	136	146	282	94	112	206	36	44	80	643 (13.4)
Pavón	41	35	76	66	63	129	42	40	82	5	5	10	297 (6.2)
Yolillal	23	24	47	29	26	55	29	30	59	4	3	7	168 (3.5)
Delicias	20	19	39	19	31	50	26	36	62	2	6	8	159 (3.3)
Sixaola	2	17	19	42	41	83	11	22	33	0	0	0	135 (2.8)
Telire	21	22	43	21	35	56	13	15	28	1	1	2	129 (2.7)
Uruca	14	4	18	21	18	39	21	7	28	11	0	11	96 (2.0)
La Cruz	9	15	24	9	12	21	15	19	34	5	6	11	90 (1.9)
Cuajiniquil	4	12	16	9	6	15	15	13	28	2	3	5	64 (1.3)
Santa Cecilia	8	11	19	9	7	16	13	13	26	0	3	3	64 (1.3)
San Isidro	4	7	11	9	21	30	9	4	13	0	1	1	55 (1.1)
Upala	7	12	19	6	10	16	6	7	13	3	2	5	53 (1.1)
San Rafael	12	12	24	7	9	16	2	8	10	0	0	0	50 (1.0)
Limón	9	6	15	10	10	20	6	3	9	0	0	0	44 (0.9)
La Garita	7	6	13	10	5	15	5	8	13	0	2	2	43 (0.9)
Pital	4	6	10	12	12	24	6	2	8	0	0	0	42 (0.9)
Río Blanco	5	8	13	8	6	14	5	6	11	0	0	0	38 (0.8)
Amparo	2	1	3	8	14	22	4	4	8	3	1	4	37 (0.8)
Río Azul	7	4	11	8	12	20	3	2	5	0	1	1	37 (0.8)
Caño Negro	5	4	9	7	8	15	5	4	9	0	0	0	33 (0.7)
Florencia	2	6	8	8	10	18	4	0	4	0	0	0	30 (0.6)
Valle La Estrella	5	2	7	2	4	6	6	4	10	0	3	3	26 (0.5)
TOTAL	509	561	1070	997	982	1979	668	702	1370	190	203	393	4812 (100.0)

Abbreviations: yo= years old

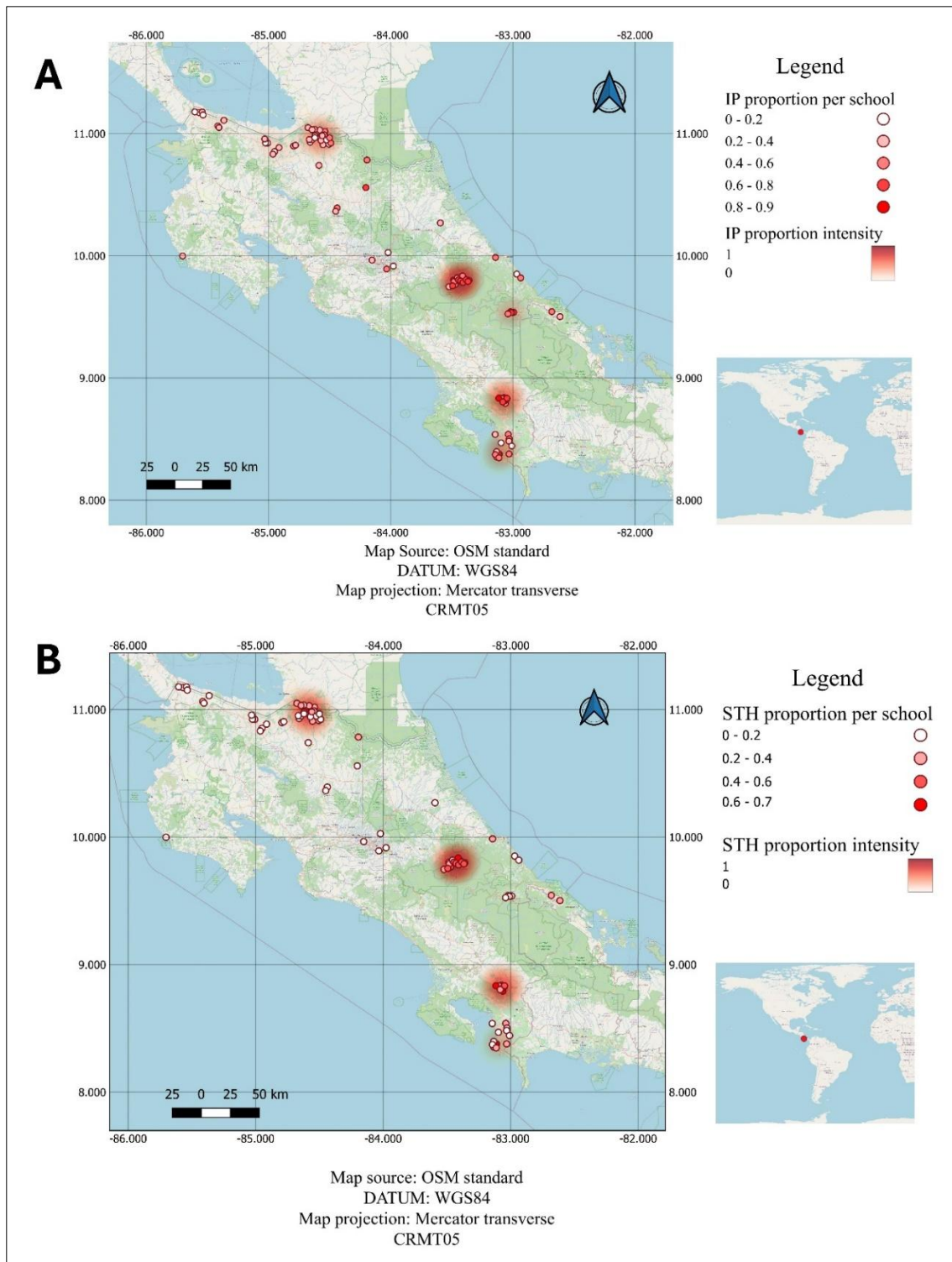


Figure 3. Proportions per school and hotspots of intestinal parasites, A) intestinal protozoan and B) soil-transmitted helminths in schoolchildren from Costa Rica (data gathered from 2008-2019). IP = Intestinal protozoans; STH = Soil-transmitted helminths.

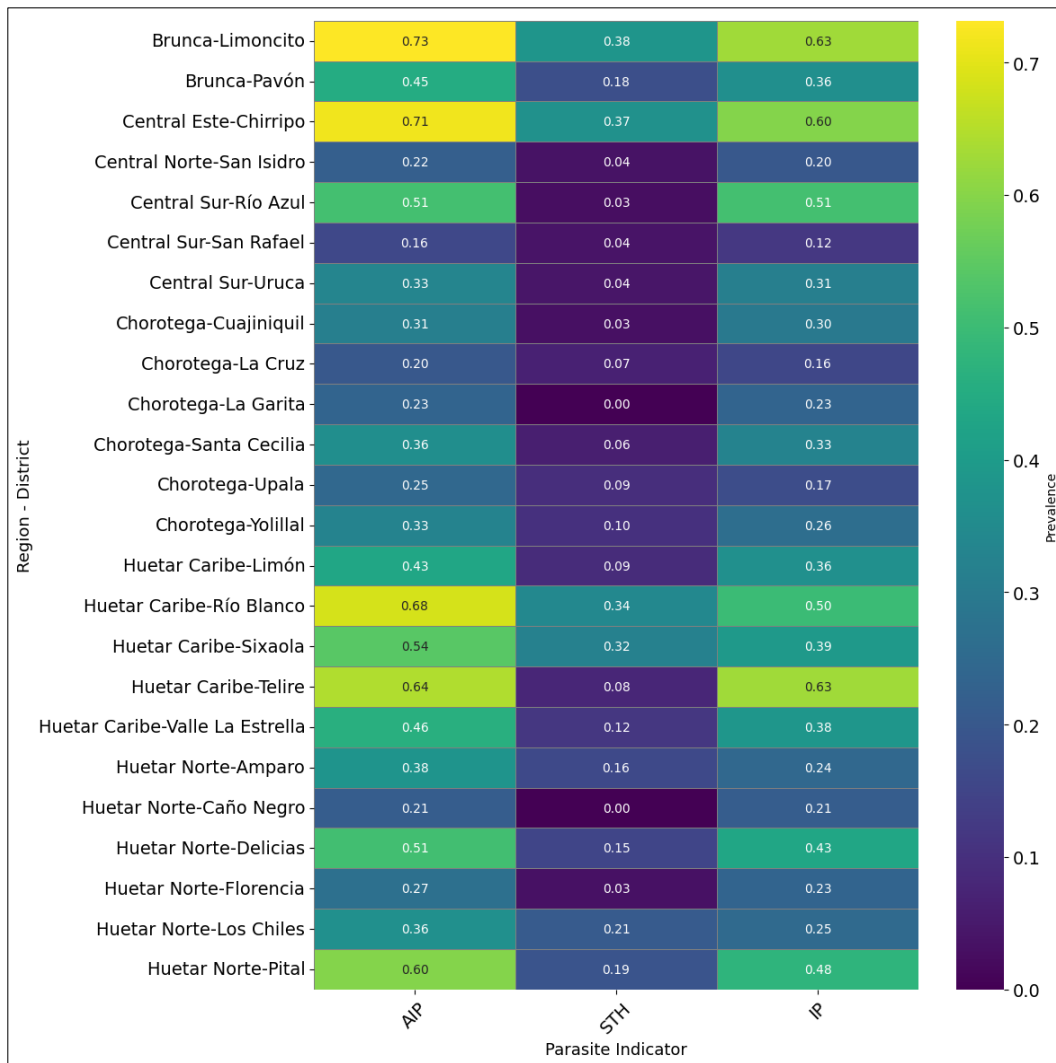


Figure 4. Heatmap showing proportions of any intestinal parasite (AIP), soil-transmitted helminths (STH), and intestinal protozoans (IP) in schoolchildren of Costa Rica in schools from different regions and districts (bootstrapped calculation using percentile 25 as minimum to balance the difference in number of observations between schools).

Proportions of parasites by sex and age group

The Overall proportion of AIP, STH, and IP infections showed no significant differences by sex (AIP $X^2 = 0.03$, $p = 0.85$; STH $X^2 = 0.56$, $p = 0.45$; IP $X^2 = 0.53$, $p = 0.47$). Differences in the overall proportion by age group can be observed in Figure 5. Younger schoolchildren (especially those 7-9 years old) had statistically significantly higher proportions of AIP, STH, and IP (AIP $X^2 = 33.25$, STH $X^2 = 21.22$, IP $X^2 = 22.44$, all p -values < 0.0001) (Figure 5).

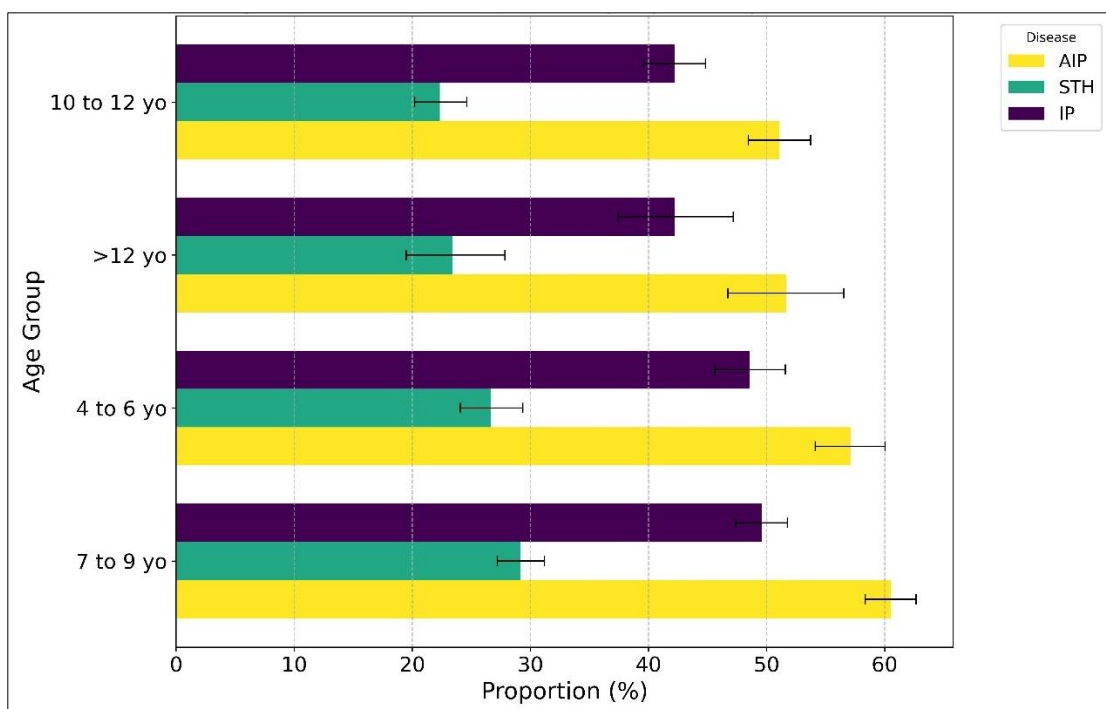


Figure 5. Proportions and 95% confidence limits of intestinal parasites in schools from vulnerable zones of Costa Rica by age group (years). AIP = any gastrointestinal parasite, STH = soil-transmitted helminths, IP = intestinal protozoan parasites (bootstrapped calculation using percentile 25 as minimum to balance the difference in number of observations between age groups).

Proportions by parasite species

Among intestinal protozoans, overall proportions show that the most frequent intestinal protozoan species are *Endolimax nana* and *Entamoeba coli*, both commensal parasites. *E. nana* had an overall proportion of 23.0% (95% CI = 21.8-24.2%), and *E. coli* had an overall proportion of 19.9 % (95% CI = 18.8-21.0%). In the case of pathogenic protozoans, *Giardia intestinalis* and *Entamoeba histolytica* had overall proportions of 13.5% (95% CI = 12.5-14.4%) and 8.7% (95% CI = 7.9-9.5%), respectively (Figure 6).

Regarding soil-transmitted helminths in schoolchildren from the study population, all districts showed hookworm infection proportions less than 10.0%. *Ascaris lumbricoides* was the most prevalent species with an overall proportion of 19.9% (95% CI = 18.4-20.7%). The overall proportions of *Trichiuris trichiura* and hookworms were lower, with proportions of 9.2% (95% CI = 8.3-10.0%) and 3.1% (95% CI = 2.6-3.6%), respectively. In Figure 6, the proportions of each parasite species compared by region and district are shown.

Again, schools of Limoncito and Chirripó showed the highest proportions for the most common intestinal parasites, with comparatively very high proportions of STH (mainly ascariasis), and IP infections by *E. nana* and *E. coli* (Figure 6).

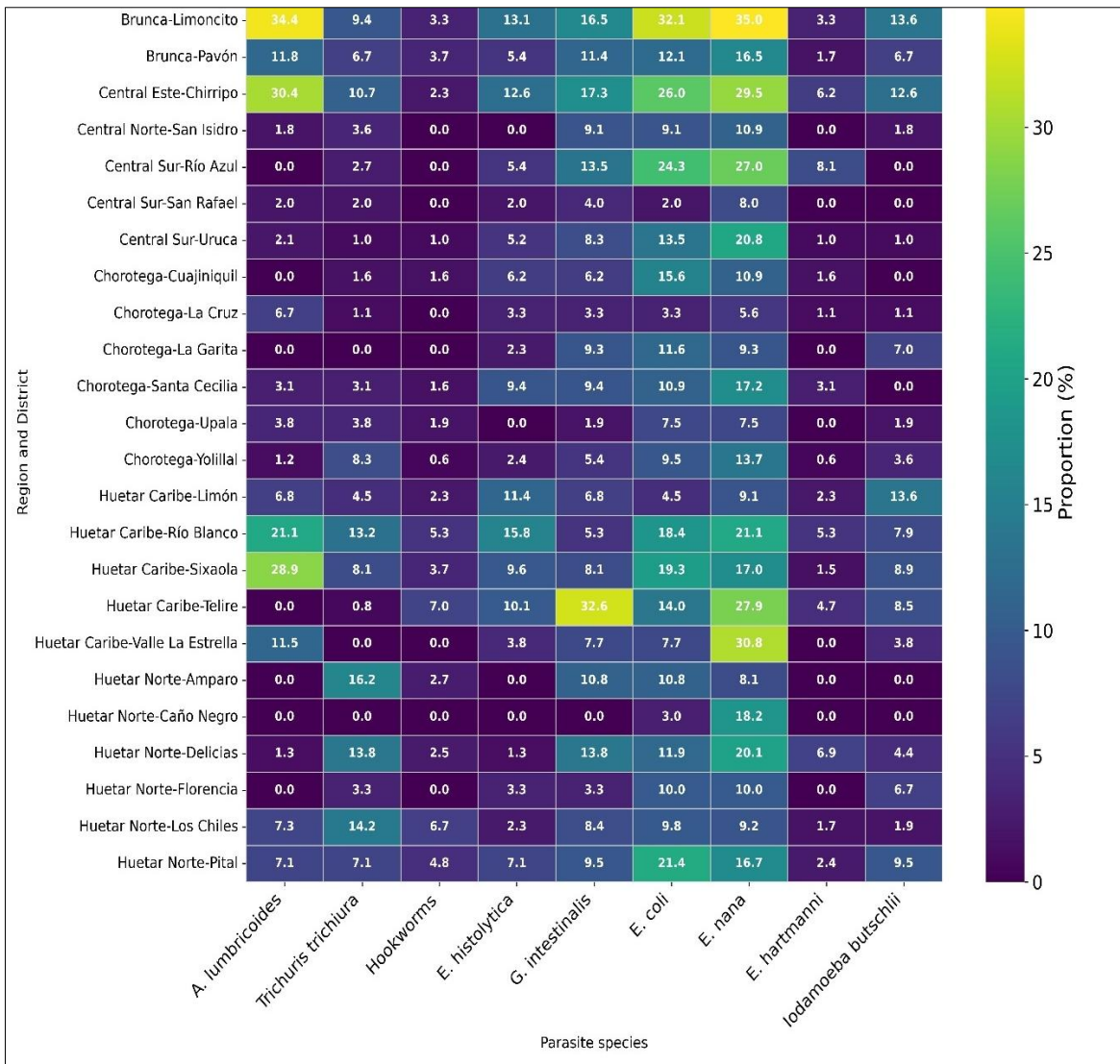


Figure 6. Heatmap showing intestinal parasite species percentages in schools from different regions and districts of Costa Rica in vulnerable places. *A. lumbricoides*=*Ascaris lumbricoides*, *T. trichiura*=*Trichuris trichiura*, *E. histolytica* = *Entamoeba histolytica*, *G. intestinalis* = *Giardia intestinalis*, *E. coli* = *Entamoeba coli*, *E. nana* = *Endolimax nana*, *E. hartmanni* = *Entamoeba hartmanni*, *I. butschlii* = *Iodamoeba butschlii*. Rare species (eg. *Chilomastix mesnili*, *Retortamonas* sp., *Taenia* sp., and others, present in very few samples, were not included in the proportion's calculations.

Logistic Regression analysis

Factors Associated with Higher Odds for Intestinal Parasitic Infections

The STH model showed that an increased index of difficulty of access (DAI) was associated with higher odds of infection (OR = 1.9, 95% CI = 1.2-3.07%, P = 0.009). Also, the land configuration score — which indicates more rural, vegetated, and low urban density areas — was highly associated with greater odds for testing positive for STH infections (OR = 16.2, 95% CI = 4.2-61.7%, p <0.001) (Figure 7).

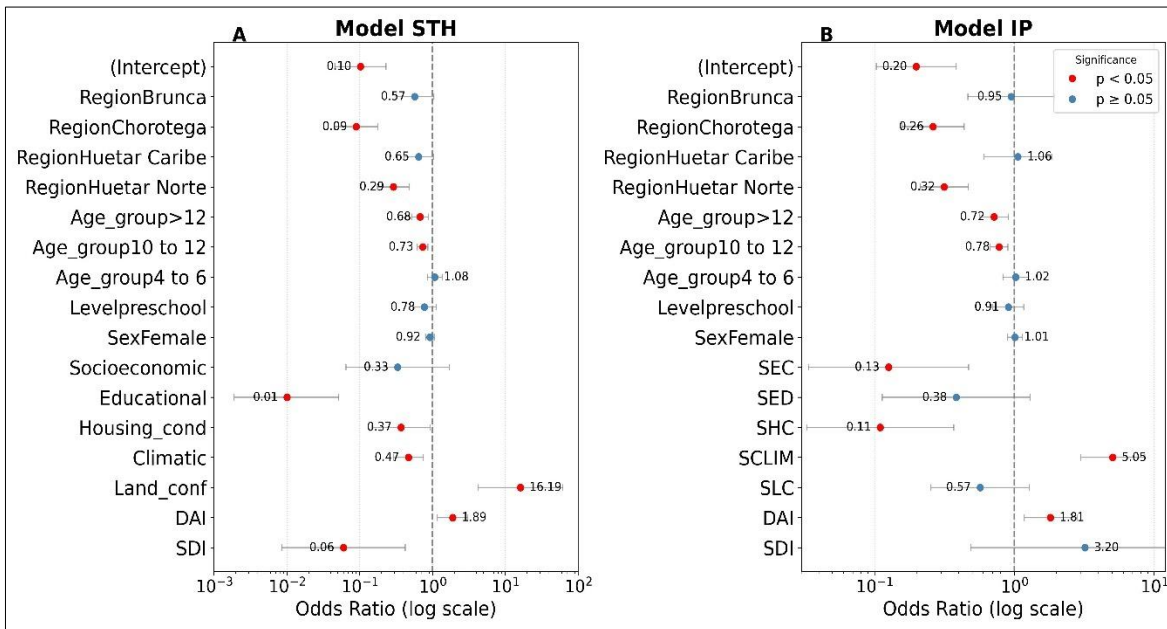


Figure 7. Forest plot visualization for logistic regression models' results. Odds ratios and 95% confidence limits. A: Model for STH infections. B: Model for intestinal protozoan infections. Age reference group is '7 to 9' years old, for educational level reference is 'primary', for Region, reference is 'Central (all central regions)', for Sex reference group is 'Male'. Abbreviations: STH= Soil transmitted helminths infections, IP= Intestinal protozoan infections; SHC= housing conditions variables score; SLC= Land configuration score; DAI=Difficult-of-Access Index; SDI = Social Development Index, SEC= socioeconomic variables score, SED= Educational score, and SCLIM= Climatic variables score.

Regarding the IP model, a higher difficulty of access was also significantly associated with greater odds of infection (OR=1.8, 95% CI = 1.2-2.8%). In addition, a higher value in the climatic score (representing conditions of higher rainfall, stronger seasonality, higher humidity, warmer temperatures, and radiation) significantly increased the odds for protozoan intestinal infections (OR=5.0, 95% CI = 3.0-8.5%, p < 0.0001) (Figure 7).

Factors Associated with Lower Odds

For STH, living in the Chorotega region (OR = 0.09, 95% CI = 0.05–0.18%, $p < 0.001$) or the Huetar Norte region (OR = 0.29, 95% CI = 0.18–0.48, $p < 0.001$) was significantly associated with reduced odds of infection compared to the Central Sur region. Children aged over 12 years (OR = 0.68, 95% CI = 0.52–0.88, $p = 0.004$) and those aged 10 to 12 (OR = 0.73, 95% CI = 0.62–0.87, $p < 0.001$) had lower odds than the 7–9 years reference group. Protective associations were also found for parasites-favorable climatic conditions (OR = 0.47, 95% CI = 0.29–0.75, $p = 0.002$), improved housing conditions (OR = 0.37, 95% CI = 0.15–0.93, $p = 0.003$), higher educational score (OR = 0.01, 95% CI = 0.002–0.05, $p < 0.001$), and higher values of the Social Development Index (SDI) (OR = 0.06, 95% CI = 0.0085–0.42, $p = 0.004$).

