Published in final edited form as:

Fam Cancer. 2019 January ; 18(1): 67–73. doi:10.1007/s10689-018-0096-0.

Modified Capture-Recapture Estimates of the Number of Families with Lynch Syndrome in Central Ohio

John Michael O. Ranola^{#1}, Rachel Pearlman^{#2}, Heather Hampel², and Brian H. Shirts¹

¹Department of Laboratory Medicine, The University of Washington, Seattle, WA

²Department of Internal Medicine and the Comprehensive Cancer Center, The Ohio State University, Columbus, OH

These authors contributed equally to this work.

Abstract

Purpose: Past methods for estimating the population frequency of familial cancer syndromes have used cases and controls ignoring the familial nature of genetic disease.

Methods: In this study we modified the capture-recapture method from ecology to estimate the number of families in central Ohio with Lynch syndrome (LS).

Results: We screened 1566 colorectal cancer cases and 545 endometrial cancer cases in central Ohio from 1999–2005 and identified 58 with LS. We screened an additional 3346 colorectal and 342 endometrial cancer cases from 2013–2016 and identified 149 with LS. We found 12 LS mutations shared between families observed in the first and second studies. We identified 3 individuals between studies who were closely related and 8 who were more distantly related. We used identified family relationships and genetic test results to estimate family size and structure. Applying a modified capture-recapture method we estimate 1693 three-generation families in the area who have 288 unique LS causing mutations. Comprehensive colorectal and endometrial cancer screening will take about 20 years to identify 50% of families with LS.

Conclusions: This is the first time that the capture-recapture method has been applied to estimate the burden of families with a specific heritable disease. Family structure reveals the potential extent of prevention and the time necessary to identify a proportion of families with LS.

Keywords

Mark-recapture; Capture-Recapture; population genetics; population structure; cascade testing; genetic counseling; prevalence; Lynch syndrome

INTRODUCTION

Estimating the population frequency of familial diseases can be challenging, as not all individuals with a genetically defined syndrome will develop symptomatic disease. In the case of Lynch Syndrome (LS) high-risk families have been used for initial estimates of

population prevalence and penetrance, but these estimates can be biased[1–6]. Subsequent estimates of LS prevalence have been made by determining the prevalence of LS among cases, multiplying that times the prevalence of the disease in the population, and then correcting for penetrance[4, 5, 7]. For accurate prevalence estimates, unbiased sampling of cases and controls is necessary[8]. However, population frequency based on counting individuals ignores the hereditary nature of familial cancer risk, with cases occurring in familial clusters. Estimates of population frequency that include family structure have not been done. LS confers autosomal dominant genetic risk of many types of cancer, most prominently colorectal and endometrial cancer. Lifetime colorectal cancer risks are 52%–82%, compared with the 4.8% risk in the general population[9]. Colorectal cancer can be prevented with regular colonoscopies, so identifying individuals with LS early can save lives[10]. Universal screening of all colorectal cancers for LS has been shown to be cost effective as cascade testing of relatives benefits relatives beyond the initial case [11–13]. Hence, the most efficient way to find all individuals with LS is to focus on families with LS[14].

Since 1999 we have been involved in two studies directed at universal LS screening in central Ohio. We realized that these two studies could be used to estimate the number and size of families with LS by modifying the capture-recapture statistical method, a well-established method from ecology. Capture-recapture methods compare ascertainment in an initial and subsequent time period noting individuals, or families in our study, that have been ascertained more than once. Since both Ohio studies gathered clinical pedigrees with identifiable individuals, we were able to apply capture-recapture principles to entire families and estimate the number of families with LS in the area, including those yet to be diagnosed. Knowing general family structure within the population may help guide prevention efforts by defining the LS distribution in terms that are consistent with cost-effective cascade testing[14, 15]. Direct estimates of the frequency of LS in the context of specific mutations and families with these mutations will give a more complete picture of population genetic heterogeneity.

