Individuals with *PMS2* pathogenic variants are at average risk of cancer before age 60

Kelly M. Schiabor Barrett[1,#], Catherine Hajek[1], James Lu[1], Kate Lynch[1], Jeremy Cauwels[2], Douglas Stoller[3], Christopher N. Chapman[4], C. Anwar A. Chahal[5,6,7,8], Daniel P. Judge[9], Douglas A. Olson[10], Joseph J. Grzymski[11,12], William Lee[1], Elizabeth T. Cirulli[1], Alexandre Bolze[1,#]

1: Helix, San Mateo, California, USA
2: Sanford Health, Sioux Falls, South Dakota, USA
3: University of Nebraska Medical Center, Omaha, Nebraska, USA
4: Department of Pathology & Laboratory Medicine, St. Luke's University Health Network, Bethlehem, Pennsylvania, USA
5: Center for Inherited Cardiovascular Diseases, WellSpan Health, York, Pennsylvania, USA
6: Department of Cardiovascular Medicine, Mayo Clinic, Rochester, Minnesota, USA
7: Barts Heart Centre, West Smithfield, London, UK
8: William Harvey Research Institute, Queen Mary University of London, Charterhouse Square, London, UK
9: Division of Cardiology, Medical University of South Carolina, Charleston, South Carolina, USA
10: HealthPartners, Minneapolis, Minnesota, USA
11: University of Nevada, School of Medicine, Reno, Nevada, USA
12: Renown Health, Reno, Nevada, USA

#: correspondence to kelly.barrett@helix.com or alexandre.bolze@helix.com

Abstract

Importance: Screening Lynch syndrome in the general population, including healthy individuals, aims to detect and prevent cancers early. Current clinical recommendations for those with pathogenic variants are based on studies of patients with cancer or strong family history. It is essential to ensure guidelines are based on accurate assumptions regarding the impact of pathogenic variants in Lynch syndrome genes.

Objective: To determine the risk of cancer associated with pathogenic variants in *MLH1*, *MSH2*, *MSH6*, or *PMS2* in the general population.

Design, Setting, and Participants: This retrospective case-control study utilizes Helix Research Network™ data from 144,852 participants across seven US health systems sequenced between 2018 and 2024.

Main outcomes and measures: An automated pipeline based on the ACMG-AMP guidelines was developed for variant interpretations. Clinical diagnoses were identified from electronic health records for 11 cancer types associated with Lynch syndrome including colorectal and endometrial cancers.

Results: Individuals with pathogenic variants in *MLH1*, *MSH2*, and *MSH6* were at significantly increased risk for Lynch syndrome-associated cancers with Hazard Ratios (HR) of 16.5 (95% CI: 8.9-30.8) for *MLH1*, 17.3 (7.8-38.6) for *MSH2* and 4.7 (3.2-6.9) for *MSH6*. No significant risk was associated with *PMS2* pathogenic variants when considering all 11 cancers combined. *PMS2* pathogenic variants were only associated with colorectal cancer (HR of 4.3, 1.6-11.4); however, this risk was observed only after the age of 60, 10 years after clinical guidelines recommend starting colonoscopies for the average population. Up to age 60, 0.6% of individuals with *PMS2* pathogenic variants were diagnosed with colorectal cancer, similar to 0.4% in the general population but lower than *MLH1* (41.0%), *MSH2* (17.9%) or *MSH6* (4.5%) pathogenic variants.

Conclusion and relevance: The findings underscore the benefits of screening the entire population for *MLH1*, *MSH2* and *MSH6*. They also highlight significantly lower cancer risk for those harboring *PMS2* pathogenic variants. This study provides data to support tailored

surveillance and prevention strategies by gene and highlights the importance of deriving clinical recommendations from relevant populations.

Introduction

Lynch syndrome is associated with increased risk and earlier onset of colorectal cancer, endometrial cancer, as well as other types of cancers. The mismatch repair genes associated with Lynch syndrome include *MLH1*, *MSH2* , *MSH6*, and *PMS2*, each of which were discovered diagnostically through clinical sequencing of patients with aforementioned cancers [1]. *EPCAM* is often included in the list of Lynch syndrome genes despite not being a mismatch repair gene because *EPCAM* is physically next to *MSH2* and large deletions in *EPCAM* can extend into the *MSH2* promoter leading to *MSH2* silencing. In recent years there has been a push to identify unaffected individuals with pathogenic variants in Lynch syndrome genes. Identifying such individuals can enable options for prevention and early detection. For example, surveillance strategies recommended by the National Comprehensive Cancer Network® (NCCN®) or the American College of Gastroenterology (ACG) for those with a *MLH1* pathogenic variant are to start high-quality colonoscopy at age 20-25y and repeat every 1-2y [2,3], and to consider using daily aspirin[2] (**Supplementary Table 1**). The existence of specific, evidence-based recommendations are why Lynch syndrome is part of the CDC Tier 1 Genomic applications alongside Hereditary Breast and Ovarian Cancer and Familial Hypercholesterolemia [4,5]. As we identify more and more individuals harboring pathogenic variants in Lynch syndrome genes in the general population, there is a need to refine our understanding of the clinical impact of these variants in this screening context. Disease risk estimates derived from cohorts of patients already diagnosed with cancer or with a suspicion of cancer are biased and may not be adequate when returning genetic interpretations to a healthy individual in a population screening context [6,7]. The aim of our study is to measure the prevalence of cancer diagnoses and determine the risk associated with pathogenic variants in *MLH1*, *MSH2*, *MSH6* or *PMS2* in the general population and intersect this data with current surveillance recommendations.