For the IP model, protective factors included residence in the Chorotega (OR = 0.26, 95% CI = 0.16–0.44, $p < 0.001$) or Huetar Norte (OR = 0.32, 95% CI = 0.21–0.47, $p < 0.001$) regions, as well as older age: children over 12 years (OR = 0.72, 95% CI = 0.57–0.90, $p = 0.005$) and those aged 10 to 12 (OR = 0.78, 95% CI = 0.67–0.90, $p < 0.001$). Additionally, higher socioeconomic score (OR = 0.13, 95% CI = 0.03–0.47, $p = 0.002$) and improved housing conditions (OR = 0.11, 95% CI = 0.033–0.37, $p < 0.001$) were significantly protective (Figure 7).

DISCUSSION

Overall proportions of intestinal parasites

This study reveals a notably high proportion of intestinal parasitic infections among schoolchildren living in vulnerable districts of Costa Rica. Of 4812 stool samples, 56.4% were positive for at least one intestinal parasite. This figure exceeds the 49.1% reported by the National-Level Nutritional and Parasitological Survey (NLNPS) conducted by INCIENSA in 2008–2009 [16]. The disparity is more pronounced for STH, with a proportion of 33.4% in this study, compared to only 2.6% in the NLNPS.

Similarly, pathogenic protozoan infections (e.g., *Giardia intestinalis* and *Entamoeba histolytica*) reached 20.5%, more than three times the rate reported in the NLNPS (6.3%). Although methodological differences between the NLNPS and this limit direct comparison,

the results highlight a significantly different scenario in schoolchildren from vulnerable underrepresented rural areas [16].

The observed differences could be linked, in part, to the fact that NLNPS did not include many of the rural and indigenous communities with a high structural vulnerability, several of which are included in the present dataset. In contrast to earlier research conducted in the Central Sur region by Solano-Barquero et al. (2018), which reported helminth prevalence of 0.7% and pathogenic protozoa at 8.5%, this study found proportions of 3.8% and 14.8%, respectively, despite analyzing the same region. These discrepancies suggest local variation and emphasize the need for continued longitudinal surveillance of these phenomena [24].

Geographical patterns

Distinct geographic patterns emerged. The highest STH and IP infections were found in schools in Chirripó, Limoncito, Sixaola, Los Chiles, and Telire, many of them in the Huetaar Caribe and Central Este regions. Schools in Río Blanco, Río Azul, Yolillal, and Delicias also showed high IP infections (>30%).

These areas share common characteristics such as high humidity, warm climates, soils suitable for the maturation of helminth eggs, and insufficient coverage of basic services. All these are factors that favor the persistence of transmission cycles, which are linked to a higher risk of parasites [10, 25, 26]. Notably, Chirripó and Limoncito, with the highest AIP proportions (71.0 and 73.0% respectively), also had the lowest SDI values (0 and 35.4, respectively) [27]. These communities face historical persistent deficits in sanitation and access to services, making them hotspots for parasitic infections [17, 28].

High prevalence is linked to high local infection intensity, which in turn is associated with greater morbidity [29, 30]. Similar environmental and socioeconomic profiles have been associated with parasitic risk in other regions [31, 32].

STH species

Ascaris lumbricoides was the most prevalent STH infection (19.9%), followed by *Trichuris trichiura* (9.2%) and hookworms (3.1%). These numbers are consistent with previous reports from other high-transmission regions in Latin America, Asia, and África [25, 33-36]. Life cycles of STH are known to be connected to wet soil, warm tropical climate, rainy localities and social determinants which drive the persistence of these parasites in the

environment [1, 37]. The high reproductive output and resilience of eggs of *A. lumbricoides* (and, though a little less, also for *T. trichiura*) contribute to the high prevalence [29, 38].

Hookworm proportion was low in all regions, peaking at 7% in schools from Telire. Since hookworm transmission occurs primarily by skin penetration, water treatment can likely reduce *A. lumbricoides* but not hookworm infections [29]. During sampling, the neighborhoods visited in the Telire district had recently implemented a new water treatment system for safer drinking water.

Trichuriasis surpassed ascariasis in some communities from the Huetar Norte region (Amparo, Los Chiles) and the Chorotega Region (Delicias, Yolillal). These areas have a flat topography, more pronounced extremes of temperature, low forest cover, low SDI values and houses that form rural town centers (not urbanized but more concentrated than in Chirripó or Limoncito). Literature suggests such conditions, along with lower humidity, may reduce STH egg and larvae survival [25, 38, 39].

Culture and habits also may influence positivity for IP and STH infections. However, this topic has been little studied in Costa Rica. A recent meta-analysis by Barboza-Solís et al. (2024) highlighted the lack of data on how cultural and ethnic differences influence disease distribution [40]. Despite the few studies about it rely on aggregated administrative data, including this one, it is notable that, while socioeconomic gradients were weak for some chronic diseases, a clear association with poverty was observed for infectious diseases [40].

IP species

The most common protozoans were the commensals *Endolimax nana* (23,0%) and *Entamoeba coli* (19.9%), followed by *Iodamoeba butschlii* and *Entamoeba hartmanni*. Their presence indicates recurrent exposure to human feces-contaminated water sources [5, 41, 42]. Although non-pathogenic, these species point to poor safe water access, bad food handling habits, and hygiene [5, 41, 42].

This study found a notably higher proportion of *Entamoeba histolytica* (8.7%) compared to NLNPS (3.0-3.5%) [16], possibly reflecting under-sampling in high-risk areas. Pathogenic protozoans were more frequent in schools in Chirripó, Limoncito, Telire, Limón, Río Blanco, Valle la Estrella, and Sixaola, reaching 33% in Telire in the case of giardiasis.

G. intestinalis and *E. histolytica* are closely linked to unsafe water consumption. Chronic giardiasis in children can cause malabsorption-derived consequences even if clinical symptoms do not appear severe [42]. This underscores the importance of these proportions in those localities in Costa Rica, which are not systematically searched because our passive epidemiological surveillance system in Costa Rica [15, 42].

Some reported *E. histolytica* cases may actually be the commensal *E. dispar*, which cannot be distinguished from *E. histolytica* by microscopy. This can lead to confusion in diagnosis and treatment. Physician judgment is essential in deciding on treatment when *E. histolytica/E. dispar* is reported [43, 44].

Complications from true *E. histolytica* infections include dysbiosis, granulomas, amoebic abscesses, and dysentery [43, 44]. These risks are heightened in communities with low piped water coverage and a high potability gap, as seen in those high-prevalence regions in this study [19,28].

Age and Sex influence

When the frequency of positive outcomes (IP or STH) was directly compared between age groups, statistically significant differences were found; schoolchildren between the ages of 7 and 9 had the largest proportion of positive outcomes (for both IP and STH). This is in line with results from other authors that indicate a higher risk of infection before adolescence [5, 9]. This has been connected to inadequate immunity and behavioral variations among age groups. As shown in other research, there were no discernible differences by sex [9-11, 44, 45].

Logistic regression models

Logistic regression models supported previous observations, identifying associations between STH/IP and variables such as age group, region, housing conditions, SDI, and access difficulty indicators. The age group of 12 and above was slightly protective for STH and IP when compared to the 7 to 9 year-old group, as also seen in other studies [5-9, 46-48].

Scores related to rurality and geographic inaccessibility showed strong associations with STH and IP. These findings align with previous research, which has linked remoteness to poorer sanitation and healthcare access [49].

Climatic condition score played contrasting roles in STH and IP infections. For STH high value of climatic score components (which include humidity, temperature, UV radiation, rainfall, and others) appeared protective instead of a favoring factor. This was not expected, as literature suggests that more humid environments (which in turn are favored by more rain) and higher temperatures are more likely to favor STH infections [10, 26].

From the climatic variables included in the score, only UV radiation aligned with existing evidence on egg inactivation, and so, it plays a role as a protective factor. However, as scores combine weighted factors, individual variable effects remain unclear [50]. Nonetheless, heavy rainfall seasons can have the effect of washing away the eggs and cysts from soil and diluting them in water sources. This could be a possible explanation for the climatic score's protective association, which warrants further exploration in studies.

The housing conditions score had a significant protective association for both types of infections. Good water access and quality, as well as adequate fecal waste final disposition systems, have been previously recognized in studies as protective factors for intestinal parasites [28], and these are components included within the score. Moreover, appropriate fecal disposal systems (and no open-air defecation) are vital factors for preventing STH transmission. Eggs can survive well in soil under good conditions, so less fecal soil contamination is protective, as previous studies have noted [29, 50].

SDI index consists of four dimensions and eleven variables [27]. The only health indicator included that is directly related to intestinal parasites is the quality of drinking water in homes; however, this is a very relevant variable because drinking fecal-contaminated water is thought to be one of the most critical factors associated with the high proportion of intestinal parasites (IP and STH) [9, 11].

According to present data, a high SDI value was protective against STH infections, but for protozoan infections, it was not significant. This could indicate that there are differences between STH and IP infections in how dimensions of the SDI index affect them. Other factors included in the SDI index include, for instance, low child weight, internet access, poor school attendance and performance, and political involvement. These factors may serve as stand-ins for more profound inequalities that impact STH transmission in Costa Rica more than IP infections. For instance, the SDI of certain urban peripheral populations is low, though not as low as that of the scenarios of Limoncito and Chirripó [27, 40].

The educational score also resulted in significant prevention of IP and STH infections. This factor has been repeatedly highlighted in literature as a key contributor to better health outcomes [10]. A higher educational level among parents and caregivers has been associated with a lower prevalence of intestinal parasitic diseases in some studies [49].

The consistent association of difficulty of access with increased risk for both STH and IP suggests that remoteness may be a proxy for multiple overlapping vulnerabilities, including poorer housing conditions and sanitation, limited health service reach, and environmental exposure. As such, targeted interventions in these high-risk, difficult-to-reach areas are urgently needed.

The explanatory power of this initial model was relatively low (R^2 0.06 for IP and 0.07 for STH), suggesting that even though structural and demographic factors have a certain weight in STH and IP infections, parasite transmission in these zones is governed by a more complex and specific combination of variables in the communities. The patterns also suggest that microgeographic variables are likely present with other behavioral, environmental, and access to basic services that must be more thoroughly analyzed to refine the models.

In conclusion, this study presents robust evidence that intestinal parasitic infections remain a significant public health concern in specific vulnerable areas of Costa Rica. Despite national progress in infectious disease control, these communities continue to face disproportionately high parasitic burdens. This reaffirms the need for targeted interventions, updated surveillance systems, and integrative approaches to address health's social, environmental, and structural determinants.

Funding

Projects TC-552 and ED-511 have been funded by Vicerrectoría de Acción Social, Universidad de Costa Rica, from 2008 to 2019.

Availability of data and materials

An anonymized, summarized database and all R and Python code and calculations can be requested from the corresponding author. Please write to the corresponding author if you are interested in such data.

Ethics approval

The Ethics Committee of Universidad de Costa Rica granted permission to develop this research through ordinary session No. 209-2021 and official document CEC-552-2021, under the modality of extended informed consent, given the complete anonymity of the data.

Competing interests

The authors declare that they have no competing interests.

Consent to participate

The participants granted consent for using the anonymized data in this study.

Conflicts of interest declaration

Authors declare they do not have any conflicts of interest

Declaration on the use of AI

Authors declare that AI (Chat GPT 4.0) was used exclusively to review the correctness and conciseness of English writing, as authors have Spanish as their maternal language.

References

1. Chen J, Gong Y, Chen Q, Li S, Zhou Y. Global burden of soil-transmitted helminth infections, 1990–2021. *Infect Dis Poverty*. 2024;13(5):68-77. <https://doi.org/10.1186/s40249-024-01238-9>
2. Pullan RL, Smith JL, Jasrasaria R, Brooker S J. Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. *Parasites Vectors*. 2014;7(1):1–19. <https://doi.org/10.1186/1756-3305-7-37>
3. Degarege A, Erko B, Negash Y, Animut A. Intestinal Helminth Infection, Anemia, Undernutrition and Academic Performance among School Children in Northwestern Ethiopia. *Microorganisms*. 2022;10(7):1353–63. <https://doi.org/10.3390/microorganisms10071353>

4. Feleke BE. Nutritional Status and Intestinal Parasite in School Age Children: A Comparative Cross-Sectional Study. *Int J Pediatr.* 2016;2016(1):1–8. <https://doi.org/10.1155/2016/1962128>
5. Hajissa K, Islam MA, Sanyang AM, Mohamed Z. Prevalence of intestinal protozoan parasites among school children in Africa: A systematic review and meta-analysis. *PLoS Negl Trop Dis.* 2022;16(2):e0009971. <https://doi.org/10.1371/journal.pntd.0009971>
6. Almeida MM, Monteiro KJL, Bacelar PAA, Santos JP dos, de Freitas SPC, Evangelista BBC, Leal, DN, da Silva DA, Cardoso, AB, do Nascimento EF, de Moraes-Neto AHA, Carvalho-Costa FA. Interactions between malnutrition, soil-transmitted helminthiasis, and poverty among children living in periurban communities in Maranhao State, Northeastern Brazil. *Rev Inst Med Trop São Paulo.* 2020;62:e73. <https://doi.org/10.1590/s1678-9946202062073>
7. Bisetegn H, Debash H, Ebrahim H, Erkihun Y, Tilahun M, Feleke DG. Prevalence and Determinant Factors of Intestinal Parasitic Infections and Undernutrition among Primary School Children in North-Central Ethiopia: A School-Based Cross-Sectional Study. *J Parasitol Res.* 2023;2023(1):2256910. <https://doi.org/10.1155/2023/2256910>
8. Ngonghala CN, Pluciński MM, Murray MB, Farmer PE, Barrett CB, Keenan DC, Bonds MH. Poverty, Disease, and the Ecology of Complex Systems. *PLoS Biol.* 2014;12(4):e1001827. <https://doi.org/10.1371/journal.pbio.1001827>
9. Ross AG, Papier K, Luceres-Catubig R, Chau TN, Inobaya MT, Ng SK. Poverty, dietary intake, intestinal parasites, and nutritional status among school-age children in the rural Philippines. *Trop Med Infect Dis.* 2017;2(4):e1001827. <https://doi.org/10.3390/tropicalmed2040049>
10. Bahmani P, Maleki A, Sadeghi S, Shahmoradi B, Ghahremani E. Prevalence of Intestinal Protozoa Infections and Associated Risk Factors among Schoolchildren in Sanandaj City, Iran. *Iran J Parasitol.* 2017;12(1):108–16. <http://ijpa.tums.ac.ir>
11. Tegen D, Damtie D, Hailegebriel T. Prevalence and Associated Risk Factors of Human Intestinal Protozoan Parasitic Infections in Ethiopia: A Systematic Review and Meta-Analysis. *J Parasitol Res.* 2020;2020(1):8884064. <https://doi.org/10.1155/2020/8884064>

12. de Bertodano I. The Costa Rican health system: low cost, high value. *Bull World Health Organ.* 2003;81(8):626–27. <https://www.scielosp.org/pdf/bwho/2003.v81n8/626-627/en>
13. Badilla-Mora A, Caravaca-Rodríguez I, Castro-Córdoba R, Chaverri-Esquivel A, Delgado-Jiménez S, Quesada-Rojas C, Jiménez-Solís AM, editors. Protocolo nacional para la vigilancia de enfermedades transmitidas por alimentos y agua. San José: Ministerio de Salud; 2024. p. 21-25. <https://hdl.handle.net/10669/90939>
14. Espinoza A. Comportamiento de la enfermedad diarreica en Costa Rica, de 1995 al 2001. *Rev Costarric Salud Pública.* 2004;13(24):50–58.
15. Sáenz MR, Acosta M, Muiser J, Bermúdez JL. The health system of Costa Rica. *Salud Publica Mex.* 2011;53 Suppl 2:s156–67.
16. INCIENSA. Encuesta Nacional de Nutrición: Fascículo 5. Parásitos intestinales. Costa Rica: Instituto Costarricense de Investigación y Enseñanza en Nutrición y Salud; 2013.
17. Mata L, Pardo V, Hernández F, Vizconti M, Mata A, Fernández R, et al. Cambios en la prevalencia de helmintos intestinales en Costa Rica, 1966–1982. In: Control and eradication of infectious diseases. Washington, DC: PAHO; 1985. p. 1.
18. Solano-Barquero A, León-Alán D. Parasitosis debidas a helmintos intestinales: una actualización sobre su abordaje y tratamiento farmacológico en Costa Rica. *Rev Col Microb Quim Clin Costa Rica.* 2018;24:24.
19. Fuentes E. Costa Rica a la luz del Censo 2011. San José, INEC; 2014.
20. Google Earth Pro. Versión 7.3.6.9345. Mount View, CA: Google LLC. 2024. <https://www.google.com/earth/> Accessed 12 Jun 2024.
21. R Core Team. R: A language and environment for statistical computing. Version 4.4.2. Vienna: R Foundation for Statistical Computing; 2024. <https://www.r-project.org>. Accessed 15 Jan 2025.
22. QGIS Development Team. QGIS Geographic Information System. Version 3.34.11. Open Source Geospatial Foundation Project; 2024. <https://qgis.org>. Accessed 15 Jan 2025.

23. Hunter JD. Matplotlib: A 2D Graphics Environment. *Comp. Sci Engin.* 2007;9(3):90-5. <https://doi.org/10.5281/zenodo.592536>
24. Solano-Barquero M, Montero-Salguero A, León-Alán D, Santamaría-Ulloa C, Mora A, Reyes-Lizano L. Prevalence of parasitosis in children aged 1 to 7 years in vulnerable condition in the South Central Region of Costa Rica. *Acta Med Costarric.* 2018;60(2):19–29
25. Bakarman MA, Hegazi MA, Butt NS. Prevalence, Characteristics, Risk Factors, and Impact of Intestinal Parasitic Infections on School Children in Jeddah, Western Saudi Arabia. *J Epidemiol Glob Health.* 2019;9(1):81. <https://doi.org/10.2991/jegh.k.190219.001>
26. Maru DS. Prevalence of Intestinal Parasitic Infections and Associated Risk factors among School children in Adigrat town, Northern Ethiopia. *Int J Emerg Trends Sci Tech.* 2017;4(1):4943–48. <https://doi.org/10.18535/ijetst/v4i1.03>
27. Víquez S., Sáenz JO, Ortega-Rodríguez W, Marschall-Murillo C, Bargoza G, Ureña H, Lozano S. Costa Rica, Índice de Desarrollo Social 2013. San José: MIDEPLAN; 2014. p. 1-36.
28. Méndez F, Bravo O. Costa Rica, Mapas de Pobreza 2011. In INEC (Ed.), Costa Rica a la Luz del Censo 2011. San José: INEC; 2014. P.482.
29. Jourdan PM, Lamberton PHL, Fenwick A, Addiss DG. Soil-transmitted helminth infections. *Lancet.* 2018;391(10117):252–65. [https://doi.org/10.1016/S0140-6736\(17\)31930-X](https://doi.org/10.1016/S0140-6736(17)31930-X)
30. Torgerson PR, Devleesschauwer B, Praet N, Speybroeck N, Willingham AL, Kasuga F, Rokni MB, Zhou XN, Fèvre EM, Sripta B, Gargouri N, Fürst T, Budke CM, Carabin H, Kirk MD, Angulo F J, Havelaar A, de Silva N. World Health Organization Estimates of the Global and Regional Disease Burden of 11 Foodborne Parasitic Diseases, 2010: A Data Synthesis. *PLOS Med.* 2015;12(12):e1001920. <https://doi.org/10.1371/journal.pmed.1001920>
31. Daryani A, Sharif M, Nasrolahei M, Khalilian A, Mohammadi A, Barzegar G. Epidemiological survey of the prevalence of intestinal parasites among schoolchildren in Sari, northern Iran. *Trans R Soc Trop Med Hyg.* 2012;106(8):455–59. <https://doi.org/10.1016/j.trstmh.2012.05.010>

32. Muñoz-Antoli C, Pavón A, Marcilla A, Toledo R, Esteban JG. Prevalence and risk factors related to intestinal parasites among children in Department of Rio San Juan, Nicaragua. *Trans R Soc Trop Med Hyg.* 2014;108(12):774–82. <https://doi.org/10.1093/trstmh/tru160>
33. Muñoz-Antoli C, Gozalbo M, Pavón A, Pérez P, Toledo R, Esteban JG. Enteroparasites in Preschool Children on the Pacific Region of Nicaragua. *Am J Trop Med Hyg.* 2018;98(2):570–75. <https://doi.org/10.4269/ajtmh.17-0551>
34. Pullan RL, Kabatereine NB, Quinnell RJ, Brooker S. Spatial and Genetic Epidemiology of Hookworm in a Rural Community in Uganda. *PLoS Negl Trop Dis.* 2010;4(6):e713. <https://doi.org/10.1371/journal.pntd.0000713>
35. Sarkari B, Hosseini G, Motazedian MH, Fararouei M, Moshfe A. Prevalence and risk factors of intestinal protozoan infections: a population-based study in rural areas of Boyer-Ahmad district, Southwestern Iran. *BMC Infect Dis.* 2016;16(1):703. <https://doi.org/10.1186/s12879-016-2047-4>
36. Wördemann M, Polman K, Menocal Heredia LT, Junco-Diaz R, Collado-Madurga A, Núñez-Fernández FA, Cordovi-Prado RA, Ruiz-Espinosa A, Pelayo-Duran L, Bonet-Gorbea M, Rojas Rivero L, Gryseels B. Prevalence and risk factors of intestinal parasites in Cuban children. *Trop Med Int Health.* 2006;11(12):1813–20. <https://doi.org/10.1111/j.1365-3156.2006.01745.x>
37. Sturrock SL, Yiannakoulis N, Sanchez AL. The Geography and Scale of Soil-Transmitted Helminth Infections. *Curr Trop Med Rep.* 2017;4(4):245–55. <https://doi.org/10.1007/s40475-017-0126-2>
38. Maya C, Ortiz M, Jiménez B. Viability of *Ascaris* and other helminth genera non-larval eggs in different conditions of temperature, lime (pH), and humidity. *Water Sci Technol.* 2010;62(11):2616–24. <https://doi.org/10.2166/wst.2010.535>
39. Smith G, Schad GA. *Ancylostoma duodenale* and *Necator americanus* : effect of temperature on egg development and mortality. *Parasitol.* 1989;99(1):127–32. <https://doi.org/10.1017/S0031182000061102>

40. Barboza-Solís C, Herrero R, Fantin R. Health inequalities in a middle-income country: a systematic review of the Costa Rican case. *Front. Public Health.* 2024;12:1397576. <https://doi.org/10.3389/fpubh.2024.1397576>
41. Li J, Wang Z, Karim MR, Zhang L. Detection of human intestinal protozoan parasites in vegetables and fruits: a review. *Parasit Vectors.* 2020;13(1):1–19. <https://doi.org/10.1186/s13071-020-04255-3>
42. Adam RD. *Giardia duodenalis*: Biology and Pathogenesis. *Clin Microbiol Rev.* 2021;34(4):e00019–24. <https://doi.org/10.1128/CMR>
43. Baxt LA, Singh U. New insights into *Entamoeba histolytica* pathogenesis. *Curr Opin Infect Dis.* 2008;21(5):489–94. <https://doi.org/10.1097/QCO.0b013e32830ce75f>
44. Kantor M, Abrantes A, Estevez A, Schiller A, Torrent J, Gascon J, Hernandez R, Ochner C. *Entamoeba histolytica*: Updates in Clinical Manifestation, Pathogenesis, and Vaccine Development. *Can J Gastroenterol Hepatol.* 2018;2018(1):46014–20. <https://doi.org/10.1155/2018/4601420>
45. Elmonir W, Elaadli H, Amer A, El-Sharkawy H, Bessat M, Mahmoud SF, Atta MS., El-Tras WF. Prevalence of intestinal parasitic infections and their associated risk factors among preschool and school children in Egypt. *PLOS One.* 2021;16(9):e0258037. <https://doi.org/10.1371/journal.pone.0258037>
46. Rayan P, Verghese S, McDonnell P. Geographical location and age affects the incidence of parasitic infestations in school children. *Indian J Patho Microbiol.* 2010;53(3):498. <https://doi.org/10.4103/0377-4929.68292>
47. Chammartin F, Scholte RG, Guimaraes LH, Tanner M, Utzinger J, Vounatsou P. Soil transmitted helminth infections in South America: a systematic and geostatistical meta-analysis. *Lancet.* 2013;13:507-18. [https://doi.org/10.1016/S1473-3099\(13\)70071-9](https://doi.org/10.1016/S1473-3099(13)70071-9)
48. Scholte RGC, Schur N, Bavia ME, Carvalho EM, Chammartin F, Utzinger J, Vounatsou P. Spatial analysis and risk mapping of soil-transmitted helminth infections in Brazil, using Bayesian geostatistical models. *Geospatial Health.* 2013;8(1):97. <https://doi.org/10.4081/gh.2013.58>

49. Forrer A, Khieu V, Schär F, Vounatsou P, Chammartin F, Hanspeter M, Odermatt P. *Strongyloides stercoralis* and hookworm co-infection: spatial distribution and determinants in Preah Vihear Province Cambodia. *Parasites Vectors* 2018;2018:11-33. <https://doi.org/10.1186/s13071-017-2604-8>

50. Zafar A, Attia Z, Tesfaye M, Walelign S, Wordofa M, Abera D, Desta K, Tsegaye A, Ay A, Taye B. Machine learning-based risk factor analysis and prevalence prediction of intestinal parasitic infections using epidemiological survey data. *PLOS Negl Trop Dis*. 2022;16(6): e0010517. <https://doi.org/10.1371/journal.pntd.0010517>

ARTÍCULO 2

[formato: *PLOS Neglected Tropical Diseases*]

<https://journals.plos.org/plosntds/s/submission-guidelines>

A Machine Learning-Based Framework for Identifying Multidimensional Determinants of Intestinal Parasites in Costa Rican School-Age Children

Short title: Machine learning-based models for intestinal parasites in Costa Rica.

