Supplementary Materials

**sGERD prevalence model**

The sGERD prevalence model represents the probability for sGERD as a function of age, . The probability of sGERD changes with an individual’s age represented by variable  that ranges from  (birth) to the end of followup age . Parameters ,  represent the effective sGERD initiation rate (accounting for initiation and remission) earlier and later in life, respectively, and the parameter  is an estimated change-point age. The age-adjusted sGERD prevalence is fixed approximately equal to 0.2 during the 1990-2000 year time frame, but modified outside of this time frame by cohort and/or period effects during Phase 2 modeling. The sGERD prevalence model is used by the EAC incidence model to evaluate age-dependence on the rate for BE initiation due to sGERD, that is modeled as . In Phase 2 modeling, the sGERD prevalence model is used to stratify the population (for each year of age and calendar year) by decade of acquiring sGERD. For each stratum, biological parameters are set at the baseline rate prior to sGERD acquisition, and at a different rate after sGERD onset.

**EAC incidence model parameters and parameter identifiability**

The EAC incidence model includes age-dependence on the rate for BE initiation at age  due to sGERD that is modeled as . The estimated parameter  represents the baseline rate for field-conversion from normal to BE tissue,  represents a factor increasing the rate of conversion to BE given sGERD.

 All cancers are assumed to begin in BE tissue. Cancer model parameters for cell division rates are represented by death, apoptosis or differentiation rates by with subscripts *P* and *M* referring to premalignant and malignant cell types (e.g.  is the premalignant cell division rate). Mutation rates are represented bywithand representing the rate for each of the two initial hits transforming BE stem cells to premalignant cells, andthe mutation rate to generate a malignant cell. Cancer incidence is modeled using a (per cell) detection rate. Any of these cell kinetic rates may depend on time  through cohort and period effects or through dependence on sGERD or OF, but for notational simplicity this will not be shown.

**Modeling of cohort and period trends**

MLE methods were used to estimate linear and/or sigmoidal temporal trends simultaneously with rates representing biological processes. Parameters representing a process (e.g. ) that are modified by cohort trends for individuals born in year  and period trendsfor calendar year were modeled as a product of the baseline rate for the process with the cohort and period trends (e.g., the BE initiation rate was modeled as  . To represent no trend on cohort and/or period, the trend functions were set toand/or , respectively. Linear trends for cohort were modeled using a slope and intercept parameter. Sigmoidal tends include three parameters representing an offset, an inflection time, and the slope at the inflection point. All trends were constrained to keep biological rates non-negative.

Linear and/or sigmoidal temporal trends were used to represent changing rates for biological processes during Phase 1 modeling, and to represent changing sGERD or OF exposures during Phase 2 modeling. For a given mechanism  (e.g.,  for promotion, or  for trends on sGERD) the absence of a cohort trend was modeled as ; for linear trend with intercept and slope parameters (*c1*, *c2*); and for sigmoidal trend , with offset  and maximum slope controlled by occurring at reference birth cohort year . Similarly, period trends were modeled by calendar year  as for no trend,  for linear trend, and as  for a sigmoidal period trend, with maximum slope controlled by parameters occurring at reference calendar year .

sGERD cohort and period trends were constrained to maintain age-adjusted sGERD prevalence in agreement with the data from the two studies by Locke, *et al.,* who found age-adjusted sGERD prevalence of approximately 20% ([1](#_ENREF_1), [2](#_ENREF_2)) in year 2000. This was achieved by pivoting the trends around year 2000 and birth year 1940 by modeling sGERD cohort and period trends as , and , respectively. Using these formulae, sGERD prevalence was evaluated for age , birth cohort , and calendar year  as . sGERD trends by age, period and cohort are shown form men and women in Figure S1, panels b) and c), respectively .

