Published in final edited form as:

Stat Med. 2007 January 15; 26(1): 98–109. doi:10.1002/sim.2460.

A hierarchical non-homogenous Poisson model for meta-analysis of adenoma counts

Carolyn M. Rutter*, Onchee Yu, and Diana L. Miglioretti

Center for Health Studies, Group Health Cooperative, 1730 Minor Ave, Ste 1600 Seattle WA, 98101, U.S.A

Carolyn M. Rutter: rutter.c@ghc.org

SUMMARY

We use a hierarchical model for a meta-analysis that combines information from autopsy studies of adenoma prevalence and counts. The studies we included reported findings using a variety of adenoma prevalence groupings and age categories. We use a non-homogeneous Poisson model for multinomial bin probabilities. The Poisson model allows risk to depend on age and sex, and incorporates extra-Poisson variability. We evaluate model fit using the posterior predicted distribution of adenoma prevalence reported by the studies included in our analyses and validate our model using adenoma prevalence reported by more recent colonoscopy studies. For 1990, the estimated adenoma prevalence among Americans at age 60 is 40.3 per cent for men compared to 29.2 per cent for women.

Keywords

Bayesian estimation; goodness of fit; model validation

1. INTRODUCTION

Colorectal cancer is the third leading cause of cancer death in the U.S. [1]. The vast majority of colorectal cancers arise from adenomas [2]. Colorectal cancer screening focuses on the detection of both cancers and adenomas. The effectiveness of colorectal cancer screening is partly attributable to interruption of the disease process through removal of precursor lesions. The cost-effectiveness of screening strategies also depends on the distribution of adenomas across individuals because removal of adenomas, even those that would not progress to cancer, adds cost to screening. Therefore, good adenoma risk estimates are an important component of microsimulation models used to assess the impact of screening on colorectal cancer incidence and mortality as well as economic models used to estimate the costs of screening programs.

Most of the information about adenoma risk in asymptomatic individuals comes from autopsy studies (described in Section 2). Autopsy studies allow a complete and thorough

examination of the colon and rectum. Because adenomas do not cause death, and relatively few adenomas progress to cancer [3], autopsy studies provide good information about adenoma prevalence. However, because many of these studies were carried out in the 1970s and the early 1980s, there is concern that they may not represent current risk related to overall lifestyle changes such as a reduction in smoking. There is also some concern that autopsied individuals represent a biased sample of deaths. Colonoscopy studies provide a more recent alternative source of information about adenoma risk. However, colonoscopy is not a perfect test. The miss rate for adenomas has been estimated to range from 6 per cent for adenomas 10 mm and larger up to 37 per cent for adenomas under 5 mm [4, 5]. In addition, these studies are limited to individuals who are willing to participate in screening studies, and must exclude individuals who have undergone prior endoscopic evaluation.

Our primary aim is estimation of adenoma risk for North American and European populations. We combine data from 14 autopsy studies to estimate adenoma risk and use data from colonoscopy studies to assess goodness of fit. In the following section, we describe the autopsy studies included and excluded from our meta-analysis. In Section 3 we describe the model we used to combine information across studies and general approaches to assessing goodness of fit. In Section 4 we describe our results: parameter estimates, prevalence estimates, and goodness of fit. We conclude with summary remarks in Section 5.

2. DATA

We searched the PubMed database to identify autopsy studies of adenoma prevalence that were published since 1960. The key words used in the search included ‘adenoma prevalence’ or ‘adenomatous polyps’ combined with ‘autopsy’ or ‘necropsy’. The references from each potential article found in the PubMed database were then reviewed for additional relevant articles. We found a total of 28 articles that reported adenoma prevalence rates from 30 autopsy studies. (One article reported three separate autopsy studies [6].)

Of the 30 studies identified, 16 were excluded from our analyses. Most of the excluded studies examined populations that would not reflect adenoma risk in the North American and European populations [7–17]. These 11 studies primarily examined South American and Asian populations that generally have a lower incidence of colorectal cancer relative to populations in the North America and Europe. The crude annual incidence of colorectal cancer among these populations during the autopsy years ranged from 3 to 23 per 100 000 [18–20]. The reported adenoma rates ranged from 0 to 30 per cent. One study was excluded because it specifically focused on recent Japanese immigrants to the U.S. [21]. Other reasons for exclusion included reporting of previous data [6], examination that was restricted to the colon [22], and lack of information about age-specific rates [23, 24].

Data from the remaining 14 autopsy studies were included in the meta-analysis. As shown in Table I, these studies were published between 1961 and 2001, representing populations in four geographic regions: three were from the U.S. [25, 26, 29], six were from Northern Europe [6, 28, 30, 31, 33], three were from Southern or Eastern Europe [32, 35, 36], and two were from Oceania [27, 34]. Among these populations, the crude annual incidence of colorectal cancer during the autopsy years ranged from 11 to 63 per 100 000 [18–20]. Crude

annual incidence was not available for the Greek population studied by Paspatis *et al.* [36], though the estimated age standardized rate (ASR, with standardization to the world population) in the autopsy year was 13.8 per 100 000 [37]. ASR's were not available for populations studied by Blatt [25], Chapman [26], Rickert and colleagues [29] or Vatn and Stalsberg [30]. Across the remaining studies, ASR's in autopsy years ranged from 20.1 to 43.6 per 100 000. Observed adenoma prevalence rates ranged from 10.5 to 51 per cent.