Methods

Study design and participants

This is a retrospective and observational clinico-genomic analysis of a prospectively designed study. All participants were adults enrolled in the Helix Research Network™ (HRN) study, which is a protocol open to general patient populations at various U.S. healthcare

organizations. All participant data analyzed for this publication came from seven U.S. Health Systems participating in the HRN study. The studies included under the HRN protocol are: ImagineYou (Sanford Health), DNA Answers (St. Luke's University Health Network), the Genetic Insights Project (Nebraska Medicine), the Healthy Nevada Project (Renown Health), In Our DNA SC (Medical University of South Carolina), myGenetics (HealthPartners), and the Gene Health Project (WellSpan Health). Study protocols were reviewed and approved by their respective Institutional Review Boards (projects 956068-12 and 21143). All participants provided written informed consent prior to participation, and direct identifiers were removed from the research dataset to protect participant privacy. For this analysis, data from 144,852 participants with linked Electronic Health Records (EHR) and Exome+® sequencing data were included.
For one experiment, we also used the UK Biobank dataset. The UKB study was approved by the North West Multicenter Research Ethics Committee, UK.

Genetics and Variant interpretation

Genetic data

Saliva or blood samples were collected from participants and underwent Exome+® sequencing at Helix between February 2018 and June 2024. The Exome+® assay includes a clinical exome, which is used for return of clinical results for CDC Tier 1 genes, including Lynch syndrome genes, as previously described[5]. Variants in exons 11-15 of *PMS2* are more difficult to call due to the existence of a pseudogene. Helix developed a tailored clinically-validated bioinformatics pipeline to identify variants in these exons, which are then confirmed by an orthogonal assay. To maximize consistency and reproducibility, we opted to exclude exons 11-15 of *PMS2* from our analysis for the following reasons: (i) some older samples were not analyzed with this pipeline, and (ii) most published studies looking at *PMS2* do not include variants in these exons. At the time of our analysis, only 1 pathogenic large deletion in *EPCAM* going into the promoter of *MSH2* was identified and confirmed. This large deletion affected 1 participant, and we decided to not include it in the analysis given it was a N of 1. Genotype processing for Helix data was performed in Hail 0.2.115-10932c754edb.

Variant Interpretation

Variant interpretation for the four mismatch repair Lynch syndrome genes (*MLH1*, *MSH2*, *MSH6*, and *PMS2*) were completed for the entire HRN cohort (N=144,852) using a two-step approach. First, a variant was considered pathogenic if it carried a known and well-established

clinical pathogenic interpretation (i.e., no VUS or benign interpretations present in ClinVar across high volume laboratories, using search strings ['ClinGen', 'Quest', 'Sema4', 'Natera', 'Invitae', 'All of Us', 'Baylor', 'GeneDx', 'Ambry', 'LapCorp', 'Color', 'Myriad', 'Brigham'] and/or a likely pathogenic or pathogenic interpretation by the InSiGHT Hereditary Colorectal Cancer/Polyposis Expert Panel (InSiGHT VCEP). For all remaining variants, ACMG-AMP variant interpretations were completed programmatically following the gene-specific scoring recommendations from the InSiGHT VCEP. Data from case studies as well as patient-specific information such as presenting symptoms or family history were not considered for these interpretations. For validation, a subset of the samples (313 pathogenic and VUS interpretations) were compared to clinical interpretations from an independent clinical laboratory. High sensitivity (95.3%) and specificity (100%) was observed between the resulting variant interpretations from each method across these variants (**Supplementary Table 2** for gene-level results). All variants seen in HRN, relevant annotations, scoring by data category, and resulting interpretation based off of point totals (pathogenic[>5], higher scoring VUS[3-5], lower scoring VUS[-1 to 2], benign[< -1]) are available in **Supplementary Table 3.** The number of HRN participants harboring pathogenic variants in each gene is presented in **Supplementary Table 4**. Variant annotations were made based on the MANE transcript for each gene *(MLH1*: NM_000249.4, *MSH2*: NM_000251.3, *MSH6*: NM_000179.3, *PMS2*: NM_000535.7) and leveraged the following tools: VEP-104[8], GnomADv3[9], REVEL[10], SpliceAI[11], ClinVar database (accessed: 11/20/2024), and for *MSH2* functional scores from MAVE (urn:mavedb:00000050-a)[12], and case-control (PS4) data was obtained from systematic variant-level association tests internally-calculated using clinicogenomic data from UK Biobank and All of Us cohorts (phenotypes leveraged from phecodeX map include: CA_101.41, CA_106.21 for colorectal and endometrial cancers, respectively[13]).