**Model equations and boundary conditions**

Let  represent the mean number of stem cells in BE tissue. We model carcinogenesis as beginning in BE tissue using a vector of random variables  where  is an indicator that represents the absence or presence of BE at time , and random variables  represent the number of cells in BE tissue with a single hit or mutation, the number of premalignant cells (two hits), and the number of malignant cells, respectively, at time . Let  be an indicator representing survival, or cancer diagnosis, respectively by time. For we define a generating functions  that represents the entire cancer process at time beginning with only normal cells and no observations at time , and generating functions that begin with BE or a single cell of typeat time ,

 ,

 

 where , and is the Kronecker delta which equals  if  and  otherwise, and . The generating function satisfies the Kolmogorov backward equations,

 ,

 ,

 ,

 

 ,

 

 

The survival is calculated as with boundary conditionsand .

To calculate the hazard, the equations above are differentiated with respect to time ,

 ,

 ,

 ,

 

 ,

 

 .

with the hazard calculated as . The hazard calculation is simplified slightly by defining  that satisfies the equation

 with  .

Holding the observation time  fixed, the Kolmogorov backward and time-derivative equations are solved simultaneously working backward in time from  to . This is conveniently done through a change of variable (with ). To simplify notation, let , , , ,, , , , , .

Then the solution for the EAC survival and hazard (with and ) is found by simultaneously solving the system of ordinary differential equations (ODE)s for ranging from  to ,

 

 

  

 

 

 

 

 

 with boundary conditions . The survival at time is and the hazard is .

 The inclusion of birth-cohort and period trends in the hazard was done for each model by solving the equations above with varying rates for one or more of the rate parameters based on the modeled linear or sigmoidal trends assumed by a given model, as described above. Thus, for example, in the case of birth-cohort and period trends influencing premalignant promotion, changing rates for and as a function of age  were included using the cohort and period trends in the model during solving of the equations above to generate hazards  for each age, period, and cohort, indexed with variables  respectively, with cohort calculated as .

**EAC model parameter identifiability**

Only specific parameter combinations in the model are mathematically identifiable, including combinations  (with and called the net proliferation rates for premalignant or malignant cells, respectively), the combination  and the combination  ([3](#_ENREF_3), [4](#_ENREF_4)). However,  and the combination are numerically only weakly identifiable. Thus without loss of generality, these identifiability constraints allow us to equate the background rates for the first two mutation (defining ) and choose numerical values for a few parameters (e.g., fixing , , ) while setting  to represent malignant size detection occurring on average at order  malignant cells. Other numerical choices for these fixed parameters would rescale the estimated parameters without changing the model fit. After making these fixed parameter choices, the EAC incidence model include 6 estimated parameters representing baseline biological rates. In addition, a variable number of additional parameters are used to represent cohort and period trends (Phase 1) or trends on OF and sGERD (phase 2). All mutation, division, death, and detection rates are in units of (per cell per year).

**Model likelihood**

Maximum likelihood methods were used to fit to EAC incidence data from SEER by single years for ages 20 through 84 and calendar years 1975 through 2009. The Poisson likelihood is calculated as , where  is the number of observed EAC incident cases for age  and period . In this equation,  is the expected number of EAC cancers at age , period (calendar year) , and birth cohort ,  is the number of person years of age  and period , and the birth-cohort specific hazard for Phase 1 modeling is . During Phase 2 modeling, the hazard was calculated as , where sGERD duration is stratified by decade of sGERD onset and represented by  (with index  representing never-sGERD and  representing the last age decade at end of followup) using weights representing the change in sGERD prevalence in each decade, and with hazards  calculated using no-sGERD exposure until the midpoint of the decade of sGERD onset, and with a changed rate afterwards. For example, beginning at  (end of followup), the premalignant net cell proliferation rate after onset of sGERD is increased by parameter  with the rate calculated as  (including OF cohort and period effects) then continued after the sGERD onset age using the rate  until  (age ). Similar formulas as above were used for the premalignant cell division rate  while substituting  for . Then the premalignant cell death rate was calculated as  during solution of the system of Kolmogorov equations for the hazard.