All 14 studies included in our analysis reported adenoma prevalence stratified by age. The age categorizations used to report adenoma rates varied across studies (e.g. <55, 55–74, 75+; and 30–49, 50–69, 70–89). Because studies did not provide information about the age distribution of their sample, we used the Human Life-Table Database (<http://www.lifetable.de>) to estimate the median age of decedents in each age group. For each study, we used life tables in the average autopsy year and the country corresponding to the study population. We estimated the average autopsy year for each study by assuming uniform data collection across the reported study period.

As shown in Table I, three studies contributed count information by age and sex; four studies contributed count data by age; six studies contributed prevalence data by age and sex; and one study contributed prevalence data by age but not sex. Because we were particularly interested in estimating the frequency of multiple adenomas, we chose to use count information when it was available, even if it meant that we could not use sex information. The seven studies that contributed count data reported results using various categorizations: (0, 1, 2+)[26, 33, 35], (0, 1, 2, 3, 4, 5+)[28, 29, 31], (0, 1, 2–4, 5–9, 10–11) [32]. A total of 5183 individuals were included in the meta-analysis.

3. META-ANALYTIC MODEL FOR ADENOMA RISK

We combine information across studies using a hierarchical model that assumes a multinomial distribution for adenoma counts with bin probabilities based on an individual-level Poisson model. Accurate modelling of count data required inclusion of subject-specific random effects to build extra-Poisson variability into the model.

Let y_{ijk} be the (partially observable) number of adenomas found in the i th individual in the j th age-sex group in the k th study. We assume that y_{ijk} follows a Poisson distribution with mean r_{ijk} . Let x_{ijk} be the observed grouped adenoma outcome based on y_{ijk} . The level I model describes the distribution of x_{ijk} :

$$x_{ijk} \sim \text{Multinomial}(p_{ijk}^{(1)}, p_{ijk}^{(2)}, \dots, p_{ijk}^{(m_k)}, 1)$$

where m_k is the number of adenoma count groups for the k th study. The multinomial probabilities, $p_{ijk}^{(n)}$, are based on the Poisson distribution. For example, for a study that reported counts in groupings of 0, 1, and 2 or more adenomas, the multinomial probabilities are: $p_{ijk}^{(1)} = \exp(-r_{ijk})$, $p_{ijk}^{(2)} = r_{ijk} \exp(-r_{ijk})$ and $p_{ijk}^{(3)} = 1 - (p_{ijk}^{(1)} + p_{ijk}^{(2)})$.

The level II model describes the distribution of the individual-level risks underlying the multinomial probabilities based on a non-homogeneous Poisson process [38]. These risks depend on age and sex and include a study-specific random effect. We include a time effect which is defined as the average autopsy year, to allow adenoma risk to change over time as a result in changes in population risk (e.g. declines in rates of smoking). We include three indicators of geographic region to allow risk to differ from U.S. populations (our reference group), i.e. indicators for Northern Europe, Southern or Eastern Europe, and Oceania. Under this model, the risk of developing an adenoma for an individual in the k th study at age $_{jk}$ is $\exp(\theta_{0k} + \theta_1 \text{sex}_{jk} + \theta_2 \text{time}_k + \theta_3 \text{region}_k + \theta_4 \text{age}_{jk})$, where θ_3 is a 3×1 parameter vector associated with regions and region_k is a 1×3 region indicator vector. Under this model, the number of adenomas at age $_{jk}$ has a Poisson distribution with mean

$\int_0^{\text{age}_{jk}} \exp(\theta_{0k} + \theta_1 \text{sex}_{jk} + \theta_2 \text{time}_k + \theta_3 \text{region}_k + \theta_4 u) du$ which is equal to

$$\exp(\theta_{0k} + \theta_1 \text{sex}_{jk} + \theta_2 \text{time}_k + \theta_3 \text{region}_k) (e^{\text{age}_{jk} \theta_4} - 1) / \theta_4 \quad (1)$$

We model these Poisson means using a log-Normal distribution:

$$\ln(r_{ijk}) \stackrel{\text{iid}}{\sim} N(\theta_{0k} + \theta_1 \text{sex}_{jk} + \theta_2 \text{time}_k + \theta_3 \text{region}_k + \ln\{(e^{\text{age}_{jk} \theta_4} - 1) / \theta_4\}, \sigma_1^2)$$

where σ_1 models random between-individual variability, beyond systematic effects of age, sex, and the time and region of study. Recall that age $_{jk}$ are estimated median ages for decedents (in years) in the j th study's k th age group. The covariate time_k is the midpoint of the study recruitment period (in years), centred around 1980. The covariate sex_{jk} was coded as -1 for females and 1 for males. We included studies that reported overall rates, by coding $\text{sex}_{jk} = 0$. Any overall differences in the sex distribution of these studies is accounted for in the study-level random effect.

The level III model describes the distribution of study-specific random effects:

$$\theta_{0k} \stackrel{\text{iid}}{\sim} N(\Theta_0, \sigma_2^2)$$

where Θ_0 is the mean and σ_2 is the standard deviation of the study-specific random effects.