Alberto Solano-Barquero¹ Javier Mora¹, Juan José Romero², Bernardo Vargas-Leitón², Alicia Rojas¹.

1. Centro de Investigación en Enfermedades Tropicales, Universidad de Costa Rica
2. Programa de Investigación en Medicina Poblacional, Escuela de Medicina Veterinaria, Universidad Nacional, Costa Rica.

Corresponding autor: Alberto Solano-Barquero

Abstract

Background: Intestinal parasitic diseases have essential implications in public health, and their determinants are complex. **Methodology:** Using a comprehensive dataset that integrated 68 socioeconomic, educational, housing conditions, climatic, land configuration, and topographic variables, a sequenced pipeline combining multivariate analysis, machine learning processes (Random Forest and Elastic-Net Regression), and Linear Generalized Mixed Models (GLMM) was developed. Such a pipeline was used to model two outcome variables, soil-transmitted helminths (STH) and intestinal protozoan (IP) infections in schoolchildren from Costa Rica. **Principal findings:** For STH infections, the most influential predictors included seasonal rainfall variation, urban construction dispersion, minimal average temperatures, and indicators of geographic inaccessibility. For IP infections, the profile of key predictors included those related to structural poverty (Unsatisfied Basic Needs Indicator), total rainfall, and age. Final GLMM models using schools nested within districts as random effects demonstrated good calibration and discrimination ($AUC > 0.70$), with substantial variance explained in the STH model (R^2 conditional = 0.50) and moderate explanatory power in the IP model (R^2 conditional = 0.17). **Conclusions:** These results show that while both infection types are shaped by structural issues related to poverty and

environmental backgrounds, helminths are more susceptible to ecological and spatial factors, whereas protozoan infections are more closely linked to low socio-economic indicators. Such results highlight the need for targeted, context-attentive interventions incorporating social, environmental, and health access dimensions to reduce parasitic disease burdens in endemic settings.

Keywords: *Intestinal parasites, associated factors, statistical modelling, machine learning, Costa Rica, schoolchildren*

Author summary

This study evaluated data from 4812 schoolchildren in Costa Rica to understand which social, environmental, land, and housing factors are most strongly associated with two types of intestinal parasites: soil-transmitted helminths (STH) and intestinal protozoa (IP). Using machine learning approaches, we first identified the 20 most relevant variables that help predict the presence of these parasites. Subsequently, through another machine-learning approach, we further refined this list to identify the most significant factors for each parasite group. Finally, we built mathematical models that can estimate which districts in the country are at higher risk for these infections. These models can aid public health officials in better targeting areas with the highest need, use resources more effectively, and design more robust prevention policies. Our approach shows how data science can support disease control, especially in vulnerable communities, and improve children's health through more informed decision-making.

Introduction

Intestinal parasites are one of the major burdens of disease around the world, especially in poor countries. These infections affect more than 1.4 billion people globally and cause more than 6 million disease-adjusted life years (DALYs). Nearly 5 million years lived with disability globally, which are worldwide indicators of their substantial morbidity and economic significance (1,2). Schoolchildren are notably vulnerable to these diseases. The implications of these diseases for them are remarkable because they impair their physical growth and cognitive development, negatively compromising their school performance and so perpetuating poverty in the communities that suffer from them (3-6).

Reviews of research from around the world, as well as smaller local parasitological surveys conducted in both developed and undeveloped countries, have revealed how variable the frequency of certain intestinal parasites can be in different areas. Temperate and more developed regions have lower overall prevalences of intestinal parasites, whereas tropical and poor countries have much higher rates (7,8). By investigating the various determinants of these results with a deeper emphasis, it emerges that poverty and climate are important factors (7-11)

Nevertheless, poverty and climate are associated with many other complex and difficult-to-unpack factors. Malnourishment, anemia, education of caregivers, sanitation accessibility and practices, cultural idiosyncrasies, type of economic activities, or even possible genetic associations have also been found to be associated with the possibility of intestinal parasitic infection in studies (8,12-16)

Evidence thus shows the exceptionally complex nature of intestinal parasitic infections and how that complexity changes because of influencing factors that interact at several levels, from the global and regional to the extremely local scale. Such a complex nature is explained by the multidimensionality of the factors involved in generating this type of disease (17).

Machine learning is a collection of computational algorithms and tools that enable systems to automatically improve their performance on a particular task based on self-generated experience. Computers accomplish this by optimizing models based on patterns discovered in data (18). Conventional statistical approaches for prediction in biological systems (like human diseases) are based on statistical inference methods, which achieve their measurements by fitting into a particular probability model. This is how these traditional models permit us to measure the magnitude of the effect in a relationship between independent and dependent variables with assessable confidence (19).

Nevertheless, these conventional methods have disadvantages when it comes to analyzing data sets with many variables, particularly if several of them are highly correlated. Given that intestinal parasitic determinants are closely linked to poverty and environmental factors, it is understandable that such high correlations exist between these factors (20).

Machine learning approaches focus on achieving predictions and finding general patterns, taking advantage of the availability of a significant amount of data. Another benefit of Machine Learning methods is that they make minimal assumptions about the data-generating systems, which is particularly useful when data have been gathered without a carefully controlled experimental design, and in the case of possible complex non-linear relationships among the variables (19).

When creating predictive models, combining statistical inference techniques with machine learning techniques provides a potent strategy. Inferential statistics applied to a smaller set of features offer the main benefit of having a model that is not only more robust and computationally efficient but also highly interpretable. Machine learning techniques can be used to address and resolve high dimensionality, overfitting, and collinearity. For this reason, this integration may be a viable pipeline for modeling disease determinants (19,20).

In intestinal parasite epidemiology, there are only a few examples of using this novel approach to study the disease determinants (20). The pioneer study by Zafar and colleagues used a combination of inferential and machine learning methods to analyze disease determinants of intestinal parasitic disease. It was successful in identifying novel predictors of the disease in their locality, encompassing socioeconomic, demographic, and hematological factors, thereby setting a good precedent for the use of this type of strategy in parasitological epidemiological research (20).

This study aimed to identify and model some key determinants of intestinal parasitic infections in Costa Rican school-aged children through a machine learning and general linear modelling approach employing a multidimensional framework that incorporates socioeconomic, educational, climatic, geographic, topographic, and accessibility-related variables.

Methods

Data sources: coproparasitological results database

Sources of information on intestinal parasites in schoolchildren from Costa Rica

Thirty-six databases containing schoolchildren's fecal test results were revised and anonymized to create a curated database. Data came from the University of Costa Rica's official results reports on intestinal parasites in schoolchildren from seven regions of Costa

Rica and obtained information on parasite species, children's age and sex, school, and educational level for each record. Inclusion criteria were age between 4 and 16 years old, having complete information in the record, including: sex, school, age, and official result of the coprological test. Records that could not be linked to the official reports, records with incomplete information, and records from students who were sampled more than once were among the exclusion criteria. A total of 4812 records were included in the study.

Study population and sampling methods

The database's observations came from the results of coproparasitological tests conducted on schoolchildren aged 4 to 16 who attended schools in one of 24 districts located throughout the nation's seven designated health planning zones or health regions (Central Norte, Central Sur, Central Este, Huetar Caribe, Huetar Norte, Chorotega, and Brunca).

All the children were enrolled in their local school, and sample collection was part of official University of Costa Rica community projects (TC-552 and ED-511) that aimed to improve the diagnosis and treatment of intestinal parasite infections in the nation's impoverished communities.

Every child in every school was allowed to gather one fecal sample. All collected samples were transported to the Universidad de Costa Rica's Faculty of Microbiology's Laboratorio de Helminología for processing. Each sample was examined using a fecal smear (for intestinal protozoans) and a modified Kato-Katz technique, which involved spreading 60 mg of feces across a large slide, covering it with cellophane impregnated with Kato's solution, and pressing it against a hard surface. Each sample was watched for helminth eggs ten minutes later. Only the presence of parasite species is reported in this modification; eggs are not counted.

A fecal sample was obtained from a varying number of students in each school; Table SA-1 in Supplementary Materials A shows all the information regarding the sample fraction for each school. For every coprological result record, two primary binary response variables were collected: IP (intestinal protozoan infections) and STH (soil-transmitted helminth infection), where '1' denotes presence and '0' denotes absence.

Sources of information on social, economic, environmental, and geographic data (potential predictors)

All social and economic data, disaggregated at district level, were obtained from the public databases of the National Official Census of 2011, from the Instituto Nacional de Estadística y Censos (National Institute of Statistics and Census [INEC]) of Costa Rica (21). Besides, information regarding climatic conditions for the year of each observation was obtained from the Instituto Meteorológico Nacional (National Institute for Meteorology [IMN]), which provided environmental climatic analysis data, based on information obtained from the closest weather station to each school where each parasitological record originated.

Several environmental, land use, and geotopographic variables were obtained using Geographical Information Systems tools available in the open-access platform of Google Earth Pro (22), the Digital Atlas of Costa Rica 2014 by Instituto Tecnológico de Costa Rica (TEC) (23), and the free software QGIS 3.34.11 (24). Supplementary tables in supplementary material B (Table SB-1 to Table SB-6) show the names of all independent variables and their corresponding codifications for a rapid reference, and show details on collecting methodology for each of the independent or potential predictor variables. Supplementary Material C describes how each subgroup of variables was used to calculate scores of variables for educational, socioeconomic, housing conditions, land configuration, climatic, and difficulty of access.

Data analysis process

Multivariate analysis

Given the high number of potential predictor variables ($n = 68$), a Principal Component Analysis (PCA) was performed to examine variable relationships, detect multicollinearity, and guide the exclusion of redundant predictors before modeling. All variables were standardized using min–max scaling (0–1). PCA was accomplished with the *FactoMineR* package in R software (25). Clustering patterns among records were assessed using agglomerative hierarchical clustering (Ward's method) and k -means clustering, with the optimal number of clusters ($k = 4$) determined by within-cluster sum of squares and silhouette scores. The algorithm was executed in Python using `sklearn.cluster` and `sklearn.metrics`

importing `silhouette_score`. For reproducibility, `random_state=42` was used, and `n_init = 10` was used to initialize multiple times the centroid and avoid a poor local optimum. PCA loadings biplots were generated with the PCA function from `sklearn.decomposition` module in Python (26), incorporating IP and STH as response variables (27), and visualized with *Matplotlib* to illustrate bidimensional variable positioning relative to the outcomes.

Substantial correlations and potential multicollinearity among independent variables were confirmed in R software by correlation analysis of the putative predictors. To ascertain if one variable of each highly correlated pair needed to be manually excluded, the pairs of the most highly correlated variables (correlation >0.80) were examined in terms of the prediction value of each variable for intestinal parasites (based on what is known in literature). Supplementary material D displays exclusion judgments based on highly correlated pairings of some variables (Table D1). A final collection of variables was employed for the pipeline's subsequent phase, a Random Forest Classifier Analysis, after highly associated variables were manually eliminated.

Machine learning tools for variable selection

Random Forest first selection

To identify the most influential variables associated with intestinal parasite infections, Random Forest (RF) classification models were implemented in R software (v4.3.1) using the `randomForest` package, with hyperparameter tuning performed via the `caret` package. The outcome variables were binary (presence/absence of infection).

For soil-transmitted helminth (STH) infection, a strong class imbalance was observed. To address this, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to the training set only, using the `themis` package within a preprocessing pipeline to avoid data leakage. In contrast, SMOTE was not applied to intestinal protozoan (IP) infections due to balanced class distribution.

Data was partitioned into training (60%), validation (20%), and test (20%) sets. A grid search with 5-fold cross-validation was conducted on the training set to optimize key hyperparameters, including `mtry` (number of variables randomly sampled at each split), `nodesize` (minimum size of terminal nodes), and `maxnodes` (maximum number of terminal

nodes per tree). The optimal values were selected based on maximizing the area under the receiver operating characteristic curve (AUC-ROC).

Final models used $n_{tree} = 700$ to ensure stable out-of-bag error estimation and $importance = TRUE$ to compute permutation-based variable importance. The best-performing model was evaluated on the independent test set using accuracy, AUC-ROC, and recall. The top 20 most important variables were ranked by Mean Decrease in Accuracy and retained for subsequent analyses, including clustering and interpretation of exposure patterns.

Elastic-Net (Ridge plus LASSO regression) selection

To refine the set of candidate predictors identified by Random Forest, Elastic-Net penalized logistic regression was applied to the top 20 variables associated with soil-transmitted helminth (STH) and intestinal protozoan (IP) infections using the `glmnet` package in R software. Data were partitioned into training (60%), validation (20%), and test (20%) sets via stratified sampling. Predictors were standardized in the training set to ensure coefficient comparability.

A grid search over $\alpha \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and λ was performed using 10-fold cross-validation to optimize model performance. The optimal α and λ were selected based on maximum AUC-ROC on the validation set, with λ chosen at the one-standard-error rule (λ_{1se}) when it yielded a very complex model, to favor a more parsimonious one.

Variables with non-zero coefficients were retained as final predictors. Model performance was evaluated on the independent test set using AUC, accuracy, recall, and precision. Residual multicollinearity was assessed using the Variance Inflation Factor (VIF) on the original training data, with all values < 3 indicating acceptable collinearity.

Final modelling

Generalized linear mixed model (GLMM)

For both response variables, a generalized linear model (GLM) was run using the Elastic-Net variables that were previously chosen. A generalized linear mixed model was then carried out with the final selected variables for both response variables using the `lme4` package and

the glmer function in R software. The variable ‘school’ nested within the variable ‘district’ was used as a random effect. Variable coefficients, p-values, and 95% CI were calculated, as well as pseudoR² (Nagelkerke adjustment) (28), AIC, and likelihood ratio as performance indicators. For calculating pseudoR² in the mixed models, the R package MuMin was used, and marginal and conditional pseudoR² values were calculated. Finally, the Hosmer and Lemeshow goodness of fit (GOF) test was performed to evaluate the model's fitting.

Final equation

We used a logistic regression model to estimate the probability of STH infection (P) as a function of multidimensional epidemiological determinants. The model can transform the linear predictor (n) into a probability through the logistic function:

$$P = \frac{1}{1 + e^{-n}}$$

Where:

e = Euler’s number ≈ 2.71828

n = linear predictor (logit) incorporating our selected determinants

The logit (n) was defined as:

$$n = \beta_0 + \sum_{i=1}^k \beta_i x_i$$

Where β_0 is the intercept, β_i are the coefficients, and x_i represents the determinants.

Results

Multivariate analysis: Principal Component Analysis

In the PCA analysis, the first three principal components account for 61.7% of the variation in the data, while the first ten, collectively, account for 80%. Clustering among

observations from various regions is depicted in the 3-D scatter plot (Figure 1) that represents the first three components. Observations from Chirripó (Central Este region) appear distinctly separated from other observations. Schools from the Brunca (both Brunca_Limoncito and Brunca_other_district) and Huetar Norte regions also exhibited clustering patterns that were also obtained from the formal clustering by K-means analysis (Figure 1).

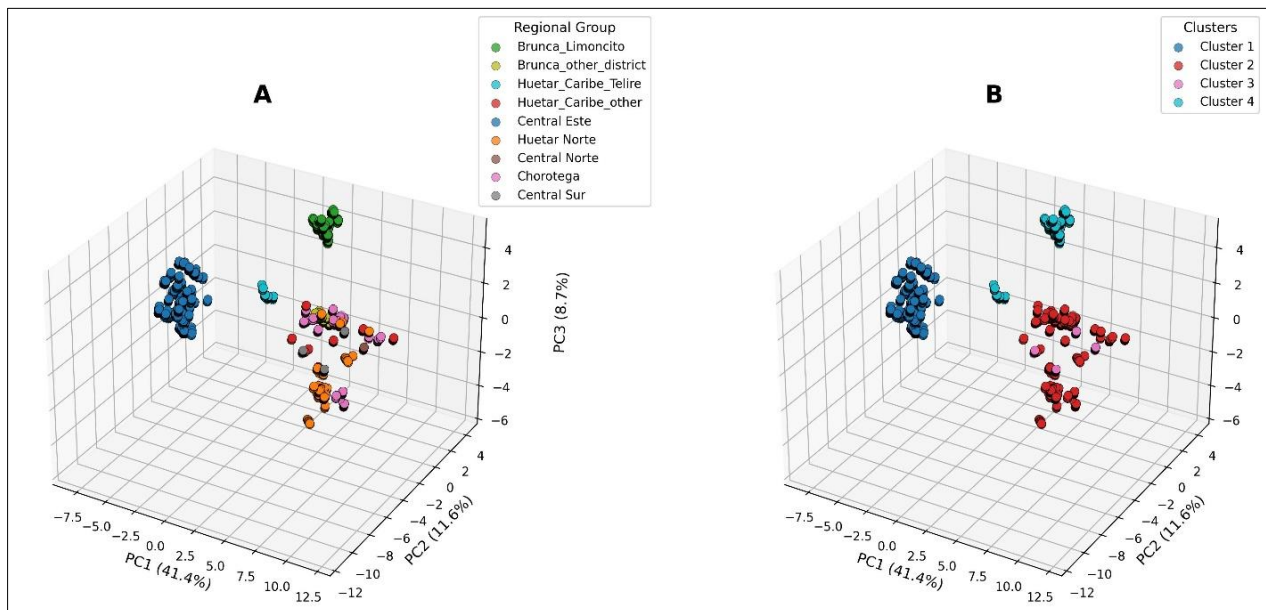


Figure 1. A) Three-dimensional principal component analysis plot shows a projection of observations from schools in each region in the three main principal components, which together explain 61.7% of total variance. Observations are colored by region. The plot reveals distinct clustering patterns, particularly for Chirripó, Brunca_Limoncito, and Huetar Norte, suggesting region-specific variation in predictor variables. B) K-means Clustering (k=4) applied on the first three components in the PCA analysis. Each point represents an observation or sample colored by the clustered assignment. A correspondence between regional patterns and data-driven cluster allocation can be seen.

PCA Biplot Interpretation

The PCA biplot, which captures the variable loadings vectors (Figure 2), indicates the contribution and orientation of each predictor variable along the first two principal components. The vectors are colored, representing the type of variable. The two response variables, IP (protozoan infections) and STH (soil-transmitted helminths), are also shown as red vectors.

In contrast, variables such as access to a sanitary sewage system (SSW), piped water (PWA), in-house potable water (PDW), higher Social Development Index (SDI), fixed salary (SAL), direct health insurance (DHI), longer duration in the educational system (GBE), and higher education attainment (UEC) were positioned opposite to protozoan infections (Figure 2).

Three variables open-air defecation (OAD), a high percentage of Indigenous population (INP), and a high percentage of the population experiencing critical shortcomings or unsatisfied basic needs (UBN), were located between helminth and protozoan infections (Figure 2).

Finally, several predictors were plotted very close to one another, suggesting a high correlation. This was demonstrated through a correlation analysis. This analysis revealed several pairs of variables with high correlations (correlation value > 0.80). This high correlation was important because, to obtain a good, fitted model later, high correlations can hamper the process. A group of variables was excluded from the analysis due to their high correlations. Each pairwise comparison was evaluated individually, and the selection of variables was based on their potential relevance to the outcomes of interest.

The decision-making process incorporated findings from the preceding multivariate analysis, as well as evidence from scientific literature regarding the role of similar variables in promoting or protecting against intestinal protozoan infections and soil-transmitted helminthiases. In supplementary material D Table D1, a detailed explanation of the variables manually excluded before modelling can be read.

Analysis of soil-transmitted helminth infection determinants

Random Forest Analysis for Soil-Transmitted helminth infections

The best hyperparameter set found for this analysis was achieved at an n of trees = 700, node size of 5, and max nodes of 30. Importance value for each variable was obtained. The model exhibited a stable predictive performance across all data sets: training, validation, and test (Table 1, Figure 3). This demonstrates the robustness and generalizability of the model. Performance metrics included AUC (0.79–0.80), accuracy (0.71–0.75), precision (0.70–

0.71), F1-score (0.73–0.75), and recall (0.77–0.80). Those metrics highlight good discriminatory ability, particularly in identifying the true positives.

Table 1. Performance metrics of the Random Forest Model for soil-transmitted helminths infection prediction across training, validation, and test datasets

Dataset	AUC	Recall	Accuracy	Precision	F1-Score
Training	0.80	0.80	0.74	0.71	0.75
Validation	0.79	0.79	0.72	0.70	0.74
Test	0.79	0.77	0.72	0.70	0.73

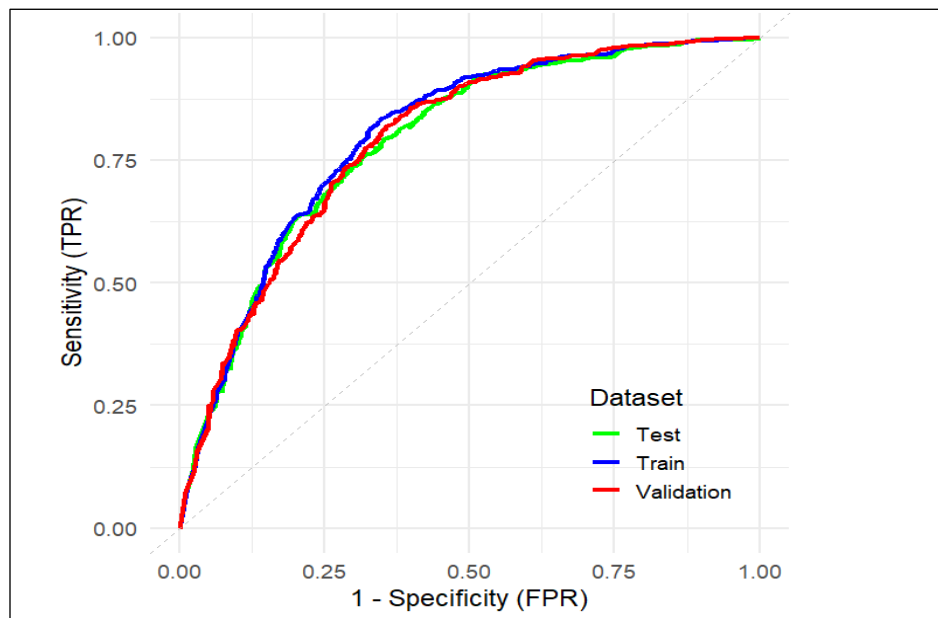


Figure 3. ROC-AUC curves for the training, validation, and testing datasets to observe the performance of the Random Forest Analysis for the response variable ‘Soil Transmitted Helminth infections. AUC=0.80.

Random Forest’s variable importance results using the ‘mean decrease in accuracy’ criterion revealed that a diverse set of predictors contributed most to the classification of STH infections. They included a combination of land configuration, climatic, socioeconomic, and geographic access variables as shown in Table 2. This supports the multifactorial nature of the disease.

