Atovaquone-Proguanil and Reduced Digestive Cancer Risk: A *Toxoplasma gondii* Connection

Ariel Israel MD, PhD[1,2]*, Sarah Israel MD[3], Abraham Weizman MD[4], Shai Ashkenazi MD, MSc[5,6], Shlomo Vinker MD, MHA[1], Eli Magen MD[7], Eugene Merzon MD, MPH[1,5]

Authors Affiliations:

[1] Leumit Research Institute, Leumit Health Services, 23 Sprinzak St., Tel Aviv, Israel

[2] Department of Epidemiology and Preventive Medicine, School of Public Health, Faculty of Medical & Health Sciences, Tel Aviv University; Ramat Aviv, Tel Aviv, Israel

[3] Department of Clinical Microbiology and Infectious Diseases, Hadassah-Hebrew University Medical Center, Faculty of Medicine, Hebrew University of Jerusalem, P.O. Box 12000, Jerusalem, Israel

[4] Research Unit, Geha Mental Health Center, Felsenstein Medical Research Center, and Sagol School of Neurosciences, Faculty of Medical & Health Sciences, Tel Aviv University, Tel Aviv, Israel

[5] Adelson School of Medicine, Ariel University, Ariel, Israel

[6] Unit of Infectious Diseases, Schneider Children's Medical Center, 14 Kaplan St., Petah Tikva, Israel

[7] Medicine A Department, Assuta Ashdod University Medical Center, Ben Gurion University of the Negev, Derech HaRofe 7, Ashdod, Israel

Corresponding Author:

Ariel Israel, MD, PhD

Director of Leumit Research Institute, Leumit Health Services, Tel-Aviv, Israel

Department of Epidemiology and Preventive Medicine, School of Public Health, Faculty of Medical & Health Sciences, Tel Aviv University, Tel-Aviv, Israel

Email: dr.ariel.israel@gmail.com

Word count: 2609

Abstract

Background: Cancers of the digestive tract, notably colorectal cancer (CRC) and pancreatic malignancies, represent a major global health burden, with microbial risk factors remaining poorly understood. Atovaquone-proguanil (A-P), an antimalarial agent, has been associated with reduced CRC incidence. We hypothesized that A-P may lower the risk of digestive cancers by eliminating *Toxoplasma gondii*, a protozoan capable of colonizing the digestive mucosa.

Methods: Using the TriNetX US Collaborative Network, we conducted propensity score-matched cohort analyses of patients aged 40-49 (mean age, 45.2 years), 50-59 (mean, 54.8), and 60-69 (mean, 64.6) who had received A-P versus any other medication. We assessed incident digestive cancers and calculated hazard ratios (HRs). Independently, we reanalyzed metagenomic data from 1,044 fecal microbiota samples (PRJEB6070; 156 individuals), quantifying *T. gondii* sequence reads among 8,425 microbial species and evaluating associations with CRC using the Mann-Whitney U test.

Results: A-P use was associated with a 49-54% lower risk of digestive cancers, with HRs of 0.51, 0.46, and 0.46 across the three age groups, respectively ($P<0.0001$). Protective associations extended to pancreatic malignancies (HRs, 0.55, 0.71, and 0.69; $P<0.05$). The reduced risk persisted more than a decade after treatment. In metagenomic analyses, *T. gondii* most strongly differentiated CRC from non-cancer samples ($P=1.8\times10^{-16}$); its presence (≥50 reads) was associated with an odds ratio of 18.2 for CRC ($P=2.3\times10^{-22}$).

Conclusions: *T. gondii* may represent a major, underrecognized microbial risk factor for digestive cancers. The long-term protective association with A-P use may reflect eradication of this parasite, suggesting a novel target for cancer prevention.

Abbreviations:

A-P- Atovaquone-Proguanil

BMI - Body mass Index

CI - Confidence Interval

CRC - colorectal cancer

DM - Diabetes Mellitus

FOBT - Fecal occult blood test

HR - Hazards ratio

SMD: Standardized Mean Difference

Introduction

Digestive cancers - including malignancies of the esophagus, stomach, small and large intestines, liver, gallbladder, and pancreas (ICD-10 codes C15-C26) - are among the leading causes of cancer-related death globally, accounting for more than 3 million deaths annually[1]. Colorectal cancer (CRC) alone causes over 900,000 deaths each year, making it the second leading cause of cancer mortality worldwide[1]. While genetic predisposition, chronic inflammation, and environmental exposures are recognized contributors, a growing body of evidence highlights the role of the microbiome in digestive tract carcinogenesis[2]. Bacterial pathogens such as *Helicobacter pylori* and *Fusobacterium nucleatum* have been implicated in gastric and colorectal cancers, respectively[3,4]. However, most microbiome-cancer studies to date have focused almost exclusively on bacteria, often overlooking protozoa-organisms that are undetectable by 16S rRNA gene sequencing and frequently underrepresented or absent in the reference databases used for shotgun metagenomic sequences recognition[5,6].

Toxoplasma gondii, a widespread protozoan parasite capable of chronically infecting the gastrointestinal mucosa[7], has been linked primarily to neurological and ocular diseases in immunocompromised or congenitally infected individuals[8]. However, its potential involvement in gastrointestinal malignancies has not been systematically studied.