METHODS

Sample

Between 1999 and 2005, the Columbus-area HNPCC study comprehensively screened unselected colorectal and endometrial cancer cases from 3 hospital systems for Lynch syndrome as previously described[16–18]. Between 2013 and 2016, the Ohio Colorectal Cancer Prevention Initiative (OCCPI; ClinicalTrials.gov identifier: NCT01850654) study screened colorectal cancer cases from 51 participating hospitals across the state of Ohio, and endometrial from Ohio State University (OSU). Institutional Review Board (IRB) approval was obtained by the individual hospitals, Community Oncology Programs, or by ceding review to the OSU IRB.

Lynch Syndrome Case Ascertainment and Genotyping

All tumors in both the first (HNPCC) and second (OCCPI) cohorts were screened for mismatch repair (MMR) deficiency by microsatellite instability (MSI) testing and/or

immunohistochemical (IHC) analysis at OSU, if not already completed in a CLIA-approved laboratory for clinical care as previously described[16–18]. All genetic results were available to researchers HH and RP. Three-generation pedigrees were ascertained for all individuals identified with LS in both studies. During pedigree ascertainment, at-risk relatives were identified for cascade testing. Through this process, it was discovered that some relatives, and even participants, had already been identified as a member of a known LS family through the prior Columbus-area and current OCCPI studies, respectively. Furthermore, potential connections between individuals with the same genotype were investigated using pedigree comparison and genealogical research, when possible. Public online family history tools, birth certificates, death certificates, census data, cemetery documents, and archived obituaries were used to construct extended pedigrees.

Capture-recapture analysis

To estimate of the number of families with LS we used a solution from ecology called the capture-recapture or mark-recapture method[20–22]. This method was classically used to estimate the number of animals in a given area. To apply this on deer, for example, deer would be caught, marked, and released noting how many were marked (m_1). At a later time, deer would again be caught and released after noting how many were previously marked (m_2) and the total caught (n_2). Then, the number of deer in the area (n) is estimated by equating the proportion of marked deer in the second capture to the proportion of marked deer in the entire population (i.e. $m_2/n_2 = m_1/n$). With three variables known, we can solve for the number of deer in the entire population: $n = (m_1 * n_2) / m_2$. This formula is known to overestimate the true size of the population for small sample sizes[23]. To minimize bias in our analysis we use the related equation $n = \{[(m_1+1)*(n_2+1)] / (m_2+1)\} - 1$ [23]. Furthermore, we used the amended 95% confidence interval $CI = N \pm 1.96 * SE$ where the standard error is $SE = \sqrt{\{(m_1+1)(n_2+1)(m_1-m_2)(n_2-m_2) / (m_2+1)^2(m_2+2)\}}$ [23].

There are several underlying assumptions to this method [20–22]. First, the probability an individual belongs to the first set (set A), the first capture period in the deer example, is identical for those who are in the second set (B) and those who are not. Next, the probability an individual is caught in a given set is equal for all individuals and independent of whether an individual has been caught before. Lastly, the probability an individual is in set A (or B) is constant over time for the duration of the study. This last assumption implies that the population is closed, i.e. there are no births, no deaths, nor migrants.

Applying capture-recapture to human families

Genetic counselors in both HNPCC and OCCPI studies collected pedigree information for each identified LS patient. We considered families to be “recaptured” in the second study if an individual with cancer had the same LS variant and was in the same pedigree as someone who was identified in the first study. As there are many definitions of family, we evaluated three sizes of families: 1) standard 3-generation families defined as relatives of the identified LS case out to first cousins, 2) extended families defined as any known or identified relationship between LS cases, and 3) genetic LS families defined as anyone with the same mutation. Extended families were identified by the research study and were often unknown to the families. The rationale for distinctions in family size is that cascade testing typically

does not extend beyond 3-generation. However, information on extended families was often available because both HNPCC and OCCPI studies collated extensive family history information from the LS patients identified. Genetic families consist of all individuals with the same LS mutation inherited from a common ancestor.