Table 2. Top 20 predictors of soil-transmitted helminth infection according to mean decrease in accuracy (Random Forest Model)

Variable code	Variable name	Mean decrease in accuracy
UCOV	Percentage of urban land coverage	22.2
MITY	Average minimum temperature in a year	20.4
DRBS	DRBS = Total rainfall in the rainy season - Total rainfall in the dry season	20.0
DRY	Number of days with rain in a year	19.7
AAD	Average of all paired distances between constructions in the surroundings* of the school	19.0
INP	The percentage of the indigenous population in the district to which the school belongs	18.7
TMXD	The absolute difference between the average maximum temperature in the year of the record and the annual average temperature	17.9
RH	Relative humidity	17.4
TRY	Total rainfall in the year (in cm)	17.1
SBD	School to bus route total distance	16.5
PSR*	Total area in m2 that is permissive to parasites transmissible forms	15.7
TRHD	Terrain rugosity of the route from the school to the nearest hospital or clinic	15.6
NNI	Nearest neighbor index for the surroundings of the school*	15.3
DAAD	Standard deviation of AAD	15.0
COV1	% of the surrounding area not covered by forest or urban constructions	14.9
GNR	General terrain rugosity for the school surroundings	14.7
ADNN	Average distance to the nearest neighbor	14.6
FCOV	% of the school's surrounding area covered by forest	14.2
ATI	Average terrain inclination in the route from school to the nearest clinic or hospital	14.1
TMMD12	The absolute difference between the average maximum temperature in the year of the record and the annual average minimum temperature	14.0

*School surroundings are a 16000 m2 area delimitation (square-shaped) around each school from which a record comes. This area permits little overlapping between schools, so it mostly harbors the houses of children from whom those records come. The top 20 variables were selected to be analyzed through the Elastic-Net regression analysis in the next step of modelling the determinants of soil-transmitted helminth infections. Abbreviations in the “variable code” column are explained in the “variant name” column; that is why they are not described here in the footnote.

Elastic-Net analysis for the soil-transmitted helminths determinants

Elastic Net Variable Selection Results

Elastic net regression was used to further refine variable selection from the set previously identified by the RF. Through 10-fold cross-validation and grid search over α and λ , the optimal mixing parameter was found to be $\alpha = 0.85$, favoring LASSO-like sparsity. The regularization parameter was selected at $\lambda = \text{lambda.1se}$ to yield a more parsimonious model given the many non-zero coefficients and great complexity of the best-lambda model that maximizes the AUC (AUC=0.87). The cross-validation plot of binomial deviance vs $\text{Log}(\lambda)$ confirmed a clear minimum and stable selection (Figure 4).

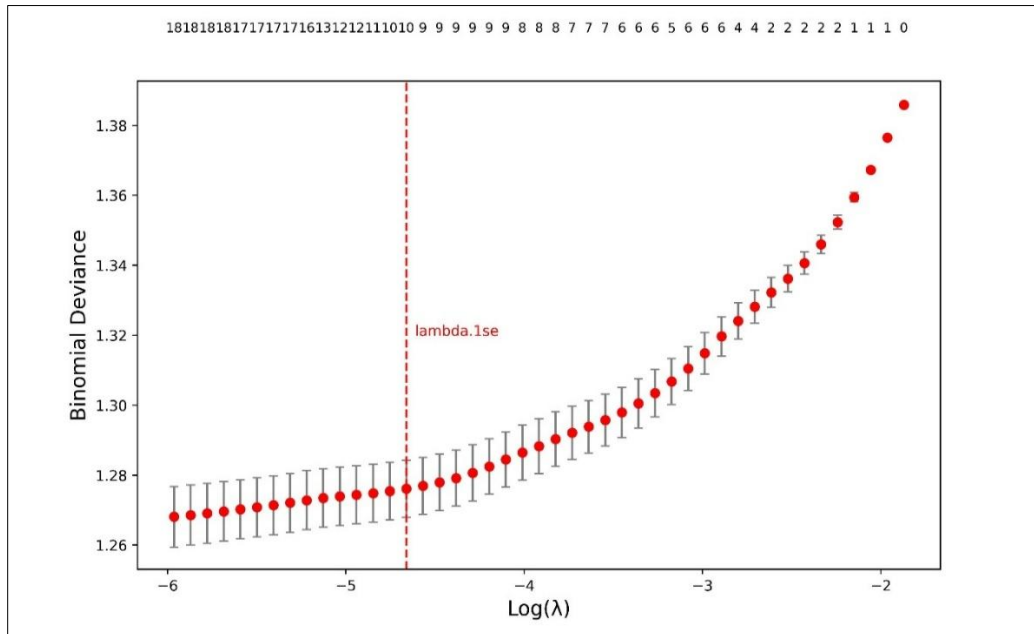


Figure 4. Elastic Net model for the selection of determinants of STH infections ($\alpha=0.85$, optimal $\lambda = 1.se$). The top numbered line represents the number of predictors (non-zero coefficient) for a particular Lambda value. The model has a good balance, which helped find the simplest model in which variables are the strongest predictors.

This approach identified ten variables with non-zero coefficients, indicating their significant contribution to predicting soil-transmitted helminth infections. The selected variables are shown in Table 3.

The magnitude and sign of the coefficients suggest that there is a high difference in the larger areas permissive to the parasites transmissible forms (PRS), the total rainfall difference between the dry and rainy seasons of the year (DRBS), a land configuration in which houses are largely separated from each other (AAD), they were the most important variables favoring

positivity for STH. On the other hand, negative relation between STH infection and predictor variables, according to the Elastic-Net regression analysis, were found mainly in the total rainfall in a year (TRY), indicators of more clustered constructions configuration shown by grater values of the nearest neighbor index (NNI) and annual average minimum temperatures (MITY) (table 3).

Table 3. Non-zero coefficients of the selected variables according to Elastic-Net regression analysis in the process of STH infections determinants modelling.

Variable code	Variable Name	Elastic-Net coefficient value
(Intercept)	Intercept of the model	-2.31
PSR	Total area in m2 that is permissible for parasites transmissible forms	1.19
DRBS	Total rainfall difference between rainy and dry seasons	1.06
AAD	Average of all paired distances between constructions in the surroundings of the school*	1.05
INP	The percentage of the indigenous population in the district to which the school belongs	0.68
SBD	Total distance in m to the nearest bus route	0.53
TRHD	Terrain rugosity of the route from the school to the nearest hospital or clinic	0.36
COV1	% of the surrounding area not covered by forest or urban constructions	0.02
TRY	Total rainfall in the year (in cm)	-0.44
NNI	Nearest neighbor index for the surroundings of the school*	-0.40
MITY	Average minimum temperature in a year	-0.26

*School surroundings are a 16,000 km² area delimitation (square-shaped) around each school from which a record comes. This area permits little overlapping between schools, so it mostly harbors the houses of children from whom those records come. Abbreviations in the “variable code” column are explained in the “variable name” column; that is why they are not described here in the footnote

Generalized Linear Mixed Model (GLMM) and random effects: school and district

Soil-transmitted helminth infections model

VIF analysis results showed no multicollinearity problems with any of the variables selected by the Elastic Net regression model. Subsequently, a generalized linear model (GLM) and a generalized linear mixed model (GLMM) were fitted using previously selected predictors (Table 3). Table 4 shows the conjunction of result values for the GLMM model for STH infection determinants in schoolchildren.

Table 4. Results of a fitted generalized linear mixed model for the determinants of STH infections in schoolchildren of Costa Rica.

Variable (code in equation)	Name	Coefficient value	Standard error	p-value
Intercept (B ₀)	Intercept	-4.05	0.70	<0.0001
MITY (X ₁)	Average minimum temperature	-1.98	0.41	<0.0001
DRBS (X ₂)	Total rainfall difference between rainy and dry seasons	2.65	0.42	<0.0001
AAD (X ₃)	Average of all pairs of distances between constructions in the surroundings of the school*	0.52	0.58	0.37
INP (X ₄)	The percentage of indigenous population	1.40	0.70	0.04
TRY (X ₅)	Total annual rainfall in cm in the year and the area of the record	-2.46	0.61	<.0001
SBD (X ₆)	Total distance in m from the school to the nearest bus route	1.31	0.51	0.01
PSR (X ₇)	Mean area permissive to parasites transmissible forms in the school surroundings*	3.63	0.97	<0.001
TRHD (X ₈)	Terrain rugosity of the route from the school to the nearest hospital or clinic	-0.28	0.31	0.35
NNI (X ₉)	Nearest neighbor index	-1.63	0.81	0.04
COV1 (X ₁₀)	% of the surrounding area not covered by forest nor by urban constructions	-0.39	0.32	0.23

*The school's surroundings were defined in QGIS 3.4 as a square area of 16,000 km², being each school at the center of that area. Such an area had the best balance between having the least overlap and closeness between each area around each school. Abbreviations in the “variable (code in equation)” are explained in the “name” column, that is why they are not explained here in the footnote. Numbered X's are a reference for the final equation for the STH model.

As random effects for the GLMM, we used the variable school (SCH) nested within district (DTR), as, in the previous multivariate analysis, region (and probably school) affected the localization of the variables in the PCA biplot (Figure 2).

To evaluate whether the inclusion of a random effect improves model fit, we compared a generalized linear mixed model (GLMM) with a standard generalized linear model (GLM) using AIC, log-likelihood, and AUC (area under the curve in an AUC-ROC analysis). GLMM encompassed the above-mentioned random effects. The GLMM presented a lower AIC and a higher value of log-likelihood compared to the GLM non-random effects model, indicating

improved fit. A likelihood ratio test established that the improvement provided by the inclusion of random effects was statistically significant. Table 5 shows model comparisons.

Table 5. Generalized model comparisons between a model with no random effects and the GLMM model with random effects for the STH infections outcome.

Model	df	Log-Likelihood	AIC	AUC	R ²
GLM (only fixed effects)	11	-2532.22	5086.45	0.70	0.22 (marginal)
GLMM (with location random effects)	13	-2443.97	4913.94	0.74	0.50 (conditional)

Abbreviations: GLM = General Linear Model; GLMM = General Linear Mixed Model; df = Degrees of Freedom; AIC = Akaike Information Criterion; AUC = Area Under the Curve

Inclusion of a random intercept for location substantially improved model performance, as shown by the lower AIC (4913.94 vs. 5086.45), higher log-likelihood (-2443.97 vs. -2532.22), and a higher AUC (>0.70 vs. 0.74) in the GLMM compared to the fixed-effects GLM. The likelihood ratio test further confirmed that the improvement was statistically significant ($p < 0.001$), supporting the relevance of accounting for clustering by location. (Table 5).

The marginal and conditional R² values were 0.22 and 0.50, respectively, indicating that both fixed and random effects models explained a substantial proportion of the variance, but greatly improved in the fixed effects model. In the test set, the model achieved an overall accuracy of 0.66 (95% CI: 0.63–0.68), with a kappa of 0.30 and balanced accuracy of 0.69. Importantly, sensitivity increased to 0.76, and specificity was 0.62, reflecting the model's improved ability to identify positive cases of infection following class balancing correctly. The F1 score was 0.54, and the Hosmer-Lemeshow test indicated an adequate fit ($p = 0.26$).

The overall influence of each GLMM variable on the outcome can be visualized employing the adjusted odds ratios and 95% confidence intervals in the forest plot (Figure 5).

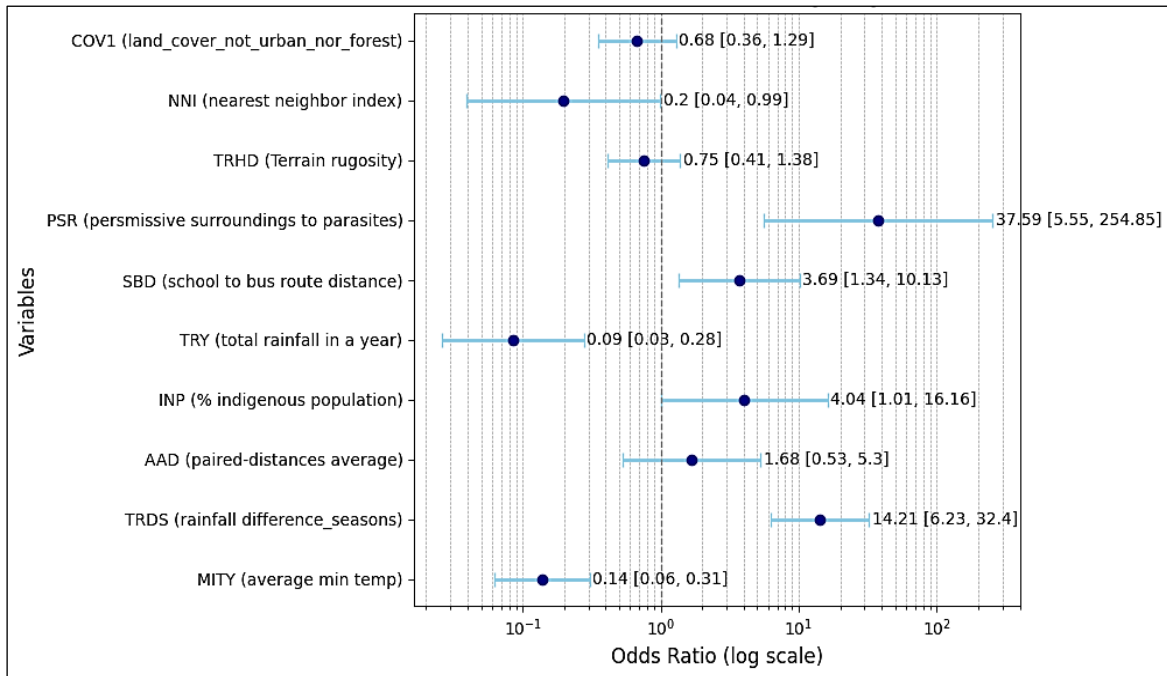


Figure 5. Forest plot showing the adjusted odds ratios to see the association of each variable in the model with the outcome variable (STH infection =positive). Blue points are the OR values, and horizontal dashed lines are the 95% confidence intervals. Significant associated variables do not cross the 1 value (10^0). Variables with the OR to the left of 1 resulted in a negative association with the outcome (protective), and variables with the OR to the right of 1 resulted in a positive association with the outcome (STH infection promoters).

Model Calibration and Fitness

The Brier Score for the model, a measure of calibration, is reasonable (BS=0.161) for a model of such a complex phenomenon. The calibration curve constructed with the predicted values vs the observed values can be seen in Figure 6.

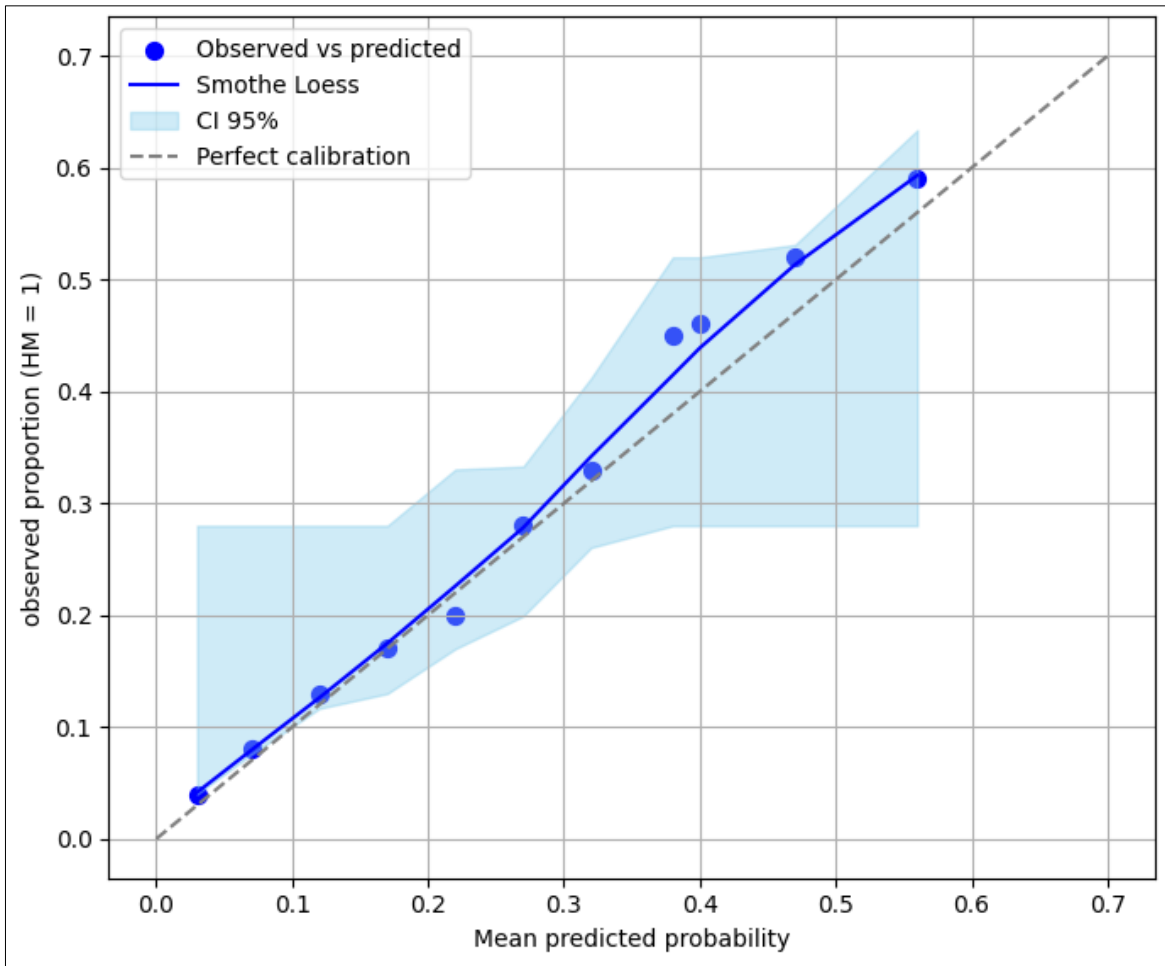


Figure 6. Calibration curve of the GLMM for STH infections. Blue points are the observed proportion values, and the dashed line marks the perfect calibration. Sky-blue shadow marks the variation of 95% intervals throughout the points. Brier score is moderate, BS=0.161, which implies a rationally good calibration of the model. LOESS = Locally estimated scatterplot smoothing; CI 95% = 95% confidence interval.

Finally, a Hosmer–Lemeshow test of goodness of fit for the model showed a value of $p=0.26$, showing no evidence of poor fit.

A formal predictive model for soil-transmitted infections in schoolchildren

We obtained a final model for the probability of STH infections,

$$P = \frac{1}{1 + e^{-n}}$$

where:

$$n = -4.05 - 1.98x_1 + 2.65x_2 + 0.52x_3 + 1.40x_4 - 2.46x_5 + 1.31x_6 + 3.63x_7 - 0.28x_8 - 1.63x_9 - 0.39x_{10}$$

Positive coefficients indicated increased odds of infection, and the negative ones suggested protective effects. Magnitude reflects the effect size of association (Table 3 for variable definitions and Figure 5 for the OR and 95% confidence intervals visualization through a forest plot).

Modelling of Determinants for intestinal protozoan infections in schoolchildren

Random Forest Analysis for potential predictors of intestinal protozoan infections

The RF final model was constructed with 700 trees, mtry = 40, node size = 5, and maxnodes = null. The model showed good consistency in performance across all data sets (training, validation, and testing). AUC results indicated a moderate discriminatory capacity (only slightly overfitting demonstrated by a small drop of ~0.05 point from training to test). In terms of sensitivity, it was robust enough to detect the true positives.

Precision and F1-score data indicated faint class imbalance (with lower precision in the minority class or positives) and a balanced performance despite data heterogeneity. These results demonstrate that the random forest model effectively distinguishes individuals with protozoan intestinal infections from those without, providing a solid basis for identifying key determinants and informing public health interventions (Table 6, Figure 7).

Table 6. Performance metrics of the Random Forest Model for intestinal protozoan parasite infection prediction across training, validation, and test datasets.

Dataset	AUC	Recall	Precision	Accuracy	F1-Score
Training	0.76	0.71	0.65	0.68	0.67
Validation	0.71	0.71	0.63	0.68	0.67
Test	0.66	0.62	0.64	0.64	0.62

AUC = area under the curve

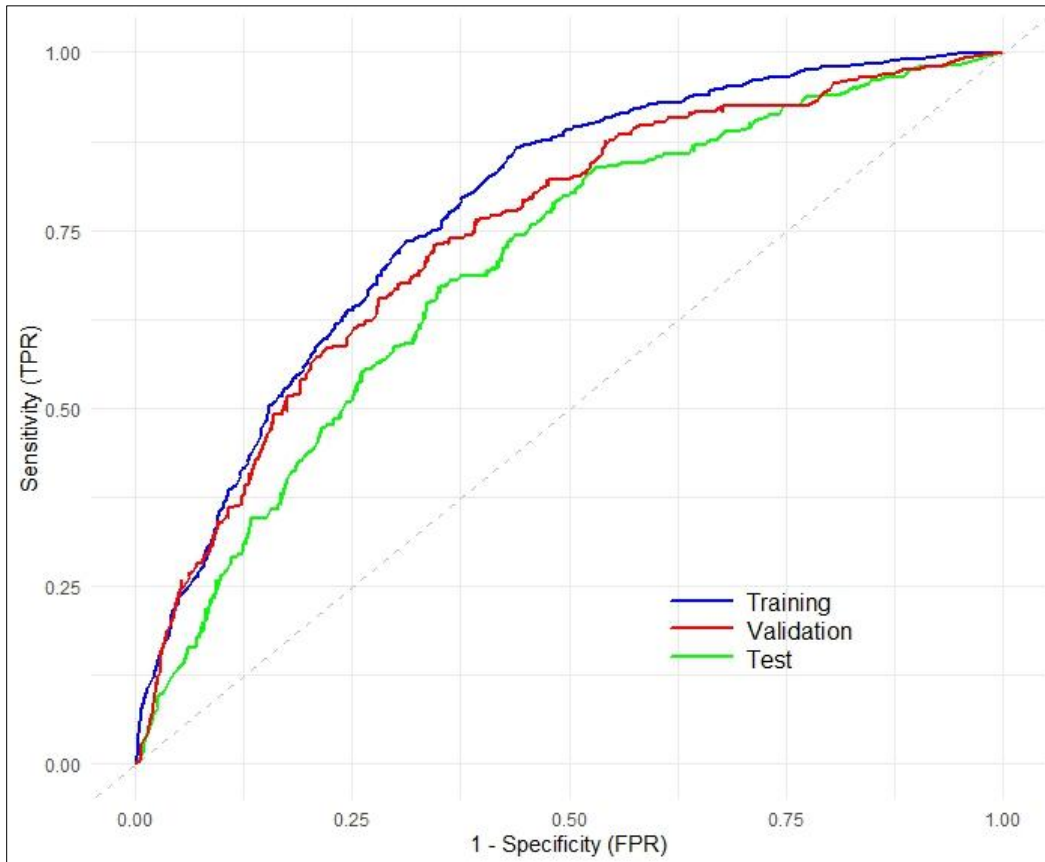


Figure 7. ROC Curves and AUC for Random Forest Model Predicting Protozoan Intestinal Infections (IP) Across Training, Validation, and Test Sets.

Variable Importance for Protozoan Intestinal Infections (IP) according to Random Forest Model.

The Mean decrease in accuracy tool was used to identify the top 20 predictors that were most strongly associated with protozoan intestinal infections (IP), agreeing with the Random Forest model. These variables included environmental, socioeconomic, educational, land configuration factors, and factors associated with the difficulty of geographic access. This reflects the multifactorial nature of intestinal protozoan infection risk.

The most influential variables were the unsatisfied basic needs indicator in the community (UBN), possessing direct health insurance (DHI), age (AGE), the distance from school to nearest population center health facilities (SNHD), permissive-to-parasite surroundings (PRS), and others (see the complete list in Table 7).

Table 7. Top 20 most important variables influencing protozoan intestinal infections in schoolchildren from Costa Rica, determined by the Random Forest Analysis, based on Mean Decrease in Accuracy.