Atovaquone-proguanil (A-P), a well-tolerated antimalarial combination used primarily for short-term chemoprophylaxis, has recently been associated with reduced cancer incidence in observational studies[9]. One proposed mechanism involves the inhibition of mitochondrial respiration and the reduction of oxidative stress in host tissues, but this does not fully explain the

apparent long-term protective effect observed years after treatment. Notably, A-P is also active against *T. gondii* [10] through inhibition of its mitochondrial electron transport chain[11], raising the possibility that elimination of this pathogen mediates some of its anticancer effect.

Here, we present a dual analytic approach to test this hypothesis. First, using the TriNetX US Collaborative Network[12], we conducted propensity score-matched cohort studies across three age groups to assess the long-term association between A-P exposure and digestive cancer incidence. Second, we reanalyzed publicly available metagenomic sequencing data to examine the presence of *T. gondii* in colorectal cancer-associated microbiota. Our findings suggest that *T. gondii* may be an underrecognized risk factor for digestive cancers and that A-P may offer a novel preventive strategy via targeted antiparasitic action.

Methods

Study design and Data Sources

We conducted a two-part investigation to evaluate the association between atovaquone-proguanil (A-P) use, *T. gondii* presence, and digestive cancer risk. First, we performed retrospective cohort analyses using the TriNetX US Collaborative Network[12], a federated health research platform aggregating deidentified electronic health record data from 70 healthcare organizations, encompassing over 120 million patients as of March 2025. Second, we reanalyzed publicly available metagenomic shotgun sequencing data (PRJEB6070) to assess *T. gondii* prevalence in colorectal cancer (CRC)-associated microbiota. PRJEB6070 comprises fecal samples from 156 individuals in France who underwent colonoscopy to diagnose colorectal neoplasia (adenomas or CRC) or confirm its absence. The dataset includes clinical annotations and microbial sequences suitable for examining microbiome-cancer associations.

TriNetX Cohort Analysis

We created three propensity score-matched (PSM) cohorts of patients receiving A-P versus any other medication, stratified by age at index date (defined as the date of first medication record): 40-49, 50-59, and 60-69 years. Inclusion criteria included at least one recorded medication purchase and no prior diagnosis of the outcome. A-P exposure was defined by the existence of a medication purchase record for both atovaquone and proguanil compounds, reflecting typical short-term use for malaria prophylaxis. Control participants received any non-A-P medication. PSM was performed 1:1 using logistic regression, matching on age, sex, race, smoking status, body mass index, and diabetes status. Follow-up extended from the index date to digestive cancer diagnosis, death, or last clinical record. Hazard ratios (HRs) were estimated using Cox proportional hazards models. Kaplan-Meier curves were used for survival analysis, with proportionality tested via Schoenfeld residuals.

Metagenomic Reanalysis

We analyzed metagenomic shotgun sequencing data from 1,044 fecal samples derived from 156 individuals (PRJEB6070), accessed through the European Nucleotide Archive. Raw reads were processed using Centrifuge[13] (v1.0.4) with the SANBAFPH database[14], which includes reference sequences for protozoa, bacteria, archaea, fungi, nematodes, and human genomes. Reads were classified by taxa according to best match, and in particular, we identified *T. gondii* sequences only if when the read uniquely matched this microorganism with ≥95% identity. Sample-level *T. gondii* presence was defined as ≥50 mapped reads; person-level presence was calculated as the mean across all samples from the same individual.

Statistical analyses

In the TriNetX cohorts, we compared the incidence of digestive cancers (ICD-10 codes C15-C26) between the A-P and control groups, including a predefined subgroup analysis for pancreatic cancer (C25). For the metagenomic analysis, we first applied the Mann-Whitney U test to evaluate the ability of each microbial taxon to distinguish CRC from non-CRC samples. *T. gondii* ranked first among 8,425 species in the SANBAFPH database by discriminatory power. Next, we defined *T. gondii* presence as ≥50 uniquely mapped reads and assessed its association with CRC status using contingency tables and Fisher's exact test to calculate odds ratios (ORs) and p-values. Diagnostic performance was evaluated using receiver operating characteristic (ROC) curves, calculating area under the curve (AUC) sensitivity and specificity across varying read-count thresholds. Analyses were conducted in Python (v3.11.9) and R (v4.4). Two-tailed p-values <0.05 were considered statistically significant.

Ethics statement

The study protocol was approved by the statutory Leumit Health Services Institutional Review Board (approval number LEU-0010-21). Informed consent was waived because this large-scale retrospective study was performed on deidentified electronic health records.

Results

TriNetX Cohort Analysis

In the propensity score-matched TriNetX cohorts, atovaquone-proguanil (A-P) use was associated with a significant reduction in the incidence of digestive cancers (ICD-10 codes C15-C26) across all age groups (Table 1, Figure 1). In the 40-49 age group (mean age, 45.2 years; n = 31,142 per group), A-P users had a 49% lower risk of digestive cancer compared to controls (hazard ratio [HR], 0.51; P < 0.0001). In the 50-59 age group (mean age, 54.8 years; n = 37,954

per group), A-P was associated with a 54% reduction in risk (HR, 0.46; P < 0.0001). Similarly, in the 60-69 age group (mean age, 64.6 years; n = 39,343 per group), the risk reduction was 54% (HR, 0.46; P < 0.0001). Kaplan-Meier curves demonstrated progressively widening separation over time for over a decade, consistent with a durable protective effect mediated by the elimination of a causative pathogen (Figure 1A, 1C, 1E).