Adjustment for cascade testing

Once an individual was identified with LS in the Columbus-area HNPCC study, their at-risk relatives were identified and underwent single-site testing for the familial mutation[7]. Relatives then received counseling and intensive cancer surveillance recommendations. This cascade testing, education, and surveillance was expected to reduce the risk that relatives would develop colorectal cancer, reducing their odds of being ascertained (recaptured) in the second (OCCPI) study period. In order to adjust for the expected lower number of families recaptured due to intensive surveillance of LS families in the HNPCC cohort, we estimated age-specific colorectal cancer incidence on published LS penetrance during the 2nd ascertainment period for individuals identified with LS [24–26]. We then estimated the number of individuals with colorectal cancer who might have been ascertained by the OCCPI had cascade testing not occurred.

Estimate of the number of 3-generation families with specific LS mutations

We combined estimates of the number of independent 3-generation families in the area with each mutation yet to be identified. To do this, we used the estimate of the total number of families with LS, the estimate of the total number of LS mutations, and the number of families observed with each known mutation. Starting with a histogram of the number of families for each of the mutations identified, we performed a Laplace smoothing which gave us counts for the expected mutations yet to be identified under the assumption that, if sampling continued, we would have identified one family with each new mutation. We converted the number of families to proportions of families with each of the mutations and multiplied proportions by number of families to get an estimate of the number of families with each mutation.

Family history and haplotype evaluation to infer common ancestor

Inheritance from a common ancestor is the most parsimonious explanation of a shared genetic mutation. To evaluate this assumption, we used haplotype data from a subset of OCCPI participants where multigene panel sequencing had been performed at the University of Washington. The ColoSeq and BROCA panels include nonrepetitive intronic sequences, facilitating haplotype analysis. Haplotype analyses were performed for *MLH1* p.K461X, c.589–2A>G, and exon 16–19 del; *MSH2* c.942+3A>T, p.A636P, exon 8 del, and exon 13–14 del; *MSH6*, p.L370S, c.3840_3846del, and c.3939_3957dup; and *PMS2* c.1831dup and p.L625X. We used population data from European populations and LDlink software to select tag SNPs with $r^2 < 0.95$ across the *MLH1*, *MSH2*, *MSH6*, and *PMS2* genes and generate population haplotype frequencies[27] (see supplemental Tables 1–4). Comprehensive haplotype evaluation of the *MSH2* exon 1–6 deletion has been previously published[28].

RESULTS

Individuals with Lynch Syndrome

Between 1999 and 2005, the Columbus-Area HNPCC study screened 1566 colorectal cancer cases and 543 endometrial cancer cases for LS. Fifty-eight individuals were identified with 23 unique LS-associated mutations. Between 2013 and 2016, the OCCPI study screened 3346 colorectal cancer cases and 342 endometrial cancer cases for LS. One hundred forty-nine individuals with 101 unique LS mutations were identified. One individual was ascertained by both the initial HNPCC study (for her endometrial cancer diagnosis) and OCCPI study (for her colorectal cancer diagnosis). This individual represents a missed opportunity for prevention of her colorectal cancer because her germline *PMS2* mutation was not able to be detected due to technology limitations at the time of her endometrial cancer diagnosis in the initial HNPCC study. Three individuals with cancer in the OCCPI study were known relatives of individuals in the HNPCC study. Work connecting pedigrees identified five OCCPI participants as distant relatives of LS patients identified by the HNPCC study in addition to the three previously known relatives. Sixteen patients from the OCCPI study were connected as biological relatives of other OCCPI patients through genotyping, pedigree comparison, and genealogy research (degree of relationship ranged from first-degree to 4th cousins). Counting either study, 99 individuals shared the same LS mutation with another individual ascertained during either study period.