Variable code	Variable name	Mean Decrease in Accuracy
UBN	Presence of 2 or more unsatisfied basic needs	91.9
DHI	Direct health insurance	45.9
AGE	Age in years	25.7
SNHD	Total distance in m from the school (from which the report came) to the nearest population center health tertiary level facilities	19.0
PSR	Mean area permissive to parasites transmissible forms in the school surroundings*	9.5
DRBS	Difference in total rainfall between dry and rainy seasons	8.4
TRY	Total annual rainfall in cm in the year and the area of the record	7.8
NNI	Nearest neighbor index	5.5
AAD	mean of all paired distances between constructions in the surroundings of the school*	5.5
MITY	The absolute difference between the average yearly temperature and the average yearly minimum temperature	5.5
RH	Average year relative humidity	5.1
UCOV	Urbanized land coverage (%)	5.1
COV1	Non-urbanized or forest land cover (%)	4.9
HI	Having health insurance	4.8
DRY	Number of days with rain in the year and the region of the record	4.2
UDEN	Urban density (houses per m2) in the school surroundings*	4.2
DAAD	Standard deviation of the AAD	3.9
TMMD	The absolute difference between the average year temperature and the average year maximum temperature	3.9
ATI	Average terrain inclination values in the route from school to the nearest population center health tertiary level facilities	3.9
GNR	General Rugosity in the school surroundings*	3.7

*Using QGIS 3.4, the school's surroundings were delineated as a square area of 16,000 km² with the school at its center. Such an area had the best balance between having the least overlap and closeness between each area around each school. Abbreviations in the “variable (code)” are explained in the “variable name” column; that is why they are not presented here in the footnote.

Elastic Net regression model for the selection of determinants of intestinal protozoan infections

An Elastic Net regression approach was applied to further refine the set of predictors for protozoan intestinal infections, combining LASSO and Ridge regularizations. For IP infections, the optimal α was 0.5, balancing L1 and L2 penalties. Here, $\lambda = \text{lambda.min}$ was selected to maximize predictive accuracy. The resulting model identified five key predictors. The model had a good balance between deviance and sparsity, being useful to identify the strongest predictors for the outcome variable IP without overfitting (Figure 8)

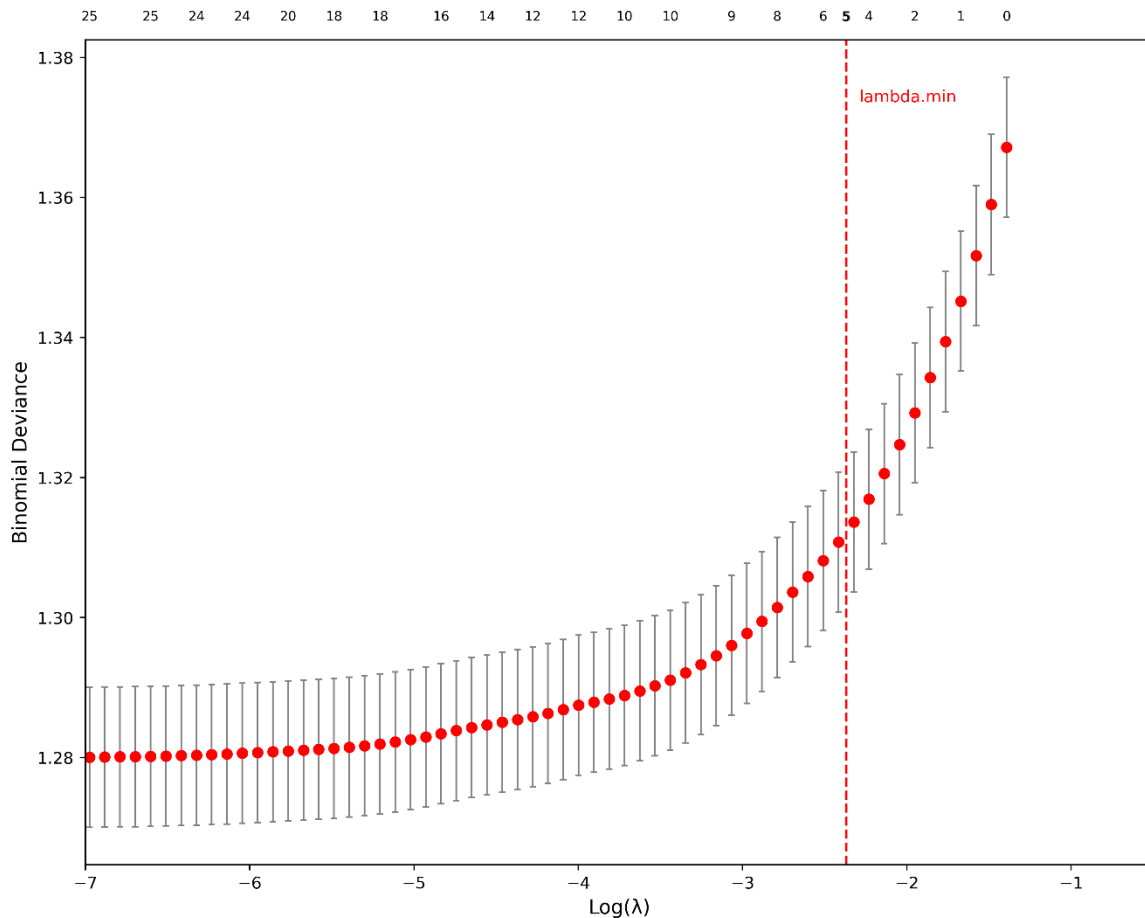


Figure 8. Elastic Net model for selecting determinants of intestinal protozoan infections, with $\alpha=0.5$ and the best lambda value (lambda.min). On the top numbered line is the number of predictors. The graphic shows the balance between binomial deviance (predictive accuracy of predictors) and regularization (lambda) strength. The model has a good balance, such as helping find the simplest model in which variables are the strongest predictors.

The resulting model identified five key predictors: UBN (unsatisfied basic needs), TRY (total annual rainfall), TMXD (temperature range), NNI (nearest neighbor index), and AGE (age in years).

A Variance Inflation Factor (VIF) analysis was performed to investigate the possibility of high residual collinearity between the selected variables, but the results of such analysis ruled it out. The magnitude and sign of the Elastic-Net Model coefficients suggested positive and negative associations between the selected variables and the positivity to intestinal protozoan infection outcomes (Table 8). Final model performance was strong, with AUC = 0.83. These selected variables represent the most robust and relevant predictors from the initial candidate set, supporting a more parsimonious and interpretable final model for IP.

The cross-validation plot of binomial deviance against the log of the penalization parameter (λ) confirmed the selection's stability and optimality. This process substantially reduced model complexity while retaining the variables with the highest predictive value.

Table 8. Non-Zero Elastic Net Coefficients of the selected variables in the IP model.

Variable code	Variable name	Elastic Net coefficient value
UBN	Unsatisfied basic needs indicator.	1.2
TRY	Total rainfall in the year and in the region from which the record came.	0.4
TMXD	Difference between the average temperature and the average maximum temperature in the year and in the region from which the record came.	0.25
NNI	Nearest neighbor index in the spatial nearest neighbor analysis.	-0.3
AGE	Age in years.	-0.04

Abbreviations in the “Variable code” column are explained in the “Variable name” column. That is why they are not described here in the footnote.

Generalized Linear Mixed Model (GLMM) for Protozoan Intestinal Infections (IP)

Variance inflation factor (VIF) analysis showed all values <5 , demonstrating no residual collinearity problems. The Elastic-Net-selected variables were used to perform a generalized linear model (GLM) and a generalized mixed linear model (GLMM) (details explained in the

methods section). Table 9 shows the results of the metrics comparing the GLM (no random effects) and GLMM (with and without random effects). Inclusion of random effects (GLMM) showed a slight improvement in the metrics. Final results for GLMM model can be seen in Table 10. All variables selected by Elastic-Net, but NNI were statistically significant. A higher value of NNI and older age had negative coefficients.

Table 9. Comparative metrics of the generalized models applied to the selected predictors for the intestinal protozoan infections as outcome.

Model	AIC	Log-likelihood	R ² Nagelkerke	R ² Marginal	R ² conditional
GLM	6165.2	-3076.6	0.13	-	-
GLMM	6111.4	-3042.3	-	0.13	0.17

Abbreviations: GLM=General Linear Model; GLMM=General Linear Mixed Model; AIC= Akaike Information Criterion

Table 10. Results of the of the GLMM model for each variable in the IP infections model.

Variable (code in equation)	Variable Name	Coefficient value	Standard error	p-value
Intercept (B ₀)	Intercept	-1.58	0.28	<0.0001
UBN (X ₁)	Unsatisfied basic needs	2.18	0.22	<0.0001
NNIS (X ₂)	Nearest neighbor index	-1.02	0.53	0.05
TMXD (X ₃)	Maximum -mean temperature difference	0.66	0.30	0.03
AGE (X ₄)	Age in years	-0.04	0.01	<0.001
TRY (X ₅)	Total rainfall (in cm) in a year	0.46	0.19	0.02

Abbreviations in the “variable (code in equation)” are explained in the “variable name” column; that is why they are not described here in the footnote. The numbered X are a reference for the final equation (IP final model). Explained variance by District resulted in negligible, so the model was run without nesting school within district, and so using only the school as a random effect. The variance explained by the inclusion of school as a random effect was 18.7% +/- 4.3%. Abbreviations already explained in the “name column”. GLMM=general linear mixed model.

GLMM with random effects (school nested within district), as deduced from Table 10, added only 3.5% to the R² parameter, improving the model's explanatory power slightly but not substantially. The same is true for the other comparative metrics. Even though the district random effect accounted for 0% of the unexplained variance in the outcome variable, the

school random effect explained up to 18.7% of that unexplained variance, justifying the inclusion of this random effect to improve the model (Table 9).

In terms of the coefficients and odds ratios found for the GLMM final model, Table 10 shows the role of each selected variable in the model. The odds of having an intestinal protozoan infection in children in Costa Rica were 8.8 times higher (95% CI: 2.0-9.6%) in children living in a population with a high percentage of people with two or more unsatisfied basic needs (Figure 9).

Higher differences between the mean and higher temperatures throughout the year, as well as higher quantities of total rain in a year, are also related to an increase in the odds of having intestinal protozoan infections. Conversely, a diminution in the probability of having an intestinal protozoan infection is associated with an older age (older age is protective against these infections) (Table 10).

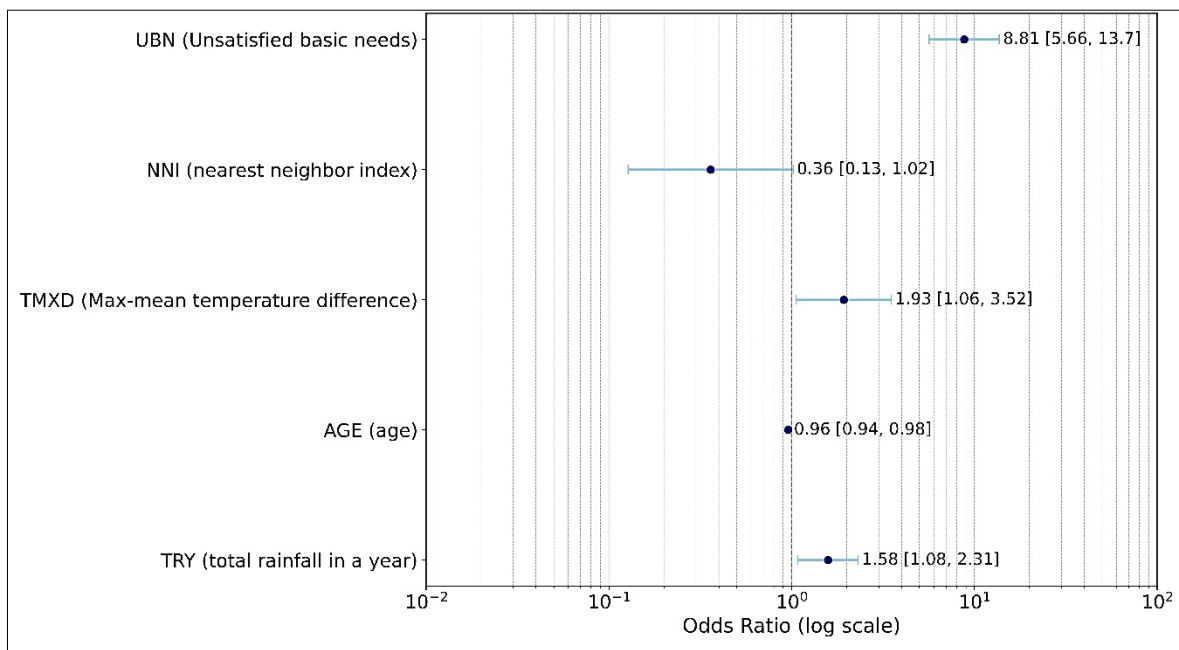


Figure 9. Forest plot for the adjusted ORs and 95% intervals of each variable in the final GLMM model for intestinal protozoan infections in Costa Rican schoolchildren. GLMM=General linear mixed model.

Model calibration and fitness

As the Hosmer-Lemeshow goodness of fit (GOF) test showed, the null hypothesis cannot be ruled out ($\chi^2=27.6$, $df=23$, $p\text{-value}=0.23$) because the p -value is greater than 0.05. The calibration curve can be seen in Figure 10, showing a good calibration of the model.

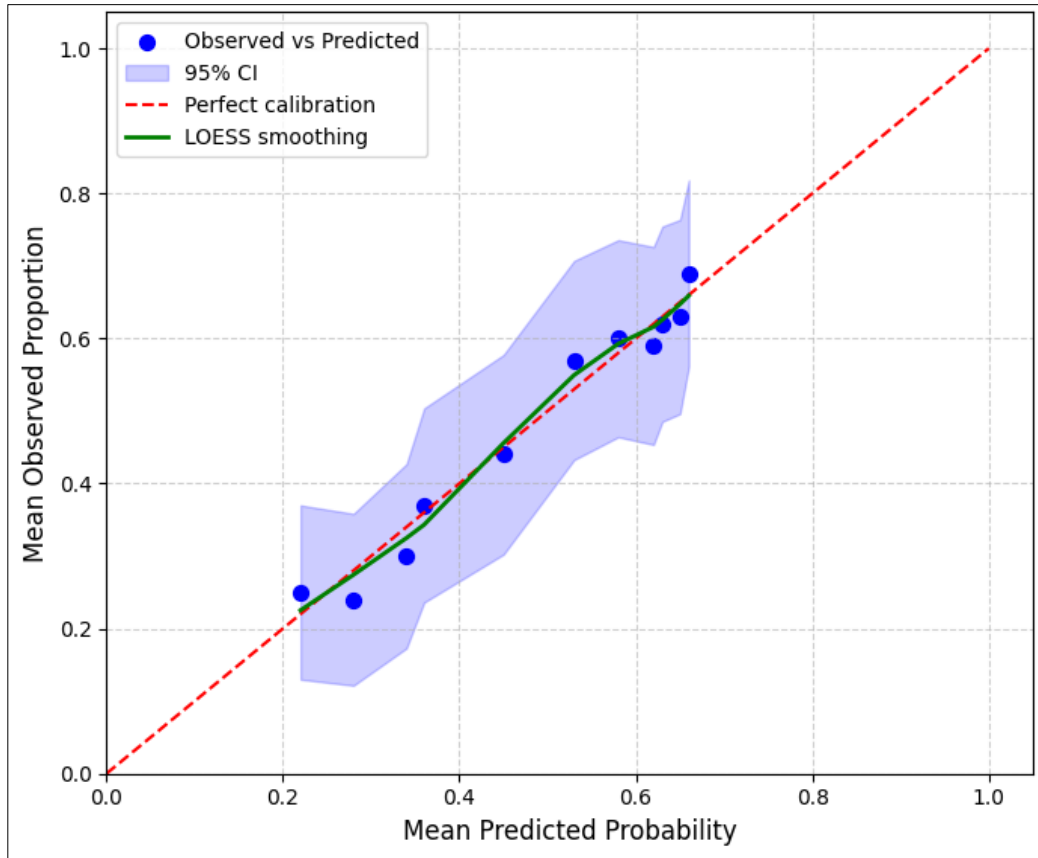


Figure 10. Calibration curve with LOESS for intestinal parasite infection (IP). Blue points are the observed proportion values, and the red dashed line marks the perfect calibration. Purple shadow marks the variation of 95% intervals throughout the points. Brier score is moderate, $BS=0.071$, which implies a good calibration of the model. LOESS = Locally estimated scatterplot smoothing; CI = confidence interval.

A formal predictive model for intestinal protozoan infections in schoolchildren of Costa Rica

Although the explanatory power of the models for intestinal protozoan infections was lower than for soil-transmitted helminths, the model had good calibration and fitness, and a reliable and not negligible explanatory value for such a complex phenomenon.

Based on the available data after all these analyses, here we propose a model for predicting intestinal protozoan infections in schoolchildren from Costa Rica, a model that can surely be improved as we will be capable of improving our epidemiological data:

Model Estimates

We obtained a final model for the probability of IP infections,

$$P = \frac{1}{1 + e^{-n}}$$

where,

$$n = -1.58 + 2.18x_1 + 1.02x_2 + 0.65x_3 - 0.04x_4 - 0.46x_5$$

Positive coefficients indicated increased odds of infection, and the negative ones suggested protective effects. Magnitude reflects the effect size of the association. Table 3 presents the variable definitions, and Figure 9 illustrates the ORs and their 95% CI visualization through a forest plot.

Discussion

PCA analysis

Schools from Chirripó appeared distinctly separated from other observations, and schools from the Brunca and Huetar Norte regions also exhibited clear clustering patterns, suggesting region-specific variation in the underlying structure of predictors captured by the PCA. STH and IP were strongly aligned with variables related to geographical access difficulty and certain environmental and housing indicators. In contrast, STH showed a closer association with other socioeconomic and educational factors. These preliminary associations suggest different drivers for protozoan versus helminth infections and highlight the potential multicollinearity between groups of predictors.

For soil-transmitted helminth infections, key associated variables included a parasite-permissive immediate environment (PSR), higher unemployment rates (UE), greater relative humidity (RH), percentage of the population working in agriculture (EAG), more rainy days

per year (DRY), low classification in the community Social Development Index (SDIQ_1), greater distances to the nearest clinic or hospital (SNHD), and the nearest bus route (SBD), as well as lower urbanization and greater distances between houses (UCOV, NNI, AAD). Conversely, higher urban population density (DENS), greater housing density (UCOV), living in informal, poorly constructed settlements (PREC), or belonging to communities with higher social development (SDIQ_3, 4, or 5) appeared to be protective (Figure 2).

Intestinal protozoan and soil-transmitted helminth infections in schoolchildren from Costa Rica, determinants analysis shows aggregated patterns among the schools analyzed. Patterns observed suggest the influence of both regional and small-scale local factors influencing those outcomes. Similar micro-epidemiological patterns have been documented in the case of urinary schistosomiasis, where the proximity of water sources in which the intermediate hosts are known to be present is associated with the infection risk and spatial clustering (29). Likewise, aggregation linked to microgeographic determinants has been observed in malaria (30), and local climatic, housing conditions, and other small-scale local determinants have been found influencing soil-transmitted helminth infections in studies in Brazil (31). As all those studies suggest, the immediate and regional factors strongly affect parasitic infectious diseases, like other diseases.

PCA biplot of loading vectors depicts the possible configuration of associations among multiple determinants of soil-transmitted helminth infections and intestinal protozoan infections in Costa Rican schoolchildren. Response variables (STH, IP) are plotted near vectors which represent social and bad housing conditions, such as the unsatisfied basic needs indicator (UBN), illiteracy (AA), female illiteracy (WAA), Non-remunerated employment (NRE), only indirect health insurance (IHI), or substandard housing (SHC). These observations tell us about important structural, socio-economic, and poverty-related factors associated with the risk of having these outcomes.

Previous studies have also found the same patterns of association between marginalization, poverty, and parasitic intestinal infections (1,7,8,11,17,31,32). The study of PCA loadings also reveals multiple potential close relationships between socioeconomic, educational, and geographical access difficulty factors that may interact in shaping the probabilities of these infections. This seems to be common in other studies (7,10,33).

A machine-learning-based pipeline to find the most influential factors associated with intestinal parasitic infections:

Random Forest initial predictor selection

A robust and stable performance was observed for the Random Forest analysis (RF) in all three data subsets for STH and IP infections, with better performance indicators for STH infections than for IP infections. High values of AUC and recall give reliability to RF models to correctly classify positive cases of STH and IP infections, even when exposed to new data. Hence, the sensitivity of the models is good. Despite STH positivity presented with class imbalance (managed through the SMOTE technique), the model maintained good accuracy, moderate precision, and F1-scores, thus highlighting its reliability and generalizability.

Random Forest has previously demonstrated a better performance in comparison to other machine-learning-based similar techniques, such as XGBoost, Support Vector Machines (SVM), for the reduction of variables associated with intestinal parasitic infections (20,34). RF can be explained as a set of classification and regression trees, which depict simple models using binary ‘decision’ splits on predictor variables to reach predictions on outcomes (34).

Random Forest Analysis makes sense in its application as a first step to determine the most influential variables. This is because these models achieve notable accuracy and possess strong classification capabilities (20,35). Also, RF is a good technique to manage a large and diverse set of predictors (multidimensional), where many of those were not linearly related to the outcome, being robust and capable of capturing that type of complexity. Also, this type of model manages the outliers, multicollinearity, and noise present in our initial data. Finally, its interpretability through “variable importance” based on Mean decreased accuracy helps identify clearly and easily the most influential variables as an initial step (20,34-36).

RF models provided an initial selection of predictors. That selection revealed different profiles of determinants for STH and IP infections in school-aged Costa Rican children. Environmental and spatial variables were the most influential variables in the case of STH infections. They included the percentage of urban land coverage (UCOV), the minimum annual temperature (MITY) in the year and locality of the observation, seasonal rainfall quantity variation (DRBS), total quantity of days with rainfall (DRY), and urban construction dispersion measure (AAD) within the 16 km² of school surroundings.

Ecological evidence shows that humidity, temperature, and land-use patterns substantially impact helminth transmission cycles (31,33). A tightly packed urban configuration can diminish the spaces where the transmission forms of soil-transmitted helminths can persist, grow, and become infective. Moreover, heavy urbanization affects soil composition and humidity, as noted in other studies, which may impact the embryonation of STH eggs (37).

Remarkably, the proportion of indigenous population (INP) in the district also appeared as a significant predictor, maybe explained by an underlying structural condition of vulnerability that has been previously associated with elevated STH burden in marginalized populations (8,11,17,38).

The RF model for IP infections identified a broader range of key determinants with strong predictive influence. Among the socioeconomic and health access predictors, the most relevant were deficit domains (presence of two or more unsatisfied basic needs in a high percentage of the population), health insurance coverage (DHI, HI), and age.

These findings are consistent with existing knowledge that protozoan infections are closely related to contaminated water sources, poor sanitation, bad hygiene, and limited access to healthcare services (16,39,40). While the RF model for STH shared certain variables with the model for IP infections, many predictors were unique to each, which is consistent with the distinct biological cycles and ecological niches of these pathogens.

In conclusion, RF analysis highlights that although both infection types share some ecological and structural risk factors, among the studied predictors, those affecting IP infections have more weight in the socio-economic factors. In contrast, a heavier impact can be seen in more ecologically related factors for STH infections. These differences emphasize the importance of adapted pathogen-specific intervention strategies that integrate environmental management and social protection measures to reduce transmission of both parasites effectively.

Elastic-Net analysis for further predictor selection

RF models are easily interpretable, but do not tell us an accurate and precise measure of the magnitude and direction of the influence of a given variable. Also, Random Forest can select many variables, making the model extremely complex. Other techniques can be

applied to reduce the dimensions further and find the most highly influential predictors of the outcome variables.

Elastic-Net is a penalized regression technique that combines the properties of the Least Absolute Shrinkage and Selection Operator (LASSO) and the better prediction capabilities of the Ridge Regression. This is accomplished by utilizing a mixing parameter called alpha, which permits the management of issues where there are high levels of collinearity between predictors and a high number of variables persist (41).

The EN analysis permitted the construction of parsimonious models for both STH and IP infection determinants. Predictive profiles were, though, different for both outcomes. For STH, EN analysis retained 10 variables, underscoring the importance of climatic factors (seasonal rainfall, total rainfall and average minimal annual temperature), land configuration (constructions dispersion measurements and type of land coverage) and of geographic access difficulty (total distance to the nearest bus route and the terrain rugosity in the route to the nearest population center's health facilities).