Pancreatic cancer (ICD-10 C25) was also analyzed as a specific outcome. In the 40-49 cohort (n = 30,909 per group), A-P use was associated with a 45% reduction in risk (HR, 0.55; P = 0.0315). In the 50-59 group (n = 35,851 per group), the risk was reduced by 29% (HR, 0.71; P = 0.0315), and in the 60-69 group (n = 39,786 per group), by 31% (HR, 0.69; P = 0.001). Kaplan-Meier analyses for pancreatic cancer showed less pronounced separation, but supported a consistent protective association across age groups (Figure 1B, 1D, 1F).

Metagenomic Analysis

We reanalyzed 1,044 fecal microbiota samples from 156 individuals in the PRJEB6070 dataset, categorized by diagnosis: 53 patients with colorectal cancer (330 samples), 27 with small adenomas (171 samples), 15 with large adenomas (86 samples), and 61 with normal findings (457 samples). In the original study[15], microbial DNA was extracted and sequenced using shotgun metagenomics; however, protozoan sequences were apparently not included in the sequence recognition database, limiting detection of non-bacterial organisms. Patient characteristics by diagnostic category are summarized in Table 2. Cancer patients were slightly older (mean age, 67.3 years) compared to those with normal findings (mean age, 62.2 years). The body mass index did not differ significantly across groups (P = 0.53).

To assess *T. gondii* and other microorganisms, we used Centrifuge[13] (v1.0.4) with the SANBAFPH database[14], which includes unique sequences for Sar (Stramenopiles, Alveolates, and Rhizaria), Nematodes, Bacteria, Archaea, Fungi, Protozoa, and Human reference and representative genomes. Among 8,425 microbial species identified from the SANBAFPH database[14], *T. gondii* showed the strongest discriminatory power between cancer and non-cancer samples (Mann-Whitney U test: U = 101,183; $P = 1.8\times10^{-1}$), ranking above the leading bacterial association with *Fusobacterium* (U = 97,964; $P = 5.6\times10^{-1}$).

On a sample-wise level, *T. gondii* (≥50 uniquely mapped reads) was detected in 177 of 1,044 samples - 22.1% of cancer samples versus 1.53% of normal samples (Fisher's exact test; odds ratio [OR] = 18.20; $P = 2.3\times10^{-22}$). At the individual level, *T. gondii* presence was observed in 16 of 156 patients - 22.5% of cancer patients and 1.64% of those with normal findings (OR = 17.2; $P < 0.001$). Cancer patients also had significantly higher *T. gondii* read counts (Table 2; Figure 2A). A stacked bar plot of read count ranges further demonstrated a dose-response relationship (Figure 2B): 15.1% of cancer patients had 50-999 reads compared to 1.6% of controls, and 7.6% had >1,000 reads compared to none in controls. The relative abundance of *T. gondii* (defined as the ratio of *T. gondii* reads to total reads per sample) was also elevated in cancer samples.

Screening Potential

We evaluated *T. gondii* read counts as a potential biomarker for colorectal cancer (CRC) screening. A receiver operating characteristic (ROC) curve using individual-wise *T. gondii* read counts to distinguish cancer from normal samples yielded an area under the curve (AUC) of 0.720, a discriminatory power comparable to that of fecal occult blood testing (FOBT) (Figure 3). Sensitivity and specificity at selected read count thresholds were as follows: 90% and 38% at

1.3 reads; 80% and 43% at 2.3 reads; 70% and 49% at 3.8 reads; 50% and 85% at 13.5 reads; and 30% and 98% at 44.0 reads. These findings, derived from a noisy and nonspecific set of reads with potential for both missed and mismatched sequences, suggest that detection of *T. gondii* DNA using optimized primers or immunohistochemical tests targeting *T. gondii* antigens may hold promise as a non-invasive marker for CRC risk, particularly at higher thresholds where specificity is high.

Discussion

Our study provides compelling evidence that *T. gondii* may be an unrecognized risk factor for digestive cancers, with atovaquone-proguanil (A-P) offering a protective effect potentially through *T. gondii* depletion. Using the TriNetX database, we observed a consistent 49-54% reduction in digestive cancer incidence (ICD-10 C15-C26) across three age cohorts (40-49, 50-59, 60-69) among A-P users compared to controls, with hazard ratios of 0.51, 0.46, and 0.46, respectively. Kaplan-Meier curves demonstrated increasing separation over more than a decade post-treatment, supporting a long-term effect consistent with pathogen elimination rather than a transient metabolic change. Notably, A-P also reduced pancreatic malignancy risk by 29-45% across the same cohorts (HRs, 0.55-0.71), suggesting a broader protective effect across digestive organs.

Our metagenomic reanalysis of 1,044 fecal microbiota samples from 156 patients confirmed a strong association between *T. gondii* and colorectal cancer (CRC), with an odds ratio of 18.2 ($P = 2.3\times10^{-22}$) at the sample level and 17.2 ($P < 0.001$) at the individual level. *T. gondii* emerged as the top discriminator among 8,425 microorganism species, with cancer patients exhibiting significantly higher read counts compared to normal, small adenoma, and large adenoma groups.