Effect of Cascade testing

Cascade testing performed as part of the Columbus-area HNPCC study identified 108 relatives of 58 probands with LS, 83 of who lived in Ohio. Based on published colon and endometrial cancer risks, an additional 3.6 instances of colon or endometrial cancer in Ohio are thought to have been prevented by aggressive surveillance and prevention strategies in these individuals between 2013 and 2016 (6.9 using upper 95% CI for LS penetrance and 2.1 using lower 95% CI for LS penetrance[25]). Approximately 1/6 of all CRC cases in Ohio were screened for LS as part of the OCCPI study, so we estimated an extra 0, 1, or 2 individuals with LS would have presented with colon or endometrial cancer in the OCCPI study had cascade testing not been performed as part of the initial Columbus-area HNPCC study.

Number of families with LS in catchment area

The capture-recapture method estimated 1693 three-generation LS families (relatives up to first cousins) in the central Ohio catchment area. The number of extended families (any documented relationship) estimated to be in Ohio was 883. Table 1 shows these results along with the 95% confidence intervals of these estimates. Estimates assume a single additional family would have been recaptured had cascade testing not been initially performed. Supplemental Table 5 shows the sensitivity analysis of how cascade testing bias might alter the estimates of number of families and mutations with LS.

Capture-recapture estimated 288 mutations that cause LS in the catchment area. We estimate that it will take 20 years to identify 50% of individuals with LS in Ohio, and 76 years to identify 90% of individuals with LS in Ohio with universal screening of all CRC and

endometrial cancer cases and cascade testing extending all relatives out to first cousins (Table 1). We expect to observe 90% of unique LS mutations in Ohio in at least one family within the next 13 years if universal tumor screening of CRC and endometrial cancer continues.

Frequency of specific mutations in the population

Our two studies identified 109 different mutations causing LS. We predict there are about 179 unique LS mutations in central Ohio that were not identified in either study. We assumed each of these 179 unidentified mutations is present in only one 3-generation family, as were 75 of the mutations already identified, and estimated the number of families with each of the mutations (Figure 1). The *MSH2* c.942+3A>T mutation is by far the most common mutation associated with LS and is expected to be present in about 200 three-generation families in the area. *MSH2* exon 1–6 deletion is the next most common, having about 100 expected families. The substantial majority of mutations are estimated to be present in 10 or fewer 3-generation families.

Common ancestor inference using haplotypes

Unphased haplotype data for tag SNPs was available for 12 different mutations shared by a total of 48 individuals (Supplemental Tables 1–4). For most mutations investigated haplotype data was consistent with a single ancestral mutational event. There were several variants where haplotype data was not consistent with a single founder event. For the most common *MSH2* mutation, c.942+3A>T, which occurs at the end of a poly-A microsatellite track and is a mutational hotspot in tumors[29], haplotype data was available for 11 individuals. *MSH2* c.942+3A>T was present on at least 3 haplotypes, suggesting 3 unique mutational events with 8, 2, and 1 individuals sharing haplotypes. Four of five haplotyped individuals with *MSH6* p.L370S shared the most common haplotype, but one individual did not share this haplotype. This could indicate a unique mutational event, a meiosis between p.L370S and the haplotype-defining marker in this individual, or genotyping error for a haplotype tagging SNP. There were apparently two independent events for *MLH1* c.589–2A>G.

Comparison with other estimates

We compared estimates of individuals in Ohio with LS to gauge the accuracy of our estimate. While there are no direct estimates of the number of families with LS, we compared our family estimates to estimates of the number of LS individuals in Ohio who have colorectal cancer. According to cancer.gov, there are 1.3 million people who have had colorectal cancer in the United States[30]. If Ohio has a proportional share of CRC, there should be around 52,000 individuals in the state who have or have had CRC. In the two Ohio studies described here, we evaluated over 4,000 of these CRC cases and found that 3.39% have LS. This gives a prediction of 1,763 individuals in Ohio with LS and colorectal cancer. Alternatively, we can start with the 1693 families we estimated using the capture-recapture method. Based on simulations of 100 *MLH1* 3-generation families using the R Package CoSeg[31] with demographic assumptions from the United States, we expect 1.27 individuals per family to have LS and be affected with colorectal cancer. So in 1,693 3-

generation families with LS, the number of individuals with LS who have colorectal cancer would be 2,150.