Phenotypes

Electronic health records data were available for all participants included in the study. EHR data were transformed into the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) version 5.4, encompassing a mean of 12.9 years (median: 11.1 years, IQR: 11.3 years) of EHRs per patient. A total of 11 different types of cancer were analyzed for this study, based on prior literature on Lynch syndrome including a comprehensive report from the prospective Lynch syndrome database (see their Table 1 for the list of cancer types) [7]. The 11 cancer types included were: colorectal, endometrial and uterine, ovarian, kidney and ureter, small bowel, bladder, stomach, pancreas, biliary tract, brain, and prostate cancers.

All the OMOP condition concept ids used to extract data from the EHR are in **Supplementary Table 5**. A subset of 5 cancer types that are more frequently associated with each of the 4 core Lynch syndrome genes was used in several analyses in this paper; they are: colorectal, endometrial and uterine, ovarian, kidney and ureter, and small bowel cancer.

Clinical guidelines

We used the NCCN guidelines[2] as a basis to assess (i) whether current recommendations were appropriate given the risk of cancer observed, and (ii) count the number of colonoscopies that would be added or subtracted if guidelines were to be changed. We focused on colorectal and endometrial cancers, the two main cancers associated with Lynch syndrome. The recommendations for surveillance and prevention strategies for colorectal cancer are summarized in **Supplementary Table 1**. For endometrial cancer, the recommendations are less actionable and less likely to directly impact the care of a patient. The recommendations emphasize education surrounding symptoms and speak to possible prevention opportunities such as hysterectomies, biopsies, and ultrasounds that can be considered.

Statistical analysis

Kaplan Meier survival curves were done using the KaplanMeierFitter function from the Lifelines python library. The lifelines package was used for time to event analyses including cumulative incidence plots, log rank test, and cox proportional hazard calculations[14]. For time to event analyses, the earliest age at relevant diagnosis or current age (in 2024) was determined for each participant.

Results

0.31% of the population has a pathogenic variant in a Lynch syndrome gene

We studied Helix Research Network™ data from 144,852 participants across seven health systems. For each participant, genetic data from Exome+® sequencing as well as phenotype data from electronic health records were available (on average, EHR data had a 12.9

years lookback). The average age in 2024 of participants was 53.3 years old, and 39.3% (n=56,933) were 60 or older at the time of analysis (**Table 1**). Variant interpretation for mismatch repair Lynch syndrome genes – *MLH1*, *MSH2*, *MSH6* and *PMS2* – was done using well-established clinical interpretations and according to ACMG-AMP criteria and following recommendations from InSiGHT Hereditary Colorectal Cancer/Polyposis Expert Panel (see **Methods**). The list of all variants identified in these genes, the annotations, evaluations of each criterion and final score and pathogenicity assignments are provided in **Supplementary Table 3**. We found that 448 participants (0.31% or 1 in 320) had one pathogenic (P/LP) variant in one of the 4 genes. The majority had either a pathogenic variant in *MSH6* (n=201) or *PMS2* (n=181), while fewer had a pathogenic variant in *MLH1* (n=41) or *MSH2* (n=25) (**Supplementary Table 4**).

Risk of diagnosis of a Lynch syndrome related cancer is gene-dependent and lowest for *PMS2* pathogenic variants

To investigate the impact of pathogenic variants in the different genes, we extracted the date of first diagnosis recorded for any of 11 cancers known to be relevant for Lynch syndrome (**Methods**, **Supplementary Table 5**) [7]. The prevalence of 9 of these 11 cancers is very low (<0.5%) in our all-comers population. Therefore, we decided to do a time-to-first diagnosis analysis for colorectal cancer alone, for endometrial cancer alone, and two analyses where we combined cancers: (i) the 5 cancers – colorectal, endometrial/uterine, ovarian, kidney/ureter and small bowel – reported to be associated with all 4 genes in the literature, and (ii) all known 11 relevant cancers together (**Figure 1, Supplementary Table 6**)[7]. As expected, those with a pathogenic variant in *MLH1* or *MSH2* were at much higher risk to develop any of the 11 cancers compared to those without a pathogenic variant or VUS in the four genes: Hazard $\text{Ratio}_{\text{MLH1_11cancers}}$ = 16.5 (95% Confidence Interval: 8.9-30.8) and $\text{HR}_{\text{MSH2_11cancers}}$ = 17.3 (95%CI: 7.8-38.6). The increased risk was even stronger when looking more specifically at colorectal cancer, $\text{HR}_{\text{MLH1_colo}}$ = 74.5 (37.1-149.8) and $\text{HR}_{\text{MSH2_colo}}$ = 62.0 (23.2-165.9). Those with a pathogenic variant in *MSH6* had an intermediate risk for any of the 11 Lynch syndrome relevant cancers $\text{HR}_{\text{MSH6_11cancers}}$ = 4.7 (3.2-6.9). The strongest risk related to *MSH6* pathogenic variants was with endometrial/uterine cancers for female participants $\text{HR}_{\text{MSH6_endo}}$ = 15.8 (8.4-29.6) (**Figure 1**).