The level IV model describes prior distributions for remaining model parameters:

$$\begin{aligned} \Theta_0, \theta_1, \theta_2, \theta_{31}, \theta_{32}, \theta_{33} &\stackrel{\text{iid}}{\sim} N(0, 10) \\ \theta_4, \sigma_1, \sigma_2 &\stackrel{\text{iid}}{\sim} \text{Uniform}(0.001, 20) \end{aligned}$$

Note that the level II model for log-risk implies that $(e^{\text{age}_{jk} \theta_4} - 1) / \theta_4 > 0$. This condition is satisfied if either $(e^{\text{age}_{jk} \theta_4} > 1$ and $\theta_4 > 0$) or $(e^{\text{age}_{jk} \theta_4} < 1$ and $\theta_4 < 0$). If age is coded as a positive covariate, then these conditions are satisfied when $\theta_4 > 0$. Because θ_4 cannot equal zero, and because we know empirically that the risk of adenomas increases with age, we restrict θ_4 to be strictly positive as part of the level IV model.

The Normal (0, 10) distribution was chosen as a prior for Θ_0 , θ_1 , θ_2 , and components of θ_3 because it is relatively flat across the range of typical Poisson regression parameter values (and therefore relatively non-informative). The Uniform (0.001, 20) distribution was chosen as the prior for θ_4 and for the standard deviation of individual and study-specific effects (σ_1 , σ_2) because it is relatively non-informative over a positive range.

3.1. Model estimation

We estimated model parameters using Markov chain Monte Carlo (MCMC) simulation carried out using WinBugs software [39]. MCMC simulates draws from the joint posterior of model parameters given the autopsy data. We ran four chains at dispersed starting points and allowed 2500 iterations for burn-in. An additional 3500 iterations were used for estimation, with each chain thinned at five iterations. We assessed convergence using trace plots in combination with the Gelman and Rubin statistic for multiple chains [40]. Gelman and Rubin statistics ranged from 1.03 to 1.05.

3.2. Assessment of model fit

We used posterior predicted values to allow more direct interpretation of model results and to assess model fit to both the autopsy data used to estimate model parameters and colonoscopy data held out for validation. Posterior predicted values were estimated by simulating study outcomes given the simulated draws from the posterior distribution for model parameters, as described by Gilks and colleagues [41]. Specifically, let the superscript $*$ denote an MCMC simulated draw from the joint posterior distribution. For each MCMC draw after convergence, we simulated a study's outcomes by drawing one study random effect from a $N(\Theta_0^*, \sigma_2^*)$ and n_s individual random effects from a $N(0, \sigma_1^*)$ distribution, where n_s is the number of individuals in the study. We then used equation (1) with study-level information (i.e. year, country, age and sex distribution) to simulate adenoma counts. Finally, we collapsed these adenoma counts into tables reported by the study.

We estimated the posterior predicted distribution of adenoma counts for men and women in the U.S. using a similar approach. At each iteration we drew a single study-level random effect from a $N(0, \sigma_2^*)$ distribution. Because these are population estimates, we drew 500 000 individuals within each 5-year age group, each with a $N(0, \sigma_1^*)$ random effect and a randomly assigned age. We report rates separately for men and women, and assume uniform age distributions within the 5-year age groups we use for reporting.

We assessed model fit using the posterior predicted distributions, comparing the observed values for each study to posterior predicted values. We examined both plots of predicted *versus* observed rates and posterior predictive p -values [42]. p -values estimate the posterior predicted probability of seeing a value that is greater than the observed value.

3.2.1. Validation data—We also assessed model fit using observed adenoma prevalence from four recent and relatively high-profile colonoscopy studies in the U.S., comparing observed adenoma prevalence to model-predicted prevalence [43–46]. We focus on these studies because of their recency and their quality, and because they provided enough

information (marginal age and sex distributions) to allow model prediction of adenoma prevalence. These studies are not intended to represent a complete collection of colonoscopy screening studies.

Each study focused on asymptomatic patients with no recent history of sigmoidoscopy, colonoscopy, or barium enema. Exclusion of individuals with a history of endoscopic examination is important because adenomas are removed during these procedures. DiSario and colleagues [43] excluded individuals with any prior sigmoidoscopy, colonoscopy, or barium enema. Other studies excluded individuals with sigmoidoscopy, colonoscopy, or barium enema within the last 3 years [44], or last 10 years [45]. Pickhardt and colleagues [46] did not exclude patients based on prior sigmoidoscopy, but did exclude individuals who underwent barium enema within the last 5 years or colonoscopy within the last 10 years. Additional information about each study is provided in Table IV.

For model prediction, we used the reported age and sex distribution and the midpoint of the recruitment period for the time covariate associated with each study. We estimated two predicted values, with corresponding p -values, for each study. The first is the model predicted adenoma prevalence, which implicitly assumes that colonoscopy is 100 per cent sensitive. We also estimated model predicted adenoma prevalence assuming that colonoscopy is 85 per cent sensitive, with sensitivity applied at the adenoma-level. We chose 85 per cent sensitivity because colonoscopy miss rates increase from 6 per cent for adenomas that are 10 mm or larger, up to 37 per cent for small adenomas (<5 mm) [4, 5], and because most adenomas (>60 per cent) are 5 mm or smaller and relatively few (<10 per cent) are 10 mm or larger [46].