Alternatively, we can use an estimate that 1:279 individuals in the U.S. have LS [6]. With 11.6 million people in Ohio[32] we expect 41,500 to have LS. Based on simulations described above, we expect each 3 generation family to have 7.5 members at any age with LS. The capture-recapture estimate of 1,693 families would indicate 12,698 individuals with LS in Ohio. Capture-recapture based estimates are 18% above the estimate of individuals with LS and colorectal cancer and 70% below the estimate of total number of individuals with LS.

DISCUSSION

The capture-recapture method from ecology can give a meaningful, independent estimate of families with LS. Our analysis indicates that over one quarter of 3-generation LS families in this area have one of five LS mutations that are common in this population. Not all of these are common in public databases [33, 34] suggesting that a portion of LS is local and that population frequency may vary depending on which mutations are prevalent in the specific population of interest. Although there have been estimates of the number of individuals with LS in the population, to our knowledge, this is the first that directly estimates the number of families in a population with a dominant genetic disease. Estimates generated from our results were higher than other estimates of patients with LS and colorectal cancer but lower than estimates of total individuals with LS. Unfortunately, large differences in previous population estimates prevent clear conclusions about potential systematic bias in our application of the capture-recapture method. Large differences in past estimates do, however highlight the need independent methods to estimate population frequency, such as the capture-recapture strategy applied here.

Our estimates may be biased because of slight deviations from the assumptions of the model. Our study follows some assumptions of the capture-recapture method better than others, as is often the case for applications of statistical methods to real world problems. The assumption of identical probability of a family being in each ascertainment set may or may not hold as the several adult individuals in each family that are risk of developing cancer at any time may not be randomly distributed throughout the catchment areas. The independence of the capture and recapture sets also likely does not hold since cascade testing was predicted to prevent colorectal cancer in individuals with LS, reducing the probability of ascertaining those families in the second study. This could cause artificially large estimates of the total number of LS families in the population (see Sup Table 1). Because of this, we corrected for this known bias in our estimates. Almost any application of statistical methods to real populations requires similar estimates and corrections[35]. The assumption of constant population over time may fit reasonably well. Migration may create or eliminate new families from the catchment area; however, census data suggest the Ohio population has been relatively stable during the study period. *De novo* mutations creating new families with LS are extremely rare, but could alter assumptions over a much longer time frame. The assumption that probability of ascertainment is equal for all families may not hold. Clearly, larger families have more individuals who could be diagnosed with LS at

any time, and thus they have a larger probability of being ascertained during both capture and recapture phases. However, overall population averages are likely to be reasonable as family history of cancer was not an enrollment criterion for either study.

One additional difficulty with the method is the time required for human cancer ascertainment and large time gap between the catch and recatch periods. There are potential effects on our estimates of families growing and moving over time, but these effects could bias estimates either way. The likely effect of these deviations from standard assumptions would be a change in the number of families recaught and the resulting estimates would be similar to those presented in our sensitivity analysis (Supplementary Table 5). More sophisticated capture-recapture analysis that assesses validity of these assumptions will be possible with additional capture phases[36].

Use of the capture-recapture method in the context of families requires unbiased ascertainment of cases from the population and accurate pedigree collection. In our study, ascertainment of cancer cases was not contingent on known familial relationships, and we could correct for expected bias due to colorectal cancer prevention efforts because of careful documentation of individuals potentially benefitting from the cascade testing intervention in the initial HNPCC study.