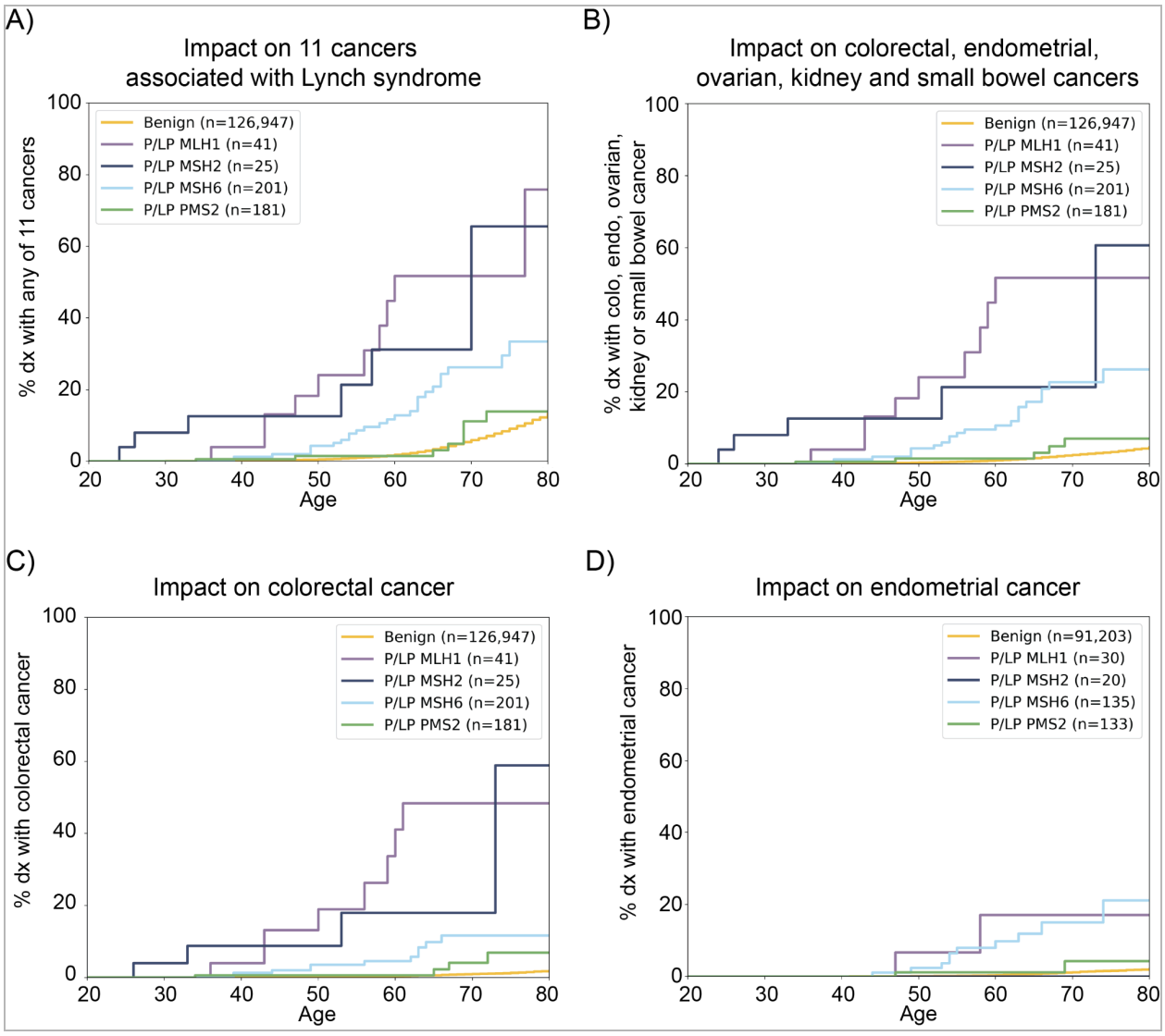
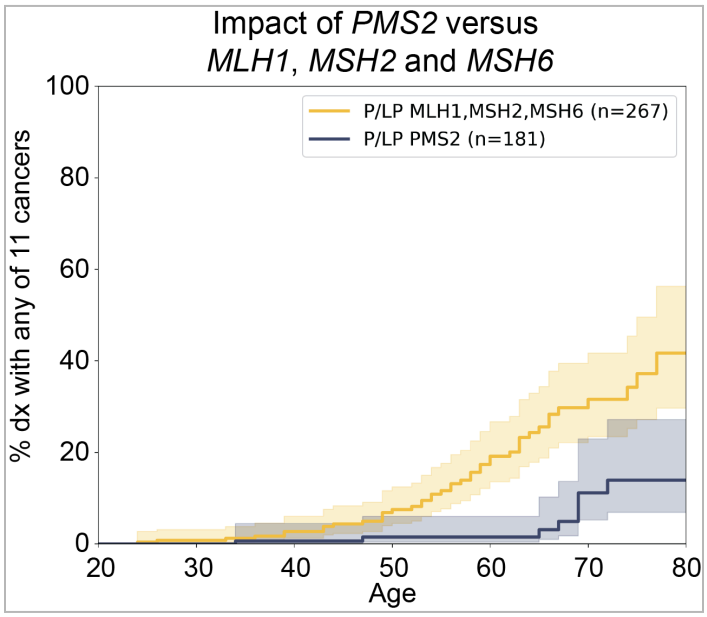