4. RESULTS

Table II shows parameter estimates with their 95 per cent credible intervals. The credible intervals for the geographic effects include zero. We retained the geographic effects (θ_{31} , θ_{32} , θ_{33}) in the model because they were considered important predictors, which may not have attained statistical significance because our analyses included few studies from each region. The credible interval for the time effect (θ_2) also includes zero. We retained the time effect in the model because without it random study effects exhibited drift over time.

Because model parameters are difficult to interpret directly, we estimate the posterior predicted distribution of adenoma counts for men and women in the U.S. ($\theta_{31} = \theta_{32} = \theta_{33} = 0$) during calendar year 1990 at ages 50, 60, and 70, shown in Table III. For both men and women, both the estimated overall probability of at least one adenoma and the expected number of adenomas increase with age. Overall, men were more likely to have an adenoma than women. For example, at age 60 the estimated adenoma prevalence for men was 40.3 per cent *versus* 29.2 per cent for women. In Appendix A we provide a table of adenoma prevalence rates in the U.S. during 1978, based on the posterior predictive distribution, that can be used as input for colorectal cancer models. (The year 1978 was chosen because this year is being used to compare CISNET models, and because it is the earliest year that U.S. colorectal cancer incidence data is available.)

4.1. Goodness of fit

Our model shows good fit to the autopsy prevalence data (Figure 1), though there is some evidence of regression to the mean in studies reporting low adenoma prevalence (below 20 per cent). This relationship holds for both men and women. There was some tendency for our model to predict higher than observed prevalence at younger ages, when prevalence is low.

Using our model we could accurately predict adenoma prevalence rates found in four recent colonoscopy studies. Table IV shows the observed adenoma prevalence rates for each study and our predicted model rates. None of the associated p -values demonstrate a lack of fit to these data. Predicted values that incorporate a 15 per cent colonoscopy miss rate are closer to observed rates than values that assume colonoscopy is 100 per cent sensitive. One study also reported that 16 per cent of individuals in their sample had two or more adenomas [43]. This was close to our model based prediction of 23.6 per cent ($p = 0.68$, based on a 15 per cent miss rate).

5. DISCUSSION

We use a hierarchical multinomial model with cell probabilities based on an underlying non-homogeneous Poisson model to combine information across studies that reported findings using a variety of adenoma prevalence groupings. Our model showed good fit, both to the autopsy data used for estimation and to more recent colonoscopy studies of prevalence used for validation. Thus, our models should provide good estimates of age- and sex-specific adenoma prevalence in a population with an intermediate to high incidence of colorectal cancer.

Both age and sex were associated with the risk of developing an adenoma. As expected, adenoma prevalence was greater for men than women. Our model specified increasing adenoma risk with age, largely because this is a well-known association. Our model enabled us to use autopsy data to estimate the magnitude of the age effect. However, goodness of fit analyses showed some potential for overestimation of adenoma risk among younger individuals (<30 years old), who tend to have very low adenoma prevalence. We also had the least information for young age groups, and age categories for younger decedents tended to be broad. Therefore, care must be taken when applying our model results to younger age groups.

We found some evidence of decreasing adenoma prevalence over time. There was also some indication that studies in U.S. populations might have a higher risk of developing an adenoma, even after controlling for the timing of the study, though there was not enough data to provide statistical evidence that U.S. populations are at higher risk. Because estimates are based on grouped outcomes, we can say little about between-individual variability. The pseudo-individual effect was included primarily as a simple way to capture extra-Poisson variability. Because we did not have each individual's age, estimates of between-individual variability incorporate variability in age within each age category.

Given the number of parameters in our model, it is not surprising that we found evidence of good model fit to the autopsy data. Because of concerns about over-fitting our data, as well as concerns about the generalizability of autopsy data and use of median age information derived from mortality data, we validated our model using independent data from four colonoscopy trials. Our model showed good fit to colonoscopy data. When we unrealistically assumed 100 per cent sensitivity for colonoscopy, our model tended to overestimate the adenoma prevalence relative to rates found on colonoscopy, though estimates were still relatively close to observed values. Model prediction improved when we assumed a more realistic 15 per cent miss rate for colonoscopy (on the adenoma-level).

The estimates of adenoma risk for the U.S. population provided by our model (as shown in Appendix A) can be used as inputs for cost-effectiveness studies and micro-simulation models of colorectal cancer screening. Microsimulation models simulate individual event histories to examine the impact of interventions, such as cancer screening, and are the focus of Cancer Intervention and Surveillance Modeling Network (CISNET), a research network sponsored by the National Cancer Institute (<http://cisnet.cancer.gov/>). Microsimulation models are becoming more important because of the difficulties inherent in clinical trials of cancer screening. Because cancer is rare, such trials require very large samples that must be followed over many years to estimate simple outcomes such as sensitivity and specificity. Effects of screening on cancer mortality require similarly large samples and very long follow-up periods. Such trials are virtually impossible to conduct once a screening modality is shown to have good sensitivity and specificity for cancer detection. The adenoma risk model we developed takes a step toward improving the inputs used in models for colorectal cancer screening.