Family size is somewhat arbitrary, and ideal definitions of family depend on the desired application or analysis. Our three definitions of family illustrate how each definition might influence results. Universal tumor screening is a public health practice designed to identify individuals with LS[37–39]. If universal screening as currently implemented in Ohio continued, it will take an additional 20 years (beyond the 10 years of the study) to identify 50% of 3- generation families in the catchment area with LS. Expanding pedigree ascertainment to larger family groups might reduce this time (see Table 1), but doing so is currently not currently feasible on such a large scale. For example, this study was able to find several distant relative pairs previously unknown to patients because of access to pedigrees from multiple individuals with the same mutation. There is no current public health mechanism to expand cascade testing beyond 3-generation pedigrees in this way. Furthermore, our haplotype analysis shows that in most cases, individuals with the same rare mutation can be linked to a genetic founder. Over one quarter of all predicted LS families were caused by the five most common mutations. Founder mutations in other populations may be different, and the magnitude of these founder effects may cause substantial changes in local LS population frequency.

Understanding population frequency with familial context illustrates the number of individuals who will benefit from improved cascade testing. This understanding may also allow better cost-benefit estimates and influence public health policy. Although the capture-recapture method as applied to families with genetic syndromes appears promising, it would benefit from additional research to validate estimates and explore the validity of underlying assumptions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The Ohio Colorectal Cancer Prevention Initiative is supported by Pelotonia, <http://pelotonia.org/>. Brian Shirts and John Ranola are supported by Damon Runyon Cancer Research Foundation (DRR-33-15) and by development funds from the Fred Hutch/University of Washington Cancer Consortium (NCI 5P30 CA015704-39).

References

1. Carayol J, Khat M, Maccario J, Bonaïti-Pellié C (2002) Hereditary non-polyposis colorectal cancer: current risks of colorectal cancer largely overestimated. *J Med Genet* 39:335–339 [PubMed: 12011152]
2. Pearlman R, Frankel WL, Swanson B, et al. (2017) Prevalence and Spectrum of Germline Cancer Susceptibility Gene Mutations Among Patients With Early-Onset Colorectal Cancer. *JAMA Oncol* 3:464–471. doi: 10.1001/jamaoncol.2016.5194 [PubMed: 27978560]
3. Yurgelun MB, Allen B, Kaldate RR, et al. (2015) Identification of a Variety of Mutations in Cancer Predisposition Genes in Patients With Suspected Lynch Syndrome. *Gastroenterology* 149:604–613.e20. doi: 10.1053/j.gastro.2015.05.006 [PubMed: 25980754]
4. Haraldsdottir S, Rafnar T, Frankel WL, et al. (2017) Comprehensive population-wide analysis of Lynch syndrome in Iceland reveals founder mutations in MSH6 and PMS2. *Nat Commun* 8:14755. doi: 10.1038/ncomms14755 [PubMed: 28466842]
5. Chen S, Wang W, Lee S, et al. (2006) Prediction of germline mutations and cancer risk in the Lynch syndrome. *JAMA* 296:1479–1487. doi: 10.1001/jama.296.12.1479 [PubMed: 17003396]
6. Win AK, Jenkins MA, Dowty JG, et al. (2017) Prevalence and Penetrance of Major Genes and Polygenes for Colorectal Cancer. *Cancer Epidemiol Biomark Prev* 26:404. doi: 10.1158/1055-9965.EPI-16-0693
7. Hampel H, de la Chapelle A (2011) The Search for Unaffected Individuals with Lynch Syndrome: Do the Ends Justify the Means? *Cancer Prev Res (Phila Pa)* 4:1–5. doi: 10.1158/1940-6207.CAPR-10-0345
8. Terdiman JP, Levin TR, Allen BA, et al. (2002) Hereditary nonpolyposis colorectal cancer in young colorectal cancer patients: high-risk clinic versus population-based registry. *Gastroenterology* 122:940–947 [PubMed: 11910346]
9. Kohlmann W, Gruber SB (1993) Lynch Syndrome In: Adam MP, Ardinger HH, Pagon RA, et al. (eds) *GeneReviews*(®). University of Washington, Seattle, Seattle (WA)
10. Møller P, Seppälä T, Bernstein I, et al. (2017) Cancer incidence and survival in Lynch syndrome patients receiving colonoscopic and gynaecological surveillance: first report from the prospective Lynch syndrome database. *Gut* 66:464–472. doi: 10.1136/gutjnl-2015-309675 [PubMed: 26657901]
11. Krawczak M, Cooper DN, Schmidtke J (2001) Estimating the Efficacy and Efficiency of Cascade Genetic Screening. *Am J Hum Genet* 69:361–370. doi: 10.1086/321973 [PubMed: 11431707]
12. Ladabaum U, Wang G, Terdiman J, et al. (2011) Strategies to identify the Lynch syndrome among patients with colorectal cancer: a cost-effectiveness analysis. *Ann Intern Med* 155:69–79. doi: 10.7326/0003-4819-155-2-201107190-00002 [PubMed: 21768580]
13. Snowsill T, Huxley N, Hoyle M, et al. (2014) A systematic review and economic evaluation of diagnostic strategies for Lynch syndrome. *Health Technol Assess Winch Engl* 18:1–406. doi: 10.3310/hta18580
14. Hampel H (2016) Genetic counseling and cascade genetic testing in Lynch syndrome. *Fam Cancer* 15:423–427. doi: 10.1007/s10689-016-9893-5 [PubMed: 26969309]
15. Ladabaum U, Wang G, Terdiman J, et al. (2011) Strategies to identify the Lynch syndrome among patients with colorectal cancer: a cost-effectiveness analysis. *Ann Intern Med* 155:69–79. doi: 10.7326/0003-4819-155-2-201107190-00002 [PubMed: 21768580]