**Phase 1 model selection: Identify biological processes driving EAC incidence trends**

The Phase 1 analysis began with likelihood-based optimization and comparison of models with a single cohort or period trend, modeled parametrically as a linear or sigmoidal trend affecting any one of five mechanisms during EAC development (See Table S1 in SM). The best likelihoods of the ten optimized 2-parameter linear trend models consisted of a model with a linear period trend on BE initiation () and a model with a linear cohort trend on premalignant promotion ( and ). These two linear models did not differ statistically in goodness of fit from each other, but were much better (with highly significant log likelihood differences of 130 points or more) than linear trends on any of the other processes, including initial mutation (), malignant transformation (), or malignant promotion (and ). Next sigmoidal cohort and period (3-parameter) trends were evaluated on each of the five biological processes, showing some improvement with a sigmoidal period trend on the development of BE (), but a highly significant (24 point log-likelihood) improvement with a sigmoidal cohort trend on premalignant promotion, compared with a linear trend, and very poor likelihoods for all other combinations of cohort or period with the other biological mechanisms. Systematic comparison of models including additional period and/or cohort trends on combinations of mechanisms (stopping with a maximum of 6 estimated trend parameters) identified two models with significant (p=1.3e-4) improvement (10.24 additional points of log-likelihood with 3 additional parameters) improvement over the sigmoidal cohort trend on premalignant promotion. These two models combined the sigmoidal cohort trend on premalignant promotion with either a sigmoidal cohort or a sigmoidal period trend affecting BE initiation. A third-ranked model (loss of 1.83 log likelihood points, with the same number of parameters compared to the best model) included cohort and period trends affecting only premalignant promotion. This promotion-only model was ~24 log likelihood points better than a BE-initiation-only model with cohort and period trends affecting only BE initiation using the same number of parameters. We also compared the likelihood for the best model against a model with sigmoidal period and cohort trends not influencing biological parameters, but instead multiplying the multistage hazard, similar to a traditional age-period-cohort (APC) model. Although not strictly comparable from a nested-model perspective, this APC model was 6.84 log likelihood points worse that the best model (with cohort trends acting on premalignant promotion and period trends on BE initiation) while requiring the same number of parameters (See Table S1).

**Phase 2 model selection: identify mechanistic actions of OF and sGERD driving EAC trends**

The models considered in Phase 2 extended Phase 1 models through introduction of age-dependent sGERD and OF acting on important biological mechanisms identified in Phase 1 modeling, and stratifying the population by decades of sGERD duration to combine hazards for these different strata with different sGERD-associated risks. First the sGERD prevalence model, including sGERD period and cohort trends, was used to stratify the SEER population for each year of age between 20 and 84 and each calendar year from 1975 through 2009 by decade of sGERD acquisition, and also for individuals never acquiring sGERD, with weights assigned to each stratum equal to the probability of acquisition of sGERD in each decade of life. sGERD and OF were applied to different combinations of promotion and BE initiation parameters in each stratum, with only OF influencing the biological parameters up to age of sGERD acquisition (modeled as the midpoint of the age decade for each stratum), followed by both sGERD and OF influencing the biological processes until age at observation.

The Phase 1 model family can be considered as a special case of the Phase 2 model family in which sGERD has no period or cohort trends and only affects BE initiation while OF are assigned period and cohort equal to the cohort and period trends optimized in Phase 1 modeling. The Phase 2 model code was tested in this limit, and verified that it produced the same likelihoods found for the Phase 1 models.

Phase 2 modeling identified highly significant cohort trends on OF as the primary driver of EAC incidence trends, especially among men, with significant sGERD period and cohort effects also making important contributions, especially for long-duration sGERD. See Table S2 for final model parameter estimates, and Tables 1-5 and Figures 1-2, and Figure S2 for attributable contributions from OF and sGERD to incidence trends for men and women.