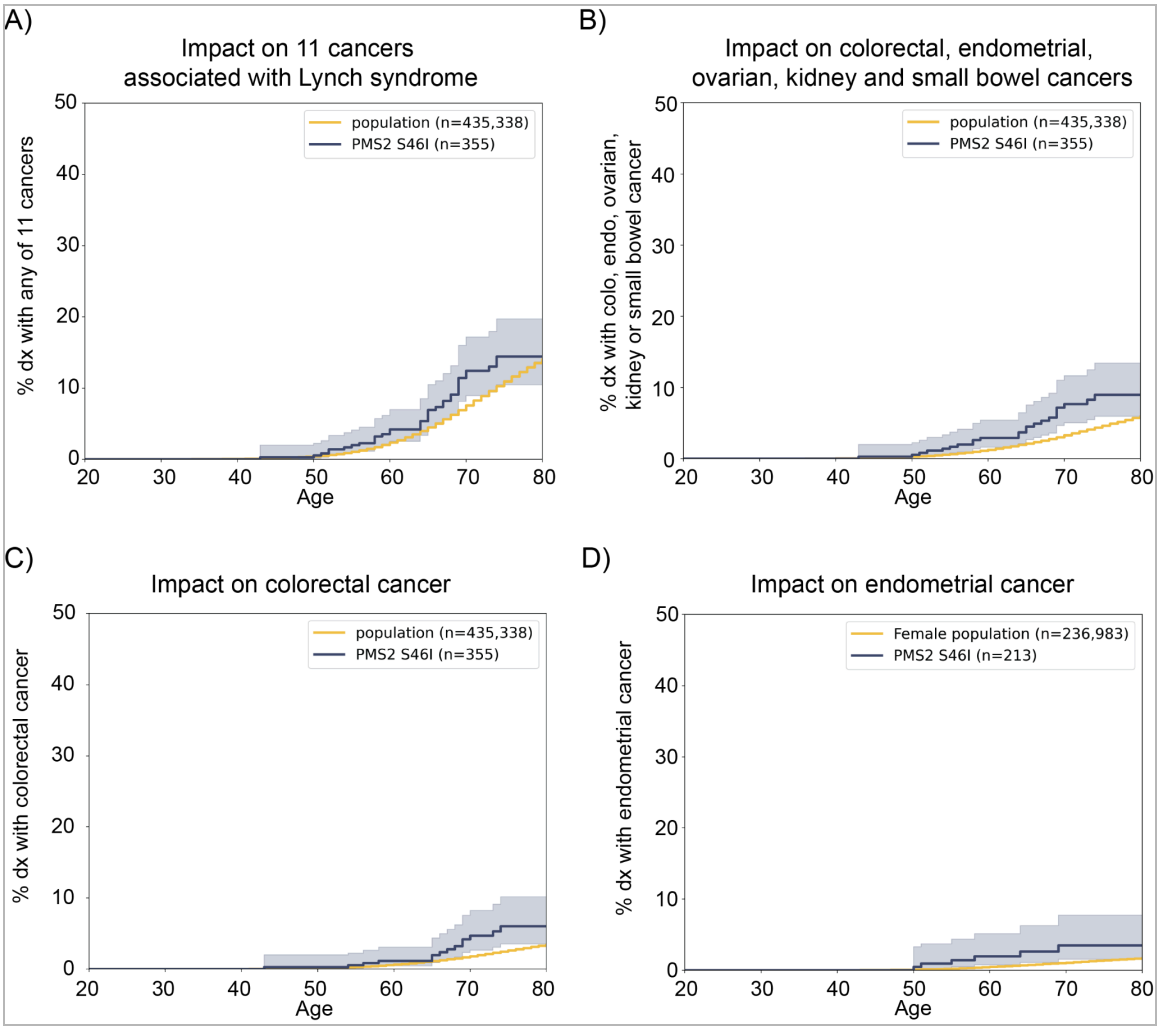
Figure 1: Gene by gene cumulative incidence of age at first Lynch syndrome-related cancer diagnosis across individuals harboring pathogenic variants from the Helix Research Network. Gene-level results for those harboring pathogenic (P/LP) variants (*MLH1* in purple, *MSH2* in blue, *MSH6* in light blue and *PMS2* in green) against those with Benign or no variant (yellow) are shown for: **A)** All 11 Lynch syndrome cancers: colorectal, endometrial and uterine, ovarian, kidney and ureter, small bowel, bladder, stomach, pancreas, biliary tract, brain, and prostate. **B)** 5 Lynch syndrome cancers with established associations with all four genes: colorectal, endometrial and uterine, ovarian, kidney and ureter, and small bowel cancer. **C)** Colorectal cancer and **D)** Endometrial cancers in female participants. Hazard ratios (with confidence intervals) and p values for each gene group against those without a Lynch syndrome pathogenic variant (yellow) are available in **Supplementary Table 6**.