Acknowledgments

Contract/grant sponsor: National Cancer Institute; contract/grant number: U01 CA97427

References

1. Jemal A, Tiwari RC, Murray T, Ghafoor A, Samuels A, Ward E, Feuer EJ, Thun MJ. Cancer Statistics. *CA-Cancer Journal for Clinicians*. 2004; 54:8–29.
2. Day DW, Morson BC. The adenoma-carcinoma sequence. *Major Probl Pathol*. 1978; 10:58–71. [PubMed: 359943]
3. Nusko G, Mansmann U, Partzsch U, Altendorf-Hofmann A, Groitl H, Wittekind C, Ell C, Hahn EG. Invasive carcinoma in colorectal adenomas: multivariate analysis of patient and adenoma characteristics. *Endoscopy*. 1997; 29:626–631. [PubMed: 9360872]
4. Hixson L, Fennerty M, Sampliner R, McGee D, Garewal H. Prospective study of the frequency and size distribution of polyps missed by colonoscopy. *Journal of the National Cancer Institute*. 1990; 82:1769–1772. [PubMed: 2231773]
5. Rex D, Cutler C, Lemmel G, Rahmani E, Clark D, Helper D, Lehman G, Mark D. Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. *Gastroenterology*. 1997; 112:24–28. [PubMed: 8978338]
6. Clark JC, Collan Y, Eide TJ, Esteve J, Ewen S, Gibbs NM, Jensen OM, Koskela E, MacLennan R, Simpson JG, Stalsberg H, Zaridze DG. Prevalence of polyps in an autopsy series from areas with varying incidence of large-bowel cancer. *International Journal of Cancer*. 1985; 36:179–186.
7. Bremner CG, Ackerman LV. Polyps and carcinoma of the large bowel in the South African Bantu. *Cancer*. 1970; 26:991–999. [PubMed: 5476799]

8. Cajucom CC, Barrios GG, Cruz L, Varin C, Herrera L. Prevalence of colorectal polyps in Filipinos. An autopsy study. *Diseases of the Colon and Rectum*. 1992; 35:676–680. [PubMed: 1611956]
9. Correa P, Duque E, Cuello C, Haenszel W. Polyps of the colon and rectum in Cali, Colombia. *International Journal of Cancer*. 1972; 9:86–96.
10. Sato E. Adenomatous polyps of large intestine in autopsy and surgical material. *Gann*. 1974; 65:295–306. [PubMed: 4426476]
11. Sato E, Ouchi A, Sasano N, Ishidate T. Polyps and diverticulosis of large bowel in autopsy population of Akita prefecture, compared with Miyagi. High risk for colorectal cancer in Japan. *Cancer*. 1976; 37:1316–1321. [PubMed: 816449]
12. Haghighi P, Nasr K, Mohallateh EA, Ghassemi H, Sadri S, Nabizadeh I, Sheikholeslami MH, Mostafavi N. Colorectal polyps and carcinoma in Southern Iran. *Cancer*. 1977; 39:274–278. [PubMed: 832242]
13. Marigo C, Correa P, Haenszel W. Cancer and 'cancer related' colorectal lesions in Sao Paulo, Brazil. *International Journal of Cancer*. 1978; 22:645–654.
14. Restrepo C, Correa P, Duque E, Cuello C. Polyps in a low-risk colonic cancer population in Colombia, South America. *Diseases of the Colon and Rectum*. 1981; 24:29–36. [PubMed: 7472099]
15. Coode PE, Chan KW, Chan YT. Polyps and diverticula of the large intestine: A necropsy survey in Hong Kong. *Gut*. 1985; 26:1045–1048. [PubMed: 4054702]
16. Lee YS. Adenomas, metaplastic polyps and other lesions of the large bowel: An autopsy survey. *Ann Acad Med Singapore*. 1987; 16:412–420. [PubMed: 3435006]
17. Chavez-Macias LG, Jessurun J, Mendez Sanchez N. Prevalence of adenomatous and hyperplastic polyps of the colon in the population of the General Hospital of Mexico. An autopsy study. *Revista de Investigacion Clinica*. 1990; 42:193–197. [PubMed: 2270365]
18. Waterhouse, J.; Muir, C.; Correa, P.; Powell, J. *Cancer Incidence in Five Continents*. Vol. III. Lyon: IARC; 1976.
19. Waterhouse, J.; Muir, C.; Shanmugaratnam, K.; Powell, J. *Cancer Incidence in Five Continents*. Vol. IV. Lyon: IARC; 1982.
20. Muir, C.; Waterhouse, J.; Mack, T.; Powell, J.; Whelan, S. *Cancer Incidence in Five Continents*. Vol. V. Lyon: IARC; 1987.
21. Stemmermann GN, Yatani R. Diverticulosis and polyps of the large intestine. A necropsy study of Hawaii Japanese. *Cancer*. 1973; 31:1260–1270. [PubMed: 4705166]
22. Arminski TC, McLean DW. Incidence and distribution of adenomatous polyps of the colon and rectum based on 1000 autopsy examinations. *Diseases of the Colon and Rectum*. 1964; 7:249–261. [PubMed: 14176135]
23. Ekelund G. On cancer and polyps of colon and rectum. *Acta Pathologica et Microbiologica Scandinavica*. 1963; 59:165–170. [PubMed: 14083758]
24. Correa P, Strong JP, Reif A, Johnson WD. The epidemiology of colorectal polyps. Prevalence in New Orleans and international comparisons. *Cancer*. 1977; 39:2258–2264. [PubMed: 870171]
25. Blatt LJ. Polyps of the colon and rectum: incidence and distribution. *Diseases of the Colon and Rectum*. 1961; 4:277–282.
26. Chapman I. Adenomatous polypi of large intestine: incidence and distribution. *Annals of Surgery*. 1963; 157:223–226. [PubMed: 14020146]
27. Hughes LE. The incidence of benign and malignant neoplasms of the colon and rectum: a post-mortem study. *Australian and New Zealand Journal of Surgery*. 1968; 38:30–35. [PubMed: 5245830]
28. Eide TJ, Stalsberg H. Polyps of the large intestine in Northern Norway. *Cancer*. 1978; 42:2839–2848. [PubMed: 728878]
29. Rickert RR, Auerbach O, Garfinkel L, Hammond EC, Frasca JM. Adenomatous lesions of the large bowel. An autopsy survey. *Cancer*. 1979; 43:1847–1857. [PubMed: 445371]
30. Vatn MH, Stalsberg H. The prevalence of polyps of the large intestine in Oslo: an autopsy study. *Cancer*. 1982; 49:819–825. [PubMed: 7055790]