The dose-response relationship between *T. gondii* read counts and CRC risk further strengthens this association. Cancer patients were more likely to have higher *T. gondii* reads, with odds increasing with higher read thresholds (OR, 6.16 for 25-49 reads; 10.48 for 50-999 reads; infinite for >1,000 reads). Interestingly, patients with large adenomas also showed elevated *T. gondii* reads, suggesting a potential role for *T. gondii* in precancerous lesions. The gradient - from normal to adenoma to cancer - hints at a progressive role for *T. gondii* in colorectal carcinogenesis.

T. gondii may promote carcinogenesis through multiple mechanisms. Chronic infection in the colonic mucosa can induce persistent inflammation, characterized by proinflammatory cytokines (e.g., IL-6, TNF-α) that activate NF-κB and STAT3 pathways, promoting epithelial cell survival and proliferation and reducing innate immune response[16]. The immune response to *T. gondii* generates reactive oxygen and nitrogen species (ROS/RNS), causing oxidative DNA damage and mutations in oncogenes (e.g., *KRAS*) or tumor suppressor genes (e.g., *TP53*)[17,18]. *T. gondii* infection has been associated with increased risk of various cancers, including colorectal cancer[19], potentially due to these inflammatory mechanisms. Additionally, *T. gondii* manipulates host cells by inhibiting apoptosis-upregulating antiapoptotic proteins (e.g., Bcl-2) and suppressing p53 activity-allowing mutated cells to survive and accumulate further genetic alterations, potentially leading to cancerous clones[20]. Interestingly, while our results support an oncogenic role for *T. gondii* in digestive cancers, some studies have reported potential anti-tumor effects of the parasite[21]. Experimental models suggest that *T. gondii* infection can modulate host immune responses and tumor-associated gene expression in ways that may inhibit certain tumor types[22]. These divergent findings underscore the complexity of host-parasite interactions and highlight the need for mechanistic studies across tissue types and contexts.

The broader reduction in digestive cancers, including pancreatic malignancies, in the TriNetX cohorts suggests that *T. gondii*'s role may extend beyond CRC. Pancreatic cancer, with a 5-year survival rate of less than 10%, is notoriously difficult to prevent, making A-P's 29-45% risk reduction particularly noteworthy. This aligns with emerging evidence linking chronic infections to pancreatic carcinogenesis, as seen with *Helicobacter pylori* and hepatitis viruses.[23]

The wide effect of A-P in preventive digestive malignancies could reflect *T. gondii*'s ability to colonize multiple gastrointestinal sites, potentially triggering localized inflammation or systemic immune responses that predispose to malignancy. The increasing separation of Kaplan-Meier curves over time aligns with a model where A-P clears a persistent pathogen like *T. gondii*, preventing long-term oncogenic processes. Atovaquone, a component of A-P, is a recognized treatment for toxoplasmosis, with established efficacy against *T. gondii* in vivo commonly combined with pyrimethamine or used alone in immunocompromised patients to treat active toxoplasmosis, targeting *T. gondii*'s mitochondrial cytochrome b.[24] The consistency of A-P's protective effect across age groups underscores its potential as a preventive agent, particularly given its established safety profile.

Our exploration of *T. gondii* read counts as a screening tool offers practical implications. Despite the challenges of using noisy and nonspecific shotgun metagenomic sequencing data, a threshold of 13.5 reads achieved 50% sensitivity and 85% specificity for CRC detection, supporting the potential for a microbiome-based diagnostic approach, with an area under the receiver operating characteristic curve of 0.720. More targeted methods, such as PCR-based detection with *T. gondii*-specific primers or antigen detection assays for *T. gondii* proteins, potentially combined with complementary biomarkers, such as FOBT, to enhance diagnostic accuracy, could provide a more cost-effective and precise screening strategy.

Limitations of our study include the observational nature of the TriNetX data, which cannot formally establish causality. The metagenomic dataset, while robust, is relatively small (156 patients), and the fecal microbiota context may not fully reflect gut dynamics. Additionally, TriNetX data lack granularity on A-P dosing duration and adherence, which could influence effect size. The higher age and male predominance in cancer patients could introduce confounding, though the strong *T. gondii* association persists across analyses. We also did not assess cancer-specific hazard ratios for all digestive malignancies, though the pancreatic cancer data suggest a broader effect.

Future research should focus on prospective studies to confirm A-P's effect on cancer incidence, alongside mechanistic studies to elucidate *T. gondii*'s oncogenic pathways, particularly its role in chronic inflammation, antiapoptotic manipulation, and precancerous lesions like adenomas, as well as its impact on other digestive cancers like pancreatic malignancies. Clarifying *T. gondii*'s tissue tropism, immune evasion strategies, and impact on precancerous lesions will also be critical for defining its role in carcinogenesis and therapeutic targeting. Larger metagenomic analyses in diverse populations are warranted. Randomized controlled trials of A-P in *T. gondii*-positive individuals could establish causality, along with optimal dosing and schedule for cancer prevention.

If validated, these findings could redefine microbial risk factors in oncology, positioning A-P as a novel preventive strategy and *T. gondii* as a target for screening and intervention.

References:

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71(3):209-49.

2. Garrett WS. Cancer and the microbiota. Science (1979) 2015;348(6230):80-6.

3. Polk DB, Peek RM. Helicobacter pylori: gastric cancer and beyond. Nat Rev Cancer 2010;10(6):403-14.

4. Kostic AD, Gevers D, Pedamallu CS, et al. Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. Genome Res 2012;22(2):292-8.