On the other hand, pathogenic variants in *PMS2* were not significantly associated with the risk of developing any of the 11 LS-relevant cancers (P=0.39, log-rank test), or any of the top 5 LS-relevant cancers (P=0.094, log-rank test). *PMS2* pathogenic variants show a comparatively small increase in the risk of developing colorectal cancer HR_{PMS2_colo} = 4.3

(1.6-11.4) (**Figure 1**). The difference between *PMS2* and the 3 other genes was statistically significant (P=2.5E-05, log-rank test) when comparing them directly for the 11 LS-relevant cancers (**Supplementary Figure 1**). This deviation in cancer risk for those harboring *PMS2* variants was also recently reported in a similar retrospective analysis of the UK Biobank, an all comers cohort of 500,000 participants from the United Kingdom[6]. As validation, we used our phenotypic definitions to assess the UK Biobank results for *PMS2* p.S46I, the most frequent pathogenic variant in *PMS2*, accounting for one-third of individuals with Lynch syndrome variants in *PMS2* and more than 10% of all individuals with Lynch syndrome variants. Matching gene-level results for pathogenic variants in *PMS2* in HRN, the hazard ratios for individuals carrying this variant were minimal compared to those without a pathogenic variant. More specifically, *PMS2* p.S46I only had a relatively small effect, HR < 3, for colorectal cancer, HR_{S46I_colo} = 2.1 (1.2-3.5), or endometrial cancer, HR_{S46I_endo} = 2.8 (1.3-6.3), and there was no association when looking at the 11 Lynch syndrome cancers together (**Supplemetary Figure 2**, **Supplementary Table 7)**.



Supplementary Figure 1: Comparison of the impact of *PMS2* pathogenic variants with pathogenic variants in other mismatch repair genes (*MLH1*, *MSH2*, and *MSH6*). Cumulative incidence for 11 Lynch syndrome cancers in those with pathogenic *PMS2* variants are shown in blue and those with either a *MLH1*, *MSH2*, or *MSH6* pathogenic variant are shown in yellow. The difference between these groups is statistically significant (P=2.5E-05, log-rank test).



Supplementary Figure 2: Clinically-established pathogenic variant *PMS2* p.S46I shows limited risk for Lynch syndrome-related cancers in the UK Biobank. Cumulative incidence plots comparing individuals harboring *PMS2* p.S46I (blue) to those without a pathogenic variant in mismatch repair genes (yellow) for: **A)** All 11 Lynch syndrome cancers: colorectal, endometrial and uterine, ovarian, kidney and ureter, small bowel, bladder, stomach, pancreas, biliary tract, brain, and prostate. **B)** 5 Lynch syndrome cancers with established associations with all four genes: colorectal, endometrial and uterine, ovarian, kidney and ureter, and small bowel cancer. **C)** Colorectal cancer (HR_{S46I_colo} = 2.1 (1.2-3.5)) and **D)** Endometrial cancers in female participants (HR_{S46I_endo} = 2.8 (1.3-6.3)).

Clinical implications of heterogenous risk profiles across Lynch syndrome genes

As a result, we asked whether *PMS2* heterozygotes would still benefit from high-risk surveillance strategy recommendations or whether the risk is likely adequately addressed by the standard of care (**Supplementary Table 1**), especially in relation to colorectal cancer which (i) showed the highest hazard ratio for *PMS2* and (ii) has the most actionable recommendations

(see **Methods**). The US Preventive Services Task Force (USPSTF) recommends colorectal cancer screening for adults 50-75 yo (grade A recommendation, substantial net benefit), and for those 45-49 yo (grade B recommendation, moderate benefit)[15]. The recommended frequency of screening varies based on the type of test; it is every 10-years for a colonoscopy. Given that the goal of screening is to identify cancers early and that the official grade A recommendation is to perform a colonoscopy every 10-years starting at age 50, we focused our analysis at age 60 (as cancers diagnosed after would or could have been identified following a regular screening regimen). We looked at the percentage of individuals harboring Lynch syndrome variants in each gene who were diagnosed with either (i) colorectal cancer or (ii) any LS-relevant cancer at or before age 60 (**Table 2**). By age 60, the risk of developing colorectal cancer was similar for those with a *PMS2* pathogenic variant (0.6%) and the average population (0.4%). Moreover, by age 60, the risk of developing any of the 11 cancers associated with Lynch syndrome was similar for those with a *PMS2* pathogenic variant (1.5%) compared to the average population (1.8%), whereas this risk is much higher for those with a *MLH1* pathogenic variant (51.7%), *MSH2* pathogenic variant (31.2%) or *MSH6* pathogenic variant (12.8%). From this perspective, a change in start of colorectal screening age would likely not mitigate excess risk for those harboring *PMS2* pathogenic variants.

Discussion

The recent availability of large biobanks with genetic and phenotype data allow us to assess the clinical impact of pathogenic variants at the population level. Our results show attenuated clinical impact for pathogenic variants in *PMS2.* Our study design controlled for two potential biases that could have led to artificially low penetrance. First, we restricted our analysis to exons 1 to 10 of *PMS2* to avoid potential false positives caused by a pseudogene that has very high sequence similarity to exons 11-15 of *PMS2.* Second, the high penetrance and increased clinical risk caused by pathogenic variants in *MLH1*, *MSH2*, and even *MSH6* in our cohort suggest that the low penetrance of *PMS2* pathogenic variants is not the result of studying a ‘healthier population’ compared to the general population. Moreover, our results were consistent with two recent studies that reported lower penetrance of pathogenic variants in *PMS2* compared to the 3 other Lynch syndrome genes in the UK Biobank[6] and in All Of Us[16], two other population biobanks where participants were not enrolled based on criteria related to Lynch syndrome or other cancers. Lastly, a large study showed that the majority of solid tumors

in patients with germline pathogenic variants in *PMS2* were found to be microsatellite stable (MSS tumor) and, in these cases, it was likely that the *PMS2* pathogenic variants were not causative of the MSS tumor[17]. Therefore, we question whether it is appropriate to return pathogenic *PMS2* interpretations, given current surveillance and prevention strategies associated with *PMS2*, to a healthy adult for screening purposes.