31. Williams AR, Balasooriya BAW, Day DW. Polyps and cancer of the large bowel: a necropsy study in Liverpool. *Gut*. 1982; 23:835–842. [PubMed: 7117903]
32. Bombi JA. Polyps of the colon in Barcelona, Spain. *Cancer*. 1988; 61:1472–1476. [PubMed: 3345499]
33. Johannsen LGK, Momsen O, Jacobsen NO. Polyps of the large intestine in Aarhus, Denmark. An autopsy study. *Scandinavian Journal of Gastroenterology*. 1989; 24:799–806. [PubMed: 2799283]
34. Jass JR, Young PJ, Robinson EM. Predictors of presence, multiplicity, size and dysplasia of colorectal adenomas. A necropsy study in New Zealand. *Gut*. 1992; 33:1508–1514. [PubMed: 1452076]
35. Szczepanski W, Urban A, Wierchowski W. Colorectal polyps in autopsy material. Part I. Adenomatous polyps. *Pat Pol*. 1992; 43:79–85. [PubMed: 1294936]
36. Paspatis GA, Papanikolaou N, Zois E, Michalodimitrakakis E. Prevalence of polyps and diverticulosis of the large bowel in the Cretan population. An autopsy study. *International Journal of Colorectal Disease*. 2001; 16:257–261. [PubMed: 11515686]
37. Black R, Bray F, Ferlay J, Parkin D. Cancer incidence and mortality in the European union: cancer registry data and estimates of national incidence for 1990. *European Journal of Cancer*. 1997; 33:1075–1107. [PubMed: 9376190]
38. Cox, DR.; Miller, HD. *The Theory of Stochastic Processes*. Chapman and Hall; New York: 1965.
39. Spiegelhalter, D.; Thomas, A.; Best, N. *WinBUGS Version 1.2 User Manual*. MRC Biostatistics Unit; 1999.
40. Gelman A, Rubin D. Inference from iterative simulation using multiple sequences. *Statistical Sciences*. 1992; 7:457–511.
41. Gilks, WR.; Richardson, S.; Spiegelhalter, DJ. *Markov Chain Monte Carlo in Practice*. Chapman and Hall; New York: 1996.
42. Carlin J, Wolfe R. A case study on the choice, interpretation and checking of multilevel models for binary outcomes. *Biostatistics*. 2001; 2:397–415. [PubMed: 12933632]
43. DiSario JA, Foutch PG, Mai HD, Pardy K, Manne RK. Prevalence and malignant potential of colorectal polyps in asymptomatic, average-risk men. *American Journal of Gastroenterology*. 1991; 86:941–945. [PubMed: 1858757]
44. Rex DK, Lehman GA, Ulbright TM, Smith JJ, Pound DC, Hawes RH, Helper DJ, Wiersema MJ, Langefeld CD, Li W. Colonic neoplasia in asymptomatic persons with negative fecal occult blood tests: influence of age, gender, and family history. *American Journal of Gastroenterology*. 1993; 88:825–831. [PubMed: 8503374]
45. Lieberman DA, Weiss DG, Bond JH, Anhen DJ, Garewal H, Chejfec G. Use of colonoscopy to screen asymptomatic adults for colorectal cancer. *New England Journal of Medicine*. 2000; 343:162–168. [PubMed: 10900274]
46. Pickhardt PJ, Choi JR, Hwang I, Butler JA, Puckett ML, Hildebrandt HA, Wong RK, Nugent PA, Mysliwiec PA, Schindler WR. Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. *New England Journal of Medicine*. 2003; 349:2191–2200. [PubMed: 14657426]

APPENDIX A: ESTIMATED ADENOMA RATES

Tables AI show posterior predictive probabilities of 0, 1, 2 and 3 or more adenomas for an American population by gender for 5-year age groups during 1978. The year 1978 was and AII chosen because it precedes diffusion of colorectal cancer screening tests and is being used to compare CISNET microsimulation models. Estimates were calculated by marginalizing over the 5-year age groups, assuming a uniform age distribution. Estimated means are shown with 95 per cent credible intervals in parenthesis.