Seasonal rainfall differences and construction dispersion showed positive associations with the infections, which coincides with previous studies that show that marked rainfall seasonality and dispersing of houses make it difficult to access basic services, hindering access to preventive medications (e.g., mass drug administration) (42). Also, more dispersed constructions and differences in seasonal rainfall can favor STH infections in other ways. More dispersed houses leave more available transmission forms to persist and the eggs to embryonate. In fact, in the PCA initial analysis, the percentage of permissive-to-parasitic environment was very correlated to urban coverage and to construction dispersion (31,33).

Soil-transmitted helminths develop better in a range of temperatures between 25-30°C. *Ascaris lumbricoides* eggs have been shown to lose their viability fast below 20 °C and beyond 35°C; *Trichuris trichiura* and hookworm eggs are even more susceptible to temperatures outside that range (43-45). On the other hand, that interrelation of variables (warmer temperatures, patchy distribution of houses, a more moderate annual precipitation value, and a higher seasonal difference in total rainfall could configure landscapes with better humidity, temperature, and spaces for the transmissible forms to survive, as these conditions modulate the soil temperature and humidity (46).

It is known that less humidity, for example, can permit a higher time of survival of STH eggs developed at $>30^{\circ}\text{C}$. It would be necessary to understand if some egg pathogens (e.g., fungal) are reduced in their impact on the eggs in such conditions. However, this could help explain why a higher difference in total rainfall between dry and rainy seasons seems to be related to higher STH infection probability (43,44). Mathematical models have been developed to explain that phenomenon, a sort of trade-off between development and survival of eggs, where periods of importance but not so heavy rains and less temperature help the eggs to disperse, prolong the survival of eggs but delay their development, and dry seasons with warmer temperatures, less rain permit the deposited eggs to accelerate their embryonation, but reduce their survival. Understanding these dynamics is very important in mass drug administration and control (46,47).

A negative association between total rainfall in a year and the probability of STH was unexpected. Several studies have found positive correlations between total rainfall in a year and the risk of ascariasis and trichuriasis (31). Others have found a negative relation and a more critical positive relation with days with rain than the total rainfall (48). The EN for STH also found the total days with rain in a year for the observation region as a selected factor favoring the STH infections. The specific role of rain in a context where many other factors are intervening is complex to interpret. Rainfall is undoubtedly crucial in modulating soil moisture, as it is a necessary condition for STH transmissible forms. However, heavy rains with brief periods of interruption can flush out these forms, reducing the likelihood of eggs maturing and reaching the human host (46).

The total distance to the bus route is related to the geographic difficulty of accessing the nearest population center's health facilities for many of the schoolchildren from vulnerable zones. This variable was associated with other geographic access difficulty variables previously selected by the RF model, which included the terrain rugosity and total distance to the nearest second or third-level health facility.

A higher representation of the indigenous population is an interesting finding that needs further analysis. It is remarkable how these populations tend to have higher prevalences of STH in Costa Rica than other rural populations. It is possible that cultural factors, not studied here, are involved. The authors have observed, for example, how some of these populations deny using a sanitary paper technique for after defecating sanitization. These indigenous

people have pointed out that they think sanitary paper for after-defecating hygiene is incomplete. They prefer searching other places near little water sources to defecate (open-air defecation, OAD) for later water-based after-defecating hygiene, as several other cultures around the world do. Doubtless, further analysis of these potentially related cultural issues is necessary for clarification.

However, available data indicate that a significant portion of the indigenous population is highly correlated with a low value in the Social Development Index (SDI), and a high percentage of the population is experiencing critical shortcomings (two or more). That is surely related to more essential indicators of poverty for this population; as mentioned above, poverty is highly associated with more risk for STH infections (1,7,8,11,17).

A higher percentage of bare soil (PSR), that is, soil that is not covered by cement, constructions, or dense vegetation, provides a more favorable environment for the survival and development of transmissible stages of parasites such as STH eggs and larvae, increasing the likelihood of their persistence and eventual transmission to human hosts. This characteristic was strongly associated with favoring STH infections, which have also been investigated as a risk factor in previous studies. Nearness to urbanized areas or living in strongly urbanized areas can also be related to better access to healthcare, to sanitation of the environment and wastewater elimination systems, but also be implied in less space for transmissible forms to survive (less bare soil and as such areas) (37).

In the case of IP infections, Elastic Net selection helped substantially reduce the complexity of the initial model for IP infections, retaining the most robust and stable predictors. The unsatisfied basic needs indicator (UBN) has a major positive relation with the IP infections. This is a multidimensional indicator that measures how many homes are deprived in essential dimensions for the well-being of their inhabitants. This indicator encompasses dimensions like housing conditions, health, and economic capacity. For this study, we used the percentage of unsatisfied basic needs in the district of the record for two or more dimensions unsatisfied (49).

Supplemental material B has a detailed description of this indicator. Still, here, it is important to mention that among the several components of this indicator, access to safe drinking water and an adequate feces elimination system are included. Also, educational and other socioeconomic variables are considered. In less robust models run before in which the

basic needs indicator was excluded, variables related to drinking water quality and feces elimination systems appeared as significant predictors, confirming their essential role in predicting STH and IP infections. This could explain why models selected UBN instead of several expected variables influencing IP infections, like those related to feces elimination, water sources, education, and other socioeconomic factors. The percentage of the indigenous population in a community is 99% correlated to UBN, reflecting that this indicator is also helpful in predicting STH infections in schoolchildren from Costa Rica.

STH and IP final models.

The STH infections GLMM final predictive model comprises a robust group of determinants with statistically significant associations. The extension of bare soil permissive to parasites transmissible forms (PSR) and seasonality of rains (DRBS) showed strong positive links to the risk of STH infections. Also, a socioeconomic variable, the percentage of the indigenous population, had a significant positive association.

This highlights the importance of a combination of socioeconomic and spatial configuration factors contributing to transmission. Additionally, accessibility indicators such as distance to the nearest official bus route and terrain rugosity in the distance to nearest population center are a set of variables that bind characteristics of challenging topography together with distance for access to healthcare and other services, increasing the odds that geographic barriers can hinder healthcare access and promote exposure to parasites.

On the contrary, a high level of total rainfall in a year (TRY), a lower average minimum temperature in a year in a region (MITY), and the pattern of houses aggregation (NNI) help to configure protecting associations as found in other studies (37,44,50,51). These findings support the notion that STH infections are strongly affected by direct climatic variables, climatic modulators, and spatial configuration factors related to accessibility. These factors showed stronger and more significant effects than many other variables tested in previous models, and they undoubtedly interact with those earlier variables.

As shown for *A. lumbricoides*, seasonal variations in rainfall and warmer temperatures influence egg humidity, which in turn affects egg survival. Optimal survival occurs when humidity is moderate under high average temperatures. In contrast, excessive humidity in

warm or hot climates may promote fungal growth on eggs or introduce other environmental factors that compromise their viability, as observed with *Ascaris* eggs (31, 43, 50). Many places in Costa Rica are already places with high level of rain, maybe, enough rain to support *Ascaris* eggs development, so, it is possible that a having a slightly bigger difference in rain between the dry and the wet season could be suitable for the eggs so they could be not exposed to excessive humidity (43, 47).

An extensive study in Brazil examined the associations between STH infections and various climatic, spatial, and socioeconomic predictors. Although they did not use the same variables, they found some interesting relations. They found a higher percentage of urban population was negatively associated with hookworm infections. The strong positive association of our variable PSR (bare soil permissive to parasite transmissible forms) could be in part related to that issue, a less crowded organization of human construction that, modulated by the type of land coverage, humid conditions, and social structural issues (like no systems for fecal waste adequate elimination), could lead to more places for the parasites transmissible forms to survive (31). Low temperatures and higher rainfall levels serve as protective factors in some scenarios, where they also facilitate excessive humidity and the growth of probable egg pathogens. Additionally, lower temperatures make egg maturation more difficult. Additionally, lower temperatures make egg maturation more difficult. Additionally, lower temperatures make egg maturation more challenging (43,47).

The sum of the effects on STH infections is found in this model. This indicates that spatial and climatic variables have a more substantial impact on STH infections than the socioeconomic variables included in the analysis. Even so, the percentage of indigenous population (INP) had a significant positive association with the parasite, as well as the distance to the nearest bus route. Both indicators are related to poverty and the difficulty of access. The INP variable exhibited high collinearity and correlation with the unsatisfied basic needs indicator in the initial multivariate analysis and correlations. This suggests that these factors also play an important role in modulating the probability of STH infections.

The Sequential modeling process selected INP, which may indicate other factors captured by this characteristic, possibly beyond the structural shortcomings in the population. Some cultural practices could have a role. Other studies have found cultural differences associated

with parasitological outcomes, although a systematic analysis of this is still lacking among the evidence available to support this notion (11,33).

In contrast, the strongest predictor for protozoan infections was structural poverty, represented by the percentage of the population with two or more critical unsatisfied basic needs (UBN), which increased the odds of infection nearly ninefold. This aligns with extensive literature linking protozoan infections such as *Giardia* or *Entamoeba* to household-level poverty and inadequate sanitation (52,53). Climatic drivers—particularly total annual rainfall and thermal amplitudes— were also positively associated with infection risk, suggesting a role for environmental persistence of protozoan cysts under moist and thermally dynamic conditions (54).

Older age was protective, likely reflecting improved hygiene and partial acquired immunity with age (54,55). The model had modest explanatory power (R^2 marginal = 0.13) but good calibration (Figure 9). It significantly improved when school-level random effects were included (R^2 conditional = 0.17), indicating within-school clustering of risk.

These models illustrate how protozoan and helminth infections respond to overlapping but distinct determinants. Protozoan risk is more tightly clustered within schools and tied to socio-demographic hardship, while broader spatial-environmental features shape STH risk. The findings underscore the need for differentiated context-sensitive control strategies that integrate landscape, climate, and social inequities to reduce parasitic disease burdens in endemic areas effectively.

Funding

Projects TC-552 and ED-511 have been funded by Vicerrectoría de Acción Social, Universidad de Costa Rica, from 2008 to 2019.

Availability of data and materials

An anonymized, summarized database and all R and Python code and calculations can be requested from the corresponding author. Please write to the corresponding author if you are interested in such data.

Declaration on the use of AI

The authors declare that AI was used exclusively for reviewing the correctness/conciseness of English writing, as the authors' maternal language is Spanish.

Ethics approval

The ethics committee at Universidad de Costa Rica granted permission to develop this research through ordinary session No. 209-2021 and official document CEC-552-2021, under the modality of extended consent, given the complete anonymity of the data.

Competing interests

The authors declare that they have no competing interests.

References

1. Chen J, Gong Y, Chen Q, Li S, Zhou Y. Global burden of soil-transmitted helminth infections, 1990–2021. *Infect Dis Poverty*. 2024 Dec 1;13(1). Available from: <https://idpjournal.biomedcentral.com/articles/10.1186/s40249-024-01238-9>
2. Pullan RL, Smith JL, Jasrasaria R, Brooker SJ. Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. *Parasit Vectors*. 2014 Jan 21;7(1):1–19. Available from: <https://parasitesandvectors.biomedcentral.com/articles/10.1186/1756-3305-7-37>
3. Hall A, Hewitt G, Tuffrey V, De Silva N. A review and meta-analysis of the impact of intestinal worms on child growth and nutrition. *Matern Child Nutr* [Internet]. 2008 Apr

- 14;4(s1):118–236. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/j.1740-8709.2007.00127.x>
4. Degarege A, Erko B, Negash Y, Animut A. Intestinal Helminth Infection, Anemia, Undernutrition and Academic Performance among School Children in Northwestern Ethiopia. *Microorganisms*. 2022 Jul 1;10(7):1353–63. Available from: <https://www.mdpi.com/2076-2607/10/7/1353>
 5. Bisetegn H, Debash H, Ebrahim H, Erkihun Y, Tilahun M, Feleke DG. Prevalence and Determinant Factors of Intestinal Parasitic Infections and Undernutrition among Primary School Children in North-Central Ethiopia: A School-Based Cross-Sectional Study. Lelo EA, editor. *J Parasitol Res* [Internet]. 2023 Mar 15;2023(1):1–10. Available from: <https://www.hindawi.com/journals/jpr/2023/2256910/>
 6. Assemie MA, Shitu Getahun D, Hune Y, Petrucka P, Abebe AM, Telayneh AT, et al. Prevalence of intestinal parasitic infection and its associated factors among primary school students in Ethiopia: A systematic review and meta-analysis. Bentwich Z, editor. *PLoS Negl Trop Dis* [Internet]. 2021 Apr 27;15(4):e0009379. Available from: <http://dx.doi.org/10.1371/journal.pntd.0009379>
 7. Ross AG, Papier K, Luceres-Catubig R, Chau TN, Inobaya MT, Ng SK. Poverty, dietary intake, intestinal parasites, and nutritional status among school-age children in the rural Philippines. *Trop Med Infect Dis*. 2017;2(4):1–10.
 8. Almeida MM, Monteiro KJL, Bacelar PAA, Santos JP dos, Freitas SPC de, Evangelista BBC, et al. Interactions between malnutrition, soil-transmitted helminthiasis and poverty among children living in periurban communities in Maranhao State, Northeastern Brazil. *Rev Inst Med Trop Sao Paulo* [Internet]. 2020;62:1–9. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0036-46652020000100234&tlng=en
 9. Maru DS. Prevalence of Intestinal Parasitic Infections and Associated Risk factors among School children in Adigrat town, Northern Ethiopia. *International journal of Emerging Trends in Science and Technology*. 2017;04(01):4943–8.

10. Hotez PJ, Woc-Colburn L, Bottazzi ME. Neglected tropical diseases in Central America and Panama: Review of their prevalence, populations at risk and impact on regional development. *Int J Parasitol*. 2014;44(9):597–603.
11. Hotez PJ. Neglected Parasitic Infections and Poverty in the United States. Aksoy S, editor. *PLoS Negl Trop Dis* [Internet]. 2014 Sep 4;8(9):e3012. Available from: <https://dx.plos.org/10.1371/journal.pntd.0003012>
12. Gutierrez-Jimenez J, Torres-Sanchez MGC, Fajardo-Martinez LP, Schlie-Guzman MA, Luna-Cazares LM, Gonzalez-Esquinca AR, et al. Malnutrition and the presence of intestinal parasites in children from the poorest municipalities of Mexico. *The Journal of Infection in Developing Countries* [Internet]. 2013 Oct 15;7(10):741–7. Available from: <https://jidc.org/index.php/journal/article/view/24129627>
13. Bakarman MA, Hegazi MA, Butt NS. Prevalence, Characteristics, Risk Factors, and Impact of Intestinal Parasitic Infections on School Children in Jeddah, Western Saudi Arabia. *J Epidemiol Glob Health* [Internet]. 2019;9(1):81. Available from: <https://link.springer.com/10.2991/jegh.k.190219.001>
14. Wördemann M, Polman K, Menocal Heredia LT, Junco Diaz R, Collado Madurga A, Núñez Fernández FA, et al. Prevalence and risk factors of intestinal parasites in Cuban children. *Tropical Medicine & International Health* [Internet]. 2006 Dec 11;11(12):1813–20. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-3156.2006.01745.x>
15. Bahmani P, Maleki A, Sadeghi S, Shahmoradi B, Ghahremani E. Prevalence of Intestinal Protozoa Infections and Associated Risk Factors among Schoolchildren in Sanandaj City, Iran. *Iran J Parasitol* [Internet]. 2017 Jan;12(1):108–16. Available from: <http://ijpa.tums.ac.ir>
16. McKenna ML, McAtee S, Bryan PE, Jeun R, Ward T, Kraus J, et al. Human Intestinal Parasite Burden and Poor Sanitation in Rural Alabama. *Am J Trop Med Hyg* [Internet]. 2017 Nov 8;97(5):1623–8. Available from: <https://www.ajtmh.org/view/journals/tpmd/97/5/article-p1623.xml>

17. Ngonghala CN, Pluciński MM, Murray MB, Farmer PE, Barrett CB, Keenan DC, et al. Poverty, Disease, and the Ecology of Complex Systems. PLoS Biol [Internet]. 2014 Apr 1;12(4):e1001827. Available from: <https://dx.plos.org/10.1371/journal.pbio.1001827>
18. Bell J. What Is Machine Learning? In: Carta S, editor. Machine Learning and the City [Internet]. New Jersey: John Wiley and Sons Inc.; 2022. p. 207–16. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/9781119815075.ch18>
19. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. Nat Methods [Internet]. 2018 Apr 3;15(4):233–4. Available from: <http://dx.doi.org/10.1038/nmeth.4642>
20. Zafar A, Attia Z, Tesfaye M, Walelign S, Wordofa M, Abera D, et al. Machine learning-based risk factor analysis and prevalence prediction of intestinal parasitic infections using epidemiological survey data. Abatih EN, editor. PLoS Negl Trop Dis [Internet]. 2022 Jun 14;16(6):e0010517. Available from: <http://dx.doi.org/10.1371/journal.pntd.0010517>
21. INEC. Censo 2011, Costa Rica. 2016 [cited 2024 Dec 9]. Censo. 2011. Indicadores de estado y acceso a servicios básicos de la vivienda, según cantón y distrito. Available from: https://admin.inec.cr/sites/default/files/media/repoblaccenso2011-09.xls_3.xlsx
22. Google. Google Earth. 202 [cited 22 March 2025]. Available from: <https://www.google.es/intl/es/earth/index.html>
23. Malavasi-Ortiz E. Atlas Digital de Costa Rica [Internet]. Cartago: ITCR; 2014. p. 1–8. Available from: <https://hdl.handle.net/2238/6749>
24. QGIS Development Team. QGIS [Internet]. 2024. Available from: <https://qgis.org>
25. Le S, Josse J. FactoMineR: An R Package for Multivariate Analysis. J Stat Softw. 2008;25(1):1–18.
26. Pedregosa F, Varoquaux G, Gramfort V, Thirion B, Grisel O, Blondel M, et al. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research [Internet]. 2011;12(2011):2825–30. Available from: <http://jmlr.org/papers/v12/pedregosa11a.html>
27. Python Software Foundation. Python. 2024. Available from

28. Nakagawa S. A general and simple method for obtaining R² from generalized linear mixed-effects models. *Meth. Ecol. Evol* [Internet]. 2012 Dec;4(2):133-142.
29. Stothard JR, Mgeni AF, Khamis S, Seto E, Ramsan M, Rollinson D. Urinary schistosomiasis in schoolchildren on Zanzibar Island (Unguja), Tanzania: a parasitological survey supplemented with questionnaires. *Trans R Soc Trop Med Hyg* [Internet]. 2002 Sep;96(5):507–14. Available from: [https://academic.oup.com/trstmh/article-lookup/doi/10.1016/S0035-9203\(02\)90421-9](https://academic.oup.com/trstmh/article-lookup/doi/10.1016/S0035-9203(02)90421-9)
30. Myers WP, Myers AP, Cox-Singh J, Lau HC, Mokuai B, Malley R. Micro-geographic risk factors for malarial infection. *Malar J*. 2009;8(1):1–8.
31. Scholte RGC, Schur N, Bavia ME, Carvalho EM, Chammartin F, Utzinger J, et al. Spatial analysis and risk mapping of soil-transmitted helminth infections in Brazil, using Bayesian geostatistical models. *Geospat Health* [Internet]. 2013 Nov 1;8(1):97. Available from: <http://geospatialhealth.net/index.php/gh/article/view/58>
32. Magalhães AR, Codeço CT, Svenning JC, Escobar LE, Van de Vuurst P, Gonçalves-Souza T. Neglected tropical diseases risk correlates with poverty and early ecosystem destruction. *Infect Dis Poverty* [Internet]. 2023 Apr 10;12(1):32. Available from: <https://doi.org/10.1186/s40249-023-01084-1>
33. Pullan RL, Kabatereine NB, Quinnell RJ, Brooker S. Spatial and Genetic Epidemiology of Hookworm in a Rural Community in Uganda. Bethony JM, editor. *PLoS Negl Trop Dis* [Internet]. 2010 Jun 15;4(6):e713. Available from: <https://dx.plos.org/10.1371/journal.pntd.0000713>
34. Speiser JL, Miller ME, Tooze J, Ip E. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst Appl* [Internet]. 2019 Nov;134:93–101. Available from: <https://doi.org/10.1016/j.eswa.2019.05.028>
35. Rouza E, Arifin F, Suprpto. Essential Advances in Soil-Transmitted Helminth Detection Using Machine Learning and Deep Learning: A Systematic Review. *Int J Adv Sci Eng Inf Technol* [Internet]. 2024 Dec 25;14(6):2001–7. Available from: <https://ijaseit.insightsociety.org/index.php/ijaseit/article/view/20691>

36. Machado G, Mendoza MR, Corbellini LG. What variables are important in predicting bovine viral diarrhoea virus? A random forest approach. *Vet Res* [Internet]. 2015 Dec 24;46(1):85. Available from: <http://dx.doi.org/10.1186/s13567-015-0219-7>
37. Candela E, Goizueta C, Sandon L, Muñoz-Antoli C, Periago MV. The Relationship Between Soil-Transmitted Helminth Infections and Environmental Factors in Puerto Iguazú, Argentina: Cross-Sectional Study. *JMIR Public Health Surveill* [Internet]. 2023 Nov 7;9(1):e41568. Available from: <https://publichealth.jmir.org/2023/1/e41568>
38. González-Ramírez LC, Robalino-Flores X, De la Torre E, Parra-Mayorga P, Prato JG, Trelis M, et al. Influence of Environmental Pollution and Living Conditions on Parasite Transmission among Indigenous Ecuadorians. *Int J Environ Res Public Health* [Internet]. 2022 Jun 4;19(11):6901. Available from: <https://www.mdpi.com/1660-4601/19/11/6901>
39. Trinies V, Garn J V., Chang HH, Freeman MC. The Impact of a School-Based Water, Sanitation, and Hygiene Program on Absenteeism, Diarrhea, and Respiratory Infection: A Matched–Control Trial in Mali. *The American Society of Tropical Medicine and Hygiene* [Internet]. 2016 Jun 1;94(6):1418–25. Available from: <https://www.ajtmh.org/view/journals/tpmd/94/6/article-p1418.xml>
40. Muñoz-Antoli C, Pavón A, Marcilla A, Toledo R, Esteban JG. Prevalence and risk factors related to intestinal parasites among children in Department of Rio San Juan, Nicaragua. *Trans R Soc Trop Med Hyg* [Internet]. 2014 Dec;108(12):774–82. Available from: <https://academic.oup.com/trstmh/article-lookup/doi/10.1093/trstmh/tru160>
41. Zou H, Hastie T. Regularization and Variable Selection Via the Elastic Net. *J R Stat Soc Series B Stat Methodol* [Internet]. 2005 Apr 1;67(2):301–20. Available from: <https://academic.oup.com/jrsssb/article/67/2/301/7109482>
42. Smits HL. Prospects for the control of neglected tropical diseases by mass drug administration. *Expert Rev Anti Infect Ther* [Internet]. 2009 Feb 10;7(1):37–56. Available from: <http://www.tandfonline.com/doi/full/10.1586/14787210.7.1.37>
43. Maya C, Ortiz M, Jiménez B. Viability of *Ascaris* and other helminth genera non larval eggs in different conditions of temperature, lime (pH) and humidity. *Water Science and Technology* [Internet]. 2010 Dec 1;62(11):2616–24. Available from:

<https://iwaponline.com/wst/article/62/11/2616/15004/Viability-of-Ascaris-and-other-helminth-genera-non>

44. Senecal J, Nordin A, Vinnerås B. Fate of Ascaris at various pH, temperature and moisture levels. J Water Health [Internet]. 2020 Jun 1;18(3):375–82. Available from: <https://iwaponline.com/jwh/article/18/3/375/74162/Fate-of-Ascaris-at-various-pH-temperature-and>
45. Smith G, Schad GA. Ancylostoma duodenale and Necator americanus: effect of temperature on egg development and mortality. Parasitology [Internet]. 1989 Aug 6;99(1):127–32. Available from: https://www.cambridge.org/core/product/identifier/S0031182000061102/type/journal_article
46. Schüle SA, Clowes P, Kroidl I, Kowuor DO, Nsojo A, Mangu C, et al. Ascaris lumbricoides Infection and Its Relation to Environmental Factors in the Mbeya Region of Tanzania, a Cross-Sectional, Population-Based Study. Keating JA, editor. PLoS One [Internet]. 2014 Mar 18;9(3):e92032. Available from: <https://dx.plos.org/10.1371/journal.pone.0092032>
47. Cooper AJ, Hollingsworth TD. The impact of seasonality on the dynamics and control of Ascaris lumbricoides infections. J Theor Biol [Internet]. 2018 Sep;453:96–107. Available from: <https://doi.org/10.1016/j.jtbi.2018.05.025>
48. Gunawardena GSA, Karunaweera ND, Ismail MM. Wet-days: are they better indicators of Ascaris infection levels? J Helminthol [Internet]. 2004 Dec 12;78(4):305–10. Available from: https://www.cambridge.org/core/product/identifier/S0022149X0070318X/type/journal_article
49. Méndez F, Bravo O. Costa Rica, Mapas de Pobreza 2011. In: INEC, editor. Costa Rica a la Luz del Censo 2011. 2014th ed. San José: INEC; 2014. p. 482.
50. Ignacio CF, Silva MEC da, Handam NB, Alencar M de FL, Sotero-Martins A, Barata MM de L, et al. Socioenvironmental conditions and intestinal parasitic infections in Brazilian urban slums: a cross-sectional study. Rev Inst Med Trop Sao Paulo [Internet]. 2017 Aug 7;59:1–10. Available from:

http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0036-46652017005000229&lng=en&tlng=en

51. Mehraj V, Hatcher J, Akhtar S, Rafique G, Beg MA. Prevalence and Factors Associated with Intestinal Parasitic Infection among Children in an Urban Slum of Karachi. Sutherland CJ, editor. PLoS One [Internet]. 2008 Nov 10;3(11):e3680. Available from: <https://dx.plos.org/10.1371/journal.pone.0003680>
52. Tegen D, Damtie D, Hailegebriel T. Prevalence and Associated Risk Factors of Human Intestinal Protozoan Parasitic Infections in Ethiopia: A Systematic Review and Meta-Analysis. J Parasitol Res [Internet]. 2020 Oct 5;2020:1–15. Available from: <https://www.hindawi.com/journals/jpr/2020/8884064/>
53. Sarkari B, Hosseini G, Motazedian MH, Fararouei M, Moshfe A. Prevalence and risk factors of intestinal protozoan infections: a population-based study in rural areas of Boyer-Ahmad district, Southwestern Iran. BMC Infect Dis [Internet]. 2016 Dec 25;16(1):703. Available from: <http://dx.doi.org/10.1186/s12879-016-2047-4>
54. Alum A, Absar IM, Asaad H, Rubino JR, Ijaz MK. Impact of Environmental Conditions on the Survival of Cryptosporidium and Giardia on Environmental Surfaces. Interdiscip Perspect Infect Dis [Internet]. 2014;2014:1–7. Available from: <http://www.hindawi.com/journals/ipid/2014/210385/>
55. Guernier V, Brennan B, Yakob L, Milinovich G, Clements ACA, Soares Magalhaes RJ. Gut microbiota disturbance during helminth infection: Can it affect cognition and behaviour of children? BMC Infect Dis [Internet]. 2017;17(1):1–6. Available from: <http://dx.doi.org/10.1186/s12879-016-2146-2>

Discusión y conclusión general

Los sistemas de salud eficaces se caracterizan por una mayor inversión en estrategias preventivas y por la adopción de enfoques diversos que priorizan la prevención por encima del tratamiento. Para alcanzar este objetivo, es necesario desarrollar una comprensión más profunda de los determinantes de la enfermedad, no solo en cuanto a sus posibles interrelaciones, sino también en relación con el peso específico que estos determinantes tienen sobre la carga global de enfermedad.