5. Mthethwa-Hlongwa NP, Amoah ID, Gomez A, et al. Profiling pathogenic protozoan and their functional pathways in wastewater using 18S rRNA and shotgun metagenomics. Science of The Total Environment 2024;912:169602.

6. Lokmer A, Cian A, Froment A, et al. Use of shotgun metagenomics for the identification of protozoa in the gut microbiota of healthy individuals from worldwide populations with various industrialization levels. PLoS One 2019;14(2):e0211139.

7. Innes EA. Toxoplasmosis: Comparative species susceptibility and host immune response. Comp Immunol Microbiol Infect Dis 1997;20(2):131-8.

8. Montoya J, Liesenfeld O. Toxoplasmosis. The Lancet 2004;363(9425):1965-76.

9. Zhang N, Sundquist J, Sundquist K, Ji J. Proguanil and atovaquone use is associated with lower colorectal cancer risk: a nationwide cohort study. BMC Med 2022;20(1):439.

10. KOVACS J. Efficacy of atovaquone in treatment of toxoplasmosis in patients with AIDS. The Lancet 1992;340(8820):637-8.

11. Doggett JS, Nilsen A, Forquer I, et al. Endochin-like quinolones are highly efficacious against acute and latent experimental toxoplasmosis. Proceedings of the National Academy of Sciences 2012;109(39):15936-41.

12. Palchuk MB, London JW, Perez-Rey D, et al. A global federated real-world data and analytics platform for research. JAMIA Open 2023;6(2).

13. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res 2016;26(12):1721-9.

14. Poursalavati A, Fall ML. Centrifuge SANBAFPH database for SOVAP.

15. Zeller G, Tap J, Voigt AY, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol Syst Biol 2014;10(11).

16. Hunter CA, Sibley LD. Modulation of innate immunity by Toxoplasma gondii virulence effectors. Nat Rev Microbiol 2012;10(11):766-78.

17. Zhuang H, Yao C, Zhao X, et al. DNA double-strand breaks in the Toxoplasma gondii-infected cells by the action of reactive oxygen species. Parasit Vectors 2020;13(1):490.

18. Smolarz B, Wilczyński J, Nowakowska D. DNA repair mechanisms and Toxoplasma gondii infection. Arch Microbiol 2014;196(1):1-8.

19. Yu Y, Guo D, Qu T, et al. Increased Risk of *Toxoplasma gondii* Infection in Patients with Colorectal Cancer in Eastern China: Seroprevalence, Risk Factors, and a Case-Control Study. Biomed Res Int 2020;2020(1).

20. Molestina RE, Payne TM, Coppens I, Sinai AP. Activation of NF-κB by *Toxoplasma gondii* correlates with increased expression of antiapoptotic genes and localization of phosphorylated IκB to the parasitophorous vacuole membrane. J Cell Sci 2003;116(21):4359-71.

21. Zhu S, Lu J, Lin Z, et al. Anti-Tumoral Effect and Action Mechanism of Exosomes Derived From Toxoplasma gondii-Infected Dendritic Cells in Mice Colorectal Cancer. Front Oncol 2022;12.

22. Song Y, Yuan H, Yang X, et al. The opposing effect of acute and chronic Toxoplasma gondii infection on tumor development. Parasit Vectors 2024;17(1):247.