Table AI

Estimated adenoma prevalence among American men.

Age	Pr (0 adenomas)	Pr (1 adenoma)	Pr (2 adenomas)	Pr (3+ adenomas)
25–29	0.81 (0.58, 0.99)	0.12 (0.03, 0.21)	0.03 (0.00, 0.08)	0.04 (0.00, 0.13)
30–34	0.78 (0.52, 0.98)	0.13 (0.05, 0.22)	0.04 (0.00, 0.10)	0.05 (0.00, 0.17)
35–39	0.74 (0.46, 0.98)	0.14 (0.06, 0.22)	0.05 (0.00, 0.11)	0.07 (0.00, 0.22)
40–44	0.71 (0.41, 0.96)	0.16 (0.07, 0.22)	0.06 (0.01, 0.11)	0.08 (0.00, 0.27)
45–49	0.67 (0.35, 0.95)	0.17 (0.08, 0.22)	0.06 (0.01, 0.12)	0.10 (0.00, 0.31)
50–54	0.63 (0.30, 0.93)	0.17 (0.09, 0.22)	0.07 (0.02, 0.12)	0.13 (0.00, 0.37)
55–59	0.59 (0.26, 0.90)	0.18 (0.11, 0.22)	0.08 (0.02, 0.12)	0.15 (0.00, 0.41)
60–64	0.55 (0.21, 0.88)	0.18 (0.11, 0.22)	0.08 (0.03, 0.13)	0.18 (0.00, 0.47)
65–69	0.51 (0.16, 0.85)	0.19 (0.12, 0.22)	0.09 (0.04, 0.13)	0.21 (0.00, 0.52)
70–74	0.47 (0.12, 0.81)	0.19 (0.12, 0.22)	0.09 (0.04, 0.13)	0.25 (0.00, 0.57)
75–79	0.43 (0.09, 0.78)	0.19 (0.12, 0.23)	0.10 (0.05, 0.13)	0.28 (0.00, 0.63)
80–84	0.40 (0.07, 0.73)	0.18 (0.11, 0.22)	0.10 (0.05, 0.13)	0.32 (0.01, 0.67)
85–89	0.36 (0.05, 0.68)	0.18 (0.09, 0.22)	0.10 (0.06, 0.13)	0.36 (0.02, 0.72)
90–94	0.32 (0.03, 0.64)	0.17 (0.08, 0.22)	0.10 (0.06, 0.13)	0.40 (0.04, 0.78)
95–99	0.29 (0.01, 0.60)	0.17 (0.07, 0.22)	0.10 (0.06, 0.13)	0.44 (0.06, 0.82)
100–104	0.26 (0.01, 0.56)	0.16 (0.06, 0.22)	0.10 (0.05, 0.13)	0.48 (0.10, 0.87)

Table AII

Estimated adenoma prevalence among American women.

Age	Pr (0 adenomas)	Pr (1 adenoma)	Pr (2 adenomas)	Pr (3+ adenomas)
25–29	0.88 (0.70, 1.00)	0.09 (0.01, 0.18)	0.02 (0.00, 0.06)	0.02 (0.00, 0.06)
30–34	0.85 (0.64, 1.00)	0.10 (0.02, 0.20)	0.03 (0.00, 0.08)	0.02 (0.00, 0.10)
35–39	0.82 (0.60, 0.99)	0.11 (0.03, 0.21)	0.03 (0.00, 0.08)	0.03 (0.00, 0.12)
40–44	0.79 (0.54, 0.99)	0.12 (0.04, 0.21)	0.04 (0.00, 0.09)	0.04 (0.00, 0.18)
45–49	0.76 (0.49, 0.98)	0.14 (0.05, 0.22)	0.04 (0.00, 0.10)	0.06 (0.00, 0.20)
50–54	0.73 (0.44, 0.97)	0.15 (0.06, 0.22)	0.05 (0.00, 0.11)	0.07 (0.00, 0.25)
55–59	0.70 (0.39, 0.96)	0.16 (0.07, 0.22)	0.06 (0.01, 0.12)	0.09 (0.00, 0.28)
60–64	0.66 (0.35, 0.95)	0.17 (0.08, 0.22)	0.06 (0.01, 0.12)	0.11 (0.00, 0.32)
65–69	0.63 (0.30, 0.93)	0.17 (0.09, 0.22)	0.07 (0.02, 0.12)	0.13 (0.00, 0.37)
70–74	0.59 (0.26, 0.91)	0.18 (0.10, 0.22)	0.08 (0.02, 0.12)	0.15 (0.00, 0.41)
75–79	0.55 (0.21, 0.88)	0.18 (0.11, 0.22)	0.08 (0.03, 0.13)	0.18 (0.00, 0.47)
80–84	0.52 (0.17, 0.85)	0.19 (0.12, 0.22)	0.09 (0.04, 0.13)	0.21 (0.00, 0.51)
85–89	0.48 (0.13, 0.82)	0.19 (0.12, 0.22)	0.09 (0.04, 0.13)	0.24 (0.00, 0.56)
90–94	0.44 (0.10, 0.78)	0.19 (0.12, 0.23)	0.10 (0.05, 0.13)	0.27 (0.00, 0.61)
95–99	0.40 (0.07, 0.74)	0.18 (0.11, 0.22)	0.10 (0.05, 0.13)	0.31 (0.01, 0.67)
100–104	0.37 (0.05, 0.70)	0.18 (0.10, 0.22)	0.10 (0.06, 0.13)	0.35 (0.02, 0.71)

Figure 1.
Observed and posterior estimated adenoma prevalence for each autopsy study included in the meta-analysis.