Además, resulta fundamental evaluar la capacidad predictiva de los factores de riesgo potenciales en relación con problemas de salud y desenlaces específicos. Esta consideración adquiere particular relevancia en el contexto de las infecciones por parásitos intestinales, cuya prevalencia ha disminuido en algunas regiones como resultado de la urbanización y los avances en saneamiento básico en áreas urbanas de países de ingresos bajos. No obstante, estas infecciones continúan siendo altamente prevalentes en otras zonas menos favorecidas.

Una mejor comprensión de dichos predictores por parte de los sistemas de salud permitiría una asignación más eficiente de esfuerzos y recursos, lo cual resulta esencial para el fortalecimiento de las estrategias de prevención y control en contextos de alta vulnerabilidad, especialmente en países con recursos limitados.

Los dos estudios aquí integrados revelan, de forma complementaria, que las infecciones parasitarias intestinales continúan siendo un importante problema de salud pública en zonas vulnerables de Costa Rica, afectando de manera significativa a la niñez de muchas de estas regiones. Los datos presentados se centran en poblaciones con condiciones de alta vulnerabilidad frente a las infecciones parasitarias, lo cual puede explicar la elevada prevalencia general de cualquier parásito intestinal (56,4 %), un valor considerablemente mayor al reportado en encuestas nacionales previas, como la Encuesta Nacional de Nutrición 2008–2009. Esto sugiere una probable subrepresentación de la carga parasitaria en comunidades rurales e indígenas dentro de los sistemas oficiales de vigilancia.

Los resultados aquí expuestos permiten cuantificar de forma más precisa la magnitud del problema que representan las infecciones por helmintos y protozoarios intestinales en este tipo de población, constituyendo un valioso insumo para las políticas de salud pública. Estos hallazgos permiten dimensionar con mayor precisión la extensión de estas enfermedades y la urgencia de abordarlas en función de sus contextos socioterritoriales específicos. Asimismo,

cuestionan la percepción de que las parasitosis intestinales en la niñez costarricense constituyen un problema superado a nivel nacional.

Ambos estudios evidencian, además, que estas infecciones no se distribuyen al azar, sino que presentan patrones geoespaciales definidos. Las altas prevalencias observadas en ciertas zonas permiten identificar distritos con prevalencia significativamente elevada, considerados como puntos críticos (*hotspots*) de infección, como Chirripó, Limoncito, Telire y Los Chiles. Estas regiones comparten características estructurales comunes, como altos niveles de pobreza, baja cobertura de servicios básicos y condiciones ambientales favorables para la transmisión de helmintos y protozoarios, entre ellas suelos húmedos, alta pluviosidad y temperaturas cálidas.

En relación con los determinantes de estas infecciones, los análisis mostraron diferencias importantes entre las infecciones por helmintos transmitidos por el suelo (STH) y por protozoarios intestinales (IP). Se observó que factores ecológicos y geográficos tuvieron un peso significativo en la predicción de estas infecciones. Entre los factores más relevantes se identificaron la dispersión de las viviendas —la cual genera regiones con casas alejadas entre sí, dificultando la provisión de agua potable y un adecuado manejo de excretas—, la estacionalidad de las lluvias, la rugosidad del terreno y la distancia a los servicios de salud.

En contraste, las infecciones por protozoarios intestinales se asociaron en mayor medida con factores estructurales de tipo socioeconómico, destacándose la pobreza multidimensional, la cual condiciona de forma integral aspectos clave como el acceso al agua potable. La edad de los escolares y el índice de necesidades básicas insatisfechas (índice de carencias críticas, aquí codificada como UBN) fueron indicadores clave en la predicción de estas infecciones, mostrando patrones de comportamiento similares, aunque no idénticos, a los observados en los modelos de helmintos transmitidos por el suelo (STH). No obstante, es probable que la complejidad de las infecciones por protozoarios no haya sido captada completamente por las variables incluidas en los modelos utilizados.

Un aporte significativo de este estudio es la generación de modelos predictivos que pueden ser aplicados al estudio de otros fenómenos epidemiológicos. El marco analítico empleado —que incluyó secuencialmente técnicas de *Random Forest*, regresión *Elastic Net* y modelos lineales mixtos generalizados— permitió identificar con solidez los principales determinantes de las parasitosis intestinales en escolares de Costa Rica.

En el caso de las STH, los modelos destacaron como principales predictores la variabilidad climática (DRBS), la dispersión de construcciones (AAD), y la presencia de suelos permisivos a parásitos (PSR). Asimismo, la proporción de población indígena emergió como una variable relevante, posiblemente relacionada con altos niveles de necesidades básicas insatisfechas y bajos índices de desarrollo social en estas comunidades.

Para las infecciones por protozoarios intestinales (IP), los factores dominantes incluyeron el indicador de dos o más necesidades insatisfechas (carencias críticas o UBN), el índice de vecinos cercanos (NNI) —que sugiere que una configuración espacial de viviendas más dispersas puede actuar como factor protector—, y las diferencias térmicas anuales (TMXDS), las cuales, junto con la pluviosidad, se han relacionado previamente con un mayor riesgo de transmisión, como se ha reportado en estudios de otras regiones. Aunque no se observaron diferencias significativas por sexo, la edad fue un factor determinante: los niños entre 7 y 9 años presentaron las tasas de infección más elevadas, lo cual concuerda con hallazgos previos a nivel internacional.

En conclusión, este estudio enfatiza la necesidad de considerar de manera integrada los determinantes microgeográficos, climáticos y sociales, dado que las infecciones parasitarias responden a una compleja interacción entre estos factores. Aunque el poder explicativo de los modelos fue mayor para las STH (R^2 condicional de hasta 0,50) que para las IP (R^2 condicional de 0,17), ambos modelos demostraron una adecuada calibración. Esto los convierte en herramientas útiles para diseñar estrategias de intervención basadas en el riesgo.

Este constituye el aporte final de este trabajo: la provisión de dos modelos predictivos que pueden ser utilizados para estimar, en distintas zonas potencialmente vulnerables del país, la probabilidad de infección por helmintos y protozoarios intestinales, lo que representa un insumo relevante para la vigilancia epidemiológica y la planificación de intervenciones en salud pública.

2.1 Recomendaciones y direcciones futuras

Los resultados aquí presentados destacan la necesidad de fortalecer una vigilancia más activa de las parasitosis intestinales, que supere el modelo actual, el cual frecuentemente se limita a conjeturas clínicas sin respaldo diagnóstico. Esta limitación impide la generación de

datos de alta calidad que permitan una comprensión más precisa de la epidemiología de estas infecciones en Costa Rica. Por ello, los sistemas de salud deben incorporar estrategias más proactivas que permitan identificar con mayor eficacia los posibles focos de transmisión, especialmente en comunidades vulnerables, indígenas y rurales.

Considerando el potencial de los modelos generados —que, aunque perfectibles, ya ofrecen información relevante— es posible diseñar intervenciones diferenciadas con un enfoque contextual más adecuado. Esto podría traducirse en una mayor eficacia e impacto de las acciones, así como en una distribución más eficiente de los recursos, tanto humanos como logísticos. Por ejemplo, los programas de desparasitación masiva podrían redefinirse para centrarse en zonas donde son prioritarios, y reorientarse en aquellas donde la evidencia indica una menor necesidad. Asimismo, la promoción de la salud podría optimizarse mediante mensajes adaptados a las características locales.

Dado que las infecciones por helmintos transmitidos por el suelo (STH) y los protozoarios intestinales (IP) responden a determinantes diferentes, las estrategias de intervención deben adaptarse al perfil de riesgo de cada tipo de infección. Así, mientras que las medidas de saneamiento ambiental y control de condiciones físicas del entorno podrían tener mayor impacto en la reducción de STH, las estrategias educativas y el mejoramiento del acceso a servicios básicos serían fundamentales para disminuir la prevalencia de IP.

Resulta conveniente incluir variables microgeográficas y microclimáticas en el diseño de políticas públicas y en la elaboración de indicadores más sensibles para la vigilancia epidemiológica. Factores como la rugosidad del terreno, la dispersión de las viviendas y el acceso a rutas de transporte deben ser considerados en la planificación territorial de las intervenciones en salud.

Los efectos aleatorios identificados en los modelos indican que existen factores locales que interactúan con los efectos fijos, lo cual refuerza la necesidad de desarrollar modelos predictivos a escala local. La incorporación de nuevos datos en el futuro permitirá continuar con procesos de validación cruzada y actualización de los modelos, lo cual favorecerá su precisión y aplicabilidad. Estos avances serán posibles gracias al uso de técnicas de aprendizaje automático, que permiten manejar un mayor volumen y complejidad de variables.

Además, se sugiere integrar nuevas dimensiones relacionadas con factores culturales y conductuales, los cuales no fueron abordados en el presente estudio debido a limitaciones en el diseño inicial, pero cuya incorporación futura, mediante metodologías de investigación adecuadas, sin duda aportará información crítica para una comprensión integral del fenómeno.

Finalmente, perseverar en el estudio de los determinantes de enfermedades en comunidades altamente vulnerables de nuestro país —como en el caso de las parasitosis intestinales— representa un acto de reconocimiento hacia estas poblaciones. Esto implica asumir un compromiso ético con su bienestar y con la superación de las condiciones estructurales de marginación y pobreza que históricamente han enfrentado.

Estos esfuerzos científicos contribuyen a visibilizar su realidad, a diseñar soluciones específicas y, sobre todo, a no seguir perpetuando un ciclo de exclusión que afecta a niños que, en el futuro, serán adultos con derecho a mejores condiciones de vida.

Supplementary material A

Table SA-1. Overview of sampled schools' data.

School	District	Month ¹	Year ²	Record analyzed			Total of children (TSY)	Sampling fraction ³	Records excluded ⁴
				Preschool records/ Total Preschool children	Primary records/ Total primary children	Total analyzed (TRA) n			
Veracruz	Caño Negro	August	2013	9/19	24/67	33	86	38.4	0
Grano de Oro	Chirripó	March	2008	7/15	44/85	51	100	51.0	2
Villa Damaris	Chirripó	March	2008	7/12	40/81	47	93	50.5	0
Kabebata	Chirripó	March	2008	4/14	38/42	42	56	75.0	1
Ñuka Kichá	Chirripó	March	2008	0/0	32/40	32	40	80.0	0
Grano de Oro	Chirripó	March	2009	7/18	80/108	87	126	69.0	0
Jokbata	Chirripó	March	2009	0/10	8/33	8	43	18.6	0
Kabebata	Chirripó	March	2009	3/7	27/38	30	45	66.7	1
Kokotsakubata	Chirripó	March	2009	1/3	36/40	37	43	86.0	0
Sikua Ditso	Chirripó	March	2009	0/0	10/35	10	35	28.6	0
Villa Damaris	Chirripó	March	2009	3/9	43/75	46	84	54.8	1
Jokbata	Chirripó	March	2010	0/9	16/34	16	43	37.2	0
Kabebata	Chirripó	March	2010	1/6	18/40	19	46	41.3	0
Kokotsakubata	Chirripó	March	2010	0/4	18/24	18	28	64.3	0
Tsipirí	Chirripó	March	2010	3/9	16/22	19	31	61.3	0
Villa Damaris	Chirripó	March	2010	0/10	15/78	15	88	17.0	0
Jokbata	Chirripó	March	2011	2/2	19/31	21	33	63.6	0
Kabebata	Chirripó	March	2011	2/10	15/39	17	49	34.7	0
Kokotsakubata	Chirripó	March	2011	0/0	20/24	20	24	83.3	2
Ñuka Kichá	Chirripó	March	2011	2/8	19/29	21	37	56.8	3
Kabebata	Chirripó	March	2015	11/11	56/61	67	72	77.8	5
Villa Damaris	Chirripó	March	2015	4/13	79/88	83	101	82.2	7
Jakui	Chirripó	March	2015	0/22	14/33	14	55	25.5	11
Yoldi-Kichá	Chirripó	March	2015	0/0	10/12	10	12	83.3	0
Jokbata	Chirripó	March	2016	2/10	15/29	17	39	43.6	0
Kabebata	Chirripó	March	2016	3/7	25/53	28	60	46.7	0
Kokotsakubata	Chirripó	March	2016	0/7	14/30	14	37	37.8	0
Ñuka Kichá	Chirripó	March	2016	5/11	21/21	26	32	81.3	2
Paso Marcos	Chirripó	March	2016	3/8	17/35	20	43	46.5	1

¹Month= month of sampling; ²Year=Year of the sampling; ³Sampling fraction=(Total records analyzed/ Total schoolchildren for the school in that year) *100 (TRA/TSY*100); ⁴Records excluded=records analyzed that don't meet the inclusion criteria of this study.

Abbreviations: TRA=Total records analyzed; TSY=Total schoolchildren for the school in that year

Table SA-1. Continued. Overview of sampled schools' data.

School	District	Month ¹	Year ²	Record analyzed			Total of children (TSY)	Sampling fraction ³	Records excluded ⁴
				Preschool records/ Total Preschool children	Primary records/ Total primary children	Total analyzed (TRA) n			
Sikua Ditso	Chirripó	March	2016	3/5	23/28	26	33	78.8	3
Sharabata	Chirripó	March	2016	5/5	17/19	22	24	91.7	0
Tsipirí	Chirripó	March	2016	1/6	12/24	13	30	43.3	0
Tsipirí-Ñak	Chirripó	March	2016	0/0	5/9	5	9	55.6	0
Villa Damaris	Chirripó	March	2016	3/13	37/85	40	98	40.8	4
Yoldi-Kichá	Chirripó	March	2016	0/0	14/15	14	15	93.3	0
Jekui	Chirripó	March	2017	4/16	16/36	20	52	38.5	0
Kabebata	Chirripó	March	2017	0/9	16/46	16	55	29.1	2
Kokotsakubata	Chirripó	March	2017	2/7	9/24	11	31	35.5	4
Ñuka Kichá	Chirripó	March	2017	6/10	20/34	26	44	59.1	0
Sharabata	Chirripó	March	2017	4/7	10/22	14	29	48.3	0
Sikua Ditso	Chirripó	March	2017	0/11	6/26	6	37	16.2	0
Tsipirí	Chirripó	March	2017	1/9	6/25	7	34	20.6	0
Jokbata	Chirripó	April	2018	0/7	11/33	11	40	27.5	4
Kabebata	Chirripó	April	2018	2/9	15/47	17	56	30.4	3
Kokotsakubata	Chirripó	April	2018	1/6	12/22	13	28	46.4	2
Sikua Ditso	Chirripó	April	2018	2/9	19/29	21	38	55.3	5
Ñuka Kichá	Chirripó	April	2018	7/14	30/38	37	52	71.2	0
Tsipirí	Chirripó	April	2018	3/5	18/25	21	30	70.0	2
Tsipirí-Ñak	Chirripó	April	2018	0/0	5/8	5	8	62.5	0
Jekui	Chirripó	February	2019	8/15	22/42	30	57	52.6	3
Ñoribata	Chirripó	February	2019	2/8	7/11	9	19	47.4	1
Paso Marcos	Chirripó	February	2019	10/10	16/35	26	45	57.8	4
Sikua Ditso	Chirripó	February	2019	1/10	8/28	9	38	23.7	0
Sharabata	Chirripó	February	2019	6/6	15/22	21	28	75.0	2
Yoldi-Kichá	Chirripó	February	2019	1/3	6/10	7	13	53.8	0
Ostional	Cuajiniquil	March	2018	10/26	54/79	64	105	61.0	5
La Cruz	Delicias	March	2017	4/15	67/64	71	79	89.9	3
Las Delicias	Delicias	March	2017	10/24	50/64	60	88	68.2	0
San Ramón	Delicias	March	2017	6/11	22/38	28	49	57.1	0
Puente Casa	Florencia	October	2019	4/32	26/85	30	117	25.6	0

¹Month= month of sampling; ²Year=Year of the sampling; ³Sampling fraction= (Total records analyzed/ Total schoolchildren for the school in that year) *100 (TRA/TSY*100); ⁴Records excluded=records analyzed that don't meet the inclusion criteria of this study.

Abbreviations: TRA=Total records analyzed; TSY=Total schoolchildren for the school in that year

Table SA-1. Continued. Overview of sampled schools' data.

School	District	Month ¹	Year ²	Record analyzed			Total of children (TSY)	Sampling fraction ³	Records excluded ⁴
				Preschool records/ Total Preschool children	Primary records/ Total primary children	Total analyzed (TRA) n			
Las Brisas	Garita	March	2018	6/11	32/36	38	47	80.9	5
San Fernando	Garita	March	2018	2/6	16/14	18	20	90	9
La Libertad	Garita	March	2018	5/11	20/18	25	29	86.2	6
San Dimas	La Cruz	March	2018	8/19	44/87	52	106	49.1	8
La Casona	Limoncito	October	2012	13/38	56/137	69	175	39.4	5
Mrusara	Limoncito	October	2012	5/5	23/26	28	31	90.3	4
Qiabdo	Limoncito	October	2012	13/20	26/110	39	130	30	1
Villa Palacios	Limoncito	October	2012	0/10	5/31	5	41	12.2	1
La Casona	Limoncito	August	2013	13/30	50/128	63	158	39.9	5
Qiabdo	Limoncito	August	2013	0/22	14/105	14	127	11.0	5
Villa Palacios	Limoncito	August	2013	1/8	16/26	17	34	50.0	0
Jonkruhora	Limoncito	March	2014	8/8	30/52	38	60	63.3	0
La Casona	Limoncito	March	2014	11/38	60/137	71	175	40.6	0
Qiabdo	Limoncito	March	2014	5/20	46/110	51	130	39.2	0
Qiabdo	Limoncito	April	2016	2/20	16/109	18	129	14.0	0
Brus Malis	Limoncito	April	2016	10/21	11/61	21	82	25.6	0
Jonkruhora	Limoncito	April	2016	4/7	11/52	15	59	25.4	0
La Casona	Limoncito	April	2016	4/24	38/131	42	155	27.1	1
Madaribotda	Limoncito	April	2016	1/5	17/45	18	50	36.0	0
Brus Malis	Limoncito	April	2017	1/10	6/59	7	69	10.1	0
Jonkruhora	Limoncito	April	2016	4/7	11/52	15	59	25.4	0
La Casona	Limoncito	April	2016	4/24	38/131	42	155	27.1	1
Madaribotda	Limoncito	April	2016	1/5	17/45	18	50	36.0	0
Brus Malis	Limoncito	April	2017	1/10	6/59	7	69	10.1	0
Jonkruhora	Limoncito	April	2017	0/11	3/45	3	56	5.4	0
La Casona	Limoncito	April	2017	6/27	28/109	34	136	25.0	0
Villa Palacios	Limoncito	April	2017	1/10	13/22	14	32	43.8	0
La Casona	Limoncito	April	2015	12/26	69/109	81	135	60.0	2
Brus Malis	Limoncito	April	2015	8/17	18/59	26	76	34.2	0
Copey	Limoncito	April	2015	0/14	18/47	18	61	29.5	0
Jonkruhora	Limoncito	April	2015	2/7	18/45	20	52	38.5	0

¹Month= month of the sampling; ²Year=Year of the sampling; ³Sampling fraction= (Total records analyzed/ Total schoolchildren for the school in that year) *100 (TRA/TSY*100); ⁴Records excluded=records analyzed that don't meet the inclusion criteria of this study.

Abbreviations: TRA=Total records analyzed; TSY=Total schoolchildren for the school in that year

Table SA-1. Continued. Overview of sampled schools' data.