23. Michaud DS. Role of bacterial infections in pancreatic cancer. Carcinogenesis 2013;34(10):2193-7.

24. Spencer CM, Goa KL. Atovaquone. Drugs 1995;50(1):176-96.

Table 1: Statistics of the TriNetX Cohorts

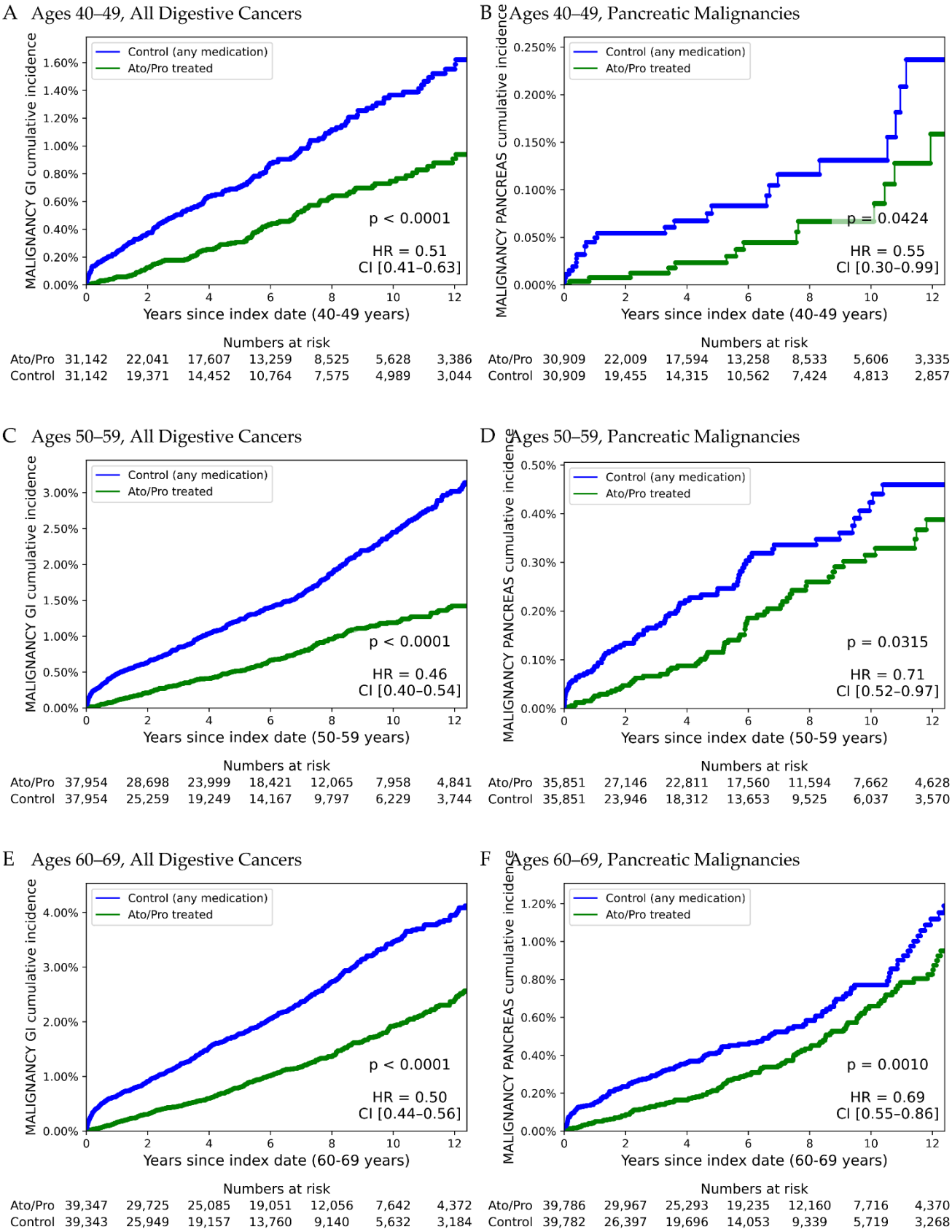
Cohort and Outcome	**40-49 Years**	**50-59 Years**	**60-69 Years**
Age at Index (mean)	45.3	54.8	64.6
Male (%)	44.8%	43.5%	43.3%
Digestive Cancers (ICD-10 C15-C26)			
Cohort Size (A-P / Control)	31,142 / 31,142	37,954 / 37,954	39,343 / 39,343
HR (A-P vs. Control)	0.51	0.46	0.46
95% CI	[0.41-0.63]	[0.40-0.54]	[0.44-0.56]
p-value	<0.0001	<0.0001	<0.0001
Pancreatic Malignancies (ICD-10 C25)			
Cohort Size (A-P / Control)	30,909 / 30,909	35,851 / 35,851	39,786 / 39,786
HR (A-P vs. Control)	0.55	0.71	0.69
95% CI	[0.30-0.99]	[0.52-0.97]	[0.55-0.86]
p-value	0.0315	0.0424	0.001

Table 2: Patient Characteristics and Metagenomic Data by Diagnosis

		Normal (N=61)	Small adenoma (N=27)	Large adenoma (N=15)	Cancer (N=53)	Odds Ratio (95% CI)	p-value
age		60.6 ± 11.4	60.3 ± 8.7	67.7 ± 8.1	66.8 ± 10.9		0.004
sex	**female**	33 (54.1 %)	9 (33.3 %)	3 (20.0 %)	24 (45.3 %)	0.70 [0.31 to 1.57]	0.453
	male	28 (45.9 %)	18 (66.7 %)	12 (80.0 %)	29 (54.7 %)	1.42 [0.64 to 3.19]	0.453
Body Mass Index (median [IQR])	**mean ± SD**	23.9 ± 5.4	25.3 ± 6.8	25.3 ± 2.9	24.6 ± 7.1		0.532
***T. gondii* reads**	**mean ± SD**	6.68 ± 10.39	7.55 ± 12.63	40.59 ± 111.31	523.11 ± 1724.05		0.021
	median [IQR]	3.88 [0.89;7.60]	3.20 [0.62;9.40]	3.00 [1.83;13.20]	13.00 [2.57;44.00]		<0.001
Total reads in sequence	**median [IQR]**	1,578,026 [1114342;1975874]	1,221,620 [858626;1961940]	1,390,726 [1129384;2472647]	1,785,875 [1132569;2234854]		0.221
***T. gondii* reads % in total reads**	**mean ± SD**	0.0044 ± 0.0074	0.0032 ± 0.0058	0.0186 ± 0.0437	0.1158 ± 0.4977		0.083
	median [IQR]	0.00101 [0.00034;0.00497]	0.00071 [0.00022;0.00246]	0.00126 [0.00048;0.01546]	0.00163 [0.00049;0.01606]		0.179
***T. gondii detection* (reads ≥ 50)**		1 (1.64 %)	1 (3.70 %)	2 (13.33 %)	12 (22.64 %)	17.21 [2.38 to 759.97]	<0.001
***T. gondii* reads range**	**0-24**	59 (96.72 %)	26 (96.30 %)	12 (80.00 %)	36 (67.92 %)	0.07 [0.01 to 0.34]	<0.001
	25-49	1 (1.64 %)	0 (0.00 %)	1 (6.67 %)	5 (9.43 %)	6.16 [0.66 to 299.96]	0.095
	50-999	1 (1.64 %)	1 (3.70 %)	2 (13.33 %)	8 (15.09 %)	10.48 [1.33 to 479.85]	0.012
	1000-	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)	4 (7.55 %)	Inf [0.78 to Inf]	0.044