Our results show that by age 60, the risk of developing colorectal cancer or any of the 11 cancers associated with Lynch syndrome is similar between those with a *PMS2* pathogenic variant (0.6% and 1.5% respectively) and the average population (0.4% and 1.8%), whereas this risk is much higher for those with a pathogenic variant in *MLH1*, *MSH2* or *MSH6*. The NCCN guidelines recommend starting high quality colonoscopy at age 30-35y for those with a *PMS2* pathogenic variant and to repeat those every 1-3y [2], while the USPSTF guidelines recommend starting colonoscopies at age 45-50y and to repeat every 10y for the general population[15]. Considering that the risk of colorectal cancer appears to be equivalent for those with a *PMS2* pathogenic variant and those without a variant in any Lynch syndrome gene until age 60y, is it worth the individual efforts (financial cost, pain to patients, risks of a procedure etc.) to start screening earlier and do 7 or more additional colonoscopies? This math amplifies when tackling this question at the population level. Given that there are at least 125 individuals per 100,000 people with *PMS2* pathogenic variants in the population, is it best to perform 875 high-quality colonoscopies on those with a *PMS2* pathogenic variant under 50y, or on another group at increased risk? In circumspect, this deviation in penetrance for colorectal and other Lynch Syndrome cancers for *PMS2* in the general population highlights the importance of aligning surveillance and prevention strategies to the natural history trends derived from the population they are intended to serve[18].

Article information

Declarations of interests:

K.M.S.B., C.H., K.L., J.L., W.L., E.T.C., and A.B. are employees of Helix Inc. No other disclosures were reported.

Ethics declaration:

The Helix Research Network protocol has been IRB-approved, enabling secondary research use of data. Approvals for the protocol were granted by the Salus IRB (reliance on Salus for all sites; approval number 21143), the WCG IRB (Western Institutional Review Board, WIRB-Copernicus Group; approval number 20224919), the MUSC Institutional Review Board for Human Research (approval number Pro00129083), and the University of Nevada, Reno Institutional Review Board (approval number 7701703417). All participants provided written informed consent prior to participation. All data used for research had direct identifiers removed to safeguard participant privacy.

Data sharing statement:

Access to Helix Research Network™ (HRN) data are available to qualified researchers subject to approval by the HRN Steering Committee and Helix. Interested researchers must enter into a Data Use Agreement, which prohibits re-identification of participants, sharing of data with third parties, and uploading data to public domains. The HRN is open to individual collaborations with scientific researchers. Considerations for data access requests include: (1) affiliation with an accredited academic institution that is committed to participant privacy and data security; (2) specificity, type and volume of data requested; (3) feasibility of the proposed research project; and (4) resource commitments from Helix and HRN member institutions required to support a collaboration.

Acknowledgements:

We thank all of the participants of Imagine You, DNA Answers, the Genetic Insights Project, the Healthy Nevada Project, In Our DNA SC, myGenetics, and The Gene Health Project. Funding was provided to the Desert Research Institute by the Renown Institute for Health Innovation and the Renown Health Foundation. Funding was provided to DRI by the Nevada Governor's Office of Economic Development. Funding was provided to the myGenetics program by HealthPartners. Funding was provided to C.A.A.C. by the European Union's Horizon

Europe, under grant agreement No 101136962. Funding was provided to C.A.A.C. by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee [grant agreements No 10098097, No 10104323]. Funding was provided to C.A.A.C. by Federal Deparment of Economic Affairs, Education and Research EAER, State Secretariat for Education, Research and Innovation SERI. We also thank all of the participants of the UK BIobank study. This research has been conducted using the UK Biobank Resource under Application Number 40436. Lastly, we acknowledge the entire Helix research, bioinformatics and lab teams for their contributions to the production of the exome sequencing pipeline as well as the research administration team for coordinating the project. We thank Dr. Hang Dai, Dr. Xiao-Fei Kong, Dr. Kevin Hughes, Dr. Raymond Kim, and Dr. Alan Yahanda for their valuable feedback and discussions related to this manuscript.

References

1. Hampel H, Hall MJ. Hereditary Aspects of Colorectal Cancer: Mismatch Repair Genes Drive Lynch Syndrome. *Journal of the Advanced Practitioner in Oncology*. 2018;9(3):311. Accessed December 18, 2024. https://pmc.ncbi.nlm.nih.gov/articles/PMC6333561/

2. National Comprehensive Cancer Network. NCCN Guidelines Version 3.2024 Lynch Syndrome. October 31, 2024. Accessed January 13, 2025. https://www.nccn.org/home

3. Syngal S, Brand RE, Church JM, et al. ACG clinical guideline: Genetic testing and management of hereditary gastrointestinal cancer syndromes. *Am J Gastroenterol*. 2015;110(2):223-262; quiz 263. doi:10.1038/ajg.2014.435

4. Murray MF, Evans JP, Angrist M, et al. A proposed approach for implementing genomics-based screening programs for healthy adults. *NAM Perspect*. Published online December 3, 2018. doi:10.31478/201812a