School	District	Month ¹	Year ²	Record analyzed			Total of children (TSY)	Sampling fraction ³	Records excluded ⁴
				Preschool records/ Total Preschool children	Primary records/ Total primary children	Total analyzed (TRA) n			
Villa Palacios	Limoncito	April	2015	1/10	12/22	13	32	40.6	0
Arco Iris	Los Chiles	April	2010	0/0	16/24	16	24	66.7	0
Cachito	Los Chiles	April	2010	0/0	10/10	10	10	100.0	0
Caño Castilla	Los Chiles	April	2010	0/9	16/34	16	43	37.2	0
El Combate	Los Chiles	April	2010	0/0	11/20	11	20	55.0	0
El Recreo	Los Chiles	April	2010	0/0	17/17	17	17	100.0	0
Escaleras	Los Chiles	April	2010	0/0	10/10	10	10	100.0	0
Hernández	Los Chiles	April	2010	2/24	16/98	18	122	14.8	0
Isla Chica	Los Chiles	April	2010	4/16	27/60	31	76	40.8	0
Los Ángeles	Los Chiles	April	2010	3/18	48/77	51	95	53.7	0
Las Nubes	Los Chiles	April	2010	0/15	39/64	39	79	49.4	0
La Trocha	Los Chiles	April	2010	3/25	30/116	33	141	23.4	0
Medio Queso	Los Chiles	April	2010	4/11	27/38	31	49	63.3	0
Primavera	Los Chiles	April	2010	0/0	6/14	6	14	42.9	0
Punta Cortés	Los Chiles	April	2010	0/8	26/26	26	34	76.5	0
Pueblo Nuevo	Los Chiles	April	2010	0/0	8/8	8	8	100.0	0
San Francisco	Los Chiles	April	2010	0/9	20/76	20	85	23.5	0
San Gerardo	Los Chiles	April	2010	2/28	34/81	36	109	33.0	0
San Ramón	Los Chiles	April	2010	0/5	22/25	22	30	73.3	0
Nidiribotda	Pavón	March	2014	8/8	24/36	32	44	72.7	0
Líder Comte	Pavón	March	2014	0/19	20/105	20	124	16.1	3
Altamira	Pavón	March	2014	2/5	15/18	17	23	73.9	1
Bahía de Pavón	Pavón	March	2014	9/18	53/69	62	87	71.3	0
El Jardín	Pavón	March	2014	0/0	11/12	11	12	91.7	0
La Escuadra	Pavón	March	2014	0/10	20/31	20	41	48.8	0
La Estrella	Pavón	March	2014	3/16	33/43	36	59	61.0	0
Kogoribta	Pavón	March	2014	8/0	15/38	23	38	60.5	2
Las Gemelas	Pavón	March	2014	0/16	27/43	27	59	45.8	0

¹Month= month of sampling; ²Year=Year of the sampling; ³Sampling fraction=(Total records analyzed/ Total schoolchildren for the school in that year)*100 (TRA/TSY*100); ⁴Records excluded=records analyzed that don't meet the inclusion criteria of this study.

Abbreviations: TRA=Total records analyzed; TSY=Total schoolchildren for the school in that year

Table SA-1. Continued. Overview of sampled schools' data.

School	District	Month ¹	Year ²	Record analyzed			Total of children (TSY)	Sampling fraction ³	Records excluded ⁴
				Preschool records/ Total Preschool children	Primary records/ Total primary children	Total analyzed (TRA) n			
Punta Banco	Pavón	March	2014	0/6	8/10	8	16	50.0	0
Punta Zancudo	Pavón	March	2014	1/6	17/24	18	30	60.0	0
Santa Clara	Pavón	March	2014	0/0	12/16	12	16	75.0	0
Vista del Mar	Pavón	March	2014	0/0	11/13	11	13	84.6	0
Chaparrón	Pital	October	2019	0/0	5/6	5	6	83.3	0
Río San Carlos Sector Este	Pital	October	2019	8/9	29/42	37	51	72.5	0
Río Azul	Río Azul	March	2013	11/27	26/71	37	98	37.8	0
Limón 2000	Río Blanco	March	2012	4/115	34/499	38	614	6.2	0
San Isidro Labrador	San Isidro, Heredia	February	2012	7/11	48/59	55	70	78.6	3
La Celia	Sixaola	March	2019	0/32	80/131	80	163	49.1	6
Arenal (Swakbli)	Telire	April	2019	5/5	19/23	24	28	85.7	0
Bajo Coén	Telire	April	2019	10/11	22/26	32	37	86.5	3
Coroma	Telire	April	2019	8/18	27/47	35	65	53.8	0
Sibödi	Telire	April	2019	3/30	23/62	26	92	28.3	5
Wawet	Telire	April	2019	4/7	8/18	12	25	48.0	0
Colonia Puntarenas	Upala	August	2013	0/20	14/78	14	98	14.3	0
El Fósforo	Upala	August	2013	9/19	24/74	33	93	35.5	0
La Palmera	Upala	August	2013	2/2	4/6	6	8	75.0	0
La Carpio	Uruca	April	2013	18/304	78/1225	96	1529	6.3	8
San Clemente	Valle La Estrella	March	2012	4/39	11/112	15	151	9.9	5
San Andrés	Valle La Estrella	March	2012	0/16	11/68	11	84	13.1	0
San Gabriel	Yolillal	February	2013	4/10	34/38	38	48	79.2	4
Las Garzas	Yolillal	February	2013	3/6	4/14	7	20	35.0	0
San Antonio	Yolillal	February	2017	16/20	39/78	55	98	56.1	0
San Jorge	Yolillal	February	2013	10/21	35/85	45	106	42.5	0
Campo Verde	Yolillal	February	2013	3/6	20/24	23	30	76.7	14

¹Month= month of the sampling; ²Year=Year of the sampling; ³Sampling fraction= (Total records analyzed/ Total schoolchildren for the school in that year) *100 (TRA/TSY*100); ⁴Records excluded=records analyzed that don't meet the inclusion criteria of this study. Abbreviations: TRA=Total records analyzed; TSY=Total schoolchildren for the school in that year

Supplementary material B

Table SB-1. Socioeconomic variables explained with source references

Variable code	Variable name	Annotations about data collection for that variable	Reference
SDI	Social Development Index	The official Social Development Index, 2013	INEC
SDI_1 to SDI_5	Quintile of Social Development Index	Based on the official data of the last published Social Development Index for Costa Rica.	INEC
UE	Unemployment	Percentage of the population in working age, unemployed	INEC
SAL	Fixed salary	Percentage of the population that has a fixed salary	
NRE	Non-remunerated employment	Percentage of the population who work but do not receive an official remuneration	INEC
ESI	Employed in services or industry	Percentage of the population employed in services or industry jobs	INEC
EAG	Employed in agriculture	Percentage of the population employed in agriculture or the primary sector	INEC
HI	Health Insurance	Percentage of the population with health insurance	INEC
DHI	Direct health insurance	Percentage of the population with direct health insurance	INEC
IHI	Indirect health insurance	Percentage of population with indirect health insurance coverage	INEC
UBN	Unsatisfied basic needs indicator	Percentage of population with at least two dimensions of basic needs not satisfied	INEC
INP	Indigenous population	Percentage of the population belonging to an indigenous ethnicity	INEC
SEC	Score of socioeconomic variables	Composite score for socioeconomic variables (see supplementary materials C)	Supplementary materials C

Abbreviations: INEC= National Institute of Statistics and Census of Costa Rica. The meaning of each abbreviation in the “variable code column” is explained in the “variant name”; that is why it is not described as a footnote.

Table SB-2. Educational variables explained and source references.

Variable code	Variable name	Annotations about data collection for that variable	Reference
AA	Analphabetism	Percentage of population with illiteracy in the district	INEC
MAA	Men analphabetism	Percentage of men in illiteracy in the district	INEC
WAA	Woman Analphabetism	Percentage of women with illiteracy in the district	INEC
REA	Regular education attendance (%)	Percentage of the population with regular attendance in general education	INEC
GBE	General basic education completed	Percentage of the population in school age who have completed the general basic education programs	INEC
HSC	High school levels completed	Percentage of the population that has completed high school	INEC
DEL	Delayed educational levels	Percentage of the population who are delayed in the completion of one or more basic educational levels	INEC
AUE	Any university education	% of the population which have received a university education	INEC
UEC	University education completed	Percentage of the population with a university title	INEC
INTA	Internet access	Percentage of population with access to the internet	INEC
ASY	Average school years	Number of years in the educational system, on average, for the district	INEC
SED	Score of educational variables	composite score of educational variables (see supplementary materials C)	Supplementary materials C

Abbreviations: INEC= National Institute of Statistics and Census of Costa Rica. The meaning of each abbreviation in the “variable code column” is explained in the “variant name”; that is why it is not described as a footnote.

Table SB-3. House conditions variables explained and source references.

Variable code	Variable name	Annotations about data collection for that variable	Reference
PD	General population density	General density of population for the district according to INEC's 2011 Census	INEC
ANPH	Average number of people per house	An indicator of crowded houses: Average number of people living within the same house for the district. INEC's 2011 Census.	INEC
BWL	Fecal waste materials final disposition through a black well or a sewage-connected latrine	Percentage of houses whose fecal waste materials are finally disposed of in a black well or any final disposition different from a sewage-connected latrine. From INEC's Census 2011	INEC
HBC	Houses in bad condition	Percentage of houses in bad condition in the district: with no regard to classification (as normal house or tugurio). Tugurio is a house that is very poorly constructed. From INEC's 2011 Census	INEC
PWA	Piped water access	Percentage of houses connected to piped water for both water supply and wastewater, regardless of drinkable quality. From INEC's 2011 Census	INEC
OAD	Open Air Defecation	% of houses in the district where people use open-air defecation: this is an indirect datum obtained from the data of people with no septic tank, no latrines, and no access to sanitary sewage systems. From INEC's 2011 Census	INEC
TUG	Tugurios	Percentage of tugurios. From INEC's 2011 Census	INEC
PDW	In-House piped drinkable water access	Percentage of houses with in-house piped drinkable water access, from the INEC's 2011 Census for the district	INEC
PREC	Precarios	Percentage of houses within "precarious". Precario is an impoverished neighborhood with houses in inferior condition. From INEC's 2011 Census	INEC
SSW	Homes connected to sanitary sewage system (%)	Percentage of homes connected to a sanitary sewage system from the INEC's 2011 Census for the district	INEC
SHC	Score of housing conditions	Composite score of housing conditions (see supplementary materials C)	Supplementary materials C
ECT	House with electricity	Percentage of houses with electricity services. From the INEC's 2011 Census for the district	INEC
INT	Internet Access	Percentage of houses with internet access. From the INEC's 2011 Census for the district	INEC

Abbreviations: INEC= National Institute of Statistics and Censuses of Costa Rica. The meaning of each abbreviation in the "variable code column" is explained in the "variant name"; that is why it is not described as a footnote.

Table SB-4. Climatic Variables explained and source references.

Variable code	Variable name	Annotations about data collection for that variable	Reference
DRY	Days with rain in a year	Sum of days in which it rained according to the nearest-to-sampling-point available climatic station.	IMN
TRY	Total rainfall in a year (cm)	Measured by the remote sensing climatic station (the nearest one to the district and school sampled, and in the year of sampling), total rainfall in cm.	IMN
DRBS	Total difference of rainfall between the rainy and the dry seasons	The arithmetic difference between the 'cm' of rainfall in the rainy season and the 'cm' of rainfall in the dry season, according to data collected by the nearest-to-sampling-point remote sensing climatic station	IMN
TRRS	Total rainfall in the rainy season(cm)	The total amount of rainfall in the rainy season, as recorded by the nearest-to-sampling-point remote sensing climatic station.	IMN
TRDS	Total rainfall in the dry season (cm)	The total measure of rainfall in the dry season according to the nearest-to-sampling-point remote sensing climatic station.	IMN
ATY	Average temperature in a year in °C	The average temperatures in C° for the year of sampling, measured in the nearest-to-sampling-point remote sensing climatic station	IMN
MATY	Average maximum temperature in a year in °C	Average of maximum temperatures in the year of sampling according to the nearest-to-sampling-point remote sensing climatic station	IMN
MITY	Average minimum temperature in a year in °C	Average of the minimum temperature in °C, in the year of sampling, according to the nearest-to-sampling point remote sensing climatic station	IMN
TMXD	Difference between maximum temperature and the average temperature	The arithmetic difference between the average value of maximum temperatures and the average mean temperatures in a year for the district. Measured by the remote sensing climatic station (the nearest to the sampling point) for the year of sampling.	Calculated
RH	Relative humidity	Average of relative humidity according to the nearest-to-sampling-point remote sensing climatic station.	IMN
RAD	Radiation in MJ/m2 (average in a year)	Average of Radiation (in MJ/m2) sensed by the nearest-to-sampling-point remote sensing climatic station	IMN
SCLIM	Score of climatic conditions	Composite score of climatic conditions, for details on the composition of this score (see supplementary materials B)	Supplementary materials C

Abbreviations: IMN= National Meteorological Institute of Costa Rica. The meaning of each abbreviation in the “variable code column” is explained in the “variant name”; that is why it is not described as a footnote.

Table SB-5. Topographic and geographic variables explained, and source references.

Variable code	Variable name	Annotations about data collection for that variable	Reference
SPCD	Distance from school to primary healthcare facility (m)	Total distance in m measured from the sampling point (school) to the nearest primary care facility (EBAIS). The measure was done in Google Earth Pro by tracing the route from the school to the EBAIS and annotating the total distance in meters.	Calculated from GIS tools
SNHD	Distance in m from the school to the nearest hospital	The distance from each school to the nearest population center with a secondary or tertiary health facility was measured in meters along the shortest available road network using the 'Path' tool in Google Earth Pro."	Calculated from GIS tools
TRHD	Terrain rugosity (topographic irregularity) in the route from school to the nearest clinic or hospital	Rugosity of terrain in the route from school to hospital was obtained by tracing the path from the school to the nearest clinic or hospital in Google Earth Pro. This trace provides an output with the inclination profile for that distance. From that profile, we extracted the values of this arithmetic difference: maximum inclination - average inclination, and the value of the difference: average inclination - minimum inclination. The sum of both positive values is multiplied by a factor obtained from the standardized (max-min method) value of the maximum difference altitude, which is also derived from the profile. That final arithmetic product is the Rugosity of terrain from school to the nearest hospital.	Calculated from GIS tools
DRH	(SNHD+GNR)/2	Arithmetic mean between the max-min standardized value of RHO AND DHO	Calculated from GIS tools
SBD	Distance from the school to the nearest official bus route (m)	Total distance from each school to the nearest public bus route in m. Official bus routes shapefile was obtained from https://aresep.go.cr/datos-abiertos/rutas-autobuses/ public database. In QGIS, we placed this shapefile over the school's points shapefile. Then, by tracing the shortest route (using roads) from each school to the point where it meets the bus route, we measured that distance in meters.	Calculated from GIS tools
GNR	General terrain rugosity in the school's surrounding area	By using Google Earth Pro, a 16000 m ² area was traced with the center being the school (this was made for each of the schools), and this was called the surrounding area, an area where neighboring schools have the minimum overlap. To trace it, we first drew a circle with the school as its center and a radius of 2 km. Then, the square was traced with the circle inscribed inside the square. The circle was eliminated. Having traced the school surrounding area, we then traced transects from one extreme to the other of the square, and going down continuously, first making horizontal transects and later making vertical transects, each separated by 500 m. After completing the grid of all squares, the elevation profile was obtained to calculate the rugosity (general rugosity) using the same equation as RHO.	Calculated from GIS tools
ATI	Average of terrain inclination	Obtained from the elevation profile of the route from school to the nearest hospital, using Google Earth Pro and tracing that route.	Calculated from GIS tools
RC_Yes, RC_No	River-crossings (no bridges) present or absent	A thorough analysis of each route between each school and the nearest hospital or clinic revealed the presence or absence of river crossings. If that condition was present, the route was categorized as RC_Yes, otherwise, it was categorized as RC_No.	Calculated from GIS tools

Table SB-5. *Continued.* Topographic and geographic variables explained, and source references.

Variable code	Variable name	Annotations about data collection for that variable	Reference
PSR	Permissive-to-parasites surroundings	Within the previously defined surrounding area (see GNR for methodology), we subdivided the square into 16 sub-squares of 1000x1000 m ² and marked the ones with houses. If all had houses, we randomly selected 10 sub-squares. Each of those sub-squares was photographed, and the photographs were uploaded to ImageJ, where we used the tool to convert images to 8 bits. Then we used the tool to adjust the image's gray scale threshold, allowing us to paint the zones where no houses, roads, or forest cover were present. Then, we measured that area with the measure tool and summed those areas to obtain the total area that was soil-dominated terrain with no concrete or dense forest cover, which is a theoretically permissive soil to maintain parasite transmission forms. That was called the permissive surroundings variable PSR.	Calculated from GIS tools
FLO_A, FLO_B or FLO_C	Classification of flooding risk within the school's surroundings definition	Classification of flooding risk within the school's surroundings. This classification was extrapolated from the 2014 Digital Atlas of Costa Rica. In QGIS, locate the 2014 Flood-risk shapefile from the Costa Rica Digital Atlas over the school's points shapefile. Schools within the highest risk zones were categorized as FLO_A, schools within the intermediate risk zones FLO_B, and schools in the least risk zones as FLO_C.	Calculated from GIS tools
PAV	Paved route from school to nearest hospital or clinic (%)	Using the Google Earth Pro route tracing tool, we traced the portion of each route from school to the nearest paved hospital or clinic and calculated the total distance to determine the percentage of the route that is paved.	Calculated from GIS tools
DAI	Difficulty (to access) index	This composite score combines these variables to index the difficulty of access between school and the nearest population center's health services, providing an indicator of access difficulty (see Supplementary materials C).	Supplementary materials C

Table SB-6. Land configuration variables explained and source references.

Variable code	Variable name	Annotations about data collection for that variable	Reference
COV1	Not urban nor forest land cover in the surrounding area %	Percentage of the surrounding area lands covered by non-urban or forest coverage. Using Google Earth Pro, a 16000 m ² area was traced, with the school as the center (this was done for each school), and this area was designated as the surrounding area. The square satellite image was cropped and converted into an 8-bit image in ImageJ. The threshold of the Image was adjusted, and colored the area corresponding to non-urban and non-forest cover was colored. Then the area was calculated, and the percentage of the total area.	Calculated from GIS tools
NNI	Nearest neighbor analysis	In Google Earth Pro, each human construction within the surrounding area (previously defined) was pinpointed for each surrounding area corresponding to each school, and a .kml file was saved and loaded in QGIS 3.4 and converted to a points shapefile. The points shapefile was used to conduct a nearest neighbor analysis in QGIS for each of the schools, pinpointing all human constructions within a 16,000 km ² surrounding area. This variable corresponds to the value of the Nearest neighbor index.	Calculated from GIS tools
FCOV	Forest cover in the surrounding area %	A similar methodology was used for the variable COV1, with the exception that for FCOV, the colored area and calculated area represented forest coverage after adjusting the threshold of the 8-bit image to color the forest coverage. The FCOV was calculated as the percentage of the total of 16000 m ² surrounding area.	Calculated from GIS tools
ADNN	Average distance to the nearest neighbor (m)	ADNN is the value of the parameter of the same name calculated by QGIS in the nearest neighbor analysis.	Calculated from GIS tools
UCOV	Urban land cover in the surrounding area	Following the same methodology as for COV1 and FCOV variables but adjusting the threshold of the 8-bit converted satellite image to color the urban constructions, we calculated the urban area and the percentage it represented from the total area.	Calculated from GIS tools
AAD	Average of all distances between constructions in the surrounding area (m)	Following the same methodology as for the NNI variable to elaborate a points shapefile of all human constructions in the surrounding area, we used the QGIS tool to measure all distances between all points to obtain the average of all distances between constructions in the surrounding area as another way to measure the level of urbanization and human concentration.	Calculated from GIS tools
UDEN	Urban density in the surrounding area per km ²	This is another way to measure urban density, providing a more precise measure than the Population density for the whole district. To measure it, we sum up the total of human constructions in the surrounding area, divide it by 0.16, and multiply it by 100 to obtain the value of # of constructions per 100 Km ² in the surrounding area.	Calculated from GIS tools
DAAD	Standard deviation of the average of all distances between constructions in the surrounding area (m)	This standard deviation indicates the dispersion of distances between human constructions in the surrounding area, serving as an indicator of their aggregation. The lower the value of DAAD, the more the aggregation.	Calculated from GIS tools
SLC	score of the land configuration variables	A composite score using all of the land configuration variables. To see details on the construction of the score, see Supplementary Materials C.	Supplementary materials C

Supplementary material C

SCORES OF VARIABLES

“*” means multiplication.

Socioeconomic variables score (SEC)

It is a composite score of socioeconomic variables. Gives more weight to IDS, having a fixed salary and health insurance, and negative weight to having indirect insurance, working in the primary sector of the economy, or being unemployed.

$$\text{SEC} = 3 * \text{SDI} - \text{UE} - \text{EAG} + \text{ESI} + 2 * \text{SAL} - \text{NRE} + 2 * \text{HI} + \text{DHI} - \text{IHI}$$

Educational variables score (SED)

SED is an educational composite score calculated as follows: it penalizes higher levels of illiteracy and low educational attainment, while giving positive weight to completed schooling and higher education levels. The formula is:

$$\text{SED} = \text{INTA} - 4 * \text{AA} - \text{MAA} - 2 * \text{WAA} + 4 * \text{ASY} + 3 * \text{REA} + \text{GBE} + 2 * \text{SECS} + 2 * \text{AUE} + 3 * \text{UEC} - \text{DEL}$$

Housing conditions score (SHC)

SHC is a composite score. It is calculated giving positive weight to structural good conditions of the house, access to clean drinking water and piped water inside the house, and negative weight to conditions of lacking adequate sewage systems, open air defecation, or use of black wells and not sanitary piping systems for feces final disposition. This is the formula:

$$\text{SHC} = \text{ECT} + 5 * \text{PWA} + 3 * \text{PDW} - \text{BWL} - 4 * \text{OAD} - \text{ANPH} - \text{HBC} - \text{TUG} - \text{PREC}$$

Climate conditions score (SCLIM)

SCLIM is a composite score. It is calculated by giving positive weight to conditions of average temperatures, more rain, more humidity, and less UV-radiation, conditions that theoretically favor parasite transmission forms in the soil.

$$\text{SCLIM} = \text{TRY} + \text{TRDS} + \text{DRY} + \text{TMXD} + \text{RH} + \text{MATY} - \text{MITY} - \text{RAD}$$

Land configuration score (SLC)

SLC is a composite score. It is calculated giving positive weight to variables that favor more urbanized and less dispersed human constructions, more forested areas, and negative weight to bare soil (surroundings that are more permissive to parasite transmission forms).

$$\text{SLC} = \text{NNI} + 2 * \text{ADNN} + \text{AAD} + 2 * \text{DAAD} - \text{PD} - \text{UDEN} + \text{COV1} + \text{FCOV} + \text{UCOV} - 2 * \text{PRS}$$

Access difficulty index (DAI)

DAI is a composite score. It is calculated giving positive weight to terrain characteristics that make more difficult the access to the places with health and other important services (“population centers or downtowns”), for example it has high positive weight for more distance, inclination and terrain irregularity (rugosity) between each school and the town centers where health facilities and other services are available. Also gives positive weight to larger distances to the nearest official bus route, also to a higher probability of flooding in the school surroundings, for the presence of rivers with no bridges that must be crossed to go from school surroundings to the nearest population center, and negative weight to the percentage of the main route to nearest population center that is paved.

$$\text{DAI} = 5 * (\text{SNHD} * \text{ATI}) + (\text{SPCD} * \text{ATI}) + 4 * (\text{TRHD} + \text{GNR}) + 3 * \text{FLO_C} + \text{SBD} + \text{R_Yes} - \text{PAV}$$

Supplementary material D

Table D1. Variables with an extremely high correlation were manually excluded at the beginning of the analysis. Note: Their exclusion does not imply a lack of importance, but rather their high correlation with other important variables, selected based on their biological implications in STH and IP infections, and their superior apparent power of explanation of the explained variance (obtained in the PCA analysis of vector loadings).

Variable Code	Name of the variable
RC_No	Absence of a river for which there is no bridge to cross in the route from school to the nearest population center's central health facility, negatively correlated to RC_Yes, the opposite variable.
PWA	Piped water access. Highly correlated to open-air defecation OAD, OAD was conserved.
BWL	Use of a black well or open air, not sewage-connected latrine (highly correlated with open air defecation and many other variables, so the others were maintained).
AA	Analphabetism in the population. Very highly correlated to illiteracy in women, in men, with the Socioeconomic score, piped in-house drinkable water indicator, and others.
WAA	Analphabetism in women has a similar profile of high correlation with other variables.
NRE	Non-remunerated work in the population. Very highly correlated to the percentage of university education and title.
SHC	Very highly correlated to several more specific variables that are biologically important.
SCLIM	Very highly correlated to more specific climatic variables, those were preferred.
SED	Very highly correlated to more specific educational and socioeconomic variables, those were preferred.
SLC	Very highly correlated to more specific land configuration variables, those were preferred.
DEL	Education level completion delay. It is highly correlated with other educational and household conditions, important variables that were controlled.
HBC	Percentage of houses in bad condition. Significantly correlated to other, more specific house condition variables that are more relevant biologically for the outcomes.

The meaning of each abbreviation in the "code" column is explained in the "name of the variable"; that is why it is not described here as a footnote.