**Annual attributable risks from sGERD and OF by cohort and period, and by mechanism**

Figure S2 shows comparisons of annual observed and model predicted age-adjusted EAC incidence rates (per 100,000 individuals) between 1975-2009 for men and women, showing contributions from sGERD versus OF, cohort versus period for OF and for sGERD, and contributions from sGERD promotion versus sGERD on BE initiation. Observed age-adjusted rates from SEER data are shown as circles, and estimated rates shown as stacked bar graphs showing contributions from different factors. Figure panels 2a and 2b show model predictions for men and women, respectively, are divided into background rates expected in the absence of sGERD and OF trends (black bar segments), contributions from direct effects from OF (green), contributions from interaction of OF and sGERD (violet), and from direct effects of sGERD (brown), with data shown by black circles. OF dominate the increase in risk for both sexes, with smaller contributions from sGERD and interactions. The absolute risk per 100,000 individuals and the increase in risk over background are both smaller for women than men. Panels 2c and 2d compare sGERD (brown) versus direct plus interaction effects from OF due to cohort (blue), period (red), and interaction between cohort and period (green) for men and women. These panels show that OF cohort effects dominate the attributable risk increase in EAC between 1975 and 2009. Similarly, panels 2e and 2f compare the effects of cohort (blue), period (red), and interaction between cohort and period (green) on sGERD and sGERD interactions, along with direct effects of OF (brown) for men and women. Period effects are more important for explaining sGERD trends, especially among women. Panels 2g and 2h compare attributable risk contributions from sGERD on BE initiation (blue), interaction of BE with promotion (green) and direct effects of premalignant (violet) and malignant (red) promotion for men and women. Malignant promotion makes only a minor contribution to risk. Premalignant promotion and interactions between promotion and BE dominate the contributions of sGERD to EAC risk trends.