5. Grzymski JJ, Elhanan G, Morales Rosado JA, et al. Population genetic screening efficiently identifies carriers of autosomal dominant diseases. *Nat Med*. 2020;26(8):1235-1239. doi:10.1038/s41591-020-0982-5

6. Fummey E, Navarro P, Plazzer JP, Frayling IM, Knott S, Tenesa A. Estimating cancer risk in carriers of Lynch syndrome variants in UK Biobank. *J Med Genet*. 2024;61(9):861-869. doi:10.1136/jmg-2023-109791

7. Dominguez-Valentin M, Haupt S, Seppälä TT, et al. Mortality by age, gene and gender in carriers of pathogenic mismatch repair gene variants receiving surveillance for early cancer diagnosis and treatment: a report from the prospective Lynch syndrome database. *EClinicalMedicine*. 2023;58(101909):101909. doi:10.1016/j.eclinm.2023.101909

8. McLaren W, Gil L, Hunt SE, et al. The ensembl variant effect predictor. *Genome Biol*. 2016;17(1). doi:10.1186/s13059-016-0974-4

9. Chen S, Francioli LC, Goodrich JK, et al. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature*. 2024;625(7993):92-100. doi:10.1038/s41586-023-06045-0

10. Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: An ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet*. 2016;99(4):877-885. doi:10.1016/j.ajhg.2016.08.016

11. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting splicing from primary sequence with deep learning. *Cell*. 2019;176(3):535-548.e24. doi:10.1016/j.cell.2018.12.015

12. Jia X, Burugula BB, Chen V, et al. Massively parallel functional testing of MSH2 missense variants conferring Lynch syndrome risk. *Am J Hum Genet*. 2021;108(1):163-175. doi:10.1016/j.ajhg.2020.12.003

13. Shuey MM, Stead WW, Aka I, et al. Next-generation phenotyping: introducing phecodeX for enhanced discovery research in medical phenomics. *Bioinformatics*. 2023;39(11):btad655.

doi:10.1093/bioinformatics/btad655

14. Davidson-Pilon C. lifelines: survival analysis in Python. *J Open Source Softw*. 2019;4(40):1317. doi:10.21105/joss.01317

15. US Preventive Services Task Force, Davidson KW, Barry MJ, et al. Screening for colorectal cancer: US Preventive Services Task Force recommendation statement: US preventive services task force recommendation statement. *JAMA*. 2021;325(19):1965-1977. doi:10.1001/jama.2021.6238

16. Park J, Karnati H, Rustgi SD, Hur C, Kong XF, Kastrinos F. Impact of population screening for Lynch syndrome insights from the All of Us data. *Nat Commun*. 2025;16(1):1-7. doi:10.1038/s41467-024-52562-5

17. Latham A, Srinivasan P, Kemel Y, et al. Microsatellite Instability Is Associated With the Presence of Lynch Syndrome Pan-Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2019;37(4). doi:10.1200/JCO.18.00283

18. Schiabor Barrett KM, Bolze A, Ni Y, et al. Positive predictive value highlights four novel candidates for actionable genetic screening from analysis of 220,000 clinicogenomic records. *Genet Med*. 2021;23(12):2300-2308. doi:10.1038/s41436-021-01293-9

Tables

Table 1: Demographics from the Helix Research Network data

Demographics table	Number of participants (% of total)
Participants with genetic and phenotype data available	144,852
Sex	F: 103,692 (71.6%) M: 41,160 (28.4%)
Mean age in 2024, y (min - max)	53.3 (18 - 106)
Genetic similarity	Africa: 6,500 (4.5%) Americas: 10,620 (7.3%) East Asia: 2,858 (2.0%) Europe: 119,979 (82.8%) South Asia: 1,014 (0.7%) Other: 3,881 (2.7%)
with a colorectal cancer dx	738 (0.51%)
with a endometrial/uterine cancer dx	491 (0.47%*)
with an ovarian cancer dx	288 (0.28%*)
with a kidney/ureter cancer dx	370 (0.26%)
with a bladder cancer dx	433 (0.30%)
with a stomach cancer dx	73 (0.05%)
with a pancreas cancer dx	193 (0.13%)
with a small bowel cancer dx	34 (0.02%)
with a biliary tract cancer dx	54 (0.04%)
with a brain cancer dx	152 (0.10%)
with a prostate cancer dx	1,895 (4.6%**)
avg EHR lookback, y	12.9

* percentage based on overall number of female participants
** percentage based on overall number of male participants

Table 2: Gene by gene cumulative risk of cancer by age 60

	average HRN population	**P/LP *MLH1* heterozygotes**	**P/LP *MSH2* heterozygotes**	**P/LP *MSH6* heterozygotes**	**P/LP *PMS2* heterozygotes**
Number	126,947	41	25	201	181
% dx with colorectal cancer by age 60	0.4%	41.0%	17.9%	4.5%	0.6%
% dx with colorectal, endometrial/uterine, ovarian, kidney/ureter or small bowel cancer by age 60	1.0%	51.7%	21.3%	10.7%	1.5%
% dx with any Lynch relevant cancer by age 60	1.8%	51.7%	31.2%	12.8%	1.5%