Figures

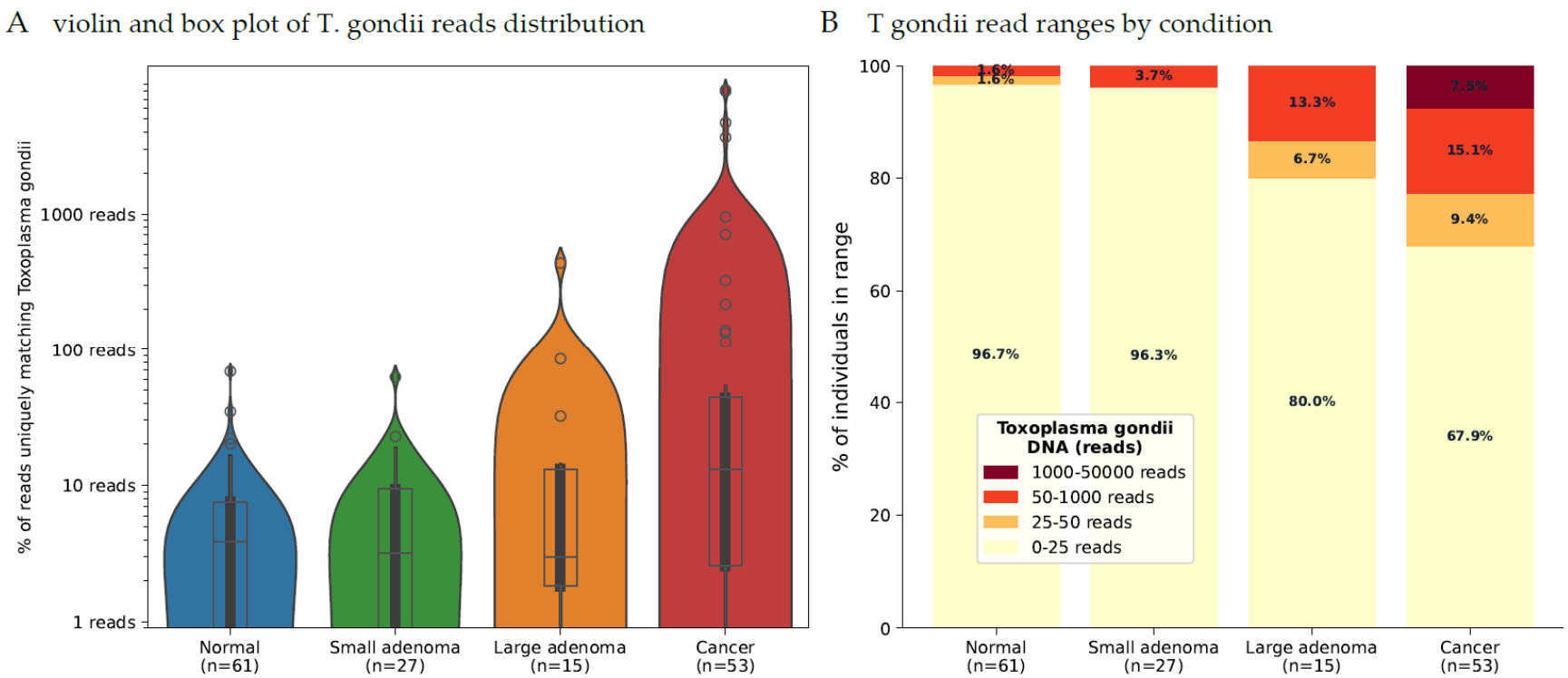
Figure 1: Cumulative Incidence of Digestive Cancers and Pancreatic Malignancies in Propensity Score-Matched TriNetX Cohorts



Kaplan-Meier curves show the cumulative incidence of digestive cancers (ICD-10 codes C15-C26) and pancreatic malignancies among users of atovaquone-proguanil (A-P, blue) versus propensity score-matched controls (green) across three age cohorts. Numbers at risk are presented at yearly intervals. P-values were derived from log-rank tests.

- **Panel A:** Digestive cancers in patients aged 40-49 years (n = 31,142 per group). A-P users showed a significantly lower incidence (hazard ratio [HR], 0.51; $P < 0.0001$).
- **Panel B:** Pancreatic malignancies in patients aged 40-49 years (n = 30,909 per group). A-P use was associated with reduced risk (HR, 0.55; $P = 0.0315$).
- **Panel C:** Digestive cancers in patients aged 50-59 years (n = 37,954 per group). A-P users exhibited a marked risk reduction (HR, 0.46; $P < 0.0001$).
- **Panel D:** Pancreatic malignancies in patients aged 50-59 years (n = 35,851 per group). A modest reduction was observed (HR, 0.71; $P = 0.0315$).
- **Panel E:** Digestive cancers in patients aged 60-69 years (n = 39,343 per group). A-P use was associated with significantly lower incidence (HR, 0.46; $P < 0.0001$).
- **Panel F:** Pancreatic malignancies in patients aged 60-69 years (n = 39,786 per group). A-P users had moderately reduced risk (HR, 0.69; $P = 0.001$).