**Table S1**

|  |
| --- |
| Phase 1 Model comparisonsLikelihoods with period and cohort factors on Barrett’s initiation and premalignant promotion |
| **BE initiation ()** | **Linear cohort ()** | **Linear period ()** | **Sigmoid cohort ()** | **Sigmoid period ()** | **Linear cohort ()****Linear period ()** | **Sigmoid cohort ()****Sigmoid period ()** |
| (-log(lkh)) | 336.50 | 53.60 | 147.01 | 52.98 | 51.41 | 25.84 |
| Shape DOF | 2 | 2 | 3 | 3 | 4 | 6 |
|  |
| **Premalignant promotion ()** | **Linear cohort ()** | **Linear period ()** | **Sigmoid cohort ()** | **Sigmoid period ()** | **Linear cohort ()****Linear period ()** | **Sigmoid cohort ()****Sigmoid period ()** |
| (-log(lkh)) | 54.32 | 184.31 | 10.24 (good) | 184.30 | 22.30 | 1.83 (better) |
| Shape DOF | 2 | 2 | 3 | 3 | 4 | 6 |
|  |
| **BE initiation ()****& premalignant promotion ()** |  |  |  |  | **Sigmoid period ()****Sigmoid cohort ()** | **Sigmoid cohort ()****Sigmoid cohort ()** |
| (-log(lkh)) |  |  |  |  | 0.00 (best model) | 0.13 (~tie best) |
| Shape DOF |  |  |  |  | 6 | 6 |
|  |
| **APC model – non-bio PC** |  |  |  |  | **Sigmoid period****Sigmoid cohort** |  |
| (-log(lkh)) |  |  |  |  | 6.84 |  |
| Shape DOF |  |  |  |  | 6 |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Table S2**  |  |  |  |  |  |  |  |
| **EAC** | **Male** | **Male** | **Male** | **Male** | **Female** | **Female** | **Female** | **Female** |
| **Model****Parameter** | **MLE** | **MCMC****median** | **MCMC****Lower CI** | **MCMC****Upper CI** | **MLE** | **MCMC****median** | **MCMC****Lower CI** | **MCMC****Upper CI** |
|  | 3.1082E-04 | 3.1468E-04 | 2.5245E-04 | 3.8137E-04 | 1.4376E-04 | 1.4354E-04 | 9.6841E-05 | 2.6451E-04 |
|  | 6.9069E-04 | 6.7344E-04 | 5.4387E-04 | 9.5365E-04 | 5.8084E-04 | 5.3778E-04 | 3.4424E-04 | 8.3027E-04 |
|  | 8.6760E-02 | 8.7356E-02 | 6.9856E-02 | 9.6273E-02 | 4.4409E-02 | 4.3328E-02 | 3.0459E-02 | 4.8761E-02 |
|  | 1.1564E-04 | 1.1910E-04 | 6.8415E-05 | 1.7235E-04 | 1.0093E-04 | 9.9172E-05 | 6.8198E-05 | 1.4670E-04 |
|  | 6.7512E-01 | 6.6721E-01 | 5.2802E-01 | 8.1773E-01 | 6.7742E-01 | 6.1769E-01 | 4.9405E-01 | 7.8001E-01 |
|  | 7.1538E-01 | 7.0479E-01 | 5.2499E-01 | 9.3236E-01 | -2.5768E-01 | -1.2354E-01 | -2.8499E-01 | 4.8365E-02 |
|  | 5.8556E-02 | 5.9942E-02 | 5.4326E-02 | 7.2086E-02 | 1.8121E-02 | 2.0126E-02 | 1.7249E-02 | 2.6835E-02 |
|  | -3.7741E-02 | -3.6785E-02 | -6.2484E-02 | 7.8880E-03 | 2.9034E+00 | 2.9094E+00 | 2.8126E+00 | 3.0513E+00 |
|  | 2.6473E-03 | 2.6620E-03 | 2.5944E-03 | 2.7321E-03 | 2.7117E-04 | 2.7050E-04 | 2.5748E-04 | 2.8705E-04 |
|  | 1.9141E+03 | 1.9140E+03 | 1.9119E+03 | 1.9157E+03 | 1.9513E+03 | 1.9517E+03 | 1.9494E+03 | 1.9651E+03 |
|  | 1.9876E+03 | 1.9879E+03 | 1.9843E+03 | 1.9911E+03 | 1.9917E+03 | 1.9916E+03 | 1.9876E+03 | 1.9949E+03 |
|  | 3.0600E+00 | 3.0486E+00 | 3.8079E-01 | 6.8961E+00 | 5.8811E+00 | 7.3403E+00 | 1.5212E+00 | 1.3933E+01 |
|  | 1.8451E+00 | 1.9174E+00 | 4.5892E-01 | 4.6722E+00 | 1.3859E+00 | 1.4285E+00 | 7.2677E-01 | 4.3558E+00 |
|  | 1.3676E+00 | 1.3476E+00 | 1.2531E+00 | 1.3926E+00 | 2.0628E+00 | 2.0773E+00 | 1.9613E+00 | 2.1709E+00 |
|  | 6.4747E-01 | 6.9367E-01 | 4.9244E-01 | 1.1596E+00 | 2.0263E+00 | 1.9115E+00 | 1.2607E+00 | 2.4881E+00 |
|  | 3.8676E-01 | 3.8312E-01 | 3.3880E-01 | 4.4912E-01 | 4.2820E-01 | 4.2232E-01 | 3.2462E-01 | 5.0083E-01 |
|  | 2.2237E+00 | 2.3015E+00 | 1.9321E+00 | 2.9443E+00 | 5.7622E+00 | 5.7775E+00 | 4.2861E+00 | 8.4305E+00 |
|  | 9.0407E-02 | 9.9360E-02 | 2.2481E-02 | 3.2349E-01 | \* | \* | \* | \* |
| **sGERD** | Male | Male | Male | Male | Female | Female | Female | Female |
| **Prevalence****Model** | MLE | MCMCmedian | MCMCLower CI | MCMCUpper CI | MLE | MCMCmedian | MCMCLower CI | MCMCUpper CI |
|  | 8.5259E-04 | 2.1748E-03 |  1.1866E-04 | 9.1875E-03 | 1.1488E-03 | 1.9205E-03 | 2.2031E-04 | 1.7428E-02 |
|  | 7.0330E-03 | 5.5393E-03 |  1.4770E-03 | 1.0617E-02 | 7.9586E-03 | 6.8473E-03 | 3.3227E-03 | 4.8441E-02 |
|  | 2.6629E+01 | 2.4198E+01 | 1.0208E+01 | 3.6612E+01 | 3.3179E+01 | 2.7799E+01 | 1.2212E+01 | 4.1103E+01 |