16. Hampel H, Frankel WL, Martin E, et al. (2005) Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). *N Engl J Med* 352:1851–1860. doi: 10.1056/NEJMoa043146 [PubMed: 15872200]
17. Hampel H, Frankel W, Panescu J, et al. (2006) Screening for Lynch syndrome (hereditary nonpolyposis colorectal cancer) among endometrial cancer patients. *Cancer Res* 66:7810–7817. doi: 10.1158/0008-5472.CAN-06-1114 [PubMed: 16885385]
18. Hampel H, Frankel WL, Martin E, et al. (2008) Feasibility of screening for Lynch syndrome among patients with colorectal cancer. *J Clin Oncol Off J Am Soc Clin Oncol* 26:5783–5788. doi: 10.1200/JCO.2008.17.5950
19. Shia J, Tang LH, Vakiani E, et al. (2009) Immunohistochemistry as first-line screening for detecting colorectal cancer patients at risk for hereditary nonpolyposis colorectal cancer syndrome: a 2-antibody panel may be as predictive as a 4-antibody panel. *Am J Surg Pathol* 33:1639–1645. doi: 10.1097/PAS.0b013e3181b15aa2 [PubMed: 19701074]
20. Petersen C GJ (1896) The Yearly Immigration of Young Plaiice Into the Limfjord From the German Sea. *Rep Dan Biol Stn* 6:
21. Lincoln F (1930) Calculating waterfowl abundance on the basis of banding returns. *US Dept Agric Circ* 118:1–4
22. Tilling K (2001) Capture-recapture methods—useful or misleading? *Int J Epidemiol* 30:12–14. doi: 10.1093/ije/30.1.12 [PubMed: 11171841]
23. Chapman DG (1951) Some properties of the hypergeometric distribution with applications to zoological sample censuses. University of California Press, Berkeley
24. Quehenberger F, Vasen HFA, van Houwelingen HC (2005) Risk of colorectal and endometrial cancer for carriers of mutations of the hMLH1 and hMSH2 gene: correction for ascertainment. *J Med Genet* 42:491–496. doi: 10.1136/jmg.2004.024299 [PubMed: 15937084]
25. Bonadona V, Bonaiti B, Olschwang S, et al. (2011) Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. *JAMA* 305:2304–2310. doi: 10.1001/jama.2011.743 [PubMed: 21642682]
26. Senter L, Clendenning M, Sotamaa K, et al. (2008) The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations. *Gastroenterology* 135:419–428. doi: 10.1053/j.gastro.2008.04.026 [PubMed: 18602922]
27. Machiela MJ, Chanock SJ (2015) LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinforma Oxf Engl* 31:3555–3557. doi: 10.1093/bioinformatics/btv402
28. Clendenning M, Baze ME, Sun S, et al. (2008) Origins and prevalence of the American Founder Mutation of MSH2. *Cancer Res* 68:2145–2153. doi: 10.1158/0008-5472.CAN-07-6599 [PubMed: 18381419]
29. Desai DC, Lockman JC, Chadwick RB, et al. (2000) Recurrent germline mutation in MSH2 arises frequently de novo. *J Med Genet* 37:646–652 [PubMed: 10978353]
30. Cancer of the Colon and Rectum - Cancer Stat Facts. <https://seer.cancer.gov/statfacts/html/colorect.html>. <https://seer.cancer.gov/statfacts/html/colorect.html> Accessed 28 Sep 2017
31. Ranola JMO, Liu Q, Rosenthal EA, Shirts BH (2017) A comparison of cosegregation analysis methods for the clinical setting. *Fam Cancer*. doi: 10.1007/s10689-017-0017-7
32. U.S. Census Ohio Population. <https://www.census.gov/quickfacts/fact/table/OH/PST045216>. <https://www.census.gov/quickfacts/fact/table/OH/PST045216>. Accessed 28 Sep 2017
33. Plazzer JP, Sijmons RH, Woods MO, et al. (2013) The InSiGHT database: utilizing 100 years of insights into Lynch syndrome. *Fam Cancer* 12:175–180. doi: 10.1007/s10689-013-9616-0 [PubMed: 23443670]
34. Landrum MJ, Lee JM, Riley GR, et al. (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 42:D980–5. doi: 10.1093/nar/gkt1113 [PubMed: 24234437]
35. McCarty DJ, Tull ES, Moy CS, et al. (1993) Ascertainment corrected rates: applications of capture-recapture methods. *Int J Epidemiol* 22:559–565 [PubMed: 8359975]
36. McCrea RS, Morgan BJT (2015) Analysis of Capture-Recapture Data, 1st ed. CRC Press, Boca Raton, FL