Autopsy studies included in meta-analysis.

Table 1

Source	Study location	Adenoma data analysed	Age range (years)	Sample size	Adenoma prevalence (%)
Blatt [25]	U.S., New York	Prevalence, by age and sex	30–100	446	38.6
Chapman [26]	U.S., New York	Counts, by age	10	443	51.0
Hughes [27]	Australia, Queensland	Prevalence, by age	*	200	20.0
Eide and Stralsberg [28]	Norway, Tromsø	Counts, by age and sex	20	280	37.1
Rickert <i>et al.</i> [29]	U.S., New Jersey	Counts, by age	20–102	518	46.9
Vam and Stalsberg [30]	Norway, Oslo	Prevalence, by age and sex	*	445	33.3
Williams <i>et al.</i> [31]	U.K., Liverpool	Counts, by age	*	365	33.2
Clark <i>et al.</i> [6]	Scotland, Aberdeen	Prevalence, by age and sex	20	200	38.2
Clark <i>et al.</i> [6]	Finland, Kuopio	Prevalence, by age and sex	20	200	10.5
Bombi [32]	Spain, Barcelona	Counts, by age and sex	8–92	212	21.7
Johannsen <i>et al.</i> [33]	Denmark, Aarhus	Counts, by age	30–89	336	22.0
Jass <i>et al.</i> [34] [†]	New Zealand	Prevalence, by age and sex	10	303	23.8
Szczepanski <i>et al.</i> [35]	Poland, Krakow	Counts, by age and sex	2–99	733	38.2
Paspatis <i>et al.</i> [36]	Greece, Crete	Prevalence, by age and sex	16–93	502	14.5

* All available autopsies were used. The observed age range was 0, or was not reported.

[†] Analysis was restricted to data on 303 Caucasians only.

Table II

Parameter estimates from the meta-analytic model.

Parameter	Mean	95% Credible interval
Θ_0 : Intercept for log-risk	-6.66	(-7.89, -5.55)
θ_1 : Difference in log-risk for men <i>versus</i> women	0.32	(0.22, 0.41)
θ_2 : Annual change in log-risk by year	-0.013	(-0.069, 0.042)
θ_{31} : Difference in log-risk for N Europe <i>versus</i> U.S. populations	-0.83	(-2.04, 0.44)
θ_{32} : Difference in log-risk for S Europe <i>versus</i> U.S. populations	-0.84	(-2.39, 0.86)
θ_{33} : Difference in log-risk for Oceania <i>versus</i> U.S. populations	-1.14	(-2.58, 0.32)
θ_4 : Increase in log-risk with age (years)	0.038	(0.031, 0.045)
θ_1 : Within-study (extra-Poisson) standard deviation	1.48	(1.36, 1.62)
θ_2 : Between study standard deviation	0.77	(0.39, 1.20)

Table III

Predicted distribution of adenoma counts by sex and age in the U.S. in 1990.

Number of adenomas	Male			Female		
	Age 50	Age 60	Age 70	Age 50	Age 60	Age 70
0	67.5	59.8	51.8	77.3	70.8	63.7
1	16.9	18.5	19.3	13.8	16.0	17.8
2	6.3	7.8	9.2	4.2	5.6	7.1
3	3.1	4.1	5.2	1.8	2.6	3.6
4	1.7	2.4	3.2	0.9	1.4	2.1
5+	4.5	7.3	11.3	1.9	3.5	5.8

Table IV

Adenoma prevalence at screening colonoscopy: observed and predicted rates with posterior predicted p -values.

	Source			
	DiSario <i>et al.</i>	Rex <i>et al.</i>	Lieberman <i>et al.</i>	Pickhardt <i>et al.</i>
Publication year	1991	1993	2000	2003
Study period	1989–1991	*	1994–1997	2002–2003
Recruit after referral for CRC screening	Yes	No [†]	Yes	Yes
Exclude if family history of CRC	Yes	Yes	No	No
Sample size	119	621	3121	1233
Percent male	100%	61%	97%	59%
Mean age	64.0	60.2	62.9	57.8
Observed adenoma prevalence	41.2%	27.2%	37.5%	28.5%
Predicted adenoma prevalence				
Colonoscopy 100% sensitive				
Predicted	43.9%	36.5%	41.8%	33.5%
95% CI	(15.9%, 70.7%)	(10.0%, 61.9%)	(12.2%, 69.2%)	(4.6%, 63.3%)
p -value	0.56	0.72	0.60	0.58
Colonoscopy 85% sensitive				
Predicted	41.0%	33.9%	39.0%	31.1%
95% CI	(13.6%, 67.3%)	(8.6%, 58.5%)	(10.4%, 66.0%)	(3.6%, 59.7%)
p -value	0.48	0.66	0.53	0.52

* Not reported, publication year (1993) used as midpoint of study period models.

[†] Rex and colleagues [44] recruited individuals by sending invitation letters to physicians, dentists, nurses, and their spouses.