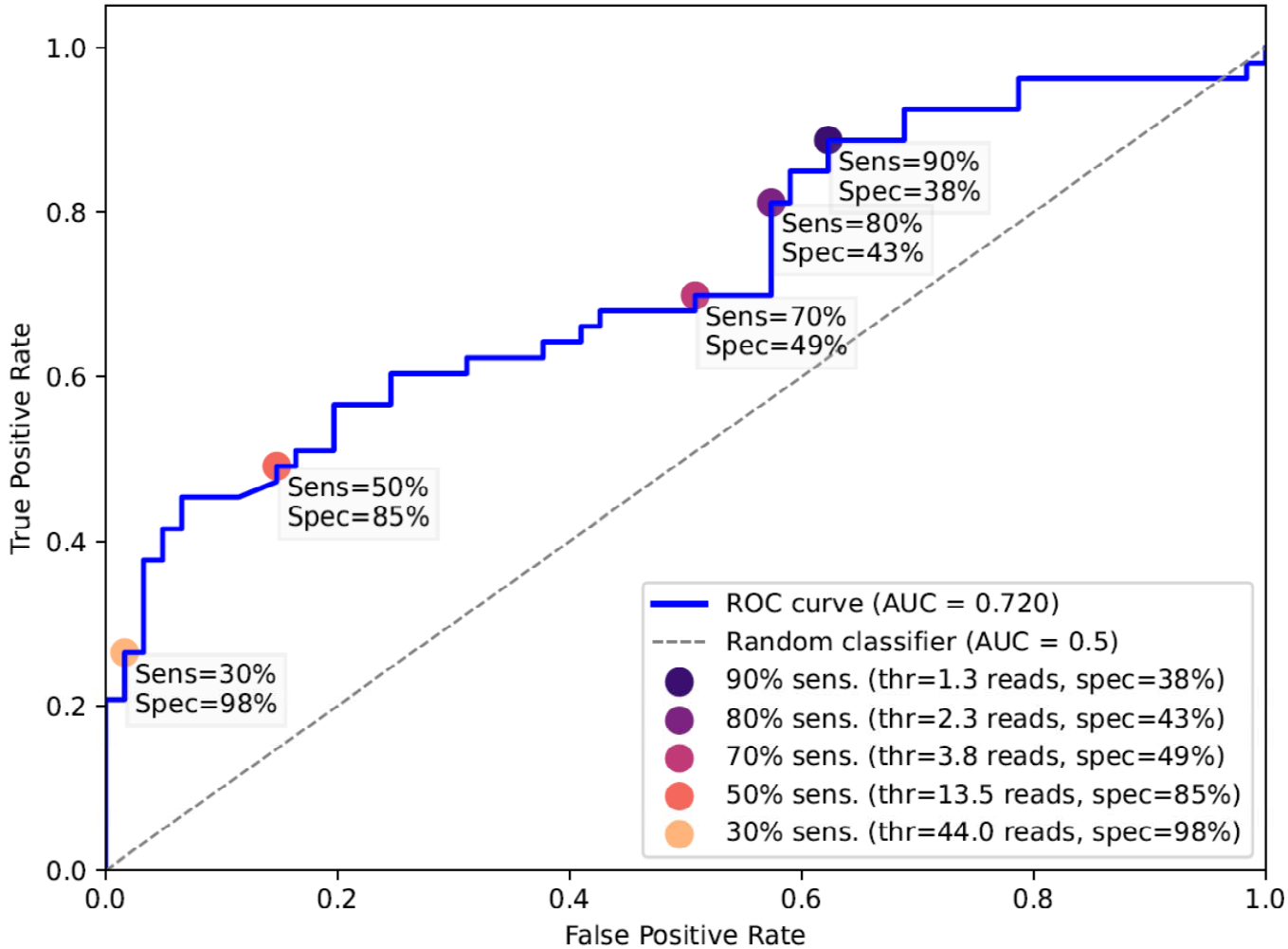
Figure 2: Distribution of *Toxoplasma gondii* Reads by Patient Condition



Panel A: Violin and box plots show the distribution of *T. gondii* reads (log scale, number of reads) in 1,044 fecal samples from 156 individuals in the PRJEB6070 dataset, grouped by diagnosis: normal (n = 61), small adenoma (n = 27), large adenoma (n = 15), and colorectal cancer (n = 53). Cancer patients exhibited significantly higher *T. gondii* read levels, with a progressive increase from normal to cancer ($P < 0.001$).

Panel B: Stacked bar plot illustrating the percentage of patients with mean *T. gondii* read counts in four ranges (0-25, 25-50, 50-1,000, and >1,000 reads) across diagnostic categories.

Figure 3: Receiver Operating Characteristic (ROC) Curve for *Toxoplasma gondii* Read Counts in Colorectal Cancer Classification.



ROC curve evaluating the ability of *T. gondii* read counts to distinguish colorectal cancer (CRC) from non-cancer cases in fecal metagenomic samples from 156 individuals in the PRJEB6070 dataset. The area under the curve (AUC) was 0.720, indicating promising discriminatory performance. Selected thresholds are marked on the curve:

- **1.3 reads:** 90% sensitivity, 38% specificity
- **2.3 reads:** 80% sensitivity, 43% specificity
- **3.8 reads:** 70% sensitivity, 49% specificity
- **13.5 reads:** 50% sensitivity, 85% specificity
- **44.0 reads:** 30% sensitivity, 98% specificity

These results suggest that *T. gondii* read abundance may serve as a potential biomarker for CRC screening.