37. Perez-Carbonell L, Ruiz-Ponte C, Guarinos C, et al. (2012) Comparison between universal molecular screening for Lynch syndrome and revised Bethesda guidelines in a large population-based cohort of patients with colorectal cancer. *Gut* 61:865–72. doi: 10.1136/gutjnl-2011-300041 [PubMed: 21868491]
38. Marquez E, Geng Z, Pass S, et al. (2013) Implementation of routine screening for Lynch syndrome in university and safety-net health system settings: successes and challenges. *Genet Med* 18:45
39. Hunter JE, Zepp JM, Gilmore MJ, et al. (2015) Universal tumor screening for Lynch syndrome: Assessment of the perspectives of patients with colorectal cancer regarding benefits and barriers. *Cancer* 121:3281–3289. doi: 10.1002/cncr.29470 [PubMed: 26036338]

Figure 1. A plot of the number of families, defined as out to first cousins, we expect to find with the given mutation.

Note that the vast majority, more than 260, of mutations we expect will be present in a small number of families. Also, the error bars given here are the 95% confidence intervals of the estimates.

Table 1.

Estimated number of families for the given family definition along with the 95% confidence intervals. Additionally, the number of years needed to identify 50% and 90% of families if the current identification rate continues.

Family Definition	Estimated Number of Families	95% Confidence Intervals	Years to 50%	Years to 90%
Standard Families (Up to 1 st Cousins)	1693	(420,2966)	20	76
Extended Families (Any Documented Relationship)	883	(397,1368)	9	38
Same Mutation	288	(191,384)	1	13