\*Female  parameter equated to male parameter estimate. MLE=maximum likelihood estimate. MCMC=Markov chain Monte Carlo. MCMC upper and lower CIs are two-sided 95% credibility intervals.

**Supplementary Table and Figure Captions**

**Table S1**. Phase 1 modeling evaluated important biological mechanisms driving the observed EAC incidence trends. Five biological mechanisms were compared using linear or sigmoidal cohort and period trends acting on one or more mechanism at a time, with all models belonging to a nested family of multiscale EAC models. The table shows selected models, including good fits for a model including only a cohort effect on premalignant promotion, and better and best models that include additional period or cohort effects on BE initiation.

 **Table S2**. Final model parameter estimates include simultaneous estimates for both biological and trend parameters based on the best model found during Phase 2 modeling, which introduced cohort and period effects on sGERD and OF that act on biological mechanisms identified during Phase 1 modeling. All parameter estimates include MLEs, MCMC medians, and 95% MCMC credibility intervals (CI)s. The estimate of was not significant for women, so this parameter was fixed to the value estimated for men.

**Figure S1**. The multiscale EAC incidence model represents the mechanistic effects of sGERD and OF in driving premalignant promotion, and to a lesser degree BE initiation, in explaining EAC incidence trends. The effects of different sGERD onset ages are represented by modeling separate strata representing decade of sGERD onset age, including no sGERD. A weighted hazard is computed for each stratum, and summed to compute the expected EAC incident cases for each year from 1975-2009 and for each age 20-84. The multiscale model, including data sources, is shown in panel a). Panels b) and c) show, for males and females respectively, the estimated effects of age, period, and cohort on sGERD. The most important mechanistic action of sGERD appears to be through increasing premalignant promotion in driving EAC trends. Note that trend lines for the latest birth cohorts include projections into the future. Panels d) and e) show estimates of the premalignant promotion rate by age and cohort, with the cohort trend on OF as the most important factors driving EAC incidence trends for both men and women, respectively. Panels f) and g) show the shapes of the estimated OF cohort trends for men and women, respectively.

**Figure S2**. Observed and expected annual age-adjusted EAC incidence (per 100,000 individuals) is shown for years 1975 through 2009 for men (left panels) and women (right panels) according to the attributable influence of direct effects and interactions due to sGERD, OF, cohort, period, and biological mechanism driving these trends.

References

1. Locke GR, 3rd, Talley NJ, Fett SL, Zinsmeister AR, Melton LJ, 3rd. Prevalence and clinical spectrum of gastroesophageal reflux: a population-based study in Olmsted County, Minnesota. Gastroenterology. 1997;112(5):1448-56. Epub 1997/05/01. PubMed PMID: 9136821.

2. Locke GR, 3rd, Talley NJ, Fett SL, Zinsmeister AR, Melton LJ, 3rd. Risk factors associated with symptoms of gastroesophageal reflux. The American journal of medicine. 1999;106(6):642-9. Epub 1999/06/23. PubMed PMID: 10378622.

3. Heidenreich WF, Luebeck EG, Moolgavkar SH. Some properties of the hazard function of the two-mutation clonal expansion model. Risk analysis : an official publication of the Society for Risk Analysis. 1997;17(3):391-9. Epub 1997/06/01. PubMed PMID: 9232020.

4. Luebeck EG, Curtius K, Jeon J, Hazelton WD. Impact of tumor progression on cancer incidence curves. Cancer research. 2013;73(3):1086-96. Epub 2012/10/12. doi: 10.1158/0008-5472.CAN-12-2198. PubMed PMID: